Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

Fisica

Ciclo XX

Settore/i scientifico disciplinari di afferenza: *FIS02-MAT07*

TITOLO TESI

NON-MARKOVIAN STOCHASTIC PROCESSES AND THEIR APPLICATIONS: FROM ANOMALOUS DIFFUSION TO TIME SERIES ANALYSIS.

Presentata da: *Antonio Mura*

Coordinatore Dottorato:

Prof. Fabio Ortolani

Relatore:

Prof. Francesco Mainardi

Esame finale anno 2008

Antonio Mura

Dipartimento di Fisica, Università e INFN
V. Irnerio 46, I-40126 Bologna, Italia, tel:051-2091109
e-mail: mura@bo.infn.it, anto.mura@gmail.com

Non-Markovian Stochastic Processes and their Applications:
from Anomalous Diffusion to Time Series Analysis.

Phd thesis: XX Ciclo
Supervisor: Prof. Francesco Mainardi
Coordinator: Prof. Fabio Ortolani

Alma Mater Studiorum, Università di Bologna.

Acknowledgments

I thank Prof. Francesco Mainardi, for the precision and competence which always punctually characterizes his scientific work, and especially, for many of the satisfactions and opportunities that he was able to give in these years.

Special thanks to Prof. Murad S. Taqqu, for the extraordinary hospitality and greetings that he gave during my stay in Boston, and his immense scientific competence to which I have drawn in the short period that we spent together.

Finally, I thank Prof. Massimo Campanino, with which it has been a pleasure, and an honour, to work together.

A special thanks to all my family for their love and support that I have ever miss.

Ringrazio il Prof. Francesco Mainardi, sia per la precisione e la competenza che sempre e puntualmente contraddistingue il suo lavoro scientifico, e soprattutto per le tante soddisfazioni ed opportunità che è stato in grado di regalarmi in questi anni.

Un ringraziamento speciale al Prof. Murad S. Taqqu, per la straordinaria ospitalità ed accoglienza che ci ha riservato durante il nostro indimenticabile soggiorno a Boston, e per la sua immensa competenza scientifica al quale ho attinto voracemente nel breve periodo che abbiamo passato assieme.

Ringrazio infine il prof. Massimo Campanino con cui è stato un piacere, ed un onore, collaborare professionalmente.

Un ringraziamento speciale a tutta la mia famiglia per l'affetto ed il supporto che non mi hanno mai fatto mancare.

Antonio Mura

Contents

1 Introduction

The concept of diffusion processes in physics is quite different from those adopted in the mathematical theory of stochastic processes. For mathematicians, diffusion processes are defined as continuous time stochastic processes that satisfy the Markov property (eq. 3.18) and for which the sample paths are (with probability one) continuous functions of time (see **Section 3.5**, Def. 3.12). In physics, diffusion processes stand for general processes of *re-distribution* in space and time of the spatial density of an extensive quantity (as mass, charge or probability) for which the total quantity is conserved and, when required, the initial sign is preserved. We will be interested in processes for which a probabilistic interpretation is possible.

A characteristic property of diffusion processes concerns the asymptotic growth of the *variance* for large times, that is the centered second moment of the spatial density function (**Section 2.2**, Def. 2.22). We call the diffusion process *normal* or *anomalous* if the asymptotic growth of the variance is linear or not linear in time, respectively. Among these, we single out *sub-diffusive* processes, which have variance asymptotically growing slower than linear, and *super-diffusive* processes, for which the variance grows faster than linear (or also it does not exist, i.e. it is infinite).

Remark 1.1 (Historical remarks)**.** Anomalous diffusion has been known since Richardson's treatise on turbulent diffusion in 1926. Within transport theory it has been studied since the late 1960s. In particular, its theoretical investigation was prompted by Scher and Montroll in their description of dispersive transport in amorphous semiconductors, a system where the traditional methods actually were failing. Today, the list of systems displaying anomalous dynamical behavior is very wide. It includes, among the others, the following systems [14]:

- (*Sub-diffusion*) Charge carrier transport in amorphous semiconductors; nuclear magnetic resonance (NMR); reptation dynamics in polymeric systems; transport on fractal geometries; diffusion of scalar tracers in convection roll arrays, bead dynamics in polymeric networks.

- (*Super-diffusion*) Special domains of rotating flows; collective slip diffusion on solid surfaces; layered velocity fields; Richardson turbulent diffusion; bulk-surface exchange controlled dynamics in porous glasses; transport in heterogeneous rocks; quantum optics; single molecule spectroscopy; transport in turbulent plasma; bacterial motion.

Many diffusion processes are governed by second-order linear parabolic equations. A typical example is provided by the one-dimensional equation

$$\frac{\partial}{\partial t}u(x,t) = C(t)\,\frac{\partial^2}{\partial x^2}u(x,t), \quad x \in \mathbb{R}, \quad t \geq 0, \tag{1.1}$$

subjected to the initial condition $u(x,0_+) = u_0(x)$, where $C(t)$ is a positive function of time called *diffusivity*, which has dimensions $[L^2\,T^{-1}]$ $(length^2/time)$.

Remark 1.2. If the diffusivity is taken constant in time, (eq. 1.1) reduces to the *standard diffusion equation*, which describes, for instance, the variation in temperature as heat diffuses through a physical medium.

From a physical point of view, (eq. 1.1) is originated from the system of two laws

$$\begin{cases} \dfrac{\partial}{\partial t}u(x,t) + \dfrac{\partial}{\partial x}F[u(x,t)] = 0, \\ F[u(x,t)] = -C(t)\,\dfrac{\partial}{\partial x}u(x,t)\,, \end{cases} \tag{1.2}$$

where F is the one-dimensional *flux*. The first law is indeed a conservation law, whereas the second one is an empirical law stating the phenomenological relationship between the flux and the gradient of the field variable.

Remark 1.3. This phenomenological relationship is met in several physical contexts with different names according to the meaning of the filed variable u. For instance, when u is the temperature it is known as *Fourier's law*, when u is the pore pressure as *Darcy's law*, when u is a concentration of particles as *Fick's law*.

From a mathematical point of view, equations like (eq. 1.1) can be for instance obtained considering the time evolution of the marginal density function (**Section 3.1**, Rem. 3.1) of stochastic diffusions, which satisfy the Fokker-Planck equation (eq. 3.33).

We observe that (eq. 1.1) can be written as a partial integral-differential equation

$$u(x,t) = u_0(x) + \int_0^t C(s)\,\frac{\partial^2}{\partial x^2}u(x,s)\,ds, \quad x \in \mathbb{R}, \quad t \geq 0\,. \tag{1.3}$$

which has the advantage to take automatically into account of the initial condition. We interpret (eq. 1.3) in the sense of generalized functions, namely we let $u(x,t)$ be a distribution (see for instance Def. 8.1.3). So that, when $u_0(x) = \delta(x)$, where $\delta(x)$ is the Dirac distribution centered at $x = 0$, the corresponding solution is termed *fundamental solution*, *Green function* or *propagator* as well.

Example 1.1. Let us consider:

$$C(t) = \alpha K_\alpha t^{\alpha-1}, \quad K_\alpha > 0, \quad \alpha > 0. \tag{1.4}$$

In this case, (eq. 1.1) can be written

$$\frac{\partial}{\partial t}u(x,t) = \alpha t^{\alpha-1} K_\alpha \frac{\partial^2}{\partial x^2}u(x,t), \quad x \in \mathbb{R}, \quad t \geq 0. \tag{1.5}$$

Namely,

$$u(x) = u_0(x) + \alpha K_\alpha \int_0^t s^{\alpha-1}\,\frac{\partial^2}{\partial x^2}u(x,s)\,ds, \quad x \in \mathbb{R}, \quad t \geq 0\,. \tag{1.6}$$

When $\alpha \neq 1$ the positive constant K_α is called *effective diffusivity* because it has dimensions $[L^2\,T^{-\alpha}]$. For any $\alpha > 0$, the fundamental solution of (eq. 1.5) turns out to be the Gaussian density

$$u(x,t) = \frac{1}{\sqrt{4\pi K_\alpha\, t^\alpha}}\, e^{\,-x^2/(4K_\alpha\, t^\alpha)}, \tag{1.7}$$

which evolves in time with variance

$$\sigma^2(t) = 2K_\alpha\, t^\alpha\,. \tag{1.8}$$

As a consequence, the diffusion process governed by (eq. 1.5) is Gaussian and anomalous. In particular, one has sub-diffusion if $0 < \alpha < 1$ and super-diffusion if $\alpha > 1$. Moreover, the normal standard diffusion is recovered when $\alpha = 1$.

Set for simplicity $K_\alpha = 1$ in (eq. 1.5). The standard *Brownian motion* $B(t)$, $t \geq 0$[1] (**Section 3.4**) is naturally associated to the standard normal diffusion equation, that is (eq. 1.5) with $\alpha = 1$, through the Fokker-Planck equation (see eq. 3.26). But, which is the corresponding stochastic process $X(t)$, $t \geq 0$, associated to (eq. 1.5)? If one has the Einstein interpretation of Brownian motion in mind, equation (1.5) could be interpreted as associated to the random movements $X(t)$ of a free particle in a frictionless medium with time dependent diffusivity given by $C(t)$. In fact, we can consider the following Langevin equation (see eq. 4.47) describing the particle motion:

$$\frac{dX(t)}{dt} = (\alpha t^{\alpha-1})^{1/2} Z(t), \tag{1.9}$$

where $Z(t)$ is a white-noise process (**Section 8.1**). That is, one is considering the stochastic differential equation (**Section 4.2.1**)

$$dX(t) = (\alpha t^{\alpha-1})^{1/2} dB(t),$$

where $B(t)$ is a standard Brownian motion. Then, the corresponding Fokker-Planck equation is actually (eq. 1.5). We observe that the above equation must be interpreted as (see **Section 4.2**)

$$X(t) = X(0) + \int_0^t (\alpha s^{\alpha-1})^{1/2} dB(s),$$

which, by setting $X(0) = 0$ and using the integration by parts formula (eq. 4.46), can be written

$$X(t) = (\alpha t^{\alpha-1})^{1/2} B(t) + \int_0^t \frac{\alpha(\alpha-1)s^{\alpha-2}}{2\sqrt{\alpha s^{\alpha-1}}} B(s) ds. \tag{1.10}$$

Then, $X(t)$ is indeed Gaussian with zero mean, variance $2t^\alpha$ and covariance $\mathbf{cov}(X(t_1)X(t_2)) = 2\min(t_1^\alpha, t_2^\alpha)$.

Remark 1.4. The process $X(t)$ of (eq. 1.10) is equal, in the sense of finite-dimensional distributions (**Chapter 3**), to a "stretched" Brownain motion $B(t^\alpha)$, as in fact (eq. 1.7) would suggest. Indeed, $B(t^\alpha)$ is Gaussian with zero mean and covariance $\mathbf{cov}(B(t_1^\alpha)B(t_2^\alpha)) = 2\min(t_1^\alpha, t_2^\alpha)$, $t_1, t_2 \geq 0$.

The process $X(t) = B(t^\alpha)$, $t \geq 0$, is actually a stochastic (Markovian) diffusion process (Def. 3.12) and it is easy to understand that the "anomalous" behavior of the variance comes from the power-like time stretching. However, the Brownian motion stationarity of the increments (Def. 3.9) is lost due just to the non-linear time scaling. One can preserve this property on the condition to drop the Markovian property. For instance, if $0 < \alpha < 2$ equation (1.7) is also the marginal density function of a standard *fractional Brownian motion* (fBm) $B_H(t)$, $t \geq 0$ of order $H = \alpha/2$ (**Section 3.7**). Such a process is Gaussian, self-similar and with stationary increments (H-**sssi**) (**Section 3.6**). It has zero mean and variance $2t^\alpha$. However, its autocovariance function is $\mathbf{cov}(B_H(t_1), B_H(t_2)) = t_1^\alpha + t_2^\alpha - |t_1 - t_2|^\alpha$, $t_1, t_2 \geq 0$, (see eq. 3.36) and thus it is automatically non-Markovian (Remark 3.22).

Remark 1.5. Taking a fractional Brownian motion $B_{\alpha/2}(t)$, $t \geq 0$, as stochastic model for the anomalous diffusion described by (eq. 1.5) has many advantages. Even if the physical interpretation provided by (eq. 1.9) is lost, the anomalous diffusion is linked now to the notion of long-range dependence (LRD) (**Chapter 6**), which is a very common characteristic of time series concerning many different contexts (**Section 6.4.1**). An H-**sssi** process is said to exhibit long-range dependence, or long-memory, if $H > 1/2$ (Prop. 6.1). Then, the fast-anomalous diffusion region of (eq. 1.5) actually corresponds to the long-range dependence domain of fractional Brownian motion.

1) With the word standard Brownian motion here we mean that $\mathbf{var}(B(1)) = 2$.

Remark 1.6. The above example is useful to understand that there exists many different stochastic processes which can serve as models for a given diffusion equation. Some have a ready made physical interpretation, others are advisable because of their mathematical properties, which can make them very useful for instance in statistical modeling.

Example 1.2. We have seen that a Gaussian anomalous diffusion can be obtained from a standard diffusion equation with time dependent diffusivity. Anomalous diffusion processes can also be obtained as Gaussian processes with time subordination. As a consequence, the probability density is no longer Gaussian. In fact, let $f(x,t)$ be the density function of a diffusive process and $G(x,t)$ a standard Gaussian density. Namely,

$$G(x,t) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right), \quad t \geq 0. \tag{1.11}$$

Let $\varphi_\beta(\tau,t) = t^{-\beta}\phi(\tau t^{-\beta})$, with $\tau \geq 0$, $t > 0$ and $\beta > 0$, be the marginal probability density function of a *self-similar* random time process. We remember that a process $X(t)$, $t \geq 0$, is said to be self-similar with self-similarity exponent H if, for all $a \geq 0$, the processes $X(at)$, $t \geq 0$, and $a^H X(t)$, $t \geq 0$, have the same finite-dimensional distributions (**Section 3.8**). We also suppose that ϕ has moments of any order. If the subordination formula

$$f(x,t) = \int_0^{+\infty} G(x,\tau)\varphi_\beta(\tau,t)\mathrm{d}\o \tag{1.12}$$

holds, then $f(x,t)$ is the marginal density function of an anomalous diffusion process. In fact,

$$\sigma_f^2(t) = \int_{-\infty}^{+\infty} x^2 f(x,t)dx = \int_{-\infty}^{+\infty} x^{2n}\left\{\int_0^{+\infty} G(x,\tau)\varphi_\beta(\tau,t)d\tau\right\}dx$$

$$= \int_0^{+\infty}\left\{\int_{-\infty}^{+\infty} x^2 G(x,\tau)dx\right\}\varphi_\beta(\tau,t)d\tau = 2\int_0^{+\infty}\tau\varphi_\beta(\tau,t)d\tau = Dt^\beta, \tag{1.13}$$

where we set

$$D = 2\int_0^{+\infty}\zeta\phi(\zeta)\,d\zeta,$$

which is finite by hypothesis. Equation (1.12) can be interpreted as the marginal density function of a *subordinated process* $X(t) = B(l_\beta(t))$, $t \geq 0$, where $B(t)$ is a standard Brownian motion and $l_\beta(t)$ is an independent (see Def. 2.20) non-negative self-similar random time process with marginal density $\varphi_\beta(\tau,t)$.

We remember that the term "subordination" was introduced by Bochner (1962) and refers to processes of the form $Y(t) = X(l(t))$, $t \geq 0$, where $X(t)$, $t \geq 0$, is a Markov process and $l(t)$, $t \geq 0$, is a (non-negative) random time process independent of X. The marginal distribution of the subordinated process is clearly:

$$f_Y(x,t) = \int_0^{\infty} f_X(x,\tau)f_l(\tau,t)d\tau, \quad t \geq 0, \quad x \in \mathbb{R}, \tag{1.14}$$

where $f_X(x,t)$ and $f_l(\tau,t)$ represent the marginal density functions of the processes X and l. Looking at equation (1.12), we are interpreting $G(x,t) = f_B(x,t)$ and $\varphi_\beta(\tau,t) = f_{l_\beta}(\tau,t)$.

Remark 1.7. Observe that the Markov property is in general lost in the process of subordination unless $l(t)$ has non-negative independent increments. If this is the case, $l(t)$ is not decreasing and the process $X(l(t))$ is still a Markov process.

When working with subordinated processes, it is natural to study separately the models for the Markovian diffusion and the random time process. Consider first a non-random process $l(t) = t$, $t \geq 0$, which depicts a non-random linear time evolution. Thus, $f_l(\tau, t) = \delta(\tau - t)$ and it is natural to interpret $f_l(\tau, t)$ as the fundamental solution of the standard forward drift equation (see eq. 7.108), namely,

$$\partial_t u(\tau, t) = -\partial_\tau u(\tau, t), \quad \tau, t \geq 0, \tag{1.15}$$

which can be written in integral form as

$$u(\tau, t) = u_0(\tau) - \int_0^t \partial_\tau u(\tau, s) ds, \quad u_0(\tau) = u(\tau, 0). \tag{1.16}$$

If we wanted a model which takes into account of memory effects, we could introduce a memory kernel in order to obtain the following Volterra type equation

$$u(\tau, t) = u_0(\tau) - \int_0^t K(t - s) \partial_\tau u(\tau, s) ds, \quad \tau, t \geq 0, \tag{1.17}$$

where $K(t)$, with $t \geq 0$, is a suitable kernel, chosen such that the fundamental solution is still a probability density function at each $t \geq 0$.

One can say that the stochastic interpretation through subordinated processes, even if not unique, is very natural. In fact, $Y(t) = X(l(t))$ can have a direct physical interpretation. For example, in equipment usage, $X(t)$ can be the state of a machine at time t and $l(t)$ the effective usage up to time t. In an econometric study, $X(t)$ may be a model for the price of a stock at time t. If $l(t)$ measures the total economic activity up to time t, the price of the stock at time t should be described by $Y(t) = X(l(t))$ rather than simply by $X(t)$. A model for the "effective" time $l(t)$ should take into account of possible memory or non-local effects which can be present in many systems as, heuristically, social behavioral systems.

Remark 1.8. Observe that, in the models obtained starting from (eq. 1.17) the memory component is just within the definition of the marginal, that is one point, probability density function. We can say that (eq. 1.17) is "non-local" (in time), because the definition of $u(\tau, t)$ involves $u(\tau, s)$ at all $0 \leq s \leq t$. Thus, if $l(t)$ is a stochastic process associated to (eq. 1.17), it is automatically non-Markovian and the process $X(l(t))$ cannot be Markovian at all. For this reasons (eq. 1.17) is called *non-Markovian forward drift equation* (**Section 9.2**).

Example 1.3. For instance, if we choose an exponential decay kernel in (eq. 1.17), namely $K(t) = e^{-at}$, $a > 0$, the fundamental solution is actually a probability density function (**Section 9.2**, Example 9.4). In this case, we obtain a model of a system for which non-local memory effects are initially negligible. In fact, $K(t) = e^{-at} \to 1$ as $t \to 0$ and thus the system appears Markovian at small times. Moreover, for large times the system relaxes becoming stationary (**Section 9.5.3**, Rem. 9.12) exponentially fast (Rem. 9.13, eq. 9.83).

We turn now back to Example (1.2). Here, we required that the random time process $l(t)$ be self-similar of order β. We can show (**Section 9.2**, Theor. 9.1) that the only one possible suitable choice of the memory kernel $K(t)$ turns out to be:

$$K(t) = t^{\beta - 1} / \Gamma(\beta), \tag{1.18}$$

where it must be $0 < \beta \leq 1$. Therefore, one gets:

$$\begin{aligned} u(\tau, t) &= u_0(\tau) - \tfrac{1}{\Gamma(\beta)} \int_0^t (t - s)^{\beta - 1} \partial_\tau u(\tau, s) ds \\ &= u_0(x) - \mathcal{I}_t^\beta \partial_\tau u(\tau, t), \quad \tau, t \geq 0. \end{aligned} \tag{1.19}$$

The integral operator $\mathcal{I}_t^\beta f(t) = \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} f(s) ds$ is the so called *Riemann-Liouville Fractional integral* of order β (**Section 7.1**, Def. 7.2). For this reason, (eq. 1.19) is called *fractional (forward) drift equation* (**Section 7.2.4**). Its fundamental solution can be written as $f_{l_\beta}(\tau, t) = \mathcal{M}_\beta(\tau, t)$ (eq. 7.121) and can be expressed by the following power series (see eq, 7.122):

$$\begin{aligned} \mathcal{M}_\beta(\tau, t) = t^{-\beta} M_\beta(\tau t^{-\beta}) \quad &= t^{-\beta} \sum_{k=0}^{\infty} \frac{(-\tau t^{-\beta})^k}{k! \Gamma\left[-\beta k + (1-\beta)\right]} \\ &= \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-r)^k}{k!} \Gamma\left[(\beta(k+1))\right] \sin\left[\pi\beta(k+1)\right], \quad r \geq 0. \end{aligned} \tag{1.20}$$

The above series defines a transcendental function (entire of order $1/(1-\beta)$), which is indeed a probability density function in $\tau \geq 0$ for any $t \geq 0$ (see Remark 9.4) and satisfies the following properties:

1. The Laplace transform of $\mathcal{M}_\beta(\tau, t)$ with respect to t is:

$$\mathcal{L}\{\mathcal{M}_\beta(\tau, t); t, s\} = s^{\beta-1} e^{-\tau s^\beta}, \quad \tau, s \geq 0, \tag{1.21}$$

2. which suggests that in the singular limit $\beta \to 1$ one has:

$$\mathcal{M}_1(\tau, t) = \delta(\tau - t), \quad \tau, t \geq 0. \tag{1.22}$$

3. Moreover, by (eq. 1.20) follows:

$$\mathcal{M}_{1/2}(\tau, t) = \frac{1}{\sqrt{\pi t}} \exp(-\tau^2/4t), \quad \tau, t \geq 0. \tag{1.23}$$

4. Finally, one can show (see **Section 7.2.4**) that, for any $\eta, \beta \in (0, 1)$:

$$\mathcal{M}_\nu(x, t) = \int_0^\infty \mathcal{M}_\eta(x, \tau) \mathcal{M}_\beta(\tau, t) d\tau, \quad \nu = \eta\beta \quad x \geq 0. \tag{1.24}$$

Remark 1.9. The property (1.22) is noteworthy, because is indeed desirable having a random time process $l_\beta(t)$ which for $\beta = 1$ gives $l_1(t) = t$. Thus, the subordinated process $B(l_\beta(t))$ for $\beta = 1$ is the usual Brownian motion, that is the standard stochastic model of a Gaussian normal diffusion.

What is in this case $f(x, t)$? By using equations (1.23) and (1.24) one finds:

$$f(x, t) = \int_0^{+\infty} G(x, \tau) f_{l_\beta}(\tau, t) d\tau = \frac{1}{2} \int_0^\infty \mathcal{M}_{1/2}(|x|, \tau) \mathcal{M}_\beta(\tau, t) d\tau = \frac{1}{2} \mathcal{M}_{\beta/2}(|x|, t),$$

which is the fundamental solution of the so called *time-fractional diffusion equation* of order $0 < \beta \leq 1$ (**Section 7.2.5**, eq. 7.140). This equation can be written in terms of the Riemann-Liouville fractional integral of order β

$$u(x, t) = u_0(x) + \mathcal{I}_t^\beta \partial_{xx} u(x, t), \tag{1.25}$$

and actually represent a generalization of the standard diffusion equation, which is indeed recovered for $\beta = 1$. The time fractional diffusion equation is deeply used to model evolution equations of sub-diffusive-processes. In fact, the variance is given by (eq. 1.13) with $\beta \leq 1$.

Remark 1.10. Observe that (eq. 1.25) involves the same integral operator of (eq. 1.19), or in other words the same memory kernel $K(t) = t^{\beta-1}/\Gamma(\beta)$. This is a general result (**Section 9.3**, Theor. 9.2), which states that the master equation of the marginal density function of the subordinated process $X(t) = B(l(t))$, involves the same memory kernel of the non-Markovian equation (eq.1.17) which governs the one point time evolution of the non-Markovian random time $l(t)$. Namely, one has that $f_X(x,t)$ solves:

$$u(x,t) = u_0(x) + \int_0^t K(t-s)\partial_{xx}u(x,s)ds. \tag{1.26}$$

Equation (1.25) can be written in terms of Riemann-Liouville fractional derivative (**Section 7.1**, Def. 7.3):

$$\frac{\partial}{\partial t}u(x,t) = K_\beta \ \mathcal{D}_t^{1-\beta}\frac{\partial^2}{\partial x^2}u(x,t), \ 0 < \beta \leq 1, \tag{1.27}$$

where $\mathcal{D}_t^\alpha \varphi(t) := D_t \mathcal{I}_t^{1-\alpha}\varphi(t)$, $0 < \alpha < 1$, and K_β is a constant *effective* diffusivity with physical dimension $[K_\beta] = L^2T^{-\beta}$.

Remark 1.11. Equation (1.27) is also equivalent to (**Section 7.2.5**)

$${}_*\mathcal{D}_t^\beta u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t), \ \ 0 < \beta \leq 1, \tag{1.28}$$

where ${}_*\mathcal{D}_t^\beta \varphi(t) := \mathcal{I}_t^{1-\beta} D_t \varphi(t)$ is the so called *Caputo-Dzherbashyan* fractional derivative (**Section 7.1.1**, Def. 7.4). This formulation is very intuitive. In fact, it can be obtained from the standard diffusion equation simply substituting the first order time derivative with a fractional derivative of order $0 < \beta < 1$.

Also equation (1.27) can be written in a conservative form as a continuity equation:

$$\begin{cases} \dfrac{\partial}{\partial t}u(x,t) + \dfrac{\partial}{\partial x}F[u(x,t)] = 0, \\ F[u(x,t)] = -K_\beta \dfrac{\partial}{\partial x}\left[\mathcal{D}_t^{1-\beta}u(x,t)\right]. \end{cases} \tag{1.29}$$

This equation could be interpreted as generalized Fick's law where non-local (memory) effects are taken into account by the time fractional derivative of order $1-\beta$.

By construction, the subordinated process $X(t) = B(l_\beta(t))$, $t \geq 0$, provides automatically a stochastic model for (eq. 1.25). Then, $X(t)$ is clearly non-Markovian (Remark 1.8) and moreover, it has an interesting physical interpretation provided that the random time $l_\beta(t)$ is chosen in a suitable way. Let us consider a *continuous-time random walk* (CTRW), which is a random-walk subordinated to a renewal process[2].

Remark 1.12. CTRWs were introduced to study random walks on a lattice and are now used in physics to model a wide range of phenomena related to anomalous diffusion. The random walks increments represent the magnitude particle jumps and the renewal epochs represent the particle jump times.

More in details, let $\{J_k := J(\delta\tau k), \ \ k \in \mathbb{N}\}$, and $\delta\tau > 0$, be non-negative independent identically distributed (iid)[3] random variables which describe the waiting times between the consecutive jumps of

2) Renewal processes are generalization of Poisson processes, which are continuous-time Markov processes on the positive integers (usually starting at zero) which have independent identically exponentially distributed waiting times at each integer i before advancing (with probability 1) to the next integer $i+1$.

3) We call such a process a *purely random process*.

the particle. We set $\{T(n) = \sum_{k=1}^{n} J_k, \quad n \geq 0\}$, with $T(0) = 0$, the process describing the time of the n-th jump. Moreover, we suppose that also the particle jumps are given by a purely random process $\{Y_k := Y(\delta tk), \quad k \in \mathbb{N}\}$, which is assumed independent of $\{J_k\}$. Then, the position of the particle after n jumps is given by $\{S(n) = \sum_{k=1}^{n} Y_k, \quad n \geq 0\}$, with $S(0) = 0$. For any $t \in \mathbb{R}_+$, let:

$$N(t) = \max\{n \geq 0; \quad T(n) \leq t\}, \tag{1.30}$$

be the *renewal process* of the number of jumps up to time t. Therefore, the position of the particle at time $t \geq 0$ is given by:

$$X(t) = S(N_t) = \sum_{k=1}^{N_t} Y_k. \tag{1.31}$$

Now, suppose that J_k belongs to the domain of attraction of some stable law with index $0 < \beta < 1$. Namely, there exist $\{b(n) > 0, \quad n \geq 1\}$ such that

$$b_n \sum_{k=1}^{n} J_k \to \xi, \quad n \to \infty,$$

where ξ is a stable random variable (Rem. 3.17) of index β. Then, it is possible to show [13] that

$$\{c([\tau/\delta\tau])T([\tau/\delta\tau])\}_{\tau\geq 0} \longrightarrow \{L_\beta(\tau)\}_{\tau\geq 0}, \quad \delta\tau \to 0,$$

in the sense of finite dimensional distributions, where $[x]$ indicates the integer part of x. The process $L_\beta(\tau)$, $\tau \geq 0$, is a strictly β-stable process with stationary independent increments, that is a *strictly stable Lévy process*. Moreover, $L_\beta(t)$ is increasing and self-similar of order $H = 1/\beta$. The "continuous limit" of the renewal process $N(t)$ turns out to be the hitting time process

$$l_\beta(t) = \inf\{\tau; \; L_\beta(\tau) > t\}, \quad t \geq 0. \tag{1.32}$$

Remark 1.13. The process $l_\beta(t)$ is just the inverse process of the strictly β-stable process $L_\beta(\tau)$. In fact, $l_\beta(L_\beta(\tau)) = \tau$ almost surely.

Moreover, $l_\beta(t)$ is increasing, self-similar of order $H = \beta$, has not stationary increments and its marginal density function is given by the M-function (eq. 1.20) [14, 13, 22]. If now one requires that Y_k be a Gaussian white noise, then heuristically, at the scaling limit, the position of the particle at time $t \geq 0$ would be given by:

$$X(t) = B(l_\beta(t)), \quad t \geq 0, \tag{1.33}$$

where $B(t)$ is a Brownian motion independent of $l_\beta(t)$.

Remark 1.14. The above construction provides only a possible choice of the random time $l_\beta(t)$. For instance, with the suitable conventions [5], $l_\beta(t)$, $t \geq 0$, can be also viewed as the local time in zero at time t of a $d = 2(1-\beta)$ dimensional Bessel process [15].

We have seen how the theory of CTRWs provides a ready made physical interpretation of the subordinated process $B(l_\beta(t))$. Such a process is self-similar of order $H = \beta/2$, non-Markovian and has not stationary increments. Moreover, it reduces to a standard Brownian motion as $\beta \to 1$.

However, as in example (1.1), one would like a stochastic model of (eq. 1.27) which preserves the Brownian motion stationarity of the increments. In other words, one could have been looking for a self-similar with stationary increments process which serves as stochastic model for the time-fractional diffusion equation of order $0 < \beta \leq 1$. An example of such process is provided by the so called *grey Brownian motion* (**Chapter 8**). It is defined through the explicit construction of the underline probability

space (**Section 8.2.1**). The same construction could be used in order to define Brownian motion (**Section 8.1.3**), which is actually a particular case of grey Brownian motion, and is mainly based on the mathematical theory of generalized grey noises (**Section 8.2**), such as *white noise* (**Section 8.1**).

The grey Brownian motion is not a subordinated type process (**Section 8.2.1**, Rem. 8.23), and therefore the physical interpretation provided by the CTRW is lost. Indeed, we can show that it can be seen merely as a fractional Brownian motion with stochastic variance (**Section 8.3.1**, Prop. 8.14).

Remark 1.15. Then, again we have two different kinds of stochastic processes associated to the same anomalous diffusion model, given by the time-fractional diffusion (eq. 1.27); one is preferable because of its physical interpretation, the other is advisable for its mathematical properties and its connection to the long-range dependence theory of H-**sssi** processes (**Chapter 6**).

Example 1.4. We have already observed that because $0 < \beta < 1$, (eq. 1.25) depicts a system with a slow-anomalous diffusion behavior. In order to study the full range (slow and fast) of anomalous diffusion, we introduce a suitable *time-stretching* $g(t) = t^{\alpha/\beta}$, $0 < \beta \leq 1$ and $\alpha > 0$. Observe that if $f(x,t)$ is a solution of (eq. 1.25), then the function $f_{\alpha,\beta}(x,t) = f(x, t^{\alpha/\beta})$ is a solution of the *stretched time-fractional diffusion equation* (**Section 9.5.2**)

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)} \frac{\alpha}{\beta} \int_0^t s^{\frac{\alpha}{\beta}-1} \left(t^{\frac{\alpha}{\beta}} - s^{\frac{\alpha}{\beta}} \right)^{\beta-1} \frac{\partial^2}{\partial x^2} u(x,s) ds, \tag{1.34}$$

with the same initial condition. Then, the fundamental solution of (eq. 1.34) is $u(x,t) = \frac{1}{2}\mathcal{M}_{\beta/2}(|x|, t^{\alpha/\beta})$ and defines a self-similar probability density function of order $H = \alpha/2$, that is (eq. 9.64),

$$u(x,t) = \frac{t^{-\alpha/2}}{2} M_{\beta/2}(|x| t^{-\alpha/2}), \quad x \in \mathbb{R}. \tag{1.35}$$

We observe that, the diffusion is indeed slow when $\alpha < 1$, standard when $\alpha = 1$ and fast when $\alpha > 1$. Moreover, if $\beta = 1$, $u(x,t)$ is the "stretched" Gaussian density (eq. 1.7), namely,

$$u(x,t) = \frac{t^{-\alpha/2}}{2} M_{1/2}(|x|/t^{\alpha/2}) = \frac{1}{\sqrt{4\pi t^{\alpha}}} e^{-\frac{x^2}{4t^{\alpha}}}, \quad t > 0.$$

In the case $\alpha = \beta$, $0 < \beta < 1$, we recover the non-Gaussian probability density $u(x,t) = \frac{1}{2}\mathcal{M}_{\beta/2}(|x|, t)$. Finally, in the general case $0 < \beta < 1$ and $\alpha > 0$, we have a non-Gaussian "full-ranged" anomalous diffusion.

Let $X(t)$, $t \geq 0$, be a self-similar stochastic process with scaling parameter $H = \alpha/2$ and marginal probability density function defined by (eq. 1.35). At this point we know that there is a whole equivalence class of such a stochastic processes. For instance, one could immediately take $X(t) = B(l_\beta(t^{\alpha/\beta}))$, $t \geq 0$, with a suitable choice of the independent random time $l_\beta(t)$. However, one could require that the process $X(t)$ be also a stationary increments process, namely $X(t)$ is H-**sssi**, with $H = \alpha/2$.

Remark 1.16. The latter requirement forces the α-parameter to be in the range $0 < \alpha < 2$ (**Section 3.6**). Moreover, it automatically excludes subordinated type processes, because we have seen that they have in general not stationary increments.

Summarizing, we can ask that the stochastic process $X(t)$, $t \geq 0$, satisfies the following requirements: let $0 < \beta \leq 1$ and $0 < \alpha < 2$, then

i. $X(t)$ is self-similar with index $H = \alpha/2$.

ii. $X(t)$ has marginal density function $f_{\alpha,\beta}(x,t) = \dfrac{t^{-\alpha/2}}{2} M_{\beta/2}(|x|t^{-\alpha/2})$.

iii. $X(t)$ has stationary increments.

Let $B_{\alpha,\beta}(t)$, $t \geq 0$, be the so called *generalized grey Brownian motion* (ggBm) (**Section 8.2.2**, Def. 8.77). Then, $B_{\alpha,\beta}(t)$ satisfies all the above properties (Prop. 8.10). It represents a generalization of Brownian motion ($\alpha = \beta = 1$), fractional Brownian motion ($\beta = 1$, $0 < \alpha < 2$) and grey Brownian motion ($\alpha = \beta$) as well (see Rem. 8.21 and Figure 8.1).

The ggBm is defined through the explicit construction of the underline probability space (**Section 8.2**). However, it can be defined in an unspecified probability space by giving explicitly all the finite dimensional probability density functions $f_{\alpha,\beta}(x_1,t_1,x_2,t_2,\ldots,x_n,t_n)$, $n \in \mathbb{N}$. The latter are constructed only by making use of the M-function (eq. 1.20) and the autocovariance function $\gamma_{\alpha,\beta}(t_i,t_j) = \frac{1}{\Gamma(1+\beta)}(t_i^\alpha + t_j^\alpha - |t_i - t_j|^\alpha)$ (**Section 8.3**, Prop. 8.13):

$$f_{\alpha,\beta}(x_1,x_2,\ldots,x_n;\gamma_{\alpha,\beta}) = \frac{(2\pi)^{-\frac{n-1}{2}}}{\sqrt{2\Gamma(1+\beta)^n \det \gamma_{\alpha,\beta}}} \int_0^\infty \frac{1}{\tau^{n/2}} M_{1/2}\left(\frac{\xi}{\tau^{1/2}}\right) M_\beta(\tau) d\tau, \tag{1.36}$$

with:

$$\xi = \left(2\Gamma(1+\beta)^{-1} \sum_{i,j=1}^{n} x_i {\gamma_{\alpha,\beta}}^{-1}(t_i,t_j) x_j\right)^{1/2}.$$

Remark 1.17. It is interesting to observe (see Remark 8.24) that the generalized grey Brownian motion turns out to be a direct generalization of a Gaussian process. Indeed, it reduces to fractional Brownian motion when $\beta = 1$, which is a Gaussian process, and, fixed β, it can be defined only by giving its covariance structure. In other words, $B_{\alpha,\beta}(t)$ provides an example of a stochastic process characterized only by the first and second moments, which is a property of Gaussian processes.

Looking at the remark above, it is not surprising that the generalized grey Brownian motion turns out to be merely a fractional Brownian motion with stochastic variance (**Section 8.3.1**, Prop. 8.14), that is $B_{\alpha,\beta}(t) = \Lambda_\beta X_\alpha(t)$, $t \geq 0$, where Λ_β is a suitable independent random variable (see eq. 8.87). Such a process is not ergodic[4], as follows by the multiplication with the random variable Λ_β. Indeed, it is impossible with a single realization of the system $B_{\alpha,\beta}(t,\omega)$, $\omega \in \Omega$, to distinguish a ggBm from a fBm with variance $2\Lambda_\beta^2(\omega)t^\alpha$, where $\Lambda_\beta(\omega)$ indicates a single realization of the random variable Λ_β.

Remark 1.18. This representation, highlights the H-**sssi** nature of the ggBm. Moreover, a number of question, in particularly those related to the distribution properties of $B_{\alpha,\beta}(t)$, can be reduced to questions concerning the fBm $X_\alpha(t)$, which are easier since $X_\alpha(t)$ is a Gaussian process.

We have seen that H-**sssi** processes, even if they may not have a direct physical interpretation, are sometimes preferable in the modeling of anomalous diffusion, mainly because they are connected to long-range dependence. For instance, the fBm in Example (1.1), the gBm in Example (1.2) and the ggBm in Example (1.4).

We have already remarked that the LRD (**Section 6**) is a quite common property of time-series related to many different fields of studies. Heuristically, a data series has long-range dependence if the observations remain, in a certain sense, correlated even if they are very far apart in time. Such a phenomenon is

4) Here we mean that the increment process, which is indeed stationary, is not ergodic. Namely, time averages do not necessary coincide with population averages.

said to possess *long-memory*, in the sense that past events continue to affect future events for large time intervals. More precisely, a stationary stochastic process exhibits long-range dependence if the autocovariance function tends to zero with a power-law and in such a way that it does not result integrable (**Section 6**, Def. 6.1).

The more popular mathematical models used to treat the long-range dependence are the *fractional Gaussian noise*, that is the increment process of fractional Brownian motion (**Section 3.7.3**, Def. 3.15), and FARIMA models (**Section 3.8.2**, Def. 3.18). These stationary processes are both Gaussian and can be characterized by the *Hurst parameter* $0 < H < 1$ (**Section 3.8**, Rem. 6.2). Then, the LRD occurs if $1/2 < H < 1$ (**Section 6**, Prop. 6.1 and Prop. 6.2 with $d = H - 1/2$). The greater the value of H, the larger the long-range dependence between the data.

Time series manifesting LRD are often observed in nature. After the pioneering works of Mandelbrot (1968), the study of long-range dependence has aroused a growing interests, followed by a fast development. Because of the variety of applications, the corresponding literature is split in a large numbers of journals concerning many different fields: *physics, mathematics, meteorology, climatology, agronomy, astronomy, chemistry, economy, engineering, hydrology, geophysics, telecommunication sciences*, and many others (**Section 6.4.1**).

Remark 1.19. At a first analysis, long-range dependence can be recognized thanks to certain qualitative characteristic of the graph, such as persistence and local cycles and trends (**Section 6.4**). Then, it can be measured studying certain quantitative features that indeed can be used in order to obtain estimations of the parameter H, and thus obtaning a measure of the LRD intensity (**Section 6.5**).

Suppose we have observed a certain time series $\{x_t\}_{t=1,\dots,n}$, which, for instance, exhibits long-memory. The observations are generated by a certain process X, termed *generator process*, unknown a priori. Assuming stationarity, one tries to model the series behavior through a certain class of parametric models, or in other words through a parametric class of stochastic processes $\{Y_\theta\}_{\theta\in\Theta}$ which are in general lacking of a direct physical interpretation.

More precisely, suppose that the model depend on a certain parameter vector θ. One tries to obtain an "optimal" estimate of θ, starting from the observed data, with the aim to forecast future observations or to obtain a confidence interval for the mean μ (**Section 6.5.2**). At this purpose, it is not necessary to establish a direct link between the estimated parameters $\widehat{\theta}$ and the series generating process X. Although useful for statistical applications, this approach is not satisfactory from a scientific point of view. In fact, a theoretical process, created in a certain statistical context, but without a direct physical connection with the specific problem analyzed, will not permit, hardly ever, to give a physical explanation for the behavior of the observed dynamical system.

One should then try to understand, from time to time, which is the physical cause of a certain observed behavior. In the case of long-range dependence, it can look strange that events took place far away in the past could have some influence on present and future events. Many physical explanation could be given (**Section 6.3.1**). For instance, long-memory in a time series can be obtained thorough the aggregation of a large number of Markov processes (**Section 6.3.1.1**). On the other hand, long-range dependence characteristics are often encountered in the study of phase transition in termodynamics (**Section 6.3.1.2**). Moreover, physical models of long-range dependence can be obtained using stochastic partial differential equations (**Section 6.3.1.3**).

However, if a realistic physical model is not available, or if it can not be given with the necessary accuracy, then a simple stationary model is often sufficient at least for practical and immediate purposes,

for instance in the case that one wants to estimate certain characteristic parameters, such as the Hurst parameter H (**Section 6.5**).

In this introduction we have tried to provide only a brief *"hors d'oeuvre"* of some of the topics which have been faced in this piece of work. Moreover, any time we have encountered a key notion, we have given the suitable cross references, which were reminding the interested reader to the devoted work-sections.

Within any chapter, we tried to give a rough list of references which can be used to go into more depth the corresponding argument. Chapters (8) and (9), as part of Chapter (7), are originals and contain material, extracted by many paper of the author, which have been published, are in press, or which are going to be published soon. All the code used all around the work to make statistical analysis or just to simulate and visualize stochastic processes trajectories, have been developed and implemented by the author in *MatLab 7*. However, in no case line-cods are reported.

The rest of the work is organized as follows:

- In **Chapter 2** we give a basic introduction to stochastic analysis. We start with *measure theory*, *measurable functions*, *integration* and *absolute continuity and densities*. Then, we introduce the notion of *random variables*, which are just measurable functions defined on probability spaces (Def. 2.14), and their *laws*. After having treated the notion of convergence for random variables sequences, we end the chapter introducing the *conditional expectation* and its properties.
- In **Chapter 3** we introduce the notion and the basic theory of *stochastic processes*. We give the notion of *filtration* and *martingales*. Then, we define *stationary processes* and *self-similarity*. After introduced the mathematical definition of Brownian motion and its properties, we focus on *Markov processes* and *stochastic diffusion processes*. Then, we study H-**sssi** processes and in particular the fractional Brownian motion and the fractional Gaussian noise. Finally, we study stationary *linear sequences* such as FARIMA processes.
- **Chapter 4** provides an introduction to *stochastic integration*. In particular we introduce the *Ito integral* and the so called *Ito calculus*. Then, we study *stochastic differential equations*, providing some examples. In this chapter, we also introduce the idea of *Ito diffusions* which are strong solution of stochastic differential equations and particular cases of stochastic diffusions.
- In **Chapter 5** we use the theory of stochastic integral to develop useful integral representations of some non-Markovian processes. First, we study integral representation of fractional Brownian motion in the time domain and in the frequencies domain. Then, we use them to find out the spectral density function of the fractional Gaussian noise. In the same way, we use integral representation of linear processes in order to obtain the spectral density function of FARIMA processes.
- In **Chapter 6** we study the theory of long-range dependence both from a probabilistic and from a statistical point of view. After having given the mathematical definitions, we briefly discuss some physical models of long-range dependence. Then, we go more depth into the statistic analysis of long-memory time series. We give several examples of real data and we study several different methods of LRD estimation. First, we study *heuristic methods*, such as the *sample mean variance*, the *correlogram* and the *periodogram* methods. Finally, we introduce *maximum likelihood methods*, providing example of estimations, and briefly discussing robustness against certain deviation from the original parametric model.
- In **Chapter 7** we introduce the *fractional calculus* which is the theory of fractional integrals and derivatives. This theory is deeply used to develope mathematical models for anomalous diffu-

sion. As example of applications, we study *fractional relaxation* and in particular *fractional relaxation of distributed order*, which is a generalization of fractional relaxation obtained using the notion of *distributed order fractional derivatives*, that is linear superpositions of fractional derivatives of different orders. Then we study the time-fractional (forward) drift equation (eq. 1.19). We obtain the fundamental solution (eq. 1.20) in terms of *Fox H-functions*, and we study many of its properties. Then, we study the time-fractional diffusion equation (eq. 1.25) and in particular its generalization obtained using distributed order fractional derivatives. Finally, we see how fractional Brownian motion can be represented as a fractional integral of a Gaussian white noise, and this actually gives reason of its name.

- In **Chapter 8** we introduce the parametric class of generalized grey Brownian motion $\{B_{\alpha,\beta}(t)\}$, with $0 < \alpha < 2$, $0 < \beta \leq 1$. This class includes non-Markovian stochastic models either for slow and fast-anomalous diffusion. After having presented and motivated the mathematical construction (*Locally-convex spaces*, *Nuclear spaces*, *characteristic functionals*, *Minlos' theorem*, *canonical noises*), in particular the development of the *grey noise space*, we show that this class is made up of *H*-sssi processes and contain either Gaussian and non-Gaussian processes, like fractional Brownian motion and grey Brownian motion. The time evolution of the $B_{\alpha,\beta}(t)$ marginal density function is described by the partial integro-differential equations (eq. 1.34), which includes the time-fractional diffusion equation. We study different characterizations of the ggBm notwithstanding the underline probability space. Then, we show that the ggBm could be represented by a process:

$$B_{\alpha,\beta}(t) = \sqrt{L_\beta} X_\alpha(t), \quad t \geq 0,$$

where L_β is suitable chosen random variable independent of the fBm $X_\alpha(t)$. This representation is useful for path simulations. Clearly, in order to simulate ggBm trajectories, we will need a method to generate the random variable L_β. So that, the end of the chapter is devoted in the development of such a method. We make use of finite difference approximation of non-local partial integro differential equation of fractional type, in order to build a suitable random walk scheme. The construction follows two steps: using the *Grünwald-Letnikov discretization* of Caputo-Dzherbashyan derivative, and then interpreting the corresponding finite difference scheme as a random walk scheme. Following this idea, we will be able to simulate generalized grey Brownian motion trajectories both in the short and long-range dependence domain.

- In **Chapter 9** we study (eq. 1.26) for different choices of the memory-kernel $K(t)$. Actually, we consider non-local, fractional and stretched modifications of the diffusion equation. These modified equations are indeed a generalization of time fractional equations, and will be called *Non-Markovian diffusion equations*, because, while they originate from a diffusion equation, the corresponding process will be typically non-Markovian. We also consider what happens if instead of the standard diffusion equation we start from a more general Fokker-Planck equation,

$$\partial_t u(x,t) = \mathcal{P}_x u(x,t),$$

where $\mathcal{P}_x$ is a linear operator independent of t acting on the variable $x \in \mathbb{R}$ associated to a certain time homogeneous one-dimensional stochastic diffusion. In other words, we consider the non-Markovian diffusion equation:

$$u(x,t) = u_0(x) + \int_0^t g'(s) K(g(t) - g(s)) \mathcal{P}_x u(x,s) ds. \tag{1.37}$$

After a brief historical remark, we study the suitability condition on the kernel $K(t)$ giving many examples. Then, we study (eq. 1.37) finding the general form of the solutions and we give many

different stochastic models for the studied equations. We end the chapter studying many different examples: we choose different suitable kernels and different Fokker-Planck operators. In all the cases, we provide exact solutions and we study the corresponding stochastic models making trajectory simulations.

- **Chapter 10** is devoted to concluding remarks.

Bibliography

[1] E. Barkai, CTRW pathways to the fractional diffusion equation, *Chem. Phys.* **284** (2002) 13–27.

[2] E. Barkai, R. Metzler and J. Klafter, From continuous time random walk to fractional Fokker-Planck equation, *Phys. Rev. E* **61** (2000) 132-138.

[3] S. Bochner, *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley, 1955.

[4] S. Bochner, Subordination of non-Gaussian stochastic processes, *Proc. Nat. Acad. Sciences, USA* **48** (1962) 19–22.

[5] C. Donati-Martin, B. Roynette, P. Vallois and M. Yor, On constants related to the choice of the local time at 0, and the corresponding Itô measure for Bessel processes with dimension $d = 2(1-\alpha)$; $0 < \alpha < 1$, *Studia Scientiarum Mathematicarum Hungarica*, to appear (2007). [E-print `http://hal.archives-ouvertes.fr/hal-00141513/en/`]

[6] R. Gorenflo and F. Mainardi, Fractional diffusion processes: Probability distributions and continuous time random walk, in: G. Rangarajan and M. Ding (Editors), *Processes with Long Range Correlations*, Springer Verlag, Berlin, 2003, pp. 148-161. [Lecture Notes in Physics, No 621]

[7] V. M. Kenkre, R. S. Knox, Generalized-master-equation theory of excitation transfer, *Physical Review B* **9** (1974) 5279-5290.

[8] T. Kolsrud, On a class of probabilistic integrodifferential equations, in: S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Editors), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, Vol I, Cambridge University Press, Cambridge, 1990, pp. 168-172.

[9] F. Mainardi, Yu. Luchko and G. Pagnini, The fundamental solution of the space - time fractional diffusion equation, *Fractional Calculus and Applied Analysis* **4** (2001) 153-192. [E-print `http://arxiv.org/abs/cond-mat/0702419`]

[10] F. Mainardi, A. Mura, R. Gorenflo and M. Stojanovic, Two forms of fractional relaxation of distributed order, *Journal of Vibration and Control*, **13**, No. 9-10, pp.1249-1268 (2007).

[11] F. Mainardi, A. Mura, G. Pagnini and R. Gorenflo, *Sub-diffusion equations of fractional order and their fundamental solution*, in K. Tas, J. A. Tenreiro-Machado and D. Baleanu (Editors): Mathematical Methods in Engineering, Springer-Verlag, Dordrecht 2007, pp. 23-55. ISBN 978-1-4020-2 (HB).

[12] F. Mainardi, A. Mura, G. Pagnini and R. Gorenflo, Time-fractional diffusion of distributed order, *Journal of Vibration and Control*, to appear (2007).

[13] M. Meersschaert, H. Scheffler, Limit theorems for continuous-time random walks with infinite mean waiting-times, *J.Apll.Prob* **41**, pp. 623-638 (2004).

[14] R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, *Phys. Reports* **339** (2000) 1-77.

[15] S.A. Molchanov and E. Ostrovskii, Symmetric stable processes as traces of degenerate diffusion processes, *Theory of Probability and its Applications* **14** (1969) 128-131.

[16] A. Mura, M.S. Taqqu and F. Mainardi, Non-Markovian diffusion equations and processes: analysis and simulations. *submitted: in revision*. [E-print `http://arxiv.org/abs/0712.0240`]

[17] A. Mura and F. Mainardi, A class of self-similar stochastic processes with stationary increments to model anomalous diffusion in physics, *submitted*. [E-print `http://arxiv.org/abs/0711.0665`]

[18] A. Mura and G. Pagnini, Characterization and simulation of a class of stochastic processes related to anomalous diffusion, *submitted* [E-print `http://arxiv.org/abs/0801.4879`].

[19] W.R. Schneider, Grey noise, in: S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Editors), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, Vol 1, Cambridge Univ. Press, Cambridge (1990), pp. 261-282.

[20] W.R. Schneider, Fractional diffusion, in: R. Lima, L. Streit and D. Vilela Mendes (Editors), *Dynamics and Stochastic Processes, Theory and Applications*, Springer Verlag, Heidelberg, 1990, pp. 276-286. [Lecture Notes in Physics No 355]

[21] I.M. Sokolov, Solutions of a class of non-Markovian Fokker-Planck equations, *Phys. Rev. E* **66** (2002) 041101/1-5.

[22] A. A. Stanislavsky, Subordinated Brownian Motion and its Fractional Fokker-Planck Equation, *Physica Scripta* **67** (2003) 265-268.

[23] M.S. Taqqu, Fractional Brownian motion and long-range dependence, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and Applications of Long-range Dependence*, Birkäuser, Boston - Basel - Berlin, 2003, pp. 5-38.

[24] G.H. Weiss, *Aspects and Applications of Random Walks*, North-Holland, Amsterdam, 1994.

2 Basics in probability

2.1 Measure Theory

The notion of measure is a straightforward generalization of the more familiar notions of length, area and volume to more abstract settings.

Let E be a non empty set. Heuristically, a measure μ on E associates to certain subsets $A \subset E$, called *measurable sets*, a positive (possibly infinite) number $\mu(A) \in [0, \infty)$, called the measure of A. By analogy with the notions of area or volume, it is natural to say that the empty set $\emptyset$ has zero measure. Also, if A and B are disjoint measurable sets, $A \cup B$ should also be measurable and its measure is naturally defined to be $\mu(A \cup B) = \mu(A) + \mu(B)$. This is the additivity property.

In order to consider limits, it is useful to extend this property to infinite sequences: if $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of disjoint measurable subsets, then:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{2.1}$$

This countable additivity property is known as σ-additivity.

Remark 2.1. Note that we have not excluded that $\mu(A) = \infty$ for some subset A, in analogy with volume. For instance, the volume of a half space is infinite.

In particular $\mu(E)$ may be finite or infinite. We suppose that $\mu(E) < \infty$. For any measurable set $A \subseteq E$, its complements A^c verifies $A \cup A^c = E$. Thus, the additivity property can be used to define the measure of A^c:

$$\mu(A^c) = \mu(E) - \mu(A). \tag{2.2}$$

Therefore, it is natural to require that for any measurable set A its complement A^c is also measurable.

We lead to the following definition:

Def. 2.1 (σ-algebra)**.** let E be a non empty set and $\mathcal{P}(E)$ be its powerset. Let $\mathcal{E} \subseteq \mathcal{P}(E)$ be a collection of subsets of E such that:

1. contains the empty set: $\emptyset \in \mathcal{E}$,
2. is "stable" under unions; i.e. for any numerable collection $\{A_n\}_{n\geq 1}$ of disjoint elements of $\mathcal{E}$, then:

$$\bigcup_{n\geq 1} A_n \in \mathcal{E}, \tag{2.3}$$

3. contains the complementary of every element; i.e. for any $A \in \mathcal{E}$ then: $A^c \in \mathcal{E}$.

Such a collection $\mathcal{E}$ is called a σ-algebra (σ-field).

Notation 2.1. We will usually denote σ-algebras by curly letters like $\mathcal{E}, \mathcal{B}, \mathcal{F}$. A *measurable* set will then be an element of the σ-algebra.

Prop. 2.1. *Let $A \subset \mathcal{P}(E)$. There exists a unique σ-algebra, denoted $\sigma(A)$, such that for any σ-algebra $\mathcal{F}$ which contains A, then $\sigma(A) \subset \mathcal{F}$.*

In fact, it is clear that if one defines:

$$\sigma(A) = \bigcap_{\substack{\mathcal{F} \supseteq A \\ \mathcal{F}\ \sigma-algebra}} \mathcal{F},$$

($\sigma(A)$ is the smallest σ-algebra which contain A) then $\sigma(A)$ is still a σ-algebra (the intersesction of σ-algebras is still a σ-algebra). Thus, $\sigma(A)$ is termed σ-algebra *generated* by A. In the same way we say that the class A is a *generator* of $\sigma(A)$.

Def. 2.2 (Bases)**.** Let $A \subseteq \mathcal{P}(E)$. If for any finite sequence $\{A_i, i = 1, 2, \ldots, n\}$ of elements of A one has

$$\bigcap_{i=1}^{n} A_i \subseteq A,$$

then A is said a *base* for $\sigma(A)$. The class A is sometimes termed a π-system.

Therefore, a generator closed under finite intersection is said a *base*. A σ-algebra that has a countable generator is said *separable*.

Example 2.1 (Borel field)**.** Suppose that E has a topology. The σ-algebra generated by the open subsets of E is called Borel σ-algebra and is denoted by $\mathcal{B}(E)$ or simply $\mathcal{B}$. An element $B \subset \mathcal{B}$ is called a Borel set.

Any open or closed set is a Borel set. Thus, defining measures on $\mathcal{B}$ ensures that all open and closed sets are measurable. The following definition will be useful later:

Def. 2.3 (Dynkin Class)**.** Let E be a non-empty set. A class $\mathcal{D} \subseteq \mathcal{P}(E)$ is said a Dynkin class if:

1. $E \in \mathcal{D}$;
2. for any increasing sequence $\{A_n\}_{n\in\mathbb{N}}$ on $\mathcal{D}$, then $A = \bigcup_n A_n \in \mathcal{D}$;
3. for any couple (A, B) on $\mathcal{D}$ with $B \subset A$, then $A \setminus B \in \mathcal{D}$.

Therefore, a Dynkin class is closed under complementation (properties 1. and 3.). The following result holds:

Theorem 2.1 (Dynkin lemma)**.** *Let $\mathcal{D}$ be a Dynkin class. If A is a class closed under intersection with $A \subseteq \mathcal{D}$, then $\sigma(A) \subseteq \mathcal{D}$.*

Having defined σ-algebras, we are now ready to properly define measures. Let $\mathcal{E}$ be a σ-algebra over a non-empty set E. The couple $(E, \mathcal{E}) = E_{\mathcal{E}}$ is called *measurable space*.

Def. 2.4 (Measures)**.** A (positive) measure on $(E, \mathcal{E})$ is a function $\mu : \mathcal{E} \longrightarrow [0, \infty)$ with the following properties:

1. $\mu(\emptyset)$=0.
2. For any disjoint sequence $\{A_n\}_{n\in\mathbb{N}}$ of $\mathcal{E}$ elements:

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n). \tag{2.4}$$

Any element A $\in \mathcal{E}$ is a measurable set with measure $\mu(A)$.

Prop. 2.2. *Consider a sequence of events $A_n \uparrow A$; that is, A is the reunion event $\bigcup_n A_n$of the increasing sequence $\{A_n\}_{n\geq 1}$. Then,*

$$P(A_n) \to P(A).$$

The same result is true if $A_n \downarrow A$. We have also the following relationships:

- $P(\liminf_n A_n) \leq \liminf_n P(A_n)$,
- $P(\limsup_n A_n) \geq \limsup_n P(A_n)$,

where:

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \left(\bigcap_{m=n}^{\infty} A_m \right), \quad \limsup_n A_n = \bigcap_{n=1}^{\infty} \left(\bigcup_{m=n}^{\infty} A_m \right) \tag{2.5}$$

Remark 2.2 (Signed measures)**.** In Definition 2.4, we required that $\mu(A)$ be positive for any measurable set A. However, one can also consider non-positive measures. For instance, if μ_+ and μ_- are two positive measures, then the measure $\mu = \mu_+ - \mu_-$ is called a *signed measure*.

Def. 2.5. Let $(E, \mathcal{E}, \mu)$ be a measurable space with a measure μ. A property on E is said to hold a.e. (almost everywhere) on a measurable set $A \in \mathcal{E}$, if it is verified for all the points of $A \backslash N$, where $N \in \mathcal{E}$ is a zero-measure set; i.e. $\mu(N) = 0$.

Example 2.2 (Lebesgue measure)**.** Let $E = \mathbb{R}^d$. The Lebesgue measure λ is defined on the measurable space $(\mathbb{R}^d, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra, and corresponds to the d-dimensional notion of volume:

$$\lambda(A) = \int_A dx, \quad A \in \mathcal{B}. \tag{2.6}$$

Example 2.3 (Densities)**.** To any positive continuous function $\rho : \mathbb{R}^d \to \mathbb{R}_+$ one can always associate a positive measure defined on $(\mathbb{R}^d, \mathcal{B})$ such that for any Borel set $A \in \mathcal{B}$:

$$\mu(A) = \int_A \rho(x)dx = \int 1_A \rho(x)dx. \tag{2.7}$$

The function ρ is called the density of the measure μ with respect to the Lebesgue measure λ. More generally, the above equation defines a positive measure for every positive measurable function ρ (see below).

Example 2.4 (Dirac measures)**.** Dirac measures are other important examples of measures. The Dirac measure δ_x associated to a point $x \in E$ is defined as follows:

$$\begin{cases} \delta_x(A) = 1, & x \in A, \\ \delta_x(A) = 0, & x \notin A. \end{cases}$$

More generally one can consider a sum of such Dirac measures. Given a countable set of points $X = \{x_i,\ i = 0, 1, 2, ...\} \subset E$ the counting measure $\mu_X = \sum_i \delta_{x_i}$ is defined in the following way: for any $A \in \mathcal{E}$:

$$\mu(A) = \sharp\{i,\ \ x_i \in A\} = \sum_{i\geq 1} 1_A(x_i), \tag{2.8}$$

where 1_A is the *indicator* function of the set A.

Remark 2.3. A measurable set may have zero measure without being empty. Going back to the analogy with length and volume, we can note that the area of a line segment is zero. The "length" of a point is also defined to be zero.

The existence of non-trivial sets of zero measure is at the origin of many subtleties in measure theory. If A is a measurable set with $\mu(A) = 0$, it is then natural to set $\mu(B) = 0$ for any $B \subseteq A$. Such sets (subsets of sets of measure zero) are called *null sets*. If all null sets are not already included in $\mathcal{E}$, one can always include them by adding all null sets to $\mathcal{E}$. The new σ-algebra is then said to be *complete*.

A measure μ, defined on a measurable space $(E, \mathcal{E})$, is said to be *integer valued* if for any measurable set $A \in \mathcal{E}$, $\mu(A)$ is a (positive) integer. An example of an integer valued measure is a Dirac measure. More generally, any counting measure is an integer valued measure.

A measure μ is said to be finite if $\mu(E) < +\infty$ (which entails that $\mu(A)$ is finite for any measurable set A). The quantity $\mu(E)$ is usually called the (total) mass of μ. For example, a Dirac measure is a finite measure with mass one. The counting measure μ_X (eq. 2.8), associated to the set $X = \{x_1, x_2, ...\}$, is finite and its mass $\mu_X(E)$ is simply the number of elements in X.

Not all measure we will encounter, will be finite measures. A well-known examples is the Lebesgue measure on $\mathbb{R}$. In this case the total mass is infinite. A more flexible notion is that of a Radon measure:

Def. 2.6 (Radon measure)**.** Let $E \in \mathbb{R}^d$. A Radon measure on $(E, \mathcal{B})$ is a measure μ such that for every compact measurable set B $\in \mathcal{B}$, $\mu(B) < \infty$.

For example the Lebesgue measure on $\mathbb{R}$ is a Radon measure: the length of any bounded interval is finite. Dirac measures and any finite linear combination of Dirac measures are also examples of Radon measures. More generally, a measure μ on $E \subset \mathbb{R}^d$ is called σ-finite if

$$E = \bigcup_{i=1}^{\infty} E_i \quad \text{with } \mu(E_i) < \infty \tag{2.9}$$

This condition holds for the examples given above.

The Lebesgue measure and Dirac measures differ in a fundamental way: while Dirac measures are concentrated on a finite number of points, the Lebesgue measure assigns zero measure to any finite set. A measure μ_0, defined on a measurable space $(E, \mathcal{E})$, is said *diffuse* if the whole space E is the reunion of zero-measure sets. When the one-point sets are measurable, the measure is diffuse if and only if it gives zero mass to any point and is also said *atomless*. The Lebesgue measure is an example of a diffuse measure. Measures with a continuous density with respect to the Lebesgue measure on $\mathbb{R}^d$ (Example 2.3) are other examples of diffuse measures.

We have te following important result:

Theorem 2.2 (Dynkin theorem)**.** *Two measures μ_1 and μ_2, defined on a measurable space $(E, \mathcal{E})$, are equal if they coincide on a base of $\mathcal{E}$.*

Proof: let $\mathcal{D}$ the class of the sets where the two measures coincide. Therefore, it is easy to see that $\mathcal{D}$ is a Dynkin class and the result follows from the Dynkin lemma. □

Remark 2.4 (Uniqueness of the Lebesgue measure)**.** The Lebesgue measure is unique. This result is just an application of the previous result. Indeed, consider the unit interval $\Omega = [0, 1]$ with the Lebesgue measure on Borel sets. Let μ be another measure on Ω satisfying $\mu[(a, b)] = b - a$, for any $0 \leq a \leq b \leq 1$. Let $I = \{(a, b), [a, b), (a, b], [a, b], \ \ 0 \leq a \leq b \leq 1\}$. I is closed under finite intersections and generates $\mathcal{B}$. Thus, the measure μ coincides with the Lebesgue measure.

2.1.1 Measurable functions

Let us consider two measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$. Let $f : E \to F$. In most applications we will be interested in evaluating a measure on a set of the form:

$$\{x \in E, \ \ f(x) \in A\} := \{f \in A\}, \ \ A \in \mathcal{F}. \tag{2.10}$$

For a given function f, there is no reason for this set to be measurable. This motivates the following definition:

Def. 2.7 (Measurable function)**.** A function $f : E \to F$ is called measurable if for any measurable set $A \in \mathcal{F}$, the set

$$f^{-1}(A) = \{x \in E, f(x) \in A\} \tag{2.11}$$

is a measurable subset of E.

In other words, a function f, between two measurable spaces, is measurable if for any measurable set $A \in \mathcal{F}$ we have that the counter-imagine $f^{-1}(A) \in \mathcal{E}$. The following results are straightforward:

1. the composition of two measurable function is measurable;
2. a constant function is measurable.

Prop. 2.3 (Generator criterium)**.** *In order for a function f to be measurable with respect to a σ-algebra $\mathcal{E}$, it is enough that the counter-imagines through f of a generator of $\mathcal{E}$ be measurable.*

Proof: let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ with $\mathcal{F} = \sigma(A)$ and $f : E \to F$ be given. If one indicates with $\mathcal{B} \subseteq \mathcal{P}(F)$ the class of measurable sets having counter-imagine measurable, then $\mathcal{B} \subseteq \mathcal{F}$. However, by ipothesis $A \subset \mathcal{B}$, than, by definition of $\sigma(A)$, we must have $\mathcal{B} = \mathcal{F}$. □

Remark 2.5. As noted above we will often consider measures on sets which already have a metric or topological structure, such as $\mathbb{R}^d$, equipped with the Euclidean metric. On such spaces the notion of continuity for functions is well defined and one can then ask whether there is a relation between continuity and measurability for a function $f : E \longrightarrow F$. In general, there is no relation between these notions. However it is desirable that the notion of measurability be defined such that all continuous functions be measurable. Using the generator criterium, this is automatically true if the Borel σ-algebra is chosen. This is because f is continuous if and only if $f^{-1}(A)$ is open for every open set A. Hence in the following, whenever the notion of continuous functions makes sense, all continuous functions will be measurable. The opposite is not true. That is there exists functions defined on topological spaces which are measurable but not continuous. For instance, the indicator function of the rational set on the real line is measurable but not continuous.

Def. 2.8. Let $f : E_{\mathcal{E}} \to F_{\mathcal{F}}$ an application between measurable spaces. We define $\sigma(f)$ as the σ-algebra generated by the counter-images of measurable sets:

$$\sigma(f) = \sigma\left(\{f^{-1}(A), \ \ A \in \mathcal{F}\}\right).$$

$\sigma(f)$ is the coarsest σ-algebra which makes f measurable and is said *σ-algebra generated by f*.

Def. 2.9 (Product space)**.** Let $\mathcal{G}$ be a family of measurable spaces

$$\mathcal{G} = \{(\Omega_1, \mathcal{F}_1), \ldots, (\Omega_j, \mathcal{F}_j)\};$$

we define *product space* the measurable space $(\Pi = \Omega_1 \times \cdots \times \Omega_j, \pi)$, where

$$\pi := \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_j,$$

is the σ-algebra generated by the projections, named *product σ-algebra*.

Notation 2.2 (Cylinders and rectangles)**.** Therefore, π is generated by the class of *cylinder*, which are defined as the counter-images of measurable sets through projections. A base for the product σ-algebra is given by finite intersection of cylinders, which are called *rectangles*.

We have the following

Theorem 2.3. *An application f defined on a measurable space $(\Omega, \mathcal{F})$, which assumes values on a product space (Π, π), is measurable if and only if its composition with each of the projections is measurable.*

Proof: the necessary condition is obvious, because the composition of measurable functions is measurable. On the other hand, we have $f : \Omega_{\mathcal{F}} \to \Pi_{\pi}$ and, because of the generator criterium, we have just to verify that the counter-images of cylinders, which generate π, through f are measurable. Let p_i be the projection of index i, if J is the cylinder $\{p_i \in H\}$, the set $\{f \in J\} = \{p_i \circ f \in H\}$ is measurable by ipothesis. □

Consider real measurable functions defined on a measurable space. Than, we have:

Prop. 2.4. *The set of real valued measurable functions form a real vector space closed under absolute value and the operation of maximum, minimum and multiplication of two functions.*

Proof: let $(E, \mathcal{E})$ be a measurable space. Let f, g be real valued $\mathcal{E}$-measurable functions. Than, for any real numbers a and b, the application $h = af + bg$ is measurable; indeed, $h = F \circ (f, g)$, where $F : \mathbb{R}^2 \to \mathbb{R}$, with $F(x, y) = ax + by$ is measurable. In the same way, also fg is measurable as composition with $G(x, y) = xy$. Moreover, $|f|$ is the composition of f with $x \to |x|$. Finally, $f \vee g = \frac{1}{2}(f + g + |f - g|)$; while, $f \wedge g = -(-f \vee -g)$. □

Def. 2.10 (Imagine measure)**.** Consider a measurable space $(E, \mathcal{E})$ with a measure μ and a measurable space $(F, \mathcal{F})$. Let $f : E \to F$ be a measurable function. Then, f induces a measure μ_f on F said *imagine measure* of μ through f. This measure is defined such that for any $A \in \mathcal{F}$:

$$\mu_f(A) = \mu\{f \in A\}. \tag{2.12}$$

2.1.2 Integration of real measurable functions

We want to define integration of real measurable functions $f : E_{\mathcal{E}} \to \mathbb{R}$.

Def. 2.11 (Simple functions)**.** Given a measurable space $(E, \mathcal{E})$, simple examples of real valued measurable functions are functions of the form:

$$f = \sum_{j=1}^{n} c_j 1_{A_j}, \tag{2.13}$$

where $\{A_j\}_{j=1,2,\dots,n}$ are measurable sets and $c_j \in \mathbb{R}$ for any j. These functions are sometimes called *simple functions*.

The integral of such a simple function with respect to a measure μ defined on $(E, \mathcal{E})$ is naturally defined as

$$\mu(f) = \sum_{j=1}^{n} c_j \mu(A_j). \tag{2.14}$$

Having defined integrals for simple functions, we extend it to any positive measurable function $f : E \to \mathbb{R}$ by setting

$$\mu(f) = \sup\{\mu(\varphi),\ \varphi \text{ simple function},\ \varphi \leq f\}. \tag{2.15}$$

The integral $\mu(f)$ is allowed to be infinite. Since any measurable function $f : E \to \mathbb{R}$ can be written as the sum of its positive and negative parts: $f = f_+ - f_-$, where $f_+ = \max(0, f)$ and $f_- = \max(0, -f)$, we can define separately $\mu(f^+)$, $\mu(f^-)$ as above. If $\mu(f^+)$, $\mu(f^-)$, are not infinite we say that f is μ-integrable and we define $\mu(f) = \mu(f^+) - \mu(f^-)$, called the integral of f with respect to μ.

When μ is the Lebesgue measure, $\mu(f)$ is simply the Lebesgue integral of f. By analogy with this case, $\mu(f)$ is denoted using the "differential" notation $\mu(f) = \int_E f(x)\mu(dx) = \int_E f d\mu$, or simply $\int f$, if the measure is implied. Finally we set:

$$\int_A f d\mu = \int_E f I_A d\mu.$$

Two measurable functions f, g on E are said to be equal *μ-almost-everywhere* (almost-surely), we write a.e (a.s), if they differ only on a null set:

$$f = g \ \ a.e. \Leftrightarrow \mu\left(\{f \neq g\}\right) = 0 \tag{2.16}$$

If f, g are integrable then

$$f = g \ \ a.e. \Rightarrow \int_E f(x)\mu(dx) = \int_E g(x)\mu(dx). \tag{2.17}$$

Remark 2.6. The latter equation states that the two functions f and g are equal a.s. with respect to the σ-algebra $\mathcal{E}$ (we say that they are equivalent on $\mathcal{E}$); i.e. for any measurable set $A \in \mathcal{E}$ one has:

$$\int_A f = \int_A g. \tag{2.18}$$

Let a measurable space $(E, \mathcal{E}, \mu)$ be given. We indicate with $L^p(E, \mu)$, $p > 0$, the vector space of the equivalence classes of measurable functions such that:

$$L^p(E, \mu) = \left\{ f \text{ measurable}, \ \ \int_E |f|^p d\mu < \infty \right\} \tag{2.19}$$

We have the following result:

Prop. 2.5. *If the measure μ is finite, i.e. $\mu(E) < +\infty$, then:*

$$L^q(E, \mu) \subset L^p(E, \mu), \quad p < q;$$

Proof: the result is straightforward because $|f|^p \leq |f|^q + 1$. □

It is known that L^p is a Banach space with the norm $\| f \|_p = (\int | f |^p)^{1/p}$. Moreover, if $p = 2$, the norm comes form a scalar product $(f, g) = \int fg$. so that, L^2 is an Hilbert space.

Remark 2.7. Given a sequence $\{f_n\}_{n \in \mathbb{N}}$ of measurable function, the almost everywhere convergence, i.e. $\mu\{f_n \to f = 1\} = 1$, does not imply the L^p convergence, i.e. $||f_n - f||_p \to 0$. The converse is also not true. However, any sequence which converges in the L^p norm admits a subsequence which converges a.s. to the same limit.

The following results hold:

Theorem 2.4 (Dominated convergence)**.** *Let $\{f_n\}_{n \in \mathbb{N}}$ a sequence of measurable functions converging a.e. to a function f. If there exists $g \in L^p$ with $|f_n| \leq g$, then $f_n \to f$ in the L^p norm. In particular,*

$$\int f_n \to \int f.$$

Theorem 2.5 (Integration with respect to the imagine measure)**.** *Given a measurable space $(E, \mathcal{E})$ with a measure μ and a second measurable space $(F, \mathcal{F})$. Let $h : E \to F$ be a measurable function on E and let $f : F \to \mathbb{R}$ be a real measurable function on F. Then,*

$$f \circ h \in L^1(E, \mu) \iff f \in L^1(F, \mu_h);$$

In this case:

$$\int_E (f \circ h)\, d\mu = \int_F f d\mu_h. \tag{2.20}$$

Def. 2.12 (Product measure)**.** Given two finite measures P defined on $(\Omega, \mathcal{F})$ and Q over $(\Omega', \mathcal{F}')$, we indicate with $P \otimes Q$ the unique measure on $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ such that to any rectangle $H \times K$ associates a measure $P(H)Q(K)$. Such a measure is termed *product measure*:

$$P \otimes Q(H \times K) = P(H)Q(K), \;\; H \times K \in \mathcal{F} \otimes \mathcal{F}' \tag{2.21}$$

To its construction is related the well know Fubini-Tonelli theorem, which reduces the integration with respect to $P \otimes Q$ to an iterated integration with respect to the factor measures P and Q.

Theorem 2.6 (Fubini Tonelli theorem)**.** *Let f be a real measurable function defined on the product space $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', P \otimes Q)$. Then, for any $\omega \in \Omega$ the function $f(\omega, \cdot)$ is $\mathcal{F}'$ measurable (and viceversa). Moreover, the function $\int_{\Omega'} f(\cdot, \omega')dQ(\omega')$ is $\mathcal{F}$-measurable (and viceversa). Besides, if f is integrable one has*

$$\begin{aligned} \int_{\Omega \times \Omega'} f(\omega, \omega') dP \otimes dQ(\omega, \omega') &= \int_\Omega \left[\int_{\Omega'} f(\omega, \omega') dQ(\omega') \right] dP(\omega) \\ &= \int_{\Omega'} \left[\int_\Omega f(\omega, \omega') dP(\omega) \right] dQ(\omega'). \end{aligned} \tag{2.22}$$

2.1.3 Absolute continuity and densities

Consider now a measurable space $(E, \mathcal{E})$ with measures μ_1 and μ_2 defined on it. How can these two measures be compared? A natural idea to compare μ_1 and μ_2 is to look at the ratio $\mu_2(A)/\mu_1(A)$ for various measurable sets A. Of course this is only possible if $\mu_2(A) = 0$ every time $\mu_1(A) = 0$. This remarks motivates the following definition:

Def. 2.13 (Absolute continuity)**.** A measure μ_2 is said to be absolutely continuous with respect to μ_1 if for any measurable set A

$$\mu_1(A) = 0 \Rightarrow \mu_2(A) = 0. \tag{2.23}$$

We write $\mu_2 \ll \mu_1$.

Example 2.5. If a measure μ has a density with respect to the Lebesgue measure λ (example 2.3), then $\mu \ll \lambda$.

Example 2.6. Dirac measures are not absolutely continuous with respect to the Lebesgue measure.

Absolute continuity can be characterized in the following way:

Theorem 2.7 (Radon-Nikodym theorem)**.** *If μ_2 is absolutely continuous with respect to μ_1 then there exists an $\mathcal{E}$-measurable function $Z : E \to [0, \infty)$ such that for any measurable set A*

$$\mu_2(A) = \int_A Z d\mu_1 = \mu_1(Z 1_A) \tag{2.24}$$

The function Z is uniquely defined, up to a zero-measure set, and it is called the density or Radon-Nikodym derivative of μ_2 with respect to μ_1 and denoted as $d\mu_2/d\mu_1$. For any μ_2-integrable function f

$$\mu_2(f) = \int_E f d\mu_2 = \mu_1(fZ) = \int_E d\mu_1 Z f \tag{2.25}$$

Therefore if μ_2 is absolutely continuous with respect to μ_1, an integral with respect to μ_2 is a weighted integral with respect to μ_1, the weight being given by the density Z.

If both μ_2 is absolutely continuous with respect to μ_1 and μ_1 is absolutely continuous wit respect to μ_2 then μ_1 and μ_2 are said to be *equivalent* measures. This is equivalent to stating that $Z > 0$.

Remark 2.8. In the case of $(\mathbb{R}, \mathcal{B})$, a measure μ has density with respect to the Lebesgue measure λ if and only if $\mu \ll \lambda$.

2.2 Random Variables

2.2.1 Random variables and probability spaces

Consider a set Ω, called the set of scenarios, equipped with a σ-algebra $\mathcal{F}$. In modeling context, each scenario $\omega \in \Omega$ being described in terms of the evolution. A probability measure on $(\Omega, \mathcal{F})$ is a positive finite measure P with total mass $P(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, P)$ is then called a *probability space*.

A measurable set $A \in \mathcal{F}$, called an event, is therefore a set of scenarios to which a probability can be assigned. A probability measure assigns a *probability* between 0 and 1 to each event $A \in \mathcal{F}$:

$$P : \mathcal{F} \longrightarrow [0,1], \quad A \longrightarrow P(A).$$

An event A with probability $P(A) = 1$ is said to occur almost surely. If $P(A) = 0$ this is interpreted by saying the event A is impossible.

If we are dealing with several probability measures defined on the same set then one should be more specific: we will then replace "almost surely" or "impossible" by "P-almost surely" or "impossible" under P. Therefore, a P-null set is a subset of an impossible event under P.

As before, we can complete $\mathcal{F}$ to include all null sets. This means that we assign probability zero to subsets of impossible events, which is intuitively reasonable. Unless stated otherwise, we shall consider complete versions of all σ-algebras. As in (def. 2.5), we say that a property holds P-almost surely if the set of $\omega \in \Omega$ for which the property does not hold is a P-null set.

One can speak of absolute continuity and equivalence for probability measures: two probability measures P and Q on $(\Omega, \mathcal{F})$ are equivalent (or comparable) if they define the same impossible events:

$$P \sim Q \Leftrightarrow \{\forall A \in \mathcal{F}, P(A) = 0 \Leftrightarrow Q(A) = 0\}.$$

Def. 2.14 (Random variable)**.** A measurable function with values in the measurable space $(E, \mathcal{E})$,

$$X : \Omega \to E,$$

where $(\Omega, \mathcal{F}, P)$ is a probability space is called *random variable* (with values on E).

An element $\omega \in \Omega$ is called a scenario of randomness. $X(\omega)$ represents the outcome of the random variable if the scenario ω happens and is called the realization of X in the scenario ω. If X and Y are two random variables defined on the same probability space, if

$$P\{\omega \in \Omega, X(\omega) = Y(\omega)\} = P(X = Y) = 1,$$

we write

$$X = Y \ \ a.s. \ \ \textit{(almost surely)} \tag{2.26}$$

Def. 2.15 (Law)**.** The law (or distribution) μ_X of X is the imagine measure P_X of P through X on E:

$$\mu_X(A) = P\{\omega \in \Omega;\ \ X(\omega) \in A\} = P(X \in A),\ \ A \in \mathcal{E}. \tag{2.27}$$

The law of a r.v. X describes the probability distribution of the possible values of the random quantity X. If $\mu_X = \mu_Y$ then X and Y are said to be identical in law (or in distribution) and we write

$$X \overset{d}{=} Y. \tag{2.28}$$

The law-equivalence doesn't imply the a.s. equivalence. This is the case for example of r.v.s defined in different probability space. For instance, consider the r.v. X defined on the probability space $(\Omega, \mathcal{F}, P)$ with values on $(\mathbb{R}, \mathcal{B})$, and the r.v. id defined on $(\mathbb{R}, \mathcal{B}, \mu_X)$, such that $id(y) = y$, for any $y \in \mathbb{R}$. It is easy to see that they are identical in law.

Remark 2.9. On the other hand, if $X = Y$ almost surely, then $X \overset{d}{=} Y$.

A random variable $X : \Omega \to E$ is called a real-valued random variable when $E \subset \mathbb{R}$. As in Section 2.1.2, one can define the integral of a positive random variable X with respect to P.

Def. 2.16 (Cumulative distribution function)**.** Let X be a real random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. For any $x \in \mathbb{R}$, the function:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} d\mu_X(y), \tag{2.29}$$

is said *cumulative distribution function*.

Remark 2.10. The cumulative distribution completely characterizes the probability distribution of the r.v. X. Indeed, the class $\{(-\infty, x],\ \ x \in \mathbb{R}\}$ is a base of $\mathcal{B}$ (see Dynkin's Theorem 2.2). Moreover, if X possesses density function f_X, i.e. if the law $\mu_X \ll \lambda$, then:

$$F_X(x) = \int_{-\infty}^{x} f_X(y) d\lambda(y), \tag{2.30}$$

and $f_X(x) = F_X'(x)$. Sometimes, it is useful to study the opposite question and ask how often the random variable is above a particular level. This is called the complementary cumulative distribution function

$$F_X^c(x) = P(X > x) = 1 - F_X(x),\ \ x \in \mathbb{R}. \tag{2.31}$$

In survival analysis, F_X^c is called the *survival function* and is denoted by $S_X(x)$.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. If the probability measure P can be only one and zero the σ-algebra $\mathcal{F}$ is said *degenere*. A random variable is said *trivial* if it is equal to a constant **almost everywhere**. We have the following

Prop. 2.6. *Given a probability space $(\Omega, \mathcal{F}, P)$, with $\mathcal{F}$ degenere. Then, any real random variable is trivial.*

Proof: if $A \in \mathcal{F}$, then $A = \Omega$ P-a.s.; the result is thus obvious for characteristic function. Therefore, it is true for linear finite combinations, and also for simple functions. So that, the result is proven for any positive measurable function and, by difference, for any r.v. □

Def. 2.17 (Expectation)**.** The integral of a positive r.v. X, defined on a probability space $(\Omega, \mathcal{F}, P)$, is called the *expectation* of X with respect to P and is denoted by

$$E_P[X] = \int_\Omega X(\omega) dP(\omega),$$

or simply $E(X)$ or EX. It could be either a positive number or $+\infty$. If $E_P[X] < \infty$ then X is said to be *P-integrable*.

By decomposing any real-valued random variable Y into its positive and negative parts $Y = Y_+ - Y_-$, it is clear that if $E(|Y|) < \infty$ then both Y_- and Y_+ are integrable and the expectation $E(Y) = E(Y_+) - E(Y_-)$ is well-defined.

It is sometimes useful to allow "infinite" values for positive random variables, i.e., choose $E = [0, \infty) \cup \{+\infty\}$. Of course if $Y \in L^1(\Omega, P)$ then Y is almost surely finite.

Remark 2.11. Let X be a real valued r.v. defined on $(\Omega, \mathcal{F}, P)$. Using Theorem 2.5, with $f(x) = x$, $x \in \mathbb{R}$, one has:

$$E_P(X) = \int_\Omega X dP = \int_{\mathbb{R}} x d\mu_X(x). \tag{2.32}$$

2.2.2 Independence

Consider a probability space $(\Omega, \mathcal{F}, P)$. The evaluation of the probability of random events is not a steady characteristic, but depends on the information about them. Therefore, it changes when we have new information. For instance, if one knows that an event $H \in \mathcal{F}$ occurred, then the probability P must be changed with a new probability P_H:

Def. 2.18 (Conditional probability)**.** We define *the conditional probability* with respect to an event H, the probability measure P_H such that, for any event $A \in \mathcal{F}$:

$$P_H(A) = \frac{P(A \cap H)}{P(H)}.$$

We also write $P(A|H)$ to indicate the probability of A conditioned to H.

The following properties follow directly from the definition:

1. $P_H(H^c) = 0$.
2. P_H has density $P(H)^{-1} I_H$ with respect to P.
3. Let X be a r.v. then the expectation value of X conditioned to the event H is the integral of X evaluated with P_H:
$$E(X|H) = \int X dP_H = \frac{1}{P(H)} \int_H X dP. \tag{2.33}$$

Def. 2.19 (Independent events)**.** Two events A and B are said to be independent (with respect to P) if:

$$P(A \cap B) = P(A)P(B).$$

Remark 2.12. We observe that if one of the two events has zero probability, then the events are always independent. On the other hand if $P(B) > 0$, then the definition is equal to

$$P_B(A) = P(A),$$

namely, the knowledge of B does not change the probability of A and this corresponds to the intuitive idea of independence.

Prop. 2.7. *The class of the independent events with respect to an event $H \in \mathcal{F}$ form a Dynkin class.*

Proof: indeed, Ω belong to the class. Moreover, if $B \subset A$ then:

$$P_H(A \setminus B) = P_H(A) - P_H(B) = P(A) - P(B) = P(A \setminus B).$$

Finally, if $A_n \uparrow A$, then (see prop. 2.2):

$$P_H(A) = \lim_n P_H(A_n) = \lim_n P(A_n) = P(A). \quad \square$$

Def. 2.20 (Independent random variables)**.** Two random variables X and Y, defined on the probability space $(\Omega, \mathcal{F}, P)$, with values on the measurable spaces $E_{\mathcal{E}}$ and $G_{\mathcal{G}}$, are said to be independent if the corresponding σ-algebras (Def. 2.8) are independent. That is, for any $A \in \sigma(X)$ and $B \in \sigma(Y)$, then A is independent of B.

The independence entails the fact that information about a random variable (enclosed in its σ-algebra), does not influence the law of the second random variable. Indeed, if $B \in \sigma(Y)$ has non-zero probability, then, for any $A \in \mathcal{E}$, $P_B(X \in A) = P(X \in A)$, and viceversa.

We have the following result:

Prop. 2.8. *Two random variables X and Y, defined on the probability space $(\Omega, \mathcal{F}, P)$, are independent if and only if the law of the couple (X, Y) is the product measure of the single variable laws.*

Proof: we can see the couple (X, Y) as a random variable defined on $(\Omega, \mathcal{F}, P)$ with values in the product space $(E \times G, \mathcal{E} \otimes \mathcal{G})$. For any event $A \times B \in \mathcal{E} \otimes \mathcal{G}$ the law of the couple is:

$$P_{(X,Y)}(A \times B) = P(X \in A, Y \in B) = P(\{X \in A\} \cap \{Y \in B\}).$$

The sufficient condition is obvious. Moreover, if we assume the independence, we have (Def. 2.19):

$$P_{(X,Y)}(A \times B) = P(X \in A)P(Y \in B),$$

for any event $A \times B$. $\square$

Remark 2.13. Suppose that the two random variables X and Y are real valued with density f_X and f_Y with respect, for example, to the Lebesgue measure. Therefore, the couple (X, Y) has density

$$f_{(X,Y)} = f_X f_Y. \tag{2.34}$$

Theorem 2.8. *Given two real valued integrable independent random variables X and Y, defined on $(\Omega, \mathcal{F}, P)$, then the random variable $Z = XY$ is integrable with expectation value:*

$$E(XY) = E(X)E(Y).$$

Proof: let μ indicates the law of X and ν the law of Y. Because of the result above, the couple (X, Y) has law $\mu \otimes \nu$ (Def. 2.12). Using the theorem of integration with respect to the imagine measure (Teor. 2.5), one has:

$$E(XY) = \int_{\mathbb{R}^2} xy \ \mu \otimes \nu(dx, dy) = \int_{\mathbb{R}} x\mu(dx) \int_{\mathbb{R}} y\nu(dy) = E(X)E(Y),$$

where we have used the Fubini theorem. $\square$

Theorem 2.9 (Reduction lemma)**.** *Let X and Y r.v.s and let $\mathcal{F}$ be a σ-algebra independent of Y which makes X measurable. Let f a real function defined on the product of the arrival spaces of X and Y such that:*

$$Z = f(X, Y),$$

results integrable. Set $g(x) = E[f(x, Y)]$, the r.v. $g \circ X$ is equal to Z on $\mathcal{F}$.

Proof: let I_A the indicator function of an element $A \in \mathcal{F}$. The couple (I_A, X) is independent of Y. We indicate with μ the law of (I_A, X) and ν the law of Y. We have:

$$\int_A Z = \int I_A Z = \int u \int f(v,w)\nu(dw)\mu(du,dv) = \int ug(v)\mu(du,dv) = \int I_A g(X) = \int_A g(X)$$

for any $A \in \mathcal{F}$. □

The reduction lemma allows to evaluate, for any couple of independent random variables X and Y, the expectation value of $f(X,Y)$ in two steps: one first calculates the function $g(x) = E[f(x,Y)]$, then evaluates this function on X and finally takes the expectation with respect to X.

Prop. 2.9. *Let X and Y be real independent r.v.s. We have that:*

1. *if $X+Y$ is trivial, then X and Y are trivial;*
2. *if $X+Y$ belong to L^p, then both X and Y belong to L^p.*

Proof: there exists $c \in \mathbb{R}$ such that $X + Y = c$ almost surely. Then, the random variable $c - X = Y$ almost surely. However, Y is independent of X and this implies that X is independent of itself. That is, for any $A \in \sigma(X)$ one must have:

$$P(A \cap A) = P(A)^2 = P(A) \Leftrightarrow P(A) = 0 \wedge 1;$$

thus, $\sigma(X)$ is degenere; and using Proposition 2.6 follows that X is trivial. The same for Y.

Let $|X+Y|^p$ integrable. Let ν indicates the law of Y, then

$$E(|\ X+Y\ |^p) = \int E(|\ X+y\ |^p)\nu(dy) < \infty,$$

so that, $X + y \in L^p$ ν-a.e., therefore, $X \in L^p$, and the same for Y. □

Prop. 2.10. *Let X and Y be two independent real r.v.s. We also suppose that they have density f_X and f_Y with respect to the Lebesgue measure λ. Then, the random variable $Z = X + Y$ has density*

$$f_Z(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(z-x) f_Y(x) dx. \tag{2.35}$$

Proof: we have that:

$$F_Z(z) = P(Z \le z) = P(X+Y \le z) = \int_{\mathbb{R}^2} I_{x+y\le z} f_X(x) f_Y(y) dxdy = \int_{\mathbb{R}} f_X(x) \int_{\mathbb{R}} I_{y\le z-x} f_Y(y) dydx$$

(with the change of variables $x + y = y'$)

$$= \int_{-\infty}^{z} \int_{\mathbb{R}} f_X(x) f_Y(y'-x) dxdy'. \quad \square$$

The above proposition states that the density function of the sum Z of two independent real random variables X and Y is given by the convolution $f_X * f_Y$ of the two corresponding densities. What happens if we consider products XY of real independent random variables.

Prop. 2.11. *Let X and Y be two independent real r.v.s such that $P(X \ge 0) = 1$ with densities f_X and f_Y with respect to the Lebesgue measure λ. Then, the random variable $Z = XY$ has density:*

$$f_Z(z) = (f_X \star f_Y)(z) = \int_0^\infty f_X(x) f_Y(z/x) dx/x, \quad z \in \mathbb{R}, \tag{2.36}$$

where $(f \star g)(x)$ indicates the Mellin convolution of the two integrable functions f and g evaluated in $x \in \mathbb{R}$.

Proof: the cumulative distribution of Z is:

$$F_Z(z) = P(Z \leq z) = P(XY \leq z) = \int_0^\infty dx \int_{\mathbb{R}} dy I_{xy \leq z} f_X(x) f_Y(y)$$

(with the change of variables $y' = xy$)

$$\int_{-\infty}^{z} \int_0^\infty f_X(x) f_Y(y'/x) dx/x dy'. \quad \square$$

2.2.3 Characteristic function

The characteristic function of a random variable is the Fourier transform of its distribution. Many probabilistic properties of random variables correspond to analytical properties of their characteristic functions.

Def. 2.21 (Characteristic function)**.** The *characteristic function* of a $\mathbb{R}^d$-valued random variable X is the function $\Phi_X : \mathbb{R}^d \to \mathbb{R}$ defined by

$$\Phi_X(y) = E(e^{iy \cdot X}) = \int_{\mathbb{R}^d} e^{iy \cdot x} d\mu_X(x), \quad \forall y \in \mathbb{R}^d. \tag{2.37}$$

Remark 2.14. The characteristic function of a random variable completely characterizes its law. Two variables with the same characteristic function are identically distributed.

Example 2.7 (Gaussian characteristic function)**.** Let $X = \{X_1, X_2, \ldots, X_n\}$ be a collection of Gaussian random variables with covariance matrix $\gamma(i,j) = E(X_i X_j)$ and mean vector $m = \{E(X_1), \ldots, E(X_n)\}$. Therefore, for any real vector $y = \{y_1, \ldots, y_n\}$, one has:

$$E(e^{iy \cdot X}) = \exp\left(-\frac{1}{2} \sum_{k,j=1}^{n} y_k \gamma(k,j) y_j + i \sum_{k=1}^{n} m_k y_k \right) \tag{2.38}$$

A characteristic function is always continuous and verifies $\Phi_X(0) = 1$. Additional smoothness properties of Φ_X depend on the existence of *moments* of the random variable X.

Def. 2.22 (Moments)**.** The moments of a random variable may or may not exist, depending on how fast the distribution μ_X decays at infinity:

- the n-th moment m_n of a random variable X on $\mathbb{R}$ is defined by $m_n(X) = E[X^n]$;
- the absolute moments of X are the quantities $m_n(|X|) = E[|X|^n]$;
- the n-th centered moment μ_n is defined as the n-th moment of $X - E[X]$

$$\mu_n(X) = E[(X - E[X])^n]. \tag{2.39}$$

Remark 2.15 (Characteristic function and moments)**.** The moments of a random variable are related to the derivatives at 0 of its characteristic function:

1. If $E[|X|^n] < \infty$ then Φ_X has n continuous derivatives at $y = 0$ and

$$\forall k = 1, ..., n; \qquad m_k \equiv E[X^k] = \frac{1}{i^k} \frac{\partial^k \Phi_X}{\partial y^k}(0). \tag{2.40}$$

2. If Φ_X has 2n continuous derivatives at $y = 0$ then $E[|X|^{2n}] < \infty$ and

$$\forall k = 1, ..., 2n; \qquad m_k \equiv E[X^k] = \frac{1}{i^k} \frac{\partial^k \Phi_X}{\partial y^k}(0). \tag{2.41}$$

3. X possesses finite moments of any order if $\Phi_X(y)$ is $C^\infty(\mathbb{R})$ at y = 0.

Then the moments of X are related to the derivatives of Φ_X by:

$$m_n \equiv E[X^n] = \frac{1}{i^n} \frac{\partial^n \Phi_X}{\partial z^n}(0). \tag{2.42}$$

Remark 2.16. We observe that if $\{X_i, i = 1, \ldots, n\}$ is a sequence of independent random variables, then (see Proposition 2.8) the characteristic function of $S_n = X_1 + X_2 + ... + X_n$ is the **product** of the characteristic functions of the individual variables:

$$\Phi_{S_n}(y) = \prod_{i=1}^{n} \Phi_{X_i}(y). \tag{2.43}$$

Def. 2.23 (Moment generating function)**.** The moment generating function of $\mathbb{R}^d$-valued random variable X is the function M_X defined by

$$M_X(u) = E(e^{u \cdot X}), \quad \forall u \in \mathbb{R}^d. \tag{2.44}$$

Contrarily to the characteristic function, which is always well-defined (as the Fourier transform of a probability measure), the moment generating function is not always defined: the integral in (eq. 2.44) may not converge for some (or all) values of u.

Remark 2.17. When the moment generating function M_X is well-defined, it can be formally related to the characteristic function Φ_X by:

$$M_X(u) = \Phi_X(iu). \tag{2.45}$$

If the moment generating function M_X of a random variable X on $\mathbb{R}$ is defined on a neighborhood $[-\epsilon, \epsilon]$ of zero, then in particular all (polynomial) moments of X are finite and can be recovered from the derivatives of M in the following manner:

$$m_n = \frac{\partial^n M_X}{\partial u^n}(0). \tag{2.46}$$

Let X be a random variable and Φ_X its characteristic functions. As mentioned above $\Phi_X(0) = 1$ and Φ_X is continuous in zero. One can then define a continuous version of the logarithm of Φ_X.

Def. 2.24 (Cumulant generating function)**.** There exists a unique continuous function Ψ_X defined in a neighborhood of zero such that

$$\Psi_X(0) = 0 \quad and \quad \Phi_X(y) = \exp[\Psi_X(y)]. \tag{2.47}$$

The function Ψ_X is called the *cumulant generating function* or log-characteristic function of X.

Note that if $\Phi_x(y) \neq 0$ for all y, the cumulant generating function can be extended to all $\mathbb{R}^d$. The *cumulants* or *semi-invariants* of X are defined by:

$$c_n(X) = \frac{1}{i^n} \frac{\partial^n \Psi_X}{\partial u^n}(0). \tag{2.48}$$

By expanding the exponential function at $y = 0$ and using (eq. 2.47), the n-th cumulant can be expressed as a polynomial function of the moments $m_k(X), k = 1, \ldots, n$ or the central moments $\mu_k(X)$. For instance:

- $c_1(X) = m_1(X) = E(X)$,
- $c_2(X) = \mu_2(X) = E(X^2) - E(X)^2 = m_2(X) - m_1(X)^2 = \text{Var}\,(X)$,
- $c_3(X) = \mu_3(X) - 3m_2(X)m_1(X) + 2m_1(X)^3$,
- $c_4(X) = \mu_4(X) - 3\mu_2(X)$.

Scale-free versions of cumulants can be obtained by normalizing c_n by the n-th power of the *standard deviation*:

$$s(X) = \frac{c_3(X)}{c_2(X)^{3/2}}, \quad k(X) = \frac{c_4(X)}{c_2(X)^2}. \tag{2.49}$$

- The quantity $s(X)$ is called the **skewness** coefficient of X. If $s(X) > 0$, X is said to be positively skewed.
- The quantity $k(X)$ is called the excess **kurtosis** of X. If $k(X) > 0$, X is said to be **leptokurtic** or "*fat-tailed*".

Remark 2.18. If X follows a normal distribution, Ψ_X is a second degree polynomial (eq. 2.38), so that for any $n \geq 3$, one has $c_n(X) = 0$. Therefore, $s(X)$ and $k(X)$ (or higher cumulants) can be seen as measures of deviation from normality.

Remark 2.19. By construction, the skewness and the kurtosis are scale invariant; i.e. for any $\lambda > 0$:

$$s(\lambda X) = s(X), \quad k(\lambda X) = k(X). \tag{2.50}$$

2.3 Convergence of random variables

When considering sequences or families of random variables, one can give several different meaning to the notion of convergence. In this section we define and compare these different notions of convergence which will be useful in the sequel. While we define convergence in terms of sequences (indexed by integers), our definitions also hold for continuously indexed families.

2.3.1 Almost-sure convergence

Consider a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ with values in some normed vector space E, for instance $E = \mathbb{R}^d$.

Remark 2.20. A random variable is defined as a function $X : \Omega \to E$ of the "randomness" $\omega \in \Omega$. One could then apply to random variable sequences, all the various notions of convergence which exist for function sequences.

The simplest notion of convergence for function is the *pointwise convergence*, which requires that for every $\omega \in \Omega$, the sequence $\{X_n(\omega)\}_{n\geq 1}$ converges to $X(\omega)$ in the E-norm. This notion turns out to be too strong in many cases since we are asking convergence for all samples $\omega \in \Omega$ without taking into account that many events may in fact be negligible, i.e., of probability zero. The notion of almost-sure convergence takes this fact into account and requires pointwise convergence only for realizations which have nonzero probability of occurrence.

Def. 2.25 (Almost-sure convergence)**.** A sequence $\{X_n\}_{n\geq 1}$ of random variables, defined on $(\Omega, \mathcal{F}, P)$, is said to *converge almost-surely* to a random variable X if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{2.51}$$

For the above definition to make sense, the variables $\{X_n\}_{n\geq 1}$ have to be defined on the *same* probability space $(\Omega, \mathcal{F}, P)$. Note that almost sure convergence does not imply convergence of moments: if $X_n \to X$ almost surely, $E(X_n^k)$ may be defined for all $n \geq 1$ but have no limit as $n \to \infty$.

2.3.2 Convergence in probability

While the almost-sure convergence deals with the behavior of typical samples $\{X_n(\omega)\}_{n\geq 1}$, the notion of convergence in probability puts a condition on the probability of events when $n \to \infty$.

Def. 2.26 (Convergence in probability). A sequence $\{X_n\}_{n\geq 1}$ of random variables, defined on $(\Omega, \mathcal{F}, P)$, is said to *converge in probability* to a random variable X if, for each $\epsilon > 0$

$$\lim_{n\to\infty} P\left(|X_n - X| > \epsilon\right) = 0. \tag{2.52}$$

We denote convergence in probability by

$$X_n \xrightarrow{P} X. \tag{2.53}$$

Almost sure convergence implies convergence in probability but the two notions are not equivalent. Also note that convergence in probability requires that the variables $(X_n)_{n\geq 1}$ be defined on the *same* probability space $(\Omega, \mathcal{F}, P)$.

2.3.3 Convergence in distribution

In many situations, especially in a modeling context, the random variable is not a directly observable quantity itself and the only observable quantities are expectations of various functions of this random variable, $E(f(X))$. These quantities will in fact be the same for two random variables having the same distribution. From the point of view of observation, a meaningful notion of convergence is asking that $E(f(X_n))$ converges to $E(f(X))$ for a given set of "observables" or test functions $f : E \longrightarrow \mathbb{R}$. If the set of test functions is rich enough, this will ensure the uniqueness of the law of X but will not distinguish between two limits X and X' with the same distribution μ. A commonly used choice of test functions is the set of bounded continuous functions $f : E \longrightarrow \mathbb{R}$, which we denote by $C_b(E, \mathbb{R})$.

Def. 2.27 (Convergence in distribution)**.** A sequence $\{X_n\}_{n\geq 1}$ of random variables with values in a measurable space $(E, \mathcal{E})$ is said to converge in distribution to a random variable X if, for any bounded continuous function $f : E \longrightarrow \mathbb{R}$, one has:

$$E(f(X_n)) \xrightarrow[n\to\infty]{} E(f(X)). \tag{2.54}$$

If one denotes by μ_n the distribution of X_n, then μ_n is a probability measure on $(E, \mathcal{E})$. Therefore, using (Theo. 2.5), the above equation is equivalent to:

$$\int_E d\mu_n(x) f(x) \xrightarrow[n\to\infty]{} \int_E d\mu(x) f(x), \tag{2.55}$$

for any $f \in C_b(E, \mathbb{R})$. In this form it is clear that, unlike almost-sure convergence, convergence in distribution is defined not in terms of the random variables themselves, but in terms of their distributions.

Remark 2.21. Convergence in distribution is sometimes called *weak convergence* of the measure $(\mu_n)_{n\geq 1}$ on E. We write:

$$\mu_n \Rightarrow \mu \quad or \quad X_n \xrightarrow{d} X. \tag{2.56}$$

An important feature of convergence in distribution is that, unlike the other notions of convergence mentioned above, it does not require the variables to be defined on a common probability space. Nevertheless, in the case where the variables $\{X_n\}$ are defined on the same space, convergence in distribution leads to converge of probabilities of (regular) events in the following sense:

Prop. 2.12. *If (X_n) converges in distribution to X, then for any set A with boundary ∂A, such that $\mu(\partial A) = 0$, then*

$$P(X_n \in A) = E(I_A(X_n)) \underset{n\to\infty}{\longrightarrow} E(I_A(X)) = P(X \in A). \tag{2.57}$$

Note also that, unlike other notions of convergence, convergence in distribution is not "stable under sums". This means that if $\{X_n\}_{\geq 1}$ and $\{Y_n\}_{n\geq 1}$ converge in distribution to X and Y, it is not true in general that $\{X_n + Y_n\}_{n\geq 1}$ converges in distribution to $X + Y$. However, by definition, it is readily observed that if $\{X_n\}_{n\geq 1}$ converges in distribution to X then for any continuous function g, then $\{g(X_n)\}_{n\geq 1}$ converges in distribution to $g(X)$.

The notion of weak convergence is relevant in studying numerical approximations obtained by discretizing continuous time models. The following characterization of convergence in distributions in term of pointwise convergence of characteristic function is useful in practice:

Prop. 2.13. *The sequence $\{X_n\}_{n\geq 1}$ converges in distribution to X, if and only if for every $y \in \mathbb{R}^d$*

$$\Phi_{X_n}(y) \to \Phi_X(y). \tag{2.58}$$

Remark 2.22. Note however that pointwise convergence of the characteristic functions $\{\Phi_{X_n}\}_{n\geq 1}$ does not imply the existence of a weak limit for $\{X_n\}_{n\geq 1}$. This is becasue the pointwise limit of $\{\Phi_{X_n}\}_{n\geq 1}$ is not necessarily a characteristic function.

Convergence in distribution does not necessarily entail convergence of the moments. One cannot choose f to be a polynomial in (eq. 2.55) because polynomials are not bounded. In fact, the moments of X_n and X need not even exist.

When $\{X_n\}_{n\geq 1}$ and X are defined on the same probability space, convergence in distribution is the weakest notion of convergence among the above. Almost sure convergence entails convergence in probability, which entails convergence in distribution.

2.4 Conditional expectation

Let X be a real random variable defined in the probability space $(\Omega, \mathcal{F}, P)$. For any event $B \in \mathcal{F}$, we have defined (eq. 2.33) the expectation value of X conditioned to the event B

$$E(X|B) = \frac{1}{P(B)} \int_B X dP.$$

Consider now the σ-algebra generated by B:

$$\sigma(\{B\}) = \{\emptyset, \Omega, B, B^c\} := \mathcal{G}.$$

We can define a new random variable, the expectation of X conditioned to $\mathcal{G}$, in the following way:

$$E(X|\mathcal{G})(\omega) = \begin{cases} E(X|B)\,, & \omega \in B, \\ E(X|B^c), & \omega \in B^c. \end{cases} \tag{2.59}$$

The above definition satisfies the following properties:

1. $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable;
2. $\int_A XdP = \int_A E(X|\ \mathcal{G})dP$, for any $A \in \mathcal{G}$, that is, X and $E(X|\mathcal{G})$ are equivalent on $\mathcal{G}$ (Rem. 2.6).

Moreover, $E(X|\mathcal{G})$ is the unique r.v. which satisfies the above properties.

Remark 2.23. $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable even if X may not be. One can say that (8.1) gives the best estimate of X when $\mathcal{G}$ is known.

Consider a real integrable r.v. X defined on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra contained on $\mathcal{F}$. The following proposition defines the conditional expectation of X with respect to $\mathcal{G}$.

Theorem 2.10. *Let X and $\mathcal{G}$ as above. Then, there exists a r.v. $Z : \Omega \to \mathbb{R}$ such that:*

1. *$Z \in \mathcal{L}^1(\Omega, P)$, is integrable;*
2. *Z is $\mathcal{G}$-measurable;*
3. *$\int_A ZdP = \int_A XdP, \forall A \in \mathcal{G}$;*
4. *Z is unique (almost-surely) and is said* **conditional expectation** *of $X$with respect to $\mathcal{G}$.*

Proof: for any $B \in \mathcal{G}$ we define $Q(B) = \int_B XdP$. Q is a finite measure on $(\Omega, \mathcal{G})$ and $Q \ll P$. Using the Radon-Nikodyn theorem (Theo. 2.7), there exists a unique (a.s.) random variable Z on $(\Omega, \mathcal{G})$ such that, for any $B \in \mathcal{G}$, one has $Q(B) = \int_B ZdP$. □

2.4.1 Conditional expectation properties

The conditional expectations of a real integrable r.v. X with respect to a σ-algebra $\mathcal{G}$ has the following properties:

1. If X is $\mathcal{G}$-measurable, then $E(X|\mathcal{G}) = X$.
2. If X and $\mathcal{G}$ are independent, that is $\sigma(X)$ and $\mathcal{G}$ are independent, then $E(X|\mathcal{G}) = E(X)$.
3. $E(X) = E(E(X|\mathcal{G}))$.

Moreover, we have also the following:

a. $E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})$, *linearity*;

b. if $X \leq Y$ a.s., then $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ a.s., *monotonicity*;

c. if $\{X_n\}_{n\in\mathbb{Z}}$, with $0 \leq X_n \in L^1(\Omega, P)$, is an increasing sequence which converges a.s. to X, then $E(X_n|\mathcal{G}) \to E(X|\mathcal{G})$ a.s.;

d. if Y is $\mathcal{G}$-measurable, then $E(XY|\mathcal{G}) = YE(X|\mathcal{G})$;

e. if X and Y are independent, then $E(XY|\mathcal{G}) = E(X|\mathcal{G})E(Y|\mathcal{G})$;

f. if $\mathcal{H} \subseteq \mathcal{G}$, then $E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})$.

Bibliography

[1] F. Biagini and M. Campanino, *Elementi di Probabilità e Statistica*, (Springer, Milano, 2006).

[2] R. Cont and P. Tankov, *Financial Modeling With Jump Processes*, (Chapman & Hall, CRC Financial Mathematics Series, 2004).

[3] B. De Finetti, *Theory of Probability, a critical introductory treatment*, Vol. 1, (Wiley & Sons, 1990).

[4] B. De Finetti, *Theory of Probability, a critical introductory treatment*, Vol. 2, (Wiley & Sons, 1990).

[5] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, 3-rd edn. (Wiley, New York, 1968). [1-st edn. 1957]

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2-nd edn. (Wiley, New York, 1971). [1-st edn. 1966]

[7] D. Lamberton and B. Lapeyre, *Stochastic Calculus Applied to Finance*, (Chapman & Hall, London, 1996).

[8] N. Pintacuda, *Probabilità*, (Decibel editrice, Zanichelli, 1994).

[9] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, A series of Comprehensive Studies in Mathematics, (Springer Verlag, 293, 1994).

3 Basics in stochastic processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable X is a rule for assigning to every outcome ω of an experiment Ω a *number* $X(\omega)$. A stochastic process X_t is a rule for assigning to every $\omega \in \Omega$ a *function* $X_t(\omega)$. Thus, a stochastic process could be seen as a *family of time functions* depending on the parameter ω (a collection of paths or trajectories) or, equivalently, a family of random variables depending on a time parameter t, or a function of t and ω as well.

3.1 Stochastic processes and filtrations

Def. 3.1 (Stochastic process)**.** We define real valued (one-dimensional) **stochastic process** a family of random variables $\{X_t\}_{t\in I}$ defined on $(\Omega, \mathcal{F}, P)$:

$$X_t : \Omega \to \mathbb{R}, \quad t \in I \subseteq \mathbb{R}_+.$$

We shall say that $\{X_t\}_{t\in I}$ is a *discrete-state* process if its values are countable. Otherwise, it is a *continuous-state* process. The set $S \subseteq \mathbb{R}$, whose elements are the values of the process, is called *state space*. A stochastic process could be a discrete time or a continuous time process, according as the set I is countable or continuous.

Notation 3.1. We sometimes write $X(t)$ instead of X_t. Moreover, we shall use the notations: $\{X(t)\}_{t\in I}$, $\{X(t),\ t \in I\}$, or $X(t)$, $t \in I$, to represent a stochastic process. We would use simply X or $X(t)$ if the specification of the set I were redundant. Moreover, we often refer at the parameter t as a time variable and thus we often take $I = \mathbb{R}_+ = \{t \in \mathbb{R}, \ \ t \geq 0\}$.

Def. 3.2 (Finite dimensional distributions)**.** For any natural number $k \in \mathbb{N}$ and a "time" sequence $\{t_i\}_{i=1,\dots,k} \in I$, the *finite-dimensional distributions* of the real valued stochastic process $X = \{X_t\}_{t\in I}$ are the measures $\mu_{t_1,\dots,t_k}$, defined on $\mathbb{R}^k$, such that

$$\mu_{t_1,\dots,t_k}(A_1 \times \cdots \times A_k) = P\left(\{X_{t_1} \in A_1, \dots, X_{t_k} \in A_k\}\right), \tag{3.1}$$

where $\{A_1, \dots, A_k\}$ are Borel sets on $\mathbb{R}$.

Remark 3.1 (Marginal distribution)**.** If one takes $\mu_t(A) = \mu_{t,t_2\dots,t_k}(A \times \mathbb{R} \times \cdots \times \mathbb{R})$, is the same as considering the one-dimensional (one-point) distribution $\mu_t(A) = P(X_t \in A)$, $A \in \mathcal{B}$, in (eq. 3.1). In this case, we refer to μ_t as the *marginal distribution* of the process X.

The family of all finite-dimensional distributions determines many (but not all) important properties of the process X. Conversely, given a family $\{\nu_{t_1,\dots,t_k},\ k \in \mathbb{N},\ t_i \in I\}$ of probability measures on $\mathbb{R}^k$, it is important to be able to construct a stochastic process $Y = \{Y_t\}_{t\in I}$ which have $\nu_{t_1,\dots,t_k}$ as its finite-dimensional distributions. One of Kolmogorov's famous theorems states that this can be done provided $\nu_{t_1,\dots,t_k}$ satisfies two natural consistency conditions:

Theorem 3.1 (Kolmogorov extension heorem)**.** *For all $\{t_i\}_{i=1,\dots k} \in I$, $k \in \mathbb{N}$, let $\nu_{t_1,\dots,t_k}$ be probability measures on $\mathbb{R}^k$, such that:*

1. *for all permutations π on $\{1, 2, \dots, k\}$,*

$$\nu_{t_{\pi(1)},\dots,t_{\pi(k)}}(A_1 \times \cdots \times A_k) = \nu_{t_1,\dots,t_k}\left(A_{\pi^{-1}(1)} \times \cdots \times A_{\pi^{-1}(k)}\right).$$

2. *for any* $m \in \mathbb{N}$,

$$\nu_{t_1,\dots,t_k}(A_1 \times \cdots \times A_k) = \nu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(A_1 \times \cdots \times A_k \times \mathbb{R} \times \cdots \times \mathbb{R}),$$

where of course the set on the right side as a total of $k+m$ *factors.*

Then, there exists a probability space $(\Omega, \mathcal{F}, P)$ *and a real valued stochastic process* X *defined on it, such that:*

$$\nu_{t_1,\dots,t_k}(A_1 \times \cdots \times A_k) = P\left(\{X_{t_1} \in A_1, \dots, X_{t_k} \in A_k\}\right), \tag{3.2}$$

for any $t_i \in I$, $k \in \mathbb{N}$ *and* $A_i \in \mathcal{B}$.

All the above results are still valid if one considers d-dimensional processes. Of great importance is the notion of:

Def. 3.3 (Filtration)**.** An increasing family $\mathcal{F}_t = \{\mathcal{F}_t\}_{t\in I}$ of complete sub σ-fields of $\mathcal{F}$ is said a **filtration** on $(\Omega, \mathcal{F}, P)$.

Consider a stochastic process $X = \{X_t\}_{t\in I}$ and let:

$$\mathcal{F}_t^X = \sigma\left(\{X_s;\ 0 \le s \le t\}\right) = \sigma\left(\left\{\mathcal{N} \cup \{X_s^{-1}(H);\ 0 \le s \le t,\ H \in \mathcal{B}\}\right\}\right),$$

where $\mathcal{B}$ is the Borel σ-algebra and $\mathcal{N}$ indicates the class of null-sets. Clearly if $0 \le s \le t$ one has

$$\mathcal{F}_s^X \subseteq \mathcal{F}_t^X \subseteq \mathcal{F}.$$

Therefore, $\mathcal{F}^X = \{\mathcal{F}_t^X\}_{t\in I}$ defines a filtration, termed *natural filtration* of $\{X_t\}_{t\in I}$.

Def. 3.4 (Adapted process)**.** A stochastic process $\{X(t)\}_{t\in I}$ is said *adapted* to the filtration $\{\mathcal{F}_t\}_{t\in I}$ if for each $t \ge 0$:

$$\mathcal{F}_t^X \subseteq \mathcal{F}_t.$$

In other words, for each t, the r.v. $X(t)$ is $\mathcal{F}_t$-measurable. Heuristically, $\mathcal{F}_t^X$ represents all the information for the process $X(t)$ which is available up to time t.

Example 3.1. For example, the process $Y_1(t) = X(t/2)$, $t \ge 0$, is $\mathcal{F}_t^X$-adapted, whereas the process $Y_2(t) = X(2t)$, $t \ge 0$, is not.

Obviously, if $X(t)$, $t \in I$, is an integrable $\mathcal{F}_t$-adapted process (i.e. for each $t \in I$ the r.v. X_t is integrable), we have (see Section 2.4.1):

$$E(X_t \mid \mathcal{F}_t) = X(t), \quad t \in I.$$

Consider a probability space with filtration $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. An important class of stochastic processes are the so called *martingales*.

Def. 3.5 (Martingale)**.** A stochastic process $M = \{M_t\}_{t\ge 0}$ is a **martingale** with respect to the filtration $\mathcal{F}_t$ and the measure P if, for any $t \ge 0$, one has:

1. $M_t \in L^1(\Omega, P)$;
2. $E(M_t \mid \mathcal{F}_s) = M_s$, for any $0 \le s \le t$.

This means that M_t is $\mathcal{F}_t$-adapted. Moreover, the expected value of M_t does not depend on time. Indeed,

$$E(M_t) = E\left(E(M_t \mid \mathcal{F}_0)\right) = E(M_0).$$

Example 3.2 (Gaussian process)**.** A real stochastic process $\{X_t\}_{t\in I}$ is Gaussian if and only if, for every finite sequence $\{t_1, t_2, \ldots, t_k\} \in I$,

$$\mathbf{X}_{t_1,\ldots,t_k} = (X_{t_1}, \ldots, X_{t_k})$$

has a multivariate normal distribution. That is, there are a positive semi-definite symmetric matrix $\{\gamma(t_i, t_j) = \gamma_{i,j},\ \ i, j = 1, \ldots, k\}$ and real numbers $\{\mu_i\}_{i=1,\ldots,k}$ such that:

$$E\left(\exp\left(i\sum_{\ell=1}^{k} t_\ell\, X_{t_\ell}\right)\right) = \exp\left(-\frac{1}{2}\sum_{\ell,j}\gamma_{\ell j}t_\ell t_j + i\sum_{\ell}\mu_\ell t_\ell\right), \tag{3.3}$$

where γ is the covariance matrix and μ is the vector of the means. Alternatively, the random variable $Z = f(X_{t_1}, \ldots, X_{t_k})$, where f is a linear functional, is normally distributed.

3.2 Stationary processes

Def. 3.6 (Stationary process)**.** A stochastic process $\{X_t\}_{t\geq 0}$ is said a *stationary process* if any collection $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ has the same distribution of $\{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\}$, for each $\tau \geq 0$. That is,

$$\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} \stackrel{d}{=} \{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\}. \tag{3.4}$$

The above equation (eq. 3.4) expresses the *time translational invariance* of the finite dimensional distributions of a stationary process. Let X be a stationary process, then the following elementary properties hold:

- Varying t, all the random variables X_t have the same low; i.e. $X_{t_1} \stackrel{d}{=} X_{t_2}$ for any $t_1, t_2 \geq 0$.
- All the moments, if they exist, are constant in time.
- The distribution of X_{t_1} and X_{t_2} depends only on the difference $\tau = t_2 - t_1$ (time lag).

Therefore, the autocovariance function $\gamma(t_1, t_2) = \gamma(t_2 - t_1)$ depends only on $\tau = t_2 - t_1$. We write

$$\gamma(\tau) = E(X_t - \mu)(X_{t-\tau} - \mu) = \mathrm{Cov}(X_t, X_{t-\tau}), \tag{3.5}$$

where $\mu = E(X(t))$ and $\gamma(\tau)$ indicates the autocovariance coefficient at the lag τ.

Remark 3.2 (Wide-sense stationary process)**.** If the mean and the variance are finite and constant and the autocovariance function satisfies (eq. 3.5), then the stochastic process X is said *wide-sense stationary*. No further request is given for high order moments. Clearly, stationarity always implies wide-sense stationarity, but the converse is not in general true. Observe that, a wide-sense stationary Gaussian process is strictly stationary.

Def. 3.7 (Stationary increment process)**.** A stochastic process $\{X_t\}_{t\geq 0}$ is said a stationary increment process, shortly **si**, if for any $h > 0$:

$$\{X_{t+h} - X_h\}_{t\geq 0} \stackrel{d}{=} \{X_t - X_0\}_{t\geq 0}. \tag{3.6}$$

Remark 3.3. If X is an **si**-process, than, for any $\delta t > 0$, the increment process

$$\{Y_{\delta t}(t)\}_{t\geq 0} = \{\Delta_{\delta t} X_t\}_{t\geq 0} = \{X_{t+\delta t} - X_t\}_{t\geq 0}$$

is a stationary process.

3.3 Self-similar processes

Def. 3.8. A real valued stochastic process $X = \{X_t\}_{t\geq 0}$ is said *self-similar* with index $H > 0$, shortly H-**ss**, if for any $a > 0$:

$$\{X_{at}\}_{t\geq 0} \stackrel{d}{=} \{a^H X_t\}_{t\geq 0}.$$

We observe that the transformation scales differently "space" and "time", for this reason one often prefers using the word *self-affine process*. The index H is said *Hurst's exponent* or scaling exponent of the process. In the same way, it is possible to define self-similarity for processes $\{X_t\}_{t\in\mathbb{R}}$.

Remark 3.4. Observe that, if X is an H-**ss** process, then all the finite-dimensional distributions of X in $[0,\infty[$ are completely determined by the distribution in any finite real interval.

Prop. 3.1. *Let $X = \{X_t\}_{t\geq 0}$ be a non-degenerate stationary process*[1]*, then X can not be an H-**ss** process.*

Proof: indeed, for any $a > 0$:

$$X_t \stackrel{d}{=} X_{at} \stackrel{d}{=} a^H X_t,$$

and we get a contradiction when $a \to \infty$. □

We have the following proposition that relates stationary and self-similar processes:

Prop. 3.2. *Let $\{X_t\}_t$ be an H-**ss** process; then the process*

$$Y(t) = e^{-tH} X(e^t), \quad t \geq 0, \tag{3.7}$$

is stationary. Moreover, if $Y = \{Y_t\}_{t\geq 0}$ is a stationary process; then

$$X(t) = t^H Y(\log t), \quad t \geq 0, \tag{3.8}$$

*is an H-**ss** process.*

Dim: consider a collection of real numbers $\{\theta_i, \quad i = 1,2,\ldots,n\}$ and let $\{X_t\}_t$ be an H-**ss** process. Let a collection of "times" $\{t_1, t_2, \ldots, t_n\}$ be given and consider $\{Y(t_1), Y(t_2), \ldots, Y(t_n)\}$. Then, using (eq. 3.7), for any $h \in \mathbb{R}$, we could write:

$$\sum_{j=1}^{n} \theta_j Y(t_j + h) = \sum_{j=1}^{n} \theta_j e^{-(t_j+h)H} X(e^{t_j+h}) \stackrel{d}{=} \sum_{j=1}^{n} \theta_j e^{-t_j H} X(e^{t_j}) = \sum_{j=1}^{n} \theta_j Y(t_j);$$

then the random variables

$$y_1 = \textstyle\sum_{j=1}^{n} \theta_j Y(t_j + h),$$
$$y_2 = \textstyle\sum_{j=1}^{n} \theta_j Y(t_j),$$

have the same low and the same characteristic function:

$$E\left(e^{i\xi y_1}\right) = E\left(e^{i\xi y_2}\right).$$

Putting $\xi = 1$ one finds:

$$E\left(\exp\left(i\sum_{j=1}^{n} \theta_j Y(t_j + h)\right)\right) = E\left(\exp\left(i\sum_{j=1}^{n} \theta_j Y(t_j)\right)\right).$$

1) A process $\{X_t\}_{t\geq 0}$ is said to be degenerate if for any $t \geq 0$, $X_t = 0$ almost surely.

This means that the families $\{Y(t_i), \quad i = 1, \ldots, n\}$ and $\{Y(t_i + h), \quad i = 1 \ldots, n\}$ have the same characteristic function, i.e. the same probability distribution. This proves that Y, as defined in (3.7), is a stationary process. In the same way it is possible to show (3.8). □

The above result shows how could be possible to construct self-similar processes starting from stationary processes.

3.4 Brownian motion

Def. 3.9 (Brownian motion)**.** Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a probability space with filtration. A stochastic process $B = \{B_t\}_{t \geq 0}$ is an ordinary linear (one dimensional) one-sided (because $t \geq 0$) Brownian motion[2] (Bm) if:

i) B is a Gaussian H**-ss** process with $H = 1/2$ and with stationary increments.

ii) B is $\mathcal{F}_t$-adapted;

One could ever take the natural filtration associated to B, i.e. $\mathcal{F}_t = \mathcal{F}_t^B$.

Let $B(t)$ be a one-sided Brownian motion, it is possible to show that (see next section):

1. The Bm trajectories starts in zero a.s. and are continuous.
2. Let $\sigma^2 = E(B(1)^2)$, than for any $t \geq 0$, $B(t)$ has zero mean , variance $E(B(t)^2) = \sigma^2 t$ and covariance $E(B(t)B(s)) = \sigma^2 \min(t, s)$.
3. From above follows that the Bm increments are independent. That is, the random variables $B_{t_2} - B_{t_1}$ and $B_{t_4} - B_{t_3}$ are independent for any $0 \leq t_1 < t_2 \leq t_3 < t_4$.

Remark 3.5. The independence of the increments of a Bm could be stated in terms of the filtration: for each $h > 0$, then $B_{t+h} - B_t$ is independent on $\mathcal{F}_t$.

For any $\delta t \geq 0$, the Bm increment process $Z_t = B_{t+\delta t} - B_t$, $t \geq 0$, form a stationary Gaussian process with variance $E(Z_t^2) = \sigma^2 \delta t$. Therefore, it provides an example of an **iid** (idependent identically distributed) process, also said a *purely random process*.

Another interesting Brownian motion property is entailed in the following proposition.

Prop. 3.3. *The Brownian motion $B(t)$ is an $\mathcal{F}_t$-martingale.*

Proof: for any $t \geq 0$, from the Hölder inequality[3], one has:

$$E(|B_t|)^2 \leq E(B_t^2) = \sigma^2 t,$$

then $B_t \in L^1(\Omega, P)$. Moreover, for any $0 \leq s < t$:

$$E(B_t \mid \mathcal{F}_s) = E(B_t - B_s \mid \mathcal{F}_s) + E(B_s \mid \mathcal{F}_s) = B_s, \tag{3.9}$$

because $B_t - B_s$ is independent on $\mathcal{F}_s$ and B_s is $\mathcal{F}_s$-measurable. □

2) We sometimes refer to Brownian motion as *Wiener process*, reserving the term Brownian motion to the physical phenomenon of the random movements of particles suspended in a fluid.

3) Given a probability space $(\Omega, \mathcal{F}, P)$, let $p, q \geq 1$ such that $1/p + 1/q = 1$, then Hölder inequality reads:

$$E(|XY|) \leq (E|X|^p)^{1/p} (E|X|^q)^{1/q}, \quad X \in L^p(\Omega, P), \; Y \in L^q(\Omega, P).$$

Example 3.3 (Ornstein-Ulhenbeck process)**.** The *Ornstein-Ulhenbeck process* with parameter $\gamma > 0$ is a stationary Gaussian process $\{X(t)\}_{t\geq 0}$ such that $X(0)$ is standard normally distributed, i.e. $X(0) \sim N(0,1)$, has zero mean and with covariances $E(X(s)X(t)) = e^{-\gamma|t-s|}$, for any $t, s \geq 0$.

Remark 3.6. The Ornstein-Ulhenbeck process, introduced in the example above, is the stationary process related to Brownian motion through the transformation defined by (eq. 3.7). Indeed, suppose $B(t)$ is a Brownian motion with $E(B(1)^2) = 1$, thus, for any $\gamma > 0$, the stationary process

$$X(t) = e^{-\gamma t} B(e^{2\gamma t}), \quad t \geq 0, \tag{3.10}$$

is an Ornstein-Ulhenbeck process with parameter γ.

3.5 Markov processes

We turn to a discussion of the particular class of normal densities occurring in *Markov processes*. Without loss of generality we consider only densities centered at the origin. Consider for first a Gaussian discrete time real stochastic process $\{X_k, \ k > 0\}$, defined on a certain probability space, such that:

$$E(X_k^2) = \sigma_k^2, \quad E(X_i X_j) = \sigma_i \sigma_j \rho_{ik}, \tag{3.11}$$

where $\rho_{jk} = \gamma_{jk}/\sigma_i\sigma_j$ are the *correlation coefficients*. We have the following definition:

Def. 3.10 (Normal Markov density)**.** The n-dimensional normal density of $\{X_1, \ldots, X_n\}$ is Markovian if, for $k \leq n$, the conditional density of X_k for given $X_1, \ldots X_{k-1}$ is identical with the conditional density of X_k for given X_{k-1}.

Remark 3.7. We remember that, if the couple (X, Y) of real r.v.s, defined on the same probability space, has continuous density $f_{(X,Y)}(x, y)$, then for any $h > 0$

$$Pr\{Y \leq y |\, x < X \leq x + h\} = \frac{\int_x^{x+h} \int_{-\infty}^{y} f_{(X,Y)}(\xi, \nu) d\nu d\xi}{\int_x^{x+h} f_X(\xi) d\xi}.$$

Taking $h \to 0$ one has:

$$Pr\{Y \leq y | X = x\} = \frac{1}{f_X(x)} \int_{-\infty}^{y} f_{(X,Y)}(x, \nu) d\nu. \tag{3.12}$$

Therefore, one can define the conditional density of Y given that $X = x$ being:

$$f_{(X,Y)}(y|x) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}. \tag{3.13}$$

Roughly speaking, if we know X_{k-1} (the "present") then the additional knowledge of the "past" $X_1, \ldots, X_{k-2}$ does not contribute any relevant information about the "future", that is, about X_j, with $j \geq k$.

Remark 3.8. As usual in similar situations, we apply the term Markovian interchangeably to the process X and its finite-dimensional densities.

Theorem 3.2. *For a Gaussian sequence $\{X_k, \ k > 0\}$ to be Markovian each of the following condition is necessary and sufficient.*

- *For any $k \leq n$*
$$E(X_k|X_1,\dots,X_{k-1}) = E(X_k|X_{k-1}). \tag{3.14}$$
- *For $j \leq l \leq k \leq n$*
$$\rho_{jk} = \rho_{jl}\rho_{lk}. \tag{3.15}$$

Remark 3.9. For (eq. 3.15) to hold the Gaussian condition is necessary. Moreover, it suffices that
$$\rho_{jk} = \rho_{j,k-1}\rho_{k-1,k}, \quad j \leq k. \tag{3.16}$$

Example 3.4 (Independent increment processes)**.** Let X a discrete time Gaussian stochastic process with zero mean and independent increments. Namely, for any $k > 0$, $E(X_k) = 0$ and the increment $X_k - X_j$, $j < k$, is independent of $\{X_1,\dots,X_j\}$. This implies,
$$E(X_j(X_k - X_j)) = 0,$$
that is
$$\rho_{kj} = \sigma_j/\sigma_k, \quad j < k. \tag{3.17}$$
The above equation implies (eq. 3.15). Therefore, a normal process with independent increments is automatically Markovian. Its structure is rather trite: X_k is the sum of the k mutually independent normal random variables $\{X_1, X_2 - X_1,\dots,X_k - X_{k-1}\}$.

Consider now a Gaussian stochastic process $X(t)$, $t \in I \subseteq \mathbb{R}$.

Def. 3.11 (Markovian process)**.** We say that X is Markovian if any finite collection $\{X(t_1),\dots,X(t_n)\}$, $t_i \in I$, is Markovian.

Let the process X be Gaussian and Markovian. In view of (eq. 3.14) we have the *Markov property*:
$$E(X(t+h)|\{X(s),\ s \leq t\}) = E(X(t+h)|X(t)), \tag{3.18}$$
for any $h > 0$. Moreover, by (eq. 3.15), one has:
$$\rho(s,t) = \rho(s,h)\rho(h,t), \quad s \leq h \leq t, \tag{3.19}$$
with $\rho(s,t) = E(X(s)X(t))/\sigma(s)\sigma(t)$.

Remark 3.10. A real stochastic process $\{X(t)\}_{t\in I}$, not necessarily Gaussian, is a Markov process if it satisfies the Markov property (eq. 3.18). Let X be defined in a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. For instance, suppose that $\mathcal{F}_t$ is the natural filtration of X. If we indicate with
$$E^x(X(t)) = E\left(X(t)|X(0) = x\right), \quad x \in \mathbb{R},$$
then the Markov property (eq. 3.18) reads:
$$E(X(t+h)|\mathcal{F}_t) = E^{X(t)}(X(h)), \quad h > 0. \tag{3.20}$$

Let X be a Markov process and let $f_X(x,t)$ denote the probability density function of $X(t)$. Furthermore, we indicate with $f_X(x_2,t_2|x_1,t_1)$, $t_1 < t_2$, the value at x_2 of the conditional density of $X(t_2)$ given that $X(t_1) = x_1$, namely the *transition density* from $X(t_1)$ to $X(t_2)$.

Remark 3.11. In the case of Gaussian processes, $f_X(x,t)$ is the normal density with zero expectation and variance σ_t^2, and by (eq. 3.13), the transition density is:
$$f_X(y,s|x,t) = \frac{1}{\sqrt{2\pi(1-\rho(t,s)^2)\sigma_s^2}} \exp\left(-\frac{(y - \sigma_t^{-1}\rho(t,s)\sigma_s x)^2}{2(1-\rho(s,t)^2)\sigma_s^2}\right). \tag{3.21}$$

The joint density of $\{X(t_1), X(t_2)\}$ is given by (see eq. 3.13)

$$f_X(x_1, x_2, t_1, t_2) = f_X(x_1, t_1) f_X(x_2, t_2 | x_1, t_1).$$

Moreover, the joint density of $\{X(t_1), X(t_2), X(t_3)\}$ is then:

$$f_X(x_1, x_2, x_3, t_1, t_2, t_3) = f_X(x_1, t_1) f_X(x_2, t_2 | x_1, t_1) f_X(x_3, t_3 | x_1, t_1, x_2, t_2).$$

But, in view of the Markovian property of X, the latter reduces to:

$$f_X(x_1, x_2, x_3, t_1, t_2, t_3) = f_X(x_3, t_3 | x_2, t_2) f_X(x_2, t_2 | x_1, t_1) f_X(x_1, t_1). \tag{3.22}$$

Namely, we have the following:

Prop. 3.4. *If X is a Markovian process, then all the finite dimensional probability density functions are of the form of (eq. 3.22).*

Integrating (eq. 3.22) with respect to x_2, we easily find the following corollary:

Cor. 3.1 (Chapman-Kolmogorov equation)**.** *If X is a Markovian process, then the transition probability $f_X(x_3, t_3 | x_1, t_1)$ is given by*

$$f_X(x_3, t_3 | x_1, t_1) = \int f_X(x_3, t_3 | x_2, t_2) f_X(x_2, t_2 | x_1, t_1) dx_2, \quad t_1 < t_2 < t_3. \tag{3.23}$$

Very roughly, it expresses that a transition from x_1 at time t_1 to x_3 at time t_3 takes place via an arbitrary intermediate position x_2, the transition from x_2 to x_3 being independent of the past.

Example 3.5. Because the Brownian motion $B(t)$, $t \geq 0$, is a Gaussian process with independent increments, then it is a Markovian process. Moreover, the transition densities from (x, s) to (y, t), with $0 \leq s < t$, are normal densities with mean x and variance $\sigma^2(t - s)$ (see eq. 3.21). Namely,

$$f_B(y, t | x, s) = \frac{1}{\sqrt{2\pi(t-s)\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2(t-s)} \right). \tag{3.24}$$

They depend only on $(y - x)/(t - s)$ and the Chapman-Kolmogorov equation reduces to a convolution.

A Markov process X which have transition density $f_X(y, t | x, s)$, $0 \leq s < t$, that depends only on the time difference $t - s$ rather than on the specific values of t and s, is said *homogeneous*. The Brownian motion is an example of homogeneous Markov process.

Example 3.6. Another example of homogeneous Markov process is provided by the Ornstein-Ulhenbeck process (Example 3.3), which has transition density:

$$f_X(y, t | x, s) = \frac{1}{\sqrt{2\pi(1 - e^{-2\gamma(t-s)})\sigma^2}} \exp\left(-\frac{(y - xe^{-\gamma(t-s)})^2}{2\sigma^2(1 - e^{-2\gamma(t-s)})} \right). \tag{3.25}$$

3.5.1 Stochastic diffusion processes

The transition density (eq. 3.24) of the Brownian motion is obviously a smooth function of its variables for $t > s$. Evaluating its partial derivatives explicitly, we find that they satisfy the partial differential equations:

$$\partial_t f - \tfrac{\sigma^2}{2} \partial_{yy} f = 0, \quad (s, x) \text{ fixed}, \tag{3.26}$$

and

$$\partial_s f + \tfrac{\sigma^2}{2}\partial_{xx} f = 0, \quad (t, y) \text{ fixed.} \tag{3.27}$$

The first equation (eq. 3.26) is called a *forward equation* because differentiation with respect to the final time is involved. While, for obvious reasons, the second equation (eq. 3.27) is termed *backward equation*. The transition density of the Ornstein-Ulhenbeck process satisfies related, but more complicate partial differential equations.

Remark 3.12. Observe that (eq. 3.26) is an example of a *heat equation* which describes the temperature variation as heat **diffuses** through a physical medium.

It is not surprising that the Brownian motion serves as prototype of a very rich and useful class of Markov processes termed *diffusion processes*.

Def. 3.12 (Diffusion processes)**.** A Markov process X, defined on a certain probability space, with transition density $f_X(y, t|x, s)$ is called a (linear or one dimensional) *diffusion process* if the following three limits exists for all $\epsilon > 0$, $t > s \geq 0$ and $x \in \mathbb{R}$:

$$\lim_{t \to s} \frac{1}{t - s} \int_{|y-x|>\epsilon} f_X(y, t|x, s) dy = 0, \tag{3.28}$$

$$\lim_{t \to s} \frac{1}{t - s} \int_{|y-x|<\epsilon} (y - x) f_X(y, t|x, s) dy = a(s, x), \tag{3.29}$$

$$\lim_{t \to s} \frac{1}{t - s} \int_{|y-x|<\epsilon} (y - x)^2 f_X(y, t|x, s) dy = \sigma^2(s, x), \tag{3.30}$$

where a and σ are well defined functions.

The first condition (eq. 3.28) prevents a diffusion process from having instantaneous jumps. The quantity $a(s, t)$ is called *drift* coefficient and $\sigma(s, x)$ is called *diffusion coefficient*.

Remark 3.13. The drift $a(s, x)$ is the instantaneous rate of change in the mean of the process given that $X(s) = x$. Indeed, by (eq. 3.29):

$$a(s, x) = \lim_{t \to s} \frac{1}{t - s} E\left(X(t) - X(s)|X(s) = x\right), \quad t \geq s. \tag{3.31}$$

Furthermore, $\sigma^2(s, t)$ denotes the instantaneous rate of change of the squared fluctuations of the process given that $X(s) = x$. In fact, by (eq. 3.30) one has:

$$\sigma^2(s, x) = \lim_{t \to s} \frac{1}{t - s} E\left[(X(t) - X(s))^2|X(s) = x\right], \quad t \geq s. \tag{3.32}$$

Example 3.7. In the case of Brownian motion, one has $a = 0$ and σ constant. In the case of a $\gamma = 1$ Ornstein-Ulhenbeck process, one can show that $a(s, x) = -x$ and $\sigma(s, x) = \sqrt{2}$.

When the drift a and the diffusion coefficient σ of a diffusion process X are quite well behaved, then its transition density $f_X(y, t|x, s) = u(y, t, x, s)$ satisfies partial differential equations. These are the *Kolmogorov forward equation*

$$\partial_t u + \partial_y[a(t, y)u] - \tfrac{1}{2}\partial_{yy}[\sigma^2(t, y)u] = 0, \quad (s, x) \text{ fixed,} \tag{3.33}$$

and the *Kolmogorov backward equation*:

$$\partial_s u + a(s, x)\partial_x u - \tfrac{1}{2}\sigma^2(s, x)\partial_{xx} u = 0, \quad (t, y) \text{ fixed.} \tag{3.34}$$

The first equation gives the forward evolution with respect to the final state (y, t), while the latter gives the backward evolution with respect to the initial state (x, s).

Remark 3.14 (Fokker-Planck equation)**.** The forward Kolmogorov equation is also called *Fokker-Planck* equation.

Both the equations (eq. 3.33 and 3.34) follow from the Chapman-Kolmogorov equation (eq. 3.23).

Example 3.8. A class of Markovian diffusion processes is provided by the so called *Ito diffusions*, which are strong solutions of stochastic differential equations (see Section 4.10).

3.6 H-sssi processes

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space. A stochastic process $X = \{X_t\}_{t\in I}$, $\mathcal{F}_t$-adapted, which is H-**ss** with stationary increments, is said H-**sssi** process with exponent H.

Example 3.9. We have already encountered an example of H-**sssi** process when we have introduced Brownian motion in (Section 3.4).

In the following we always suppose that $E(X_t^2) < \infty$, $t \in I$. Thus, we always consider finite variance H-**sssi** processes. H-**sssi** processes with finite variance possess many interesting properties. Then, let $X = \{X_t\}_{t\in I}$ be an H-**sssi** process with finite variance, the following properties hold:

1. $X_0 = 0$ almost surely.

   Indeed, for any $a > 0$, one has:
   $$X(0) = X(a0) \stackrel{d}{=} a^H X(0).$$
   Then, it follows that $\sigma(X(0))$ is degenere, that is it is made up of only by null-events or measure one events with respect to P. From (Prop. 2.6) follows that $X(0)$ is trivial, i.e. constant almost everywhere. But, if we set $X(0) = c$, then:
   $$E(X(0)) = c = E(X(a0)) = a^H c \Leftrightarrow c = 0.$$

2. If $H \neq 1$, then for any $t \geq 0$, $E(X_t) = 0$.

   Indeed: $E(X(2t)) = 2^H E(X(t))$. On the other hand, from the stationary increments property and from property 1:
   $$2^H E(X(t)) = E(X(2t)) = E(X(2t) - X(t)) + E(X(t)) = 2E(X(t)) \Leftrightarrow E(X(t)) = 0.$$

3. One has:
   $$X(-t) \stackrel{d}{=} -X(t);$$
   it follows from the first property and the stationarity of the incremenets:
   $$X(-t) \stackrel{a.s.}{=} X(-t) - X(0) \stackrel{d}{=} X(0) - X(t) \stackrel{a.s.}{=} -X(t).$$
   The above property allows us to extend the definition of an H-**sssi** process to the whole real line: $\{X_t\}_{t\in\mathbb{R}}$.

4. Let $\sigma^2 = E(X_1^2)$. Then,
   $$E(X_t^2) = |t|^{2H} \sigma^2. \tag{3.35}$$
   Indeed, from the third property and the self-similarity:
   $$EX(t)^2 = EX^2(|t| \operatorname{sign}(t)) = |t|^{2H} EX^2(\operatorname{sign}(t)) = |t|^{2H} E(X_1^2) = |t|^{2H} \sigma^2.$$

Notation 3.2. Sometimes, we refer to the H-**sssi** process $\{X_t\}_{t\in I}$ with the word *standard* if $\sigma^2 = 1$ or, depending on the context, if $\sigma^2 = 2$.

5. The autocovariance function of an H-**sssi** process X, with $E(X_1^2) = \sigma^2$, turns out to be:

$$\gamma_{s,t}^H = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{3.36}$$

It follows from the fourth property and the stationarity of the increments

$$E(X_s X_t) = \frac{1}{2}\left(EX_s^2 + EX_t^2 - E(X_t - X_s)^2\right).$$

6. If $X = \{X_t\}_{\in I}$ is an H-**sssi** process, then one must have $H \leq 1$.

The constraint of the scaling exponent follows directly from the stationarity of the increments:

$$2^H E\,|X_1| = E\,|X_2| = E\,|X_2 - X_1 + X_1| \leq E\,|X_2 - X_1| + E\,|X_1| = 2E\,|X_1|,$$

therefore, $2^H \leq 2 \Leftrightarrow H \leq 1$.

Remark 3.15. The case $H = 1$ corresponds a.s. to $X_t = tX_1$. Indeed, on the $L^2(\Omega, P)$ norm:

$$E(X_t - tX_1)^2 = E(X_t^2 + t^2X_1^2 - 2tX_tX_1) = \sigma^2(2t^2 - 2t^2) = 0.$$

Remark 3.16. The Brownian motion is a Gaussian H-**sssi** process with $H = 1/2$ (Section 3.4), so that its properties follow from those of a generic H-**sssi** process. In particular here we focus on the independence of the increments, which is due to (eq. 3.36) together with the definition of Gaussian process. We observe that for a generic H-**sssi** process this is not longer true. Indeed, let X_t be an H-**sssi** process with $H \leq 1$ and take:

$$X_{t_1} - X_{t_2}, \quad X_{t_3} - X_{t_4}, \quad 0 \leq t_1 < t_2 \leq t_3 < t_4,$$

then, using (eq. 3.36), one has:

$$E(X_{t_1} - X_{t_2})(X_{t_3} - X_{t_4}) = \gamma_{t_1,t_3}^H - \gamma_{t_1,t_4}^H - \gamma_{t_2,t_3}^H + \gamma_{t_2,t_4}^H =$$

$$\frac{\sigma^2}{2}\left(|t_1 - t_4|^{2H} + |t_2 - t_3|^{2H} - |t_1 - t_3|^{2H} - |t_2 - t_4|^{2H}\right), \tag{3.37}$$

which is always zero if $H = \frac{1}{2}$, but is non-zero if $H \neq \frac{1}{2}$.

The above remark states that a generic adapted H-**sssi** process X is not a martingale provided that $H \neq 1/2$. In fact, (eq. 3.9) is in general not verified.

3.6.1 On the existence of H-sssi processes

We want now briefly discuss the question related to the existence of an H-**sssi** process. We remember that:

Prop. 3.5. *Let ξ be a symmetric α-stable r.v., with scaling exponent σ, defined on the probability space $(\Omega, \mathcal{F}, P)$. Then, its characteristic function is:*

$$E(e^{it\xi}) = e^{-\sigma^\alpha |t|^\alpha}. \tag{3.38}$$

Remark 3.17 (Lèvy distribution)**.** More in general, we remember that a *Lévy skew α-stable distribution* or just *stable distribution*, is actually a family of probability distributions, characterized by four parameters $\alpha, \mu, \sigma \in \mathbb{R}$ and $-1 < \theta < 1$, such that

$$L(x; \alpha, \theta, \mu, \sigma) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) e^{-itx}\, dt,$$

where $\varphi(t)$ is given by:

$$\varphi(t) = \exp\left[it\mu - \sigma|t|^{\alpha}(1 - i\theta\, \mathrm{sign}(t)\Phi)\right], \tag{3.39}$$

with $\Phi = \tan(\pi\alpha/2)$, $\alpha \neq 1$ and $\Phi = -(2/\pi)\log|t|$ if $\alpha = 1$. The parameter μ is a shifting parameter, while θ is a measure of asymmetry. In fact, $\theta = 0$ yields a distribution symmetric about μ. Moreover, σ is a scale factor which is a measure of the width of the distribution and α is the exponent or index of the distribution and specifies the asymptotic behavior of the distribution for $\alpha < 2$.

Remark 3.18 (Strictly stable laws)**.** The stable distribution has the important property of stability. Namely, if a number of independent identically distributed random variables have a stable distribution, then a linear combination of these variables will have the same distribution, except for possibly different shift and scale parameters. To be more precise, If X_1 and X_2 are distributed according to a stable distribution $L(x; \alpha, \theta, \mu, \sigma)$, and if $Y = AX_1 + BX_2 + C$, $A, B, C \in \mathbb{R}$, then there exist $D, E \in \mathbb{R}$ such that $DY + E$ is distributed according to a stable distribution $L(DY + E; \alpha, \theta, \mu, \sigma)$ or, equivalently, Y is distributed according to a stable distribution $L(Y; \alpha, \theta, (\mu - E)/D, \sigma/D)$. If $E = 0$ for all A, B and C then Y is said to have a *strictly stable* distribution.

Since the normal distribution, the Cauchy distribution, and the Lévy distribution all have the above property, it follows that they are special cases of the stable distribution.

We have the following result:

Prop. 3.6. *Let* $0 < H \leq 1$. *The function:*

$$C_H(s,t) = |t|^{2H} + |s|^{2H} - |t-s|^{2H}, \quad s, t \in \mathbb{R},$$

is positive semi-defined.

Proof: we have to show that for any collection of rela numbers $\{t_i\}_{i=1,2,\ldots,n}$ and $\{u_i\}_{i=1,2,\ldots,n}$, one has:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} C_H(t_i, t_j) u_i u_j \geq 0.$$

Define $t_0 = 0$ and $u_0 = -\sum_{i=1}^{n} u_i$, such that:

$$\sum_{i=0}^{n} u_i = 0.$$

Therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |t_i|^{2H} u_i u_j = \sum_{i=1}^{n} |t_i|^{2H} u_i \sum_{j=1}^{n} u_j = -\sum_{i=1}^{n} |t_i|^{2H} u_i u_0 = -\sum_{i=1}^{n} |t_i - t_0|^{2H} u_i u_0.$$

Then,

$$\sum_{i,j=1}^{n} C_H(t_i, t_j) u_i u_j = \sum_{i,j=1}^{n} (|t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H}) u_i u_j = -\sum_{i=0}^{n}\sum_{j=0}^{n} |t_i - t_j|^{2H} u_i u_j.$$

We observe that, for any $c > 0$:

$$\sum_{i=0}^{n}\sum_{j=0}^{n} e^{-c|t_i-t_j|^{2H}} u_i u_j = \sum_{i=0}^{n}\sum_{j=0}^{n} (e^{-c|t_i-t_j|^{2H}} - 1) u_i u_j = -c \sum_{i=0}^{n}\sum_{j=0}^{n} |\, t_i - t_j \,|^{2H}\, u_i u_j + o(c),$$

as $c \to 0$. The proof reduces to show that:

$$\sum_{i=0}^{n}\sum_{j=0}^{n} e^{-c|t_i-t_j|^{2H}} u_i u_j \geq 0.$$

Using (Prop. 3.5), there exists a symmetric $2H$-stable r.v. ξ, with scaling exponent $\sigma = c^{1/2H}$, such that:

$$E(e^{it\xi}) = e^{-c|t|^{2H}}.$$

Then,

$$\sum_{i=0}^{n}\sum_{j=0}^{n} e^{-c|t_i-t_j|^{2H}} u_i u_j = \sum_{i=0}^{n}\sum_{j=0}^{n} E(e^{i(t_i-t_j)\xi} u_i u_j)$$

$$= E\left(\sum_{i=0}^{n} u_i e^{it_i\xi} \sum_{j=0}^{n} u_j e^{-it_j\xi}\right) = E\left(\Big|\sum_{i=0}^{n} u_i e^{it_i\xi}\Big|^2\right) \geq 0,$$

which ends the proof. □

From the proposition above follows that it is always possible to define a Gaussian H-**sssi** process, with zero mean, covariance matrix $\gamma^H(s,t)$ and with $\sigma^2 = E(X_1^2)$. In fact, a Gaussian process is entirely characterized by its first and second moments (eq. 3.3).

Remark 3.19. Given $H \leq 1$, all the Gaussian H-**sssi** processes differ by a multiplicative constant.

3.7 Fractional Brownian motion

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a probability space with filtration and let $I \subseteq \mathbb{R}$ [4]

Def. 3.13 (Fractional Brownian motion)**.** Let $B_H = \{B_H(t)\}_{t\in I}$, be a Gaussian H-**sssi** stochastic process, with $0 < H < 1$, defined on $\Omega_{\mathcal{F}}$ adapted to $\mathcal{F}_t$, with $\sigma^2 = E(B_H(1)^2)$. Such a process B_H is called *fractional Brownian motion* (shortly fBm) of order H.

Remark 3.20. Let $\{B_H(t)\}_{t\in I}$, $0 < H < 1$, be a fractional Brownian motion, then:

- if $H = 1/2$, $\{B_{1/2}(t)\}_{t\in I}$ is the ordinary Brownian motion (Def. 3.4), with auto-covariance function:
$$\gamma(s,t) = \sigma^2 \min(|\,s\,|, |\,t\,|), \quad \text{if } \frac{|\,s\,|}{s} = \frac{|\,t\,|}{t}, \quad \gamma(s,t) = 0, \quad \text{otherwise.} \tag{3.40}$$
- the case $H = 1$ won't be considered because it corresponds to the trivial case of a line with random slope (Remark 3.15):
$$B_1(t) = tB_1(1).$$

In the following we implicitly assume that the filtration is the natural filtration of the fBm.

Remark 3.21. If one has $I = \mathbb{R}$ one speaks about *two-sided* fBm. On the other hands, if $I = \mathbb{R}_+$ one speaks about *one-sided* fBm.

4) In this section we always take $I = \mathbb{R}_+$. However, it is straightforward to generalize to negative values using the property 3. of H-**sssi** processes.

3.7.1 Fractional Brownian motion characterization

The following proposition characterizes the fBm and gives a useful criterium which allows to recognize whether a given process is a fractional Brownian motion.

Prop. 3.7 (fBm characterization). *Let $X = \{X_t\}_{t\geq 0}$ be a stochastic process, defined on the probability space $(\Omega, \mathcal{F}, P)$, such that:*

i) $P(X_0 = 0) = 1$.

ii) *X is a zero-mean Gaussian process such that, for any $t \geq 0$, $E(X_t^2) = \sigma^2 t^\alpha$ for some $\sigma > 0$ and $0 < \alpha < 2$.*

iii) *X is a* **si***-process.*

Then, $\{X_t\}_{t\geq 0}$ is a (one-sided) fractional Brownian motion of order $H = \alpha/2$.

Proof: since X is a zero-mean Gaussian process, its finite-dimensional distributions are completely characterized by its autocovariance function. Given that, for any $t > 0$:

$$E(X_t^2) = \sigma^2 \mid t \mid^\alpha$$

and X has stationary increments, it follows that the autocovariance function is given by (eq. 3.36), which is the autocovriance of a fBm with $H = \alpha/2$. □

The corollary below follows immediately:

Cor. 3.2. *Let $X = \{X_t\}_{t\geq 0}$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $0 < H < 1$ and $\sigma^2 = E(X_1^2)$. The following statements are equivalent:*

1. *X is an H-***sssi** *Gaussian process.*
2. *X is a (one-sided) fractional Brownian motion with scaling exponent H.*
3. *X is Gaussian with zero mean and autocovariance function given by (eq. 3.36).*

Remark 3.22 (Non-Markovian property)**.** Fractional Brownian motion is **non-Markovian** provided that $H \neq 1/2$. In fact, if $H \neq 1/2$, (eq. 3.19) is not satisfied by (eq. 3.36).

3.7.2 Fractional Brownian motion trajectories

We focus now on the characterization of the fBm trajectories. Let $B = \{B_H(t)\}_{t\geq 0}$ be a fBm with $0 < H < 1$, defined on the probability space $(\Omega, \mathcal{F}, P)$. Therefore,

Prop. 3.8. *for any $\omega \in \Omega$, the function $t \to B_H(t, \omega)$ is continuous.*

Proof: in order to prove the above proposition, we make use of the following theorem, which guarantees that a stochastic process that satisfies certain constrains on the increment moments is (path)- continuous or, more precisely, has a *continuous version*.

Def. 3.14 (Version)**.** Let $X = \{X_t\}_{t\in I}$ be a stochastic process on the space $(\Omega, \mathcal{F}, P)$. A process $\widetilde{X}$, defined in the same probability space, is a *version* of the process X if, for any $t \in I$, $P(X_t = \widetilde{X}_t) = 1$.

Theorem 3.3 (Kolmogorov continuity theorem)**.** *A stochastic process $\{X_t\}_{t\in I}$ has a version with continuous trajectories if there exist: $p \geq 1$ and $\eta > 1$ and a constant c, such that, for any $t_1, t_2 \in I$:*

$$E \mid X_{t_2} - X_{t_1} \mid^p \leq c \mid t_2 - t_1 \mid^\eta . \tag{3.41}$$

Choose now $p > 1/H$ thus, for any $t_1, t_2 \geq 0$, using the self-similarity and the stationarity of the increments one has:

$$E\,|\,B_H(t_2) - B_H(t_1)\,|^p = E\,|\,B_H(1)\,|^p |\,t_2 - t_1\,|^{pH},$$

which, by using (Theorem 3.3), completes the proof. □

So that, Proposition 3.8 states that we can always find a version of the fBm which has continuous trajectories. We then suppose we always work with such a continuous version. The following proposition allows to better understand the fbm trajectories behavior.

Prop. 3.9. *For any $\omega \in \Omega$, the function $t \to B_H(t, \omega)$ is not-differentiable in the $L^2(\Omega, P)$-norm.*

Proof: indeed, for any $t_1, t_2 \geq 0$:

$$E\left(\frac{B_H(t_2) - B_H(t_1)}{t_2 - t_1}\right)^2 = \sigma^2\,|\,t_2 - t_1\,|^{2H-2} \to \infty,$$

as $t_1 \to t_2$. □

Depending on the qualitative behavior of the fBm trajectories, it is common the following fBm partitioning, which can be actually used to characterize any H-**sssi** process:

1. If $0 < H < 1/2$, the fBm is termed *anti-persistent*.
2. If $H = 1/2$, the fBm is termed *purely random*, or *chaotic*.
3. If $1/2 < H < 1$, the fBm is termed *persistent*.

This division is due to the behavior of the autocovariance function of the increment process. Indeed, as we are going to see in the next section (Section 3.7.3), in the first case the covariance of two consecutive increments is negative. Therefore, the fBm increments tend to have opposite sign. So that, the trajectories manifest an extremely alternating behavior (top panel in Figure 3.1). On the other hand, in the third case the covariance is always positive and one has a less "zigzaging" behavior of the paths (bottom panel in Figure 3.1). The case $H = 1/2$ corresponds to the Bm, which has independent identically distributed increments, i.e. a purely random increment process.

3.7.3 Discrete time fractional Gaussian noise

Let $(\Omega, \mathcal{F}, P)$ be a probability space and consider a fBm $B_H = \{B_H(t)\}_{t\in\mathbb{R}}$ defined on it. Consider a time grid $I_{\delta t} = \{t_k = k\delta t, \quad k \in \mathbb{Z}\}$, with $\delta t > 0$. Then, we can define a discrete time process $\{B_H(t_k)\}_{k\in\mathbb{Z}}$ defined as the restriction to $I_{\delta t}$ of the continuous time process $B_H(t)$. Thus, we indicate $B_H(k) \equiv B_H(t_k)$.

Def. 3.15 (Fractional Gaussian noise)**.** The discrete time increment process of B_H, namely the process $\{Z_t\}_{t\in\mathbb{Z}}$ such that

$$Z_k = \Delta B_H(k) = B_H(k+1) - B_H(k), \quad k \in \mathbb{Z},$$

is called (discrete time) *fractional Gaussian noise* (shortly fGn) of order H.

Fractional Gaussian noise properties

Consider for simplicity $\delta t = 1$, that is $t_k = k$ indeed. Then, let $\{Z_t\}_{t\in\mathbb{Z}}$ be a fGn. The following properties are straightforward:

1. Z_t is stationary.

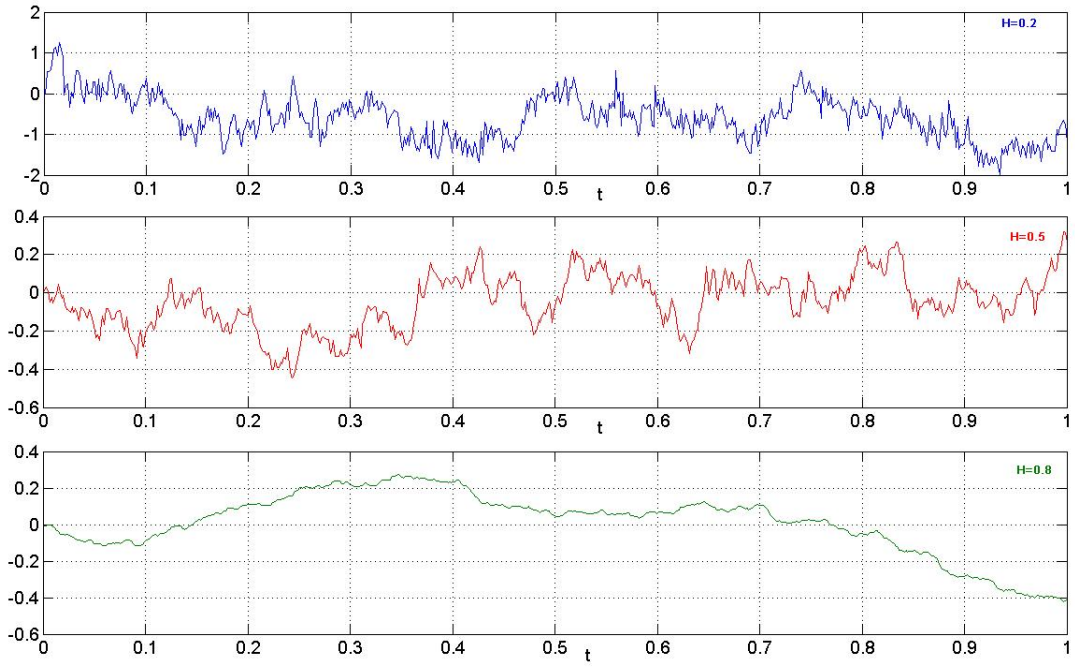

Figure 3.1: Fractional Brownian motion trajectories for $H = [0.2, 0.5, 0.8]$.

2. For any $k \geq 1$, then $EZ_k = 0$.
3. For any $k \geq 1$, then $EZ_k^2 = \sigma^2 = E(B_H(1))$.
4. the autocovariance function of the process $\{Z_t\}_{t\in\mathbb{Z}}$, is:

$$\gamma(k) = E(Z_t Z_{t+k}) = \frac{\sigma^2}{2}(|k+1|^{2H} - 2\,|k|^{2H} + |k-1|^{2H}) = \frac{\sigma^2}{2}\Delta^2\,|k|^{2H}, \tag{3.42}$$

where Δ^2 indicates the second order finite differencing operator, namely $\Delta^2 X_k = X_{k+1} - 2X_k + X_{k-1}$.

Properties of the autocovariance function

The following propositions characterize the fGn autocovariance function.

Prop. 3.10. *Let $k \neq 0$. Therefore,*

- $\gamma(k) = 0$, *if* $H = 1/2$ *(purely random region);*
- $\gamma(k) < 0$, *if* $0 < H < 1/2$ *(antipersistent region);*
- $\gamma(k) > 0$, *if* $1/2 < H < 1$ *(persistent region).*

Proof: the case $H = 1/2$ is trivial. Let $1/2 < H < 1$, the function $f(k) = k^{2H}$ is strictly convex, therefore, for any $k \geq 1$:

$$\frac{f(k+1) + f(k-1)}{2} = \frac{(k+1)^{2H} + (k-1)^{2H}}{2} > f(k) = k^{2H}.$$

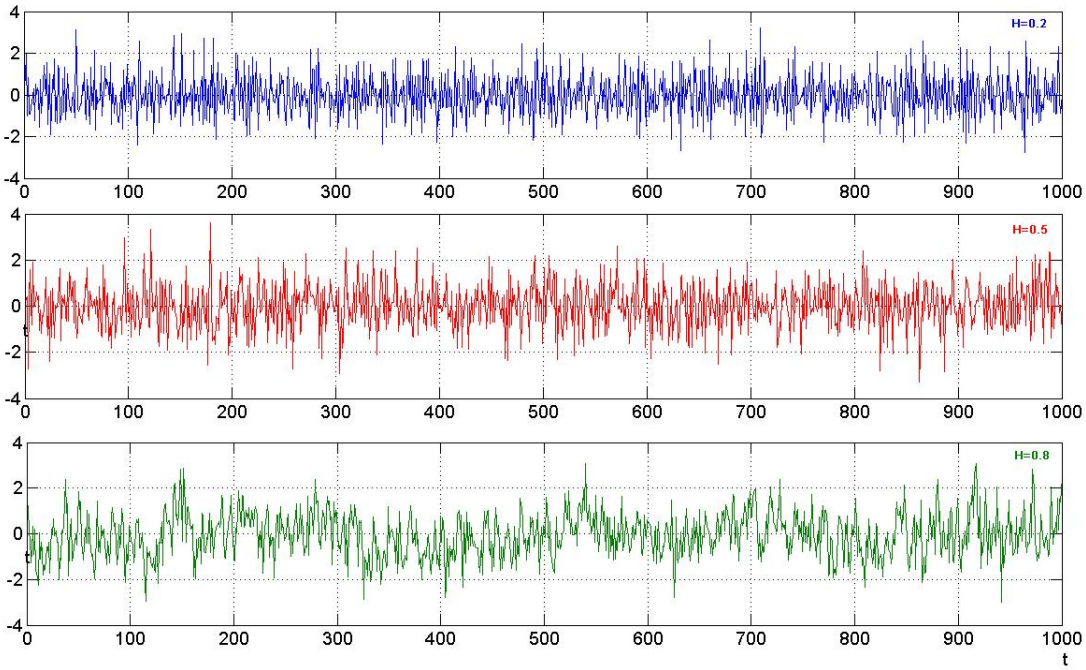

Figure 3.2: Fractional Gaussian noise for $H = [0.2, 0.5, 0.8]$ and $\delta t = 1$.

Then, $\gamma(k) > 0$. In the same way, in the case $0 < H < 1/2$ one finds $\gamma(k) < 0$. □

Prop. 3.11. *If $H \neq 1/2$, then:*

$$\gamma(k) \sim \sigma^2 H(2H-1)\,|k|^{2H-2}, \quad k \to \infty. \tag{3.43}$$

Proof: if $k \to \infty$, we have that:

$$\frac{\sigma^2}{2}\left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right) \simeq \frac{\sigma^2}{2}\frac{d^2}{dk^2}k^{2H} = \sigma^2 H(2H-1)k^{2H-2}. \ \square$$

Remark 3.23. We observe that if $H = 1/2$, the fractional Gaussian noise is a Gaussian **i.i.d** (independent identically distributed) process. Namely, a purely Gaussian random process, or (discrete time) *white-noise*.

Remark 3.24. We have defined the fGn simply as the discrete time increment process of fBm (Def. 3.15). However, the term fractional Gaussian noise, or Gaussian white noise, is also associated to the fractional Brownian motion time derivative

$$\dot{B}_H(t) = \lim_{\delta t \to 0} \frac{B_H(t+\delta t) - B_H(t)}{\delta t}, \quad t \geq 0, \tag{3.44}$$

or Brownian motion derivative $\dot{B}$. But, even if such a process could be approximated to any desired degree of accuracy by its discrete time counterpart, it cannot be defined by (eq. 3.44). Indeed, the sample paths are almost surely not differentiable anywhere (Prop. 3.9). We will come back later on this problem.

3.7.4 Aggregation of fBms

We have seen that fBm is a Gaussian H-**sssi** process with $0 < H < 1$. We can construct Gaussian self similar processes with Hurst parameter H greater than one simply taking linear averages of fBm. Moreover, the obtained process has a number of well defined time derivatives (vide infra). Let $\{B_H(t)\}_{t\geq 0}$, be a standard fractional Brownian motion; i.e. for each $t > 0$ we require $E(B_H(t)^2) = t^{2H}$. It is possible to define a process Y_H, such that:

$$Y_H(t) = \int_0^t B_H(s)ds, \quad t \geq 0. \tag{3.45}$$

Therefore, Y_H is a Gaussian process and, heuristically, it is self similar with Hurst parameter $H' = H + 1$. For instance, for any $a > 0$ and $t > 0$:

$$Y_H(at) = \int_0^{at} B_H(s)ds$$

(with the change of variables $s = as'$)

$$= a\int_0^t B_H(as')ds' \stackrel{d}{=} a^{H+1}\int_0^t B_H(s')ds'.$$

However, Y_H is no longer a stationary increment process. We want here briefly to study the covariance structure of the process Y_H.

Example 3.10. Let $H = 1/2$, in this case we have a standard Brownian motion and

$$Y(t) = \int_0^t B(s)ds.$$

In this case, it is very easy to evaluate the variance of the process. For instance, let us introduce the time grid

$$\mathcal{G}_{\delta t}^N = \{t_i = i\delta t, \quad i = 0, 1, \ldots, N-1\}, \quad \delta t = t/(N-1).$$

Thus, one could write

$$Y(t) = \lim_{N\to\infty} \sum_{i=0}^{N-1} B(t_i)\delta t.$$

Therefore,

$$E(Y(t)^2) = \lim_{N\to\infty} \sum_{i,j} \min(t_i, t_j)\delta t$$

and one is looking for the limit of:

$$\delta t^2 \sum_{j=0}^{N-1} \left(\sum_{i=0}^{j} t_i + \sum_{i=j+1}^{N-1} t_j \right) = \delta t^2 \sum_{j=0}^{N-1} \left(\frac{j(j+1)}{2}\delta t + (N-j-2)j\delta t \right)$$

$$= \frac{\delta t^3}{2} \sum_{j=0}^{N-1} \left(2Nj - j^2 - 3j \right) = \frac{\delta t^3}{2} \left(N^2(N-1) - \frac{(N-1)^3}{3} - \frac{(N-1)^2}{2} - \frac{N-1}{6} - 3\frac{N(N-1)}{2} \right).$$

Which, taking $N \to \infty$, becomes:

$$E(Y(t)^2) = \frac{t^3}{3}.$$

Remark 3.25. We could have used directly the definition:

$$E(Y(t)^2) = \int_0^t \int_0^t \min(u,s)duds = \int_0^t \left[\int_0^u sds + \int_u^t uds\right] du$$
$$= \int_0^t \left[\frac{u^2}{2} + u(t-u)\right] du = \frac{t^3}{3}.$$

In the general case $0 < H < 1$, we have:

1. $Y_H(0) = 0$ a.s.
2. $E_H(Y(t)) = 0$ for any $t \geq 0$,
3. $E(Y_H(t)^2) = \dfrac{t^{2H+2}}{2H+2}$ for any $t \geq 0$.
3. For any $t, s > 0$:

$$E(Y_H(t)Y_H(s)) = \frac{1}{2(2H+2)(2H+1)}\left[(2H+2)(ts^{2H+1} + st^{2H+1}) + |t-s|^{2H+2} - t^{2H+2} - s^{2H+2}\right]. \tag{3.46}$$

The first two are obvious, the third comes from the fourth. In order to show (eq. 3.46), we have to evaluate:

$$\frac{1}{2}\int_0^t \int_0^s \left(u^{2H} + v^{2H} - |u-v|^{2H}\right) dudv,$$

which easily gives (eq. 3.46). Summarizing: suppose to have a standard fractional Brownian motion B_H with Hurst parameter $0 < H < 1$, then the aggregating process:

$$Y_H(t) = \int_0^t B_H(s)ds$$

is Gaussian with zero mean and autocovariance given by (eq. 3.46). This means that Y_H is self-similar with parameter $H' = H + 1$ and variance $t^{2H+2}/(2H+2)$.

Example 3.11. We define the *time average* process

$$A_H(t) = \frac{1}{t}Y_H(t) = \frac{1}{t}\int_0^t B_H(s)ds, \quad t > 0. \tag{3.47}$$

Then, A_H is a Gaussian self similar process of order $0 < H < 1$, and autocovariance matrix:

$$E(A_H(t)A_H(s)) = \frac{1}{2ts(2H+2)(2H+1)}\left[(2H+2)(ts^{2H+1} + st^{2H+1}) + |t-s|^{2H+2} - t^{2H+2} - s^{2H+2}\right], \tag{3.48}$$

with $t, s > 0$.

Remark 3.26. The trajectories of Y_H are clearly continuous and also differentiable. Heuristically, we have:

$$\dot{Y}_H(t) = B_H(t), \quad t \geq 0. \tag{3.49}$$

Indeed, for any real h, the process

$$\Delta_h Y_H(t) = \frac{Y(t+h) - Y(t)}{h}, \quad t \geq 0,$$

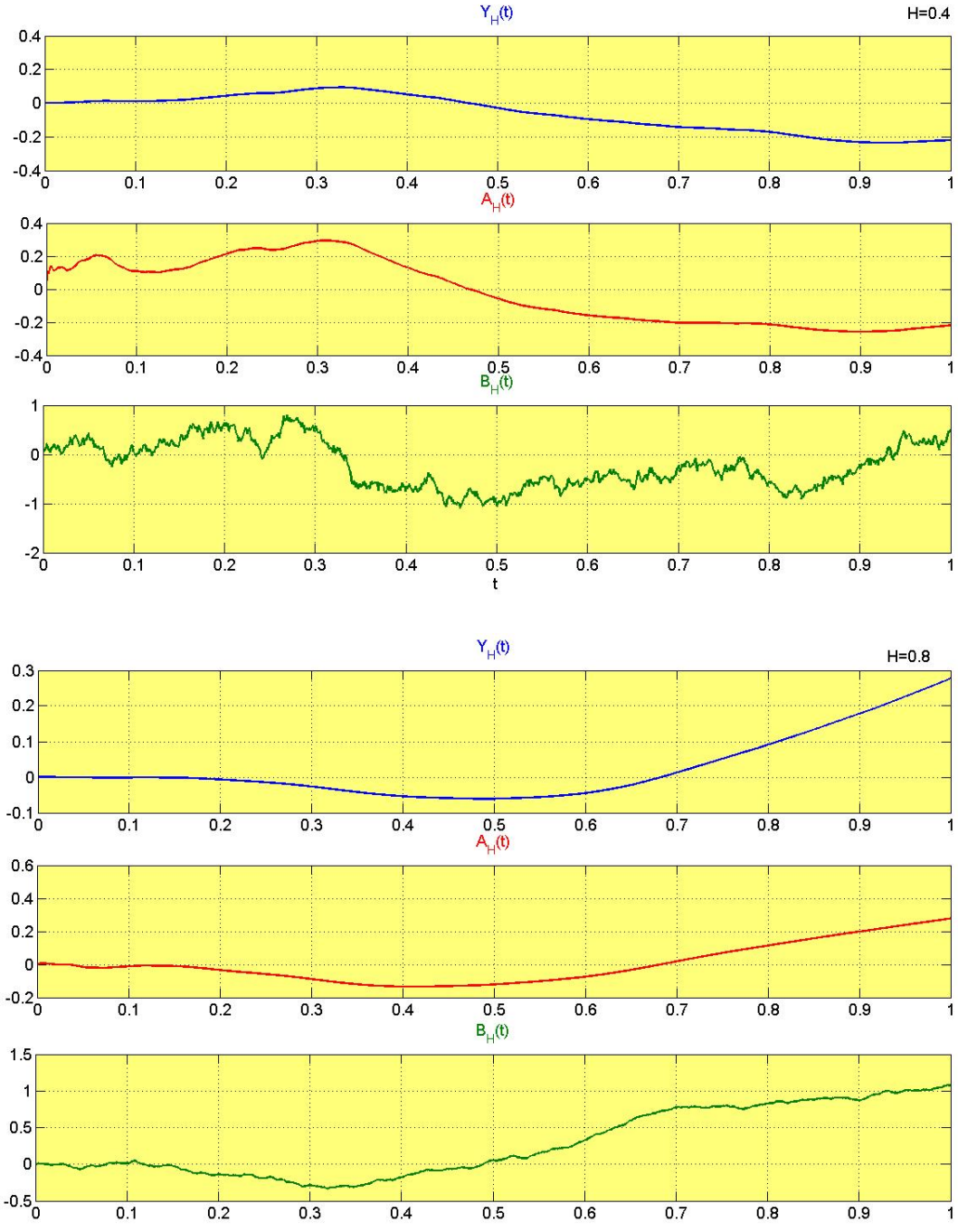

Figure 3.3: Paths of $Y_H(t)$, the time average $A_H(t)$ and the fBm $B_H(t)$ for $H = [0.4, 0.8]$.

is Gaussian with zero mean and autocovariance function (see eq. 3.46):

$$E(\Delta_h Y_H(t)\Delta_h Y_H(s))$$

$$= \frac{1}{2(2H+2)(2H+1)} \left\{ (2H+2) \frac{(t+h)^{2H+1} - t^{2H+1}}{h} + \frac{(s+h)^{2H+1} - s^{2H+1}}{h} \right\}$$
$$- \frac{1}{2(2H+2)(2H+1)} \left\{ \frac{|t-s+h|^{2H+2} - 2|t-s|^{2H+2} + |t-s-h|^{2H+2}}{h^2} \right\}.$$

Therefore, as h goes to zero, one finds (eq. 3.36) with $\sigma^2 = 1$. Thus, by Corollary 3.2, we must have (eq. 3.49).

Example 3.12. The time average process $A_H(t)$, $t \geq 0$. It is "differentiable" and we have:

$$\dot{A}_H(t) = \frac{B_H(t) - A_H(t)}{t}, \quad t \geq 0, \tag{3.50}$$

which could be seen as the "instantaneous rate of change" of $B_H(t)$ from the average up to time t, $A_H(t)$. The variance is:

$$E(\dot{A}_H(t)^2) = \frac{t^{2H-2}}{2H+2}. \tag{3.51}$$

3.8 Linear processes

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space and let $X(t)$, $t \in \mathbb{Z}$, be an adapted discrete time stochastic process defined on it. The process $X(t)$ is called a linear process if it can be expressed as a linear superposition of past and futures values of a $\mathcal{F}_t$-measurable white-noise process $\epsilon = \{\epsilon_t, \ t \in \mathbb{Z}\}$, sometimes called *innovation process*. In the following we consider only zero-mean innovation processes, we write $\epsilon_t = WN(0, \sigma^2)$. Therefore, we define:

Def. 3.16 (Linear processes)**.** The process $X(t)$ is called *linear stochastic process* if

$$X_t = \sum_{j=-\infty}^{\infty} c_{t-j}\epsilon_j = \sum_{j=-\infty}^{\infty} c_j \epsilon_{t-j}, \quad j \in \mathbb{Z}$$

where $\{c_j\}_{j\in\mathbb{Z}} \in \mathbb{R}$. In order to have a stationary process, we require that

$$\sum_{j=-\infty}^{\infty} c_j^2 < \infty. \tag{3.52}$$

The linear process X is *causal* if $c_j = 0$ for any $j < 0$. Indeed, in this case $X(t)$ is defined only by the "present" and the "past" of the innovation process. Clearly, only in this case $X(t)$ can be $\mathcal{F}_t$-measurable. The process X is Gaussian if for any $i \in \mathbb{Z}$, $\epsilon_i \sim N(0, \sigma^2)$, that is provided that the innovation process is a Gaussian white-noise.

Remark 3.27. By (eq. 3.52), a linear stochastic process is a second-order stationary process[5]. Moreover, it is stationary if it is Gaussian.

The causal process $X(t)$ is called *invertible* if the innovation ϵ_t can be expressed as a convergent sum of past and future values of X_t. Namely,

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \tag{3.53}$$

where $\sum \mid \pi_j \mid < \infty$.

5) A second order stationary process, or a wide-sense statioanry process, is a process with finite and constant mean and variance, and with covariance $\gamma(t,s)$ that depends only on $|t-s|$ (see Remark widest).

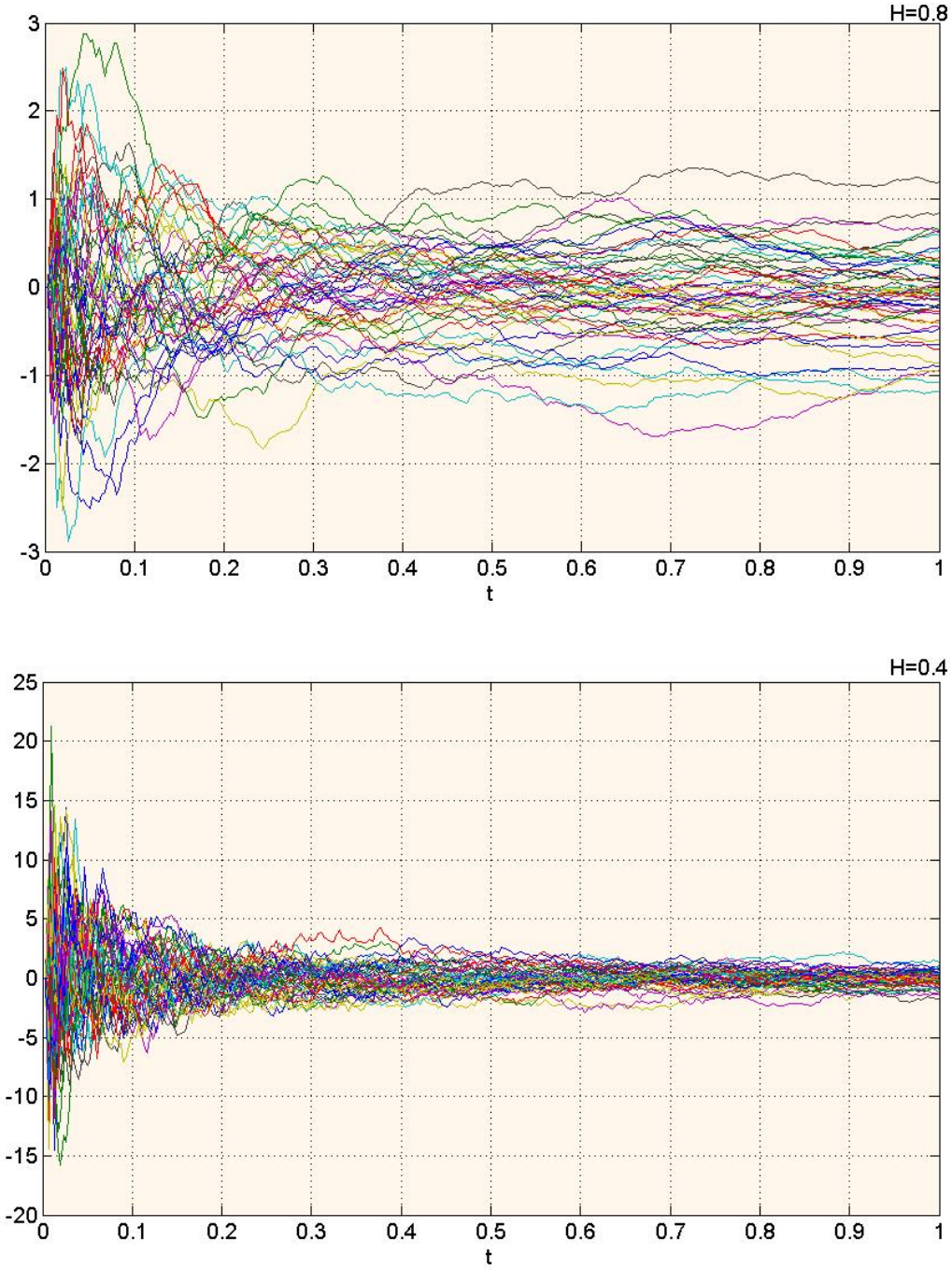

Figure 3.4: A collection of trajectories of of $\dot{A}_H(t)$ for $H = 0.8$ (top figure) and $H = 0.4$ (bottom figure). Observe the blow up of the variance for small t.

Remark 3.28. This property is actually connected to the possibility of making future forecasting. In fact, the invertible property of the linear process X allows to explicit $X(t)$ as a function of $X(s)$, $s < t$.

Example 3.13 (Linear Markov processes)**.** Consider a Gaussian causal linear process:

$$X(t) = \sum_{j=0}^{\infty} c(j)\epsilon(t-j).$$

By (eq. 3.15), one finds that X is Markovian if and only if for any $s \leq w \leq t$ one has $\rho(t-s) = \rho(t-w)\rho(w-s)$. That is, after a little algebra, if and only if

$$\sum_{j=0}^{\infty} c(j)c(j+s-t) = \frac{\sum_{j,l=0}^{\infty} c(j)c(l)c(j+w-t)c(l+s-w)}{\sum_{j=0}^{\infty} c(j)^2}. \tag{3.54}$$

For instance, choose $c(j) = \phi^j$ with $|\phi| < 1$. Then, one has

$$X(t) = \sum_{j=0}^{\infty} \phi^j \epsilon(t-j). \tag{3.55}$$

In this case, equation (3.54) is clearly satisfied, so that $X(t)$ is a Markov process.

Example 3.14 (Moving average processes)**.** A process $X(t)$, $t \in \mathbb{Z}$, is called (Gaussian) *moving average process* of order $q > 0$, shortly MA(q), if:

$$X_t = \sum_{j=0}^{q} \theta_j \epsilon_{t-j} = \Theta(B)\epsilon_t, \tag{3.56}$$

where ϵ_t is a zero mean Gaussian white-noise with variance σ_ϵ^2. The operator $\Theta(B)$ is called moving average operator of order q and is actually a polynomial function of the "backward" operator $B\epsilon_t = \epsilon_{t-1}$. The following properties are straightforward:

- X_t is stationary with $E(X_t) = 0$ and $\mathrm{Var}(X_t) = \sigma_\epsilon^2 \sum_{j=0}^{q} \theta_j^2$.
- The autocorrelation function $\frac{\gamma(k)}{\gamma(0)} = \frac{E(X_{t+k}X_t)}{\gamma(0)}$ is

$$\rho_k = \frac{\sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{\sum_{j=0}^{q} \theta_j^2}$$

  for $k \leq q$ and zero for $k > q$.

Remark 3.29. We observe that different moving average processes can have the same autocorrelation function. However, this ambiguity is avoided if we require the process to be invertible. An MA(q) process $X_t = \Theta(B)\epsilon_t$ is invertible if the roots z_i of the *characteristic function* $\Theta(z) = 0$ with $z \in \mathbb{C}$ are outside the unite circle. Namely, $|z_i| > 1$.

Example 3.15 (Autoregressive processes)**.** A process $\{X_t\}$ is called (Gaussian) *autoregressive process* of order $p > 0$, shortly AR(p), if:

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_t, \tag{3.57}$$

where ϵ_t is a Gaussian white noise as in the previous example. We have that:

- X_t is always invertible with $\epsilon_t = \Phi(B)X_t$.
- X_t is stationary if and only if the roots of the characteristic equation $\Phi(z) = 0$ are outiside the unite circle.
- The autocorrelation function is given by the so called *Yule-Walker equation* $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{p-k}$, with $k > 0$.

3.8.1 ARMA processes

The above examples are particular cases of the so called ARMA(p, q) processes.

Def. 3.17. Let $Y = \{Y_t\}_{t\in\mathbb{Z}}$ be a discrete time stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Then, Y is called (Gaussian) ARMA(p, q) process if it verifies the following equation:

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \epsilon_t - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q}, \quad t \in \mathbb{Z} \tag{3.58}$$

where $\{\phi_j\}_{j=1,\ldots,p}$ and $\{\theta_j\}_{j=1,\ldots,q}$ are real numbers, $p, q \in \mathbb{N} \cup \{0\}$ and where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a Gaussian white-noise.

If one introduces the operators:

$$\Phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$$

$$\Theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q.$$

Thus, the above equation reads:

$$\Phi_p(B)Y_t = \Theta_q(B)\epsilon_t. \tag{3.59}$$

ARMA(p,q) properties

Let $Y = \{Y_t\}_{t\in\mathbb{Z}}$ be an ARMA(p,q) process, then:

1. Y is causal and $\mathcal{F}_t$-adapted.
2. Y is stationary if the solutions of the characteristic equation $\Phi_p(z) = 0$ are outside the unit circle.
3. Y is invertible if the solutions of the characteristic equation $\Theta_q(z) = 0$ are outside the unit circle.
4. The autocovariance function tends to zero with a behavior depending of the nature of the roots of the autoregressive polynomial $\Phi_p(z)$; the decay is fast if the autoregressive component is stationary.

The last property is entailed in the following result [1]:

Prop. 3.12. *Let Y be a stationary $ARMA(p,q)$. Then, the autocovariance function $\gamma(k)$ tends exponentially to zero as $k \to \infty$:*

$$\gamma(k) \sim cs^k, \quad k \to \infty \tag{3.60}$$

where $c > 0$ and $0 < s < 1$.

Example 3.16 (Short memory)**.** The AR(1) process

$$X(t) = \phi X(t-1) + \epsilon_t,$$

sometimes used as the prototype of a Markovian process, is a stationary process provided that $|\phi| < 1$. It actually corresponds to the process defined in (eq. 3.55). In fact, one can write:

$$(1 - \phi B)X_t = \epsilon_t \Longrightarrow X_t = (1 - \phi B)^{-1}\epsilon_t = \sum_{j=0}^{\infty} \phi^j B^j \epsilon_t$$

which is (eq. 3.55). In this case, the autocorrelation function is simply $\rho(k) = \gamma(k)/\gamma(0) = \phi^k$. This exponential behavior, which is (asymptotically) a characteristic of the whole stationary ARMA class processes (see Prop. 3.12), is usually used to explain the idea of *short-memory* processes. Otherwise, the so called *long-memory* processes have a power-decay correlation function (see Chapter 6).

3.8.2 FARIMA processes

The so called FARIMA processes are Gaussian linear causal processes specified by three real parameters

$$\text{FARIMA}(p,d,q), \quad p,q \in \mathbb{Z} \cup \{0\}, \quad d \in \mathbb{R}.$$

We start defining FARIMA(0,*d*,0) processes.

Def. 3.18. A discrete stochastic process $X = \{X_t\}_{t\in\mathbb{Z}}$, defined on the probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, is called FARIMA(0,*d*,0) if satisfies:

$$\Delta^d X_t = (1 - B)^d X_t = \epsilon_t, \tag{3.61}$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a Gaussian white-noise with variance σ^2 and B is the lag (or backward) operator.

Remark 3.30. We observe that if d were an integer number, the process X could heuristically be obtained integrating d times the innovation process. Such processes are indeed called integrated ARMA processes, shortly ARIMA. For instance the ARIMA(0,1,0) process can be seen just as a "digital" Brownian motion. If one let d be real, then one could speak about fractional integrated ARMA processes, and from that actually derives the FARIMA acronym.

Let now $X = \{X_t\}_{t\in\mathbb{Z}}$ be a FARIMA(0,*d*,0) process. Then, it has the following properties:

1. X is causal and $\mathcal{F}_t$-adapted.
2. X is stationary if and only if $d < 1/2$.
3. X is invertible if and only if $d > -1/2$.
4. If $X = \{X_t\}_{t\in\mathbb{Z}}$ is a FARIMA(0,*d*,0) with $-1/2 < d < 1/2$, thus

$$\gamma(0) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}. \tag{3.62}$$

5. Furthermore,

$$\gamma(k) = \sigma^2 \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(d)\Gamma(1-d)}. \tag{3.63}$$

In fact, let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a solution of (eq. 3.61). Then,

$$X_t = (1-B)^{-d}\epsilon_t = \sum_{j=0}^{\infty} b_j B^j \epsilon_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j},$$

where $b_0 = 1$ and b_j are the generalized binomial coefficients

$$b_j = (-1)^j \frac{\Gamma(1-d)}{\Gamma(j+1)\Gamma(1-j-d)} = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{k=1}^{j} \frac{k-1+d}{k}, \quad j \in \mathbb{N}, \tag{3.64}$$

which is easily obtained using the Euler reflection formula for the gamma function

$$\Gamma(1-z) = \frac{\pi}{\sin \pi z \Gamma(z)}. \tag{3.65}$$

In order to have stationarity we must have $\{b_j\}_{\in\mathbb{N}} \in l^2$. Using Stirling's formula,

$$\Gamma(p) \sim \sqrt{2\pi} e^{-p+1}(p-1)^{p-1/2}, \quad p \to \infty$$

on has:

$$b_j \sim \frac{j^{d-1}}{\Gamma(d)}, \quad j \to \infty.$$

Then, it is necessary that $\sum_j j^{2(d-1)} < \infty$. That is, $2(d-1)+1 < 0$ and thus $d < 1/2$.

Moreover, the process X is invertible if it is convergent the series:

$$\Delta^d X_t = \epsilon_t = \sum_{j=0}^{\infty} \pi_j B^j X_t,$$

where:

$$\pi_j = \prod_{k=1}^{j} \frac{k-1-d}{k} = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad j \in \mathbb{N}. \tag{3.66}$$

If one imposes that $\{\pi_j\}_{\in\mathbb{N}} \in l^2$, one finds $d > -1/2$.

Finally, in order to show (eq. 3.62) and (eq. 3.63) we cross-refer to (Remark 5.16).

One of the noteworthy property of FARIMA processes is the asymptotic behavior of the autocovariance function. In fact, let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a FARIMA(0,*d*,0) process with $-1/2 < d < 1/2$, then:

$$\gamma(k) \sim c\,|k|^{2d-1}, \quad k \to \infty, \tag{3.67}$$

where

$$c = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}.$$

The above equation follows from (eq. 3.63) and from Stirling's formula.

Remark 3.31. Observe that the asymptotic behavior (eq. 3.67) of the FARIMA autocovariance has the same powek-like behavior of the fGn autocovariance (eq. 3.43) provided that $H = d + 1/2$.

The above formal identification of the FARIMA parameter d with H is entailed in the following result which relate FARIMA processes to fBm [12].

Prop. 3.13. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a FARIMA(0,d,0) process with $0 < d < 1/2$, then*

$$\frac{1}{n^H}\sum_{j=1}^{[nt]} X_j \to B_H(t), \quad n \to \infty, \tag{3.68}$$

where $H = d + 1/2$, $[x]$ is the integer part of x and $\{B_H(t)\}_{t\in\mathbb{R}}$ is a fBm defined in the same probability space of X. The convergence has to be meant in the sense of finite dimensional distributions.

Finally, a general FARIMA(p,d,q) is defined below.

Def. 3.19. A stochastic process $X = \{X_t\}_{t\in\mathbb{Z}}$, defined in the filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, is called FARIMA(p,d,q) if satisfies:

$$\Phi_p(B)\Delta^d X_t = \Theta_q(B)\epsilon_t; \tag{3.69}$$

where $Y_t = \Delta^d X_t$ is an ARMA(p,q) process.

As we shall see later, a FARIMA(p,d,q) process combines the long-memory power-like asymptotic behavior of the correlation function of the fractional component FARIMA(0,d,0) with the flexibility in modeling short range correlation of an ARMA(p,q) process. We end this section summarizing the principal properties of a FARIMA(p,d,q) process.

Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a FARIMA(p,d,q) process, then:

1. X is stationary if and only if $d < 1/2$ and the roots of the characteristic autoregressive polynomial $\Phi_p(z)$ are outside the unit circle.
2. X è invertibile if and only if $d > -1/2$ and the roots of the characteristic moving average polynomial $\Theta_q(z)$ are outside the unit circle.
3. The X covariance structure depends on the value of the d parameter and on the value of the roots of the characteristic polynomials.

Bibliography

[1] P.J. Brockwell, R.A. Davis, *Time series: theory and methods*, Springer-Verlag, second edition, 1991.

[2] R. Cont and P. Tankov, *Financial Modeling With Jump Processes*, (Chapman & Hall, CRC Financial Mathematics Series, 2004).

[3] J.L. Doob, *Stochastic Processes*, (Wiley & Sons. 1953).

[4] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, 3-rd edn. (Wiley, New York, 1968). [1-st edn. 1957]

[5] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2-nd edn. (Wiley, New York, 1971). [1-st edn. 1966]

[6] C.W.J. Granger and R. Joyeux, An introduction to long memory time series models and fractional differencing, *J. Time Series Analysis* **1** (1980) 15-29.

[7] P.E. Kloeden and E. Platen, *Numerical solutions of Stochastic differential Equations*, (Springer-Verlag, Berlin Heidelberg New York, 1999).

[8] D. Lamberton and B. Lapeyre, *Stochastic Calculus Applied to Finance*, (Chapman & Hall, London, 1996).

[9] B.B. Mandelbrot, Note on the definition and stationarity of fractional Gaussian noise, *Journal of Hydrology* **30** (1976) 407-409.

[10] B.B. Mandelbrot, Gaussian Self-Affinity and Fractals, *Springer-Verlag, New York, 2002*, pp. 654.

[11] B.B. Mandelbrot and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10** (1968) 422-433. [X]

[12] M.S. Taqqu, Fractional Brownian motion and long-range dependence, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and Applications of Long-range Dependence*, Birkäuser, Boston - Basel - Berlin, 2003, pp. 5-38.

[13] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, A series of Comprehensive Studies in Mathematics, (Springer Verlag, 293, 1994).

[14] B.J. West and P. Grigolini, Fractional differences, derivatives and fractal time series, in: R. Hilfer (Editor), *Applications of Fractional Calculus in Physics* (World Scientific, Singapore, 2000), pp. 171-201.

4 Introduction to stochastic integration

If we introduce some randomness in the coefficients of a differential equation, we could obtain a more realistic mathematical model. For instance, consider the simple population growth model:

$$\frac{dN(t)}{dt} = r(t)N(t), \quad N(0) = N_0, \tag{4.1}$$

where $N(t)$ is the size of the population at time $t \geq 0$ and $r(t)$ is the relative rate of growth. It might happen that $r(t)$ is not completely known but subjected to some random environmental effects. So that, we have:

$$r(t) = r_0(t) + \text{ noise}, \quad t \geq 0, \tag{4.2}$$

where we do not know the exact behavior of the noise term but only its probability distribution. The function $r_0(t)$ is assumed non-random. Then, one obtains the following equation:

$$\frac{dN(t)}{dt} = (r_0(t) + \text{ noise}) \cdot N(t), \quad t \geq 0. \tag{4.3}$$

More generally, we consider equations of the form:

$$\frac{dX_t}{dt} = a(t, X_t) + \sigma(t, X_t) \cdot Z_t, \quad t \geq 0. \tag{4.4}$$

Based on many situations, for example in engineering, one is led to assume that the process $Z(t)$, $t \geq 0$, defined on certain probability space $(\Omega, \mathcal{F}, P)$, has, at least approximately, the following properties:

1. for any $t_1 \neq t_2$ the r.v.s $Z(t_1)$ and $Z(t_2)$ are independent.
2. $Z(t)$ is a stationary process.
3. $EX(t) = 0$.

However, it turns out that there does not exist any "reasonable" stochastic process satisfying the above properties. Nevertheless, as we shall see later, it is possible to represent $Z(t)$ as a *generalized stochastic process* called *white noise* process (see for instance Remark 3.24 concerning Gaussian white noise).

Here, we will avoid this kind of construction and rather we will try to rewrite (eq. 4.4) in a form that suggests a replacement of $Z(t)$ by a proper stochastic process. Consider a time grid $I_{\delta t} = \{t_k = k\delta t, \ k \geq 0\}$, with $\delta t > 0$, then we can take a discrete version of (eq. 4.4):

$$X_{k+1} - X_k = a(t_k, X_k)\delta t + \sigma(t_k, X_k)Z_k\delta t. \tag{4.5}$$

We replace $W_k\delta t$ by $\Delta V_k = V_{k+1} - V_k$, where $\{V_t, \ t \geq 0\}$ represents some suitable stochastic process. The conditions 1, 2 and 3 on Z_t suggest that V_t have stationary independent increments with zero mean. However, the only such process which has continuous paths is the Brownian motion $B(t)$, $t \geq 0$ [3]. We obtain:

$$X_k = X_0 + \sum_{j=0}^{k-1} a(t_j, X_j)\delta t + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta B_j. \tag{4.6}$$

Is it possible to show that the limit of the right hand side of (eq. 4.6) exists indeed, "in some sense", when $\delta t \to 0$. Then, by applying the usual integration notation, we can write:

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \text{“} \int_0^t \sigma(s, X_s)dB_s\text{”}. \tag{4.7}$$

In the following we shall prove the existence of "$\int_0^t f(s,\omega)dB_s(\omega)$", $\omega \in \Omega$, for some suitable function $f : [0,\infty) \times \Omega \to \mathbb{R}$. Suppose $0 < S < T$ and let $f(t,\omega)$ be given. We want to define:

$$\int_S^T f(t,\omega)dB_t(\omega).$$

It is reasonable to start with a definition for a simple class of functions. Then, we will extend this definition by using some approximation procedure. Let $\{e_k,\ k \geq 0\}$ be a sequence of real random variables, and consider a function of the form

$$\varphi(t,\omega) = \sum_{k\geq 0} e_k(\omega)1_{[k2^{-n},(k+1)2^{-n})}(t),\ \ n \in \mathbb{N}. \tag{4.8}$$

In this case it is reasonable to define:

$$\int_S^T \varphi(t,\omega)dB_t(\omega) = \sum_{k\geq 0} e_k(\omega)\left(B(t_{k+1}) - B(t_k)\right)(\omega), \tag{4.9}$$

where:

$$t_k = t_k^{(n)} = \begin{cases} k2^{-n},\ \ S \leq k2^{-n} \leq T, \\ S,\ \ k2^{-n} < S, \\ T,\ \ k2^{-n} > T. \end{cases} \tag{4.10}$$

As shown in the example below, we will need some hypothesis on the r.v.s $\{e_k\}$ in order to avoid many problems.

Example 4.1. Let

$$\phi_1(t,\omega) = \sum_{k\geq 0} B_k(\omega)1_{[k2^{-n},(k+1)2^{-n})}(t)$$
$$\phi_2(t,\omega) = \sum_{k\geq 0} B_{k+1}(\omega)1_{[k2^{-n},(k+1)2^{-n})}(t).$$

With these choices one has:

$$E\left(\int_0^T \phi_1 dB_t\right) = \sum_{k\geq 0} E(B_k(B_{k+1} - B_k)) = 0,$$

whereas:

$$E\left(\int_0^T \phi_2 dB_t\right) = \sum_{k\geq 0} E(B_{k+1}(B_{k+1} - B_k)) = T.$$

Therefore, even if both ϕ_1 and ϕ_2 appear to be very reasonable approximations of $f(t,\omega) = B_t(\omega)$, their stochastic integrals (eq. 4.9) are not close to each other no matter how large n is taken.

Remark 4.1. This only reflects the fact that the variations of the paths of B_t are too big to enable us to define the integral in the Riemann-Stieltjes sense. In fact, the paths $t \to B_t$ of Brownian motion are nowhere differentiable, almost surely. In particular, the total variation of the paths is infinite, a.s (see [6]).

In general is natural to approximate a given function $f(t,\omega)$ by

$$f(t,\omega) = \sum_{k\geq 0} f(t_k^*,\omega)1_{[t_k,t_{k+1})}(t),$$

where the $t_k^* \in [t_k, t_{k+1})$, then one can define the integral $\int_S^T f(t,\omega)dB_t(\omega)$ as a suitable limit of $\sum f(t_k^*,\omega)(B_{k+1} - B_k)(\omega)$. However, the example above shows that, unlike the Riemann-Stieltjes integral, the definition depends on the choice of the points t_k^*. The following two choices have turned out to be the most useful ones:

1. $t_k^* = t_k$ (left end point), which leads to the *Ito integral*.
1. $t_k^* = \frac{t_{k+1}+t_k}{2}$ (mid point), which leads to the *Stratonovic integral*.

Notation 4.1. We shall denote the Ito integral simply by:

$$\int_S^T f(t,\omega)dB_t(\omega), \tag{4.11}$$

and the Stratonovic integral by:

$$\int_S^T f(t,\omega) \circ dB_t(\omega). \tag{4.12}$$

Remark 4.2. In any case, one must restrict oneself to a special class of functions $f(t,\omega)$ in order to obtain a reasonable definition of the integral, even if they have the particular form of (eq. 4.8).

Therefore, let us consider the filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, where $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration of the Bm $B(t)$, $t \geq 0$. We introduce the following class of functions.

Def. 4.1. Let $\mathcal{V}(S,T)$ be the class of real **measurable** functions $f(t,\omega)$, defined on $[0,\infty) \times \Omega$, such that:

1. $f(t,\omega)$ is $\mathcal{F}_t$-adapted.
2. $E\left(\int_S^T f(t,\cdot)^2 dt\right) < \infty$.

4.1 Ito integral

Let $f \in \mathcal{V}(S,T)$. We want to define the Ito integral of f in the interval $[S,T)$. Namely:

$$\mathcal{I}(f)(\omega) = \int_S^T f(t,\omega)dB_t(\omega), \tag{4.13}$$

where B_t is a standard ($E(B(1)^2) = 1$) one dimensional Brownian motion. We begin defining the integral for a special class of functions:

Def. 4.2 (Simple functions)**.** A function $\varphi \in \mathcal{V}(S,T)$ is called *simple* function (or *elementary*), if it can be expressed as a superposition of characteristic functions.

$$\varphi(t,\omega) = \sum_{k\geq 0} e_k(\omega)1_{[t_k,t_{k+1})}(t), \tag{4.14}$$

where t_k is defined as in (eq. 4.10) and $\{e_k\}_{k\geq 0}$ is a sequence of r.v.s defined on $(\Omega, \mathcal{F}, P)$, such that $P(e_k = e_{k-1}) = 0$ for any $k \geq 0$.

Remark 4.3. In order to have $\varphi \in \mathcal{V}$ we must have that e_k be $\mathcal{F}_{t_k}$-measurable. Indeed, $\varphi(t_k, \omega) = e_k(\omega)$. For instance, in (Exa. 4.1) the function Φ_1 is simple, while Φ_2 is not.

Remark 4.4. Moreover, observe that $e_k \in L^2(\Omega, P)$ and one has:

$$E\left(\int_S^T \varphi(t,\cdot)^2 dt\right) = \sum_{k\geq 0} E(e_k^2)(t_{k+1} - t_k). \tag{4.15}$$

Let $\varphi \in \mathcal{V}(S,T)$ be a simple function of the form of (eq. 4.14), then we define the stochastic integral according to (eq. 4.9):

$$\int_S^T \varphi(t,\omega) dB_t = \sum_{k\geq 0} e_k(\omega)(B_{t_{k+1}} - B_{t_k})(\omega); \tag{4.16}$$

Example 4.2. One has:

$$\int_S^T dB_t = B_T - B_S.$$

Indeed, it follows from the definition by setting $\varphi = 1_{[S,T)}$.

We have the noteworthy result:

Lemma 4.1 (Ito isometry)**.** *Let $\varphi \in \mathcal{V}(S,T)$ be a simple function, then:*

$$E\left(\left(\int_S^T \varphi(t,\cdot) dB_t\right)^2\right) = E\left(\int_S^T \varphi(t,\cdot)^2 dt\right). \tag{4.17}$$

Proof: by definition:

$$E\left(\left(\int_S^T \varphi(t,\cdot) dB_t\right)^2\right) = E\left(\sum_{k\geq 0}\sum_{h\geq 0} e_k e_h \Delta B_k \Delta B_h\right)$$

$$= E\left(\sum_{k\geq 0} e_k^2 \Delta B_k^2\right) + E\left(2\sum_{k<h} e_k e_h \Delta B_k \Delta B_h\right) = \sum_{k\geq 0} E(e_k^2)(t_{k+1} - t_k),$$

because e_k is independent of ΔB_k and $e_k e_h \Delta B_k$ is independent of ΔB_h when $k < h$. Then, using (eq. 4.15) one has the result. □

Remark 4.5. Observe that (eq. 4.17) is indeed an isometry. In fact, it can been written as equality of norms in L^2 spaces:

$$\left\| \int_S^T \varphi(t,\cdot) dB_t \right\|_{L^2(\Omega,P)} = \|\varphi\|_{L^2([S,T]\times\Omega)}.$$

We have the following important proposition.

Prop. 4.1. *Let $f \in \mathcal{V}$, then there exists a sequence of simple functions $\phi_n \in \mathcal{V}$, $n \in \mathbb{N}$, which converges to f in the L^2-norm. Namely,*

$$\lim_{n\to\infty} \int_S^T E\left((f(t,\cdot) - \phi_n(t,\cdot))^2\right) dt = \lim_{n\to\infty} ||f - \phi_n||^2_{L^2([S,T]\times\Omega)} = 0. \tag{4.18}$$

Proof: We give te proof when $f(t,\omega)$ is continuous on $[S,T]$ for each $\omega \in \Omega$. In this case consider:

$$\phi_n(t,\omega) = \sum_{k\geq 0} f(\xi_k, \omega) 1_{[t_k, t_{k+1})}(t), \tag{4.19}$$

with t_k given by (eq. 4.10) and where

$$\xi_k \equiv \xi_k^{(n)} = \lambda t_{k+1} + (1-\lambda)t_k, \;\; 0 \leq \lambda \leq 1. \tag{4.20}$$

Then, $\phi_n(t,\omega)$ converges as $n \to \infty$ to $f(t,\omega)$ in the $L^2([S,T]\times\Omega)$ norm. In fact, since $f(\cdot,\omega)$ is uniformly continuous[1] for each $\omega \in \Omega$, for any $\epsilon > 0$ there exists $n_\epsilon > 0$ such that for any $n > n_\epsilon$

$$\int_S^T (f(t,\omega) - \phi_n(t,\omega))^2 dt = \sum_{k\geq 0} \int_{t_k}^{t_k+1} (f(t,\omega) - f(\xi_k,\omega))^2 dt < \epsilon(T-S).$$

Hence, $E(\int_S^T (f-\phi_n)^2 dt) \to 0$ as $n \to \infty$ by bounded convergence. If we require $\phi_n \in \mathcal{V}$, then we must have $\lambda = 0$ and this is indeed the Ito condition (see Remark 4.3). For the general case see [5]. □

Given $f \in \mathcal{V}(S,T)$, the proposition above, together with Ito isometry, implies that the sequence $\left\{\int_S^T \phi_n(t,\omega)dB_t(\omega), \;\; n \in \mathbb{N}\right\}$ is Cauchy on $L^2(\Omega,P)$. So that, it converges to a limit in $L^2(\Omega,P)$. We call this limit the Ito integral of f. Summarizing:

Def. 4.3 (Ito integral)**.** Let $f \in \mathcal{V}(S,T)$. The Ito integral from S to T of f is defined as the $L^2(\Omega,P)$ limit:

$$I(f) = \int_S^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dB_t(\omega), \tag{4.21}$$

where $\phi_n \in \mathcal{V}$, $n \in \mathbb{N}$, is a sequence of simple functions which converges to f in $L^2([S,T]\times\Omega)$.

Remark 4.6. Observe, in view of (eq. 4.18), that the definition above does not depend on the actual choice of $\{\phi_n, \;\; n \in \mathbb{N}\}$.

By definition, we have that Ito isometry holds for Ito integrals:

Cor. 4.1 (Ito isometry for Ito integrals)**.** *Let $f \in \mathcal{V}(S,T)$, then:*

$$E\left(\left(\int_S^T f(t,\cdot)dB_t\right)^2\right) = E\left(\int_S^T f(t,\cdot)^2 dt\right). \tag{4.22}$$

Moreover,

Cor. 4.2. *If $f_n(t,\omega) \in \mathcal{V}(S,T)$ converges to $f(t,\omega) \in \mathcal{V}(S,T)$ as $n \to \infty$ in the $L^2([S,T]\times\Omega)$-norm, then:*

$$\int_S^T f_n(t,\cdot)dB_t \to \int_S^T f(t,\cdot)dB_t, \tag{4.23}$$

in the $L^2(\Omega,P)$-norm.

Example 4.3. Consider $f(t,\omega) = B_t(\omega)$. Then, one has

$$\int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{t}{2}. \tag{4.24}$$

In fact, the sequence of simple functions $\phi_n(t,\omega)$, such that:

$$\phi_n(t,\omega) = \sum_{k\geq 0} B_k(\omega) 1_{[t_k,t_k+1)}(t), \tag{4.25}$$

1) $f(\cdot,\omega), \omega \in \Omega$, is continuous on a compact then, using Heine-Cantor theorem, it is also uniformly continuous.

converges in the L^2-norm to $B_t(\omega)$. Thus,

$$\int_0^t B_s dB_s = \lim_{n\to\infty} \int_0^t \phi_n(s,\cdot) dB_s = \lim_{\epsilon\to 0} \sum_{k\geq 0} B_k \Delta B_k,$$

where $\epsilon = 2^{-n}$ (see eq. 4.10). We observe that:

$$\Delta(B_k^2) = \Delta B_k^2 + 2B_k \Delta B_k \Rightarrow B_t^2 = \sum_{k\geq 0} \Delta B_k^2 + 2\sum_{k\geq 0} B_k \Delta B_k.$$

Therefore,

$$\int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{1}{2} \lim_{\epsilon\to 0} \sum_{k\geq 0} \Delta B_k^2,$$

and the result follows from the fact that the so called *quadratic variation* of B_t is

$$\langle B \rangle_t := \lim_{\epsilon\to 0} \sum_{k\geq 0} \Delta B_k^2 = t, \text{ a.s.} \tag{4.26}$$

In fact,

$$E\left[\Big(\sum_{k\geq 0} \Delta B_k^2 - t \Big)^2 \right] = \sum_{k\geq 0} E\left[(\Delta B_k^2 - \epsilon)^2 \right] + 2\sum_{k<j} E\left[(\Delta B_k^2 - \epsilon)(\Delta B_j^2 - \epsilon) \right],$$

the right hand side is zero for independence, whereas the left hand side goes to zero when $\epsilon \to 0$. In fact,

$$\sum_{k\geq 0} E\left[(\Delta B_k^2 - \epsilon)^2 \right] = \sum_{k\geq 0} E(\Delta B_k^4) - \epsilon t = 2t\epsilon.$$

Remark 4.7. Observe that if in the previous example we had taken the *right end point* approximation of B_t we would have obtained:

$$\lim_{\epsilon\to 0} \sum_{k\geq 0} B_{k+1} \Delta B_k = \frac{B_t^2}{2} + \frac{t}{2}. \tag{4.27}$$

Looking at Example 4.1, the latter corresponds to ϕ_2 (or $\lambda = 1$ in (eq. 4.20)), while the choice in Example 4.3 corresponds to ϕ_1 ($\lambda = 0$).

The extra term in (eq. 4.24) shows that the Ito stochastic integral does not behave like an ordinary integral. In the following we will establish the Ito formula, which explains this result and which makes easy the evaluation of many stochastic integrals.

4.1.1 Properties of the Ito integral

First we observe the following:

Prop. 4.2. *Let $f, g \in \mathcal{V}(0,T)$ and let $0 \leq S < U < T$. Then:*

i. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t.$

ii. *For some constant $a \in \mathbb{R}$,* $\int_S^T (af + g) dB_t = a\int_S^T f dB_t + \int_S^T g dB_t.$

iii. $E\left[\int_S^T f dB_t \right] = 0.$

iv. $\int_S^T f dB_t$ *is* $\mathcal{F}_T$*-measurable.*

v. The process $M_t(\omega) = \int_0^t f(t,\omega) dB_s(\omega)$*, where* $f(t,\omega) \in \mathcal{V}(0,t)$ *for any* $t > 0$*, is a* **martingale** *with respect to* $\mathcal{F}_t$*.*

Proof: These properties clearly holds for all simple functions, so that, by taking limits, we obtain them for all $f, g \in \mathcal{V}(0,T)$. Let us check just the martingale property. M_t is $\mathcal{F}_t$-measurable. Moreover, M_t is also integrable. Indeed, using Hölder inequality:

$$E(|M_t|)^2 \leq E(M_t^2) = E\left(\int_0^t f^2(s,\cdot) ds\right) < \infty,$$

where we have used the Ito isometry and the fact that $f \in \mathcal{V}$. Let $0 < s < t$. Then,

$$E(M_t \mid \mathcal{F}_s) = E(M_s \mid \mathcal{F}_s) + E\left(\int_s^t f(\tau,\cdot) dB_\tau \mid \mathcal{F}_s\right) = M_s,$$

because M_s is $\mathcal{F}_s$-measurable, whereas:

$$E\left(\int_s^t f(\tau,\cdot) dB_\tau \mid \mathcal{F}_s\right) = \lim \sum_{k \geq 0} E(f(t_k,\cdot)(B_{t_{k+1}} - B_{t_k}) \mid \mathcal{F}_s)$$

(being $f(t_k,\cdot)$ and $B_{t_{k+1}} - B_{t_k}$ independent)

$$= \lim \sum_k E(f(t_k,\cdot) \mid \mathcal{F}_s) E(B_{t_{k+1}} - B_{t_k} \mid \mathcal{F}_s) = 0$$

because $B_{t_{k+1}} - B_{t_k}$ is independent of $\mathcal{F}_s$ and $s \leq t_k$ for any k. □

4.1.2 Extensions of Ito integral

The construction of the Ito Integral can be extended to a class of function $f(t,\omega)$ which satisfies a weak integration condition. This generalization is indeed necessary because it is not difficult to find functions which do not belong to $\mathcal{V}$. For instance, take a function of Bm which increase rapidly $f(t,\omega) = \exp(B_t(\omega)^2)$. Therefore, we introduce the following class of functions:

Def. 4.4. Let $\mathcal{W}(S,T)$ be the class of real measurable functions $f(t,\omega)$, defined on $[0,\infty) \times \Omega$, such that

- 1. $f(t,\omega)$ is $\mathcal{F}_t$-adapted.
- 2. $P\left(\int_S^T f(t,\cdot)^2 dt < \infty\right) = 1$.

Remark 4.8. Clearly, $\mathcal{V} \subset \mathcal{W}$.

In the construction of stochastic integrals for the class of functions belonging to $\mathcal{W}$ we can no longer use the L^2 notion of convergence, but rather we have to use convergence in probability (Section 2.3.2). In fact, for any $f \in \mathcal{W}$, one can show that there exists a sequence of simple functions $\phi_n \in \mathcal{W}$ such that

$$\int_S^T |\phi_n(t,\cdot) - f(t,\cdot)|^2 dt \to 0 \tag{4.28}$$

in probability. For such a sequence one has that the sequence $\left\{\int_S^T \phi_n(t,\omega) dB_t(\omega),\ \ n \in \mathbb{N}\right\}$ converges in probability to some random variable. Moreover, the limit does not depends on the approximating sequence ϕ_n. Thus, we may define:

Def. 4.5 (Ito integral II)**.** et $f \in \mathcal{W}(S,T)$. The Ito integral from S to T of f is defined as the limit in probability:

$$\int_S^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty}\int_S^T \phi_n(t,\omega)dB_t(\omega), \tag{4.29}$$

where $\phi_n \in \mathcal{W}$, $n \in \mathbb{N}$, is a sequence of simple functions which converges to f in probability.

Remark 4.9. Note that this integral is not in general a martingale. However, it is a local martingale (see [5]).

4.1.3 Other stochastic integrals

We have seen that the Ito integral of a function $f \in \mathcal{V}$ is equal to the mean-square limit

$$\lim_{n\to\infty}\sum_{k\geq 0} f(\xi_k^{(n)},\omega)\Delta B_k, \tag{4.30}$$

where $\xi_k^{(n)} = t_k^{(n)} = k2^{-n}$, $n \in \mathbb{N}$ (left end point approximation). We have also pointed out that different choices of the evaluation points $\xi_k^{(n)} \in [t_k^{(n)}, t_{k+1}^{(n)})$ are possible, but generally lead to different random variables in the limit. Suppose that, like in (eq. 4.20), these evaluation points are chosen such that for some $0 \leq \lambda \leq 1$:

$$\xi_k^{(n)} = (1-\lambda)t_k^{(n)} + \lambda t_{k+1}^{(n)}. \tag{4.31}$$

Then, we shall denote the corresponding stochastic integral with:

$$(\lambda)\int_S^T f(t,\omega)dB_t(\omega), \quad f \in \mathcal{V}(S,T). \tag{4.32}$$

Remark 4.10. The case $\lambda = 0$ corresponds to Ito integral. The other cases $0 < \lambda \leq 1$ differ in that the process they define is no longer a martingale.

Suppose that the integrand f has continuously differentiable sample paths we obtain

$$f\left((1-\lambda)t_k + \lambda t_{k+1},\omega\right) = (1-\lambda)f(t_k,\omega) + \lambda f(t_{k+1},\omega) + O(\epsilon). \tag{4.33}$$

Therefore, we see that the (λ)-integrals could be evaluated as the mean-square limit:

$$\lim_{n\to\infty}\sum_{k\geq 0}\left[(1-\lambda)f(t_k,\omega) + \lambda f(t_{k+1},\omega)\right]\Delta B_k. \tag{4.34}$$

In th general case the (λ)-integrals are usually defined in terms of the sum above rather than by (eq. 4.30).

Remark 4.11. When $\lambda = 0$ we recover Ito integral, whereas the Stratonovic integral corresponds to the choice $\lambda = 1/2$.

Using (eq. 4.24) and (eq. 4.27), one can easily find:

$$(\lambda)\int_0^t B_s dB_s = \frac{1}{2}B_t + \left(\lambda - \frac{1}{2}\right)t. \tag{4.35}$$

Unlike any of the others the case $\lambda = 1/2$, namely the Stratonovic integral, doest not contain a term in addition to that given by classical integration. As suggested by the previous argument, the Stratonovic integral obeys the transformation rules of classical calculus, and this is a major reason for its use. In fact,

let $h : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function and consider the Stratonovic integral of $h(B_t)$. Then, we have:

$$\int_0^t h(B_s) \circ dB_s = \lim \left\{ \sum_{k \geq 0} h(B_k) \Delta B_k + \frac{1}{2} \sum_{k \geq 0} h'(B_k) [\Delta B_k]^2 \right\}$$

which converges in the L^2-sense to:

$$\int_0^t h(B_t) \circ dB_t = \int_0^t h(B_s) dB_s + \frac{1}{2} \int_0^t h'(B_s) ds. \tag{4.36}$$

Let now $U(x) = U(0) + \int_0^x h(y) dy$. Applying Ito's formula (see Section 4.2) to the process $Y_t = U(B_t)$, we obtain

$$dU_t = h(B_t) dB_t + \frac{1}{2} h'(B_t) dt,$$

that is using (eq. 4.36),

$$\int_0^t h(B_s) \circ dB_s = U(B_t) - U(B_0),$$

as in classical calculus. Then, the major advantage of the Stratonovic integral is that it obeys the usual transformation rules of calculus. For this reason it is often used to formulate stochastic differential equations on manifolds such as circle or sphere, as is required, for instance, for a stability analysis of stochastic dynamical systems. Equation (4.36) is also useful for the explicit evaluation of certain Ito integrals. Stochastic processes defined through Stratonovic integrals are not, however, martingales as the Ito counterparts are.

4.1.4 Stochastic measures

Let $B = \{B_t\}_{t \geq 0}$ be a Brownain motion defined on the filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Consider $(\mathbb{R}, \mathcal{B}, m)$ Euclidean space with the Borel sets and the Lebesgue measure. Let us define:

$$\mathcal{B}_0 = \{A \in \mathcal{B};\ \ m(A) < \infty\}.$$

Then,

Def. 4.6. For any $A \in \mathcal{B}_0$, set:

$$M(A) = \int_{\mathbb{R}} 1_A(x) dB(x) = \int_A dB(x). \tag{4.37}$$

Therefor, varying $A \in \mathcal{B}_0$, the family of r.v.s $M = \{M(A),\ \ A \in \mathcal{B}_0\}$ defines a Gaussian real stochastic process with the following properties:

4.1.4.1 Random measures properties

1. $EM(A) = 0$, for any $A \in \mathcal{B}_0$.
2. $E\left(M(A)\overline{M(B)}\right) = 0$, if $A \cap B = \emptyset$ and $A, B \in \mathcal{B}_0$.
3. $E\left(|M(A)|^2\right) = m(A)^2$, for any $A \in \mathcal{B}_0$;

2) We use complex notation looking for a future generalization.

4. M is σ - additive. That is, for any disjoint collection $\{A_i\}_{i=1,2,\ldots}$ such that $\bigcup_i A_i \in \mathcal{B}_0$; then,

$$M\Big(\bigcup_i A_i\Big) = \sum_i M(A_i).$$

Properties 1. and 2. follow directly from the zero mean property of Ito integral. Moreover, 3. is a consequence of Ito isometry. Finally, 4. follows from linearity.

The process M is called (real Gaussian) *random measure*.

Notation 4.2 (Control measure)**.** We use the following notation:

$$E\,|M(dx)|^2 = m(dx),\quad E(M(dx)\overline{M(dy)}) = 0,\quad dx \neq dy.$$

The Lebesgue measure m is called *control measure* of the random measure M.

More generally, let $(E, \mathcal{E}, m)$ be a measurable space and $\mathcal{E}_0 = \{A \in \mathcal{E};\ m(E) < \infty\}$.

Def. 4.7 (Random easure)**.** We define random measure (generally complex valued) a stochastic process $M = \{M(A),\ A \in \mathcal{E}_0\}$, defined on $(\Omega, \mathcal{F}, P)$, such that it verifies the properties above. The measure m is the control measure of M.

Example 4.4 (Complex Gaussian random measure)**.** Let M_1 and M_2 be two real Gaussian independent random measures such that:

$$EM_1(dx)^2 = EM_2(dx)^2 = m(dx)/2,$$

then the $M = M_1 + iM_2$ defines a random measure and it is called *complex Gaussian random measure*.

4.1.4.2 Stochastic integration with random measures

By the time we have defined random measures, it is possible, following the steps of the previous sections, to construct a definition of stochastic integration. We limit ourself to the case of deterministic integrands. Then, let f be a complex measurable function defined on the measurable space $(E, \mathcal{E}, m)$ such that $\varphi \in L^2(E, m)$. We suppose that φ is a simple function, that is:

$$\varphi(\xi) = \sum_{i=1}^{n} c_i 1_{A_i}(\xi),\quad \xi \in E,$$

with $\{c_i\}_{i=1,\ldots,n} \in \mathbb{C}$ and $\{A_i\}_{i=1,\ldots,n}$ is a disjoint collection of $\mathcal{E}_0$ elements. In this case we define

$$I(\varphi) = \sum_i^n c_i M(A_i). \tag{4.38}$$

The following properties are straightforward consequence of the random measure definition:

1. $E(I(\varphi)) = 0$, for any simple function $\varphi \in L^2(E, m)$.
2. $E\,|I(\varphi)|^2 = \int_E |\varphi(x)|^2\, m(dx) = \|\varphi\|^2_{L^2(E,m)}$ (isometry).
3. $EI(\varphi)\overline{I(\phi)} = \int_E \varphi(x)\overline{\phi(x)} m(dx)$, for any simple functions φ, ϕ.

We know that for any $f \in L^2(E)$ there exists a sequence of simple functions $\phi_n \in L^2(E, m)$, $n \in \mathbb{N}$, which converges to f in the $L^2(E, m)$-norm.

Prop. 4.3. *The sequence $I(f_n)$, $n \in \mathbb{N}$, converges in $L^2(\Omega, P)$.*

Proof: indeed, the sequence ϕ_n is Cauchy on $L^2(E, m)$. Using the isometry property, for each $n, m \to \infty$:

$$E \mid I(\phi_n) - I(\phi_m) \mid^2 = \|\phi_n - \phi_m\|^2_{L^2(E,m)} \to 0;$$

Thus, the sequence $I(\phi_n)$ is Cauchy on $L^2(\Omega, P)$. Given that $L^2(\Omega, P)$ is a complete space, the sequence of the integrals converges. □

For any $f \in L^2(E, m)$, we define the stochastic integral as the limit:

$$I(f) = \int_E f M(dx) = \lim_{n\to\infty} I(\phi_n). \tag{4.39}$$

Such object is well defined and the properties of the stochastic integral of simple functions still hold.

Example 4.5 (Brownian motion)**.** Let M be a real Gaussian random measure define on $\mathbb{R}$, with control measure the Lebesgue measure m. For any $A \in \mathcal{E}_0$:

$$M(A) \sim N(0, m(A)).$$

Moreover, for any sequence $\{f_1, f_2, \ldots, f_n\} \in L^2(\mathbb{R})$, then the family of r.v.s $\{I(f_1), I(f_2), \ldots, I(f_n)\}$ has Gaussian distribution with zero mean and covariance matrix:

$$\Sigma_{ij} = E(I(f_i)I(f_j)) = \int_{\mathbb{R}} f_i(x) f_j(x) dx.$$

The process

$$B(t) = \int_0^t M(dx), \quad t \geq 0,$$

is a standard Brownian motion. In fact, it is Gaussian with autocovariance function given by $E(B(t)B(s)) = \min(t, s)$.

4.2 One dimensional Ito formula

Example (4.3) shows that the Ito integral basic definition is not very useful if one wants to evaluate a given stochastic integral. This is very like to what happens in classical calculus. In fact, here, if we want explicitly to evaluate some integral, we do not use the basic definition, but rather we use the fundamental theorem of calculus and the chain rule. In stochastic calculus, however, we have not a differentiation theory, but it is possible to define an Ito version of the chain rule called *Ito formula*. We introduce the following class of processes:

Def. 4.8 (Ito processes)**.** Let X_t be a stochastic process, defined on $(\Omega, \mathcal{F}, P)$, such that for any $t \geq 0$:

$$X_t = X(0) + \int_0^t u_s ds + \int_0^t v_s dB_s, \tag{4.40}$$

where $u, v \in \mathcal{W}$. Then, X_t is called (one-dimensional) *Ito process*.

Notation 4.3. Let X_t be an Ito process, then (eq. 4.40) is usually formally written in the shorter differential form:

$$dX_t = u_t dt + v_t dB_t \tag{4.41}$$

For instance, (eq. 4.24) can be written:

$$d\left(\frac{B_t^2}{2}\right) = \frac{1}{2}dt + B_t dB_t. \tag{4.42}$$

We have the following theorem:

Theorem 4.1 (Ito formula)**.** *Let $g(t,x) \in C^2(\mathbb{R}_+ \times \mathbb{R})$ and let X_t be an Ito process of the form:*

$$dX_t = u_t dt + v_t dB_t.$$

Then, the process

$$Y_t = g(t, X_t), \quad t \geq 0,$$

is again an Ito process, and the following Ito formula holds:

$$dY_t = dg(t, X_t) = \left(\partial_t g(t, X_t) + u_t \partial_x f(t, X_t) + \frac{1}{2} v_t^2 \partial_{xx} f(t, X_t)\right) dt + v_t \partial_x f(t, X_t) dB_t, \tag{4.43}$$

or equivalently:

$$dg(t, X_t) = \partial_t g(t, X_t) dt + \partial_x g(t, X_t) dX_t + \frac{1}{2} \partial_{xx} g(t, X_t) d\langle X \rangle_t, \tag{4.44}$$

where $\langle X \rangle_t = \int_0^t v_s^2 ds$ is the quadratic variation of the Ito diffusion.

Example 4.6. Chosing $g(t,x) = \frac{1}{2}x^2$, we immediately recover (eq. 4.42).

Example 4.7. Suppose $g(t,x) = tx$ and $Y_t = tB_t$. Then one has:

$$dY_t = B_t dt + t dB_t,$$

which means that:

$$\int_0^t s dB_s = tB_t - \int_0^t B_s ds. \tag{4.45}$$

The latter equation reminds to the integration by parts formula of the classical calculus, and in fact we have:

Theorem 4.2 (Integration by parts)**.** *Suppose f_t is continuous with bounded variation*[3]*, then:*

$$\int_0^t f_s dB_s = f_t B_t - \int_0^t B_s df_s. \tag{4.46}$$

4.2.1 Stochastic differential equations

The inclusion of random effects in differential equations lead to two distinct classes of equations, for which the solution process have differentiable and non-differentiable paths, respectively.

The first, and simpler, class arises when an ordinary differential equation has random coefficients, a random initial value or is "forced" by a fairly regular stochastic process, or when some combination of these holds. These equations are called *random differential equations* and are solved sample path by sample path as ordinary differential equations. The sample paths of the solution process are indeed differentiable.

The second class occurs when the "forcing" process is an irregular stochastic process such as White noises (see eq. 4.4). These equations are written symbolically as stochastic differential (see Notation 4.3), but have to be interpreted as stochastic integrals. They are called *stochastic differential equations* (SDE). In many applications such equations result for the incorporation of either endogenous or exogenous random fluctuations in the dynamical description of a system.

Example 4.8 (Langevin equation)**.** For example consider the continuous molecular bombardment of a speck of dust on a water surface, which result to Brownian motion. The intensity of this bombardment does not depend on the state variables, for instance position and velocity of the speck. In the one dimensional case, Langevin wrote the equation so termed *Langevin equation*:

$$\frac{dX_t}{dt} = -aX_t + \sigma Z_t \tag{4.47}$$

for the acceleration of the particle. This is just the sum of a retarding frictional force depending linearly on the velocity and the molecular force represented by a white noise process Z_t, with intensity σ independent of the velocity.

We now interpret the Langevin equation (eq. 4.47) symbolically as a stochastic differential equation:

$$dX_t = -aX_t dt + \sigma dB_t, \tag{4.48}$$

namely, in terms of Ito integral:

$$X_t = X_0 - \int_0^t aX_s ds + \int_0^t \sigma dB_s. \tag{4.49}$$

Remark 4.12. Typically, the random forcing of equation (4.47) or (4.48) is called *additive noise*. Example of *multiplicative noise* is given by the equation

$$dX_t = aX_t dt + \sigma X_t dB_t. \tag{4.50}$$

3) A process has bounded variation if it has zero quadratic variation.

Both the cases of the above remark are included in the general form of a (non-autonomous[4]) stochastic differential equation, that is:

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dB_t. \tag{4.51}$$

Consider then (eq. 4.51) in the interval $[t_0, T]$. This equation has to be interpreted as

$$X_t = X_{t_0} + \int_{t_0}^{t} a(s, X_s)ds + \int_{t_0}^{t} \sigma(s, X_s)dB_s, \quad t_0 \le t \le T, \tag{4.52}$$

where the first integral is a Lebesgue (or Riemann) integral for each sample path and the second is an Ito integral.

As with deterministic ordinary and partial differential equations, it is important to know whether a given SDE has a solution, and whether or not it is unique.

Remark 4.13. A solution $\{X(t), \quad t \in [t_0, T]\}$ of the above equation must have properties which ensure that these integrals are meaningful.

Def. 4.9 (Strong and weak solutions)**.** If the version B_t of Brownian motion defined in the filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is given in advance and the solution X_t constructed from it is $\mathcal{F}_t$-adapted, the solution is called a *strong solution*. If we are only given the functions $a(t, x)$ and $\sigma(t, x)$ and ask for a pair of processes (X_t, B_t) such that (eq. 4.52) holds, then the solution X_t (or more precisely (X_t, B_t)) is called a *weak solution*.

A strong solution X_t can be roughly though of as a functional of the initial value X_{t_0} and the given Bm B_s over the subinterval $t_0 \le s \le t$. Given a specific initial value X_{t_0}, the *uniqueness* of solutions of (eq. 4.51) refers to the almost surely equivalence of the processes that satisfy the stochastic integral equation (eq. 4.52). If there is a solution, then there will be a separable version which has, almost surely, continuous sample paths. We consider only these kind of solutions.

The hypothesis of an existence and uniqueness theorem are usually sufficient but not necessary, conditions. Some are quite strong, but can be weakened in several ways. Most of the assumptions concern the coefficients $b, \sigma : [t_0, T] \times \mathbb{R} \to \mathbb{R}$.

Existence and uniqueness conditions

A1. *Measurability*: $a(t, x)$ and $\sigma(t, x)$ are L^2-measurable in $[t_0, T] \times \mathbb{R}$.

A2. *Lipschitz condition*: there exists a constant $K > 0$ such that for any $\in [t_0, T]$ and $x, y \in \mathbb{R}$:

$$|a(t, x) - a(t, y)| \le K|x - y|, \tag{4.53}$$

and

$$|\sigma(t, x) - \sigma(t, y)| \le K|x - y|. \tag{4.54}$$

A3. *Linear growth bound*: there exists a constant $K > 0$ such that for any $\in [t_0, T]$ and $x, y \in \mathbb{R}$:

$$|a(t, x)|^2 \le K^2(1 + |x|^2), \tag{4.55}$$

and

$$|\sigma(t, x)|^2 \le K^2(1 + |x|^2). \tag{4.56}$$

4) The coefficients $a(t, X_t)$ and $\sigma(t, X_t)$ depend on the time t.

A4. Initial value: X_{t_0} is $\mathcal{F}_{t_0}$-measurable with $E(|X_{t_0}|^2) < \infty$.

Theorem 4.3. *Under assumptions A1-A4 the stochastic differential equation (eq. 4.51) has a unique strong solution X_t on $[t_0, T]$.*

Variations of the above theorem are possible with weakened assumptions. The most obvious is to drop the requirement that the initial value X_{t_0} satisfies condition *A4*.

4.2.2 Ito diffusions

Actually, the most important and attractive property of the solutions of stochastic differential equations is that they are usually Markov processes. In fact, in many cases, they are stochastic diffusion processes (Def. 3.12). Under assumptions *A1-A4* the solution X_t of a stochastic differential equation is indeed a Markov process in the interval $[t_0, T]$. For an *autonomous* SDE, namely

$$dX_t = a(X_t)dt + \sigma(X_t)dB_t, \tag{4.57}$$

the solutions X_t are homogeneous Markov processes.

In general, the solution of the stochastic differential equation (eq. 4.51) are diffusion processes, with transition densities satisfying the conditions of (eq. 3.28–3.30) for the drift $a(t,x)$ and the diffusion coefficient $\sigma(t,x)$. So that, we have the following

Theorem 4.4. *Assume that $a(t,x)$ and $\sigma(t,x)$ are continuous functions which satisfy the conditions A2–A4 . Then, for any fixed value X_{t_0}, the unique solution $X(t)$, $t_0 \leq t \leq T$, of the SDE (eq. 4.51) is a diffusion process on $[t_0, T]$ with drift $a(t,x)$ and diffusion coefficient $\sigma(t,x)$.*

Def. 4.10 (Ito diffusion)**.** Such a solution X_t, or more precisely the family of solutions $X_t^{0,x}$ with initial value $X_0^{0,x} = x$ a.s. is often called *Ito diffusion*.

Therefore, the transition densities of an Ito diffusion satisfies the Fokker-Planck equation (eq. 3.33).

Remark 4.14. Observe that, if one let $u(t,y)$ indicate the marginal density function of the Ito diffusion X_t, starting almost surely in $X_0 = x$, then:

$$\partial_t u + \partial_y(au) - \frac{1}{2}\partial_{yy}(\sigma^2 u) = 0, \tag{4.58}$$

with $u(y,0) := u_0(y) = \delta(y - x)$.

We have seen that an Ito diffusion is a stochastic diffusion. But, what about the converse problem. That is, given a diffusion process Y_t, with transition densities satisfying a certain Fokker-Planck equation, is Y_t the solution of some stochastic differential equation? In general the answer is negative provided that the coefficients $a(t,x)$ and $\sigma(t,x)$ satisfy certain smoothness and boundedness conditions. Heuristically, if Y_t is a stochastic diffusion or even if only the coefficients a and σ are given, one can always take the SDE

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dB_t. \tag{4.59}$$

If the coefficients satisfy *A1–A3* and if the initial value is chosen $X_{t_0} = Y_{t_0}$, then there will be a solution X_t for any Wiener process B_t equivalent to the given diffusion Y_t in the sense of finite dimensional distributions. However, in general they will not be sample path equivalent. To guarantee this we must choose the Brownian motion carefully and we must have additional regularity conditions on the coefficients (see [5,2] for more details).

4.2.3 Examples of SDEs

Let $B(t)$, $t \geq 0$, be a one dimensional standard Brownian motion, defined on a certain probability space $(\Omega, \mathcal{F}, P)$.

Example 4.9 (Linear additive-noise inhomogeneous SDEs with constant coefficients)**.** Let $\mu, \sigma, a \in \mathbb{R}$ be given. Consider the following stochastic differential equation:

$$dX_t = (aX_t + \mu)dt + \sigma dB_t, \quad t \geq 0. \tag{4.60}$$

The associated Fokker-Planck equation is:

$$\partial_t u(x,t) + (ax + \mu)\partial_x u(x,t) - \frac{\sigma^2}{2}\partial_{xx} u(x,t) + au(x,t) = 0. \tag{4.61}$$

The process:

$$X_t = e^{at}\left(X_0 + \frac{\mu}{a}(1 - e^{-at}) + \sigma \int_0^t e^{-as} dB_s\right). \tag{4.62}$$

solves (eq. 4.60). In fact, let $y_t = e^{-at}X_t$, then, using Ito formula, one has:

$$dy_t = -ay_t dt + e^{-at}dX_t = \mu e^{-at}dt + \sigma e^{-at}dB_t,$$

which integrated gives the result. We have the following noteworthy cases:

1. When $a = 0$ and $\mu, \sigma \in \mathbb{R}$ we find a *Brownian motion with drift* μ, namely,

$$X_t - X_0 := B^{(\mu,\sigma)}(t) = \mu t + \sigma B_t, \quad t \geq 0. \tag{4.63}$$

   That is, X_t may describe a component of the position of a speck dust in a frictionless medium with starting velocity μ.

2. When $\mu = 0$ and $a = -\alpha$ with $\alpha > 0$ and $\sigma \in \mathbb{R}$ we have:

$$X_t = e^{-\alpha t}\left(X_0 + \sigma \int_0^t e^{\alpha s} dB_s\right), \tag{4.64}$$

   which is the solution of the Langevin equation (eq. 4.48) and may describe the velocity of a particle in a medium with a friction force depending linearly on the velocity.

3. The case $\sigma = 0$ is moot in this context, because reduces to classical equations.

Example 4.10 (Linear multiplicative-noise homogeneous SDEs with constant coefficients)**.** Let $\mu, \sigma \in \mathbb{R}$ be given. We consider the following stochastic differential equation:

$$dX_t = \mu X_t dt + \sigma X_t dB_t, \quad t \geq 0. \tag{4.65}$$

The solution of the above SDE is called *Geometric Brownian motion* and turns out to be:

$$X(t) := S(t) = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right], \quad t \geq 0. \tag{4.66}$$

Its marginal density function satisfies the Fokker-Planck equation:

$$\partial_t u(x,t) = (\sigma^2 - \mu)u(x,t) + (2\sigma^2 - \mu)x\partial_x u(x,t) + \frac{\sigma^2}{2}x^2\partial_{xx} u(x,t), \quad t \geq 0. \tag{4.67}$$

It is easy to show (eq. 4.66). In fact, let $y_t = \log X_t$, then:

$$dy_t = \frac{1}{X_t} dS_t - \frac{1}{2X_t^2} d\langle X \rangle_t = \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dB_t,$$

which integrated gives (eq. 4.66).

Particular cases are:

1. When $\mu = \sigma^2/2$, with have the Ito exponential stochastic differential equation, which, by setting $\sigma = 1$, reads:
$$dX_t = \frac{1}{2} X_t dt + X_t dB_t, \quad X_t = X_0 \exp(B_t), \quad t \geq 0. \tag{4.68}$$
2. Setting $\mu = 0$ and $\sigma = 1$, we have the *drift-free* SDE:
$$dX_t = X_t dB_t, \quad X_t = X_0 \exp\left(B_t - \frac{1}{2} t \right), \quad t \geq 0. \tag{4.69}$$

For further examples we refer the interested reader to [2].

Bibliography

[1] J.L. Doob, *Stochastic Processes*, (Wiley & Sons. 1953).

[2] P.E. Kloeden and E. Platen, *Numerical solutions of Stochastic differential Equations*, (Springer-Verlag, Berlin Heidelberg New York, 1999).

[3] F.B. Knight, Essentials of Brownian Motion, *American Math. Soc* 1981.

[4] D. Lamberton and B. Lapeyre, *Stochastic Calculus Applied to Finance*, (Chapman & Hall, London, 1996).

[5] B. Oksendal, *Stochastic differential equations*, Universitext, Springer-Verlag, Berlin, sixth ed., 2003.

[6] A. Pascucci, *Calcolo stocastico per la finanza*, Springer-Verlag (2007).

[7] M.S. Taqqu, Fractional Brownian motion and long-range dependence, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and Applications of Long-range Dependence*, Birkäuser, Boston - Basel - Berlin, 2003, pp. 5-38.

[8] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, A series of Comprehensive Studies in Mathematics, (Springer Verlag, 293, 1994).

5 Integral representations of some non-Markovian processes

Using the theory of Ito integral introduced in the previous chapter, we want to study some fractional Brownian motion integral representations based on it. As it turns out, a fBm may be seen as a sort of "weighted sum" with respect to an ordinary Brownian motion. Finally, we will also study some useful integral representations of FARIMA processes. Thanks to these representations, we will be able to prove (eq. 3.63) and to evaluate the spectral density function of FARIMA processes.

5.1 Moving average representation of fBm

Let $\{B(t),\ t \in \mathbb{R}\}$ be a Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. We introduce the following notation: $x_+ = \max(x, 0)$. The proposition below provides a first representation of fractional Brownian motion in terms of stochastic integrals of Brownian motion.

Prop. 5.1 (Moving average representation)**.** *Let $\{B_H(t),\ t \geq 0\}$ be a standard one-sided fractional Brownian motion with $0 < H < 1$. Then, for any $t \geq 0$*

$$B_H(t) = \int_{\mathbb{R}} f_t(x) dB(x) = \frac{1}{C_1(H)} \int_{\mathbb{R}} \left((t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}} \right) dB(x), \tag{5.1}$$

where:

$$C_1(H) = \left(\int_0^\infty \left((1+x)^{H-\frac{1}{2}} - x^{H-\frac{1}{2}} \right)^2 dx + \frac{1}{2H} \right)^{1/2} \tag{5.2}$$

$$= \frac{\Gamma(H+1/2)}{(\Gamma(2H+1)\sin \pi H)^{1/2}}. \tag{5.3}$$

This representation is called **moving average representation**.

Remark 5.1. We observe that for $H = 1/2$ one obtains,

$$B(t) = \int_0^t dB_x, \quad t \geq 0,$$

that is a standard Bm (see Exa. 4.2).

The function $f_t(x)$ is called representation *kernel*. Moreover, because the integration is carried up to time $t \geq 0$, representation (5.1) is called *causal*.

Proof: in order for (eq. 5.1) to be well defined is necessary that, for any $t \geq 0$, $f_t(x) \in L^2(\mathbb{R})$. Then, consider $H \neq 1/2$, one has:

- $f_t(x) \sim (H - 1/2)(-x)^{H-3/2}$, for $x \to -\infty$;
- $f_t(x) \sim (t-x)_+^{H-1/2}$, for $x \to t$, and the same for $x \to 0$.

Therefore, for any $0 < H < 1$, $\int_{\mathbb{R}} f_t^2(x) dx < \infty$. Namely, $f_t(x) \in L^2(\mathbb{R})$.

We have now to show that a process $\{X(t),\ t \geq 0\}$, defined by (5.1), is indeed a standard fBm. We use (**Prop.** 3.7):

i) Obviously, $X_0 = 0$ a.s.

ii) Moreover, $X(t)$, $t \geq 0$, is a Gaussian process with zero mean and variance:

$$E(X(t)^2) = \frac{1}{C_1^2(H)} \left[\int_{-\infty}^{0} \left((t-x)^{H-1/2} - (-x)^{H-1/2} \right)^2 dx + \int_0^t (t-x)^{2H-1} dx \right]$$

(with the change of variables $x = tx'$)

$$= \frac{t^{2H}}{C_1^2(H)} \left[\int_{-\infty}^{0} \left((1-x)^{H-1/2} - (-x)^{H-1/2} \right)^2 dx + \int_0^1 (1-x)^{2H-1} dx \right] = t^{2H}.$$

iii) Finally, X_t is an **si**-process. In order to show this, we observe that for any $t, s \geq 0$, one has

$$E(X_t - X_s)^2 = \frac{1}{C_1^2(H)} \int_{\mathbb{R}} \left((t-x)_+^{H-1/2} - (s-x)_+^{H-1/2} \right)^2 dx$$

$$= \frac{1}{C_1^2(H)} \int_{\mathbb{R}} \left((t-s-x)_+^{H-1/2} - (-x)_+^{H-1/2} \right)^2 dx = E(X(t-s))^2;$$

thus,

$$E(X_t X_s) = \frac{1}{2} \left[t^{2H} + s^{2H} - |t-s|^{2H} \right]. \tag{5.4}$$

This implies that $X(t)$ has stationary increments. In fact, we have to show that, for any real vector $\{y_1, \ldots, y_n\}$ and $h > 0$

$$\exp \left[i \sum_{k=1}^{n} y_k \{ X(t_k + h) - X(h) \} \right] = \exp \left[i \sum_{k=1}^{n} y_k X(t_k) \right].$$

Because that X is a zero-mean Gaussian process we have (see eq. 3.3)

$$\exp \left[\sum_{ij} y_i y_j E[X(t_i + h) - X(h)][X(t_j + h) - X(h)] \right] = \exp \left[\sum_{ij} y_i y_j E(X(t_i) X(t_j)) \right],$$

which is verified because, using (eq. 5.4), one finds

$$E(X_{t_i+h} X_{t_j+h}) - E(X_{t_i+h} X_h) - E(X_{t_j+h} X_h) + E(X_h)^2 = \frac{1}{2} \left[t_i^{2H} + t_j^{2H} - |t_i - t_j|^{2H} \right] = E(X_{t_i} X_{t_j}).$$

Then, $X(t)$, $t \geq 0$ is a standard fBm. □

Remark 5.2. Heuristically, we observe that the process $\{X(t), \quad t \geq 0\}$ defined by (eq. 5.1), is indeed self-similar with scaling exponent H. In fact,

$$X(at) = \frac{1}{C_1(H)} \int_{\mathbb{R}} \left((at-x)_+^{H-1/2} - (-x)_+^{H-1/2} \right) dB(x)$$

(with the change of variables $x = ax'$)

$$= \frac{1}{C_1(H)} \int_{\mathbb{R}} \left((at-ax)_+^{H-1/2} - (-ax)_+^{H-1/2} \right) dB(ax)$$

$$\stackrel{d}{=} \frac{a^{H-1/2} a^{1/2}}{C_1(H)} \int_{\mathbb{R}} \left((t-x)_+^{H-1/2} - (-x)_+^{H-1/2} \right) dB(x) = a^H X_t.$$

We can provide several fBm representations. In fact,

Prop. 5.2. *For any couple of real numbers a, b, the stochastic integral:*

$$\int_{\mathbb{R}} \left[a \left((t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}} \right) + b \left((t-x)_-^{H-\frac{1}{2}} - (-x)_-^{H-\frac{1}{2}} \right) \right] dB(x), \tag{5.5}$$

where $x_- = -\min(x, 0)$, defines (up to a multiplicative constant) a fractional Brownian motion representation.

Remark 5.3. Setting $a = \dfrac{1}{C_1(H)}$ and $b = 0$, we recover the moving average representation.

Wheno $a = b = 1$, (eq. 5.5) reduces to the so called *well-balanced* representation:

$$B_H(t) = \int_{\mathbb{R}} \left(|t-x|^{H-1/2} - |x|^{H-1/2} \right) dB(x). \tag{5.6}$$

5.2 Spectral representation of fBm

Let f_t be the kernel of the moving average representation. Namely,

$$f_t(x) = \frac{1}{C_1(H)} \left((t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}} \right), \quad t \geq 0.$$

Since $f_t \in L^2(\mathbb{R})$, one can evaluate its Fourier transform:

$$\widetilde{f}_t(\nu) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ix\nu} f_t(x) dx; \quad f_t(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ix\nu} \widetilde{f}_t(\nu) d\nu. \tag{5.7}$$

Because $\widetilde{f}_t \in L^2(\mathbb{R})$, it is possible to define an integral representation of fBm based on $\widetilde{f}_t(\nu)$:

$$\widetilde{X}_t = \int \widetilde{f}_t(\nu) \widetilde{M}(d\nu), \tag{5.8}$$

where $\widetilde{M}$ is a suitable complex Gaussian random measure, which satisfies:

$$\widetilde{M}(d\nu) = \overline{\widetilde{M}(-d\nu)}, \tag{5.9}$$

$$E\, |\widetilde{M}(d\nu)|^2 = d\nu. \tag{5.10}$$

Remark 5.4. The previous conditions, together with the hermitian property of the Fourier transform, i.e. $\overline{\widetilde{f}_t(\nu)} = \widetilde{f}_t(-\nu)$, ensure that the process $\widetilde{X}_t$ is a real valued process.

Remark 5.5. We observe that, if $\widetilde{M}$ is a complex Gaussian measure as above, then for any $h > 0$,

$$\widetilde{M}_*(d\nu) = e^{i\nu h} \widetilde{M}(d\nu),$$

is another Gaussian measure which still verifies (eq. 5.9) and (eq. 5.10).

In order to define the measure $\widetilde{M}$, we introduce two independent Brownian motions $B_t^{(1)}$ e $B_t^{(2)}$, and we define:

$$M^{(i)}(A) = \frac{1}{\sqrt{2}} \int_A dB^{(i)}(\nu), \quad A \subset \mathbb{R}_+;$$

while, for any $A \subset \mathbb{R}_-$:

$$M^{(1)}(A) = M^{(1)}(-A), \quad M^{(2)}(A) = -M^{(2)}(-A).$$

Then, we define:

$$\widetilde{M} = M^{(1)} + iM^{(2)}. \tag{5.11}$$

The following proposition holds

Prop. 5.3 (Spectral representation of fBm)**.** *Let $\widetilde{M}$ be defined as above and let $\widetilde{f}_t$ be the Fourier transform of a fBm representation kernel f_t, for instance (eq. 5.1). Then, the process*

$$\widetilde{X}_t = \int_{\mathbb{R}} \widetilde{f}_t(\nu)\widetilde{M}(d\nu), \tag{5.12}$$

is a fractional Brownian motion.

This proposition introduces the so called *spectral representation* of fBm.

Proof: $\widetilde{X}_t$ is a real valued (Remark 5.4) zero mean Gaussian process. Furthermore, let $X_t = \int f_t(x)dB(x)$, $t \geq 0$, be a fBm of order H, then by using Parseval theorem:

$$E(X_{t_1}X_{t_2}) = \int_{\mathbb{R}} f_{t_1}(x)f_{t_2}(x)dx = \int_{\mathbb{R}} \widetilde{f}_{t_1}(\nu)\overline{\widetilde{f}_{t_2}}(\nu)d\nu = E(\widetilde{X}_{t_1}\widetilde{X}_{t_2}).$$

Therefore, $\widetilde{X}$ is a zero mean Gaussian process with the covariance of a fBm of order H. Then, by Corollary 3.2, $\widetilde{X}_t$ is a fBm. □

The following proposition provides an explicit example of $\widetilde{f}_t$ which is obtained starting from the moving average representation (see Remark 7.46 with $\kappa = H - 1/2$).

Prop. 5.4. *Let $0 < H < 1$ and let $B_H(t)$, $t \geq 0$, be a standard fBm. Thus,*

$$B_H(t) = \int \widetilde{f}_t(\nu)\widetilde{M}(d\nu) = \frac{1}{C_2(H)} \int_{\mathbb{R}} \frac{e^{i\nu t} - 1}{i\nu} \, |\nu|^{-(H-\frac{1}{2})} \, \widetilde{M}(d\nu); \tag{5.13}$$

where

$$C_2^2(H) = 8 \int_0^\infty \frac{\sin^2(\nu/2)}{\nu^{2H+1}} d\nu = \frac{\pi}{H\Gamma(2H)\sin \pi H}. \tag{5.14}$$

Proof: $\widetilde{f}_t(\nu) \in L^2(\mathbb{R})$. In fact,

- $\widetilde{f}_t(\nu) \sim |\nu|^{1/2-H}$, as $\nu \to 0$;
- $\widetilde{f}_t(\nu) \sim |\nu|^{-1/2-H}$, as $\nu \to \infty$.

That is, for any $0 < H < 1$, $\int_{\mathbb{R}} |\widetilde{f}_t(\nu)|^2 \, d\nu < \infty$.

Let $\widetilde{X}(t)$, $t \geq 0$, be a process defined by (eq 5.13). Then, $\widetilde{X}(t)$ is real valued, Gaussian and has zero mean. Moreover, it is easy to see that (eq. 5.14) makes the process standard. We want to show that $X(t)$ is self-similar of index H. That is, for any t_j and t_l greater than zero and $a > 0$,

$$E\left(X(at_j)X(at_l)\right) = a^{2H}E\left(X(t_j)X(t_l)\right).$$

Then, by definition

$$\int_{\mathbb{R}} \frac{\left(e^{iνat_j}-1\right)\left(e^{-iνat_l}-1\right)}{|ν|^{2H+1}}dν = a^{2H}\int_{\mathbb{R}} \frac{\left(e^{iνt_j}-1\right)\left(e^{-iνt_l}-1\right)}{|ν|^{2H+1}}dν,$$

which is verified after a simple change of variables. In the same way, in order to show the stationarity of the increments, we just have to verify that, for any $t_i, t_j \geq 0$ and $h > 0$:

$$E\left[(X(t_i+h)-X(h))(X(t_j+h)-X(h))\right] = E\left[X(t_i)X(t_j)\right].$$

But, this is obviously true because

$$X(t_i+h)-X(h) = e^{iνh}\frac{1}{C_2(H)}\int_{\mathbb{R}} \frac{e^{iνt_i}-1}{iν}|ν|^{-(H-\frac{1}{2})}\widetilde{d}M(ν).$$

Therefore, we have shown that $\widetilde{X}(t)$, $t \geq 0$, is actually a standard fBm of order H. □.

The following proposition provides the relationship between $C_1(H)$ and $C_2(H)$:

Prop. 5.5. *One has:*

$$\frac{C_1(H)}{C_2(H)} = \frac{\Gamma(H+1/2)}{\sqrt{2π}}. \tag{5.15}$$

Remark 5.6. We observe that, setting $\widetilde{M}(dν) = dB_1(ν) + idB_2(ν)$ with:

$$dB_1(-ν) = dB_1(ν), \;\; dB_2(-ν) = -dB_2(ν),$$

then, (eq. 5.13) is equal to:

$$\frac{2}{C_2(H)}\left[\int_0^\infty \frac{\sin νt}{ν}\,|ν|^{-H+\frac{1}{2}}\,dB_1(ν) - \int_0^\infty \frac{1-\cos νt}{ν}\,|ν|^{-H+\frac{1}{2}}\,dB_2(ν)\right]. \tag{5.16}$$

Remark 5.7 (Brownian motion spectral representation)**.** The spectral representation of Bm is obtained from (eq. 5.13) by setting $H = 1/2$:

$$B_t = \frac{1}{\sqrt{2π}}\int_{\mathbb{R}} \frac{e^{itν}-1}{iν}d\widetilde{B}(ν), \tag{5.17}$$

where $d\widetilde{B}(ν) = dB_1(ν) + idB_2(ν)$.

5.2.1 Spectral density of fractional Brownian motion

The representation (5.13) is proper to Gaussian stationary increments processes. Indeed, suppose that $Y = \{Y_t\}_{t\geq 0}$ is a Gaussian real valued stationary process. Moreover, let $d\widetilde{B}(ν)$ be a complex Gaussian measure such as in (Remark 5.6). A possible spectral representation of Y_t could be:

$$Y_t = \int_{\mathbb{R}} e^{itν}\widetilde{g}(ν)d\widetilde{B}(ν), \;\; t \geq 0, \tag{5.18}$$

where $\overline{\widetilde{g}(\nu)} = \widetilde{g}(-\nu)$. It is indeed easy to show that such a process is stationary. In fact, the covariance matrix $E(Y_t Y_s)$ depends only on the difference $t - s$.

Lemma 5.1. *The process*

$$X_t = \int_0^t Y_s ds, \quad t \geq 0,$$

has stationary increments.

Proof: X is Gaussian with zero mean starting almost surely in zero. Moreover,

$$X(t+h) - X(h) = \int_h^{t+h} Y_s ds = \int_0^t Y_{s+h} ds.$$

Therefore, for any $t_1, t_2 \geq 0$, one has:

$$E\left[(X(t_1+h) - X(h))(X(t_2+h) - X(h))\right] = E \int_0^{t_1} Y_{s_1+h} ds_1 \int_0^{t_2} Y_{s_2+h} ds_2$$

$$= \int_0^{t_1} \int_0^{t_2} E(Y_{s_1+h} Y_{s_2+h}) ds_1 ds_2 = \int_0^{t_1} \int_0^{t_2} E(Y_{s_1} Y_{s_2}) ds_1 ds_2 = E(X_{t_1} X_{t_2}). \ \square$$

Using the representation (5.18) of Y, one finds:

$$X_t = \int_0^t \left[\int_{\mathbb{R}} e^{is\nu} \widetilde{g}(\nu) d\widetilde{B}(\nu) \right] ds = \int_{\mathbb{R}} \left[\int_0^t e^{is\nu} ds \right] \widetilde{g}(\nu) d\widetilde{B}(\nu)$$

$$= \int_{\mathbb{R}} \frac{e^{it\nu} - 1}{i\nu} \widetilde{g}(\nu) d\widetilde{B}(\nu). \tag{5.19}$$

We introduce the following definition.

Def. 5.1. The function

$$S(\nu) = \left| \frac{\widetilde{g}(\nu)}{i\nu} \right|^2 = \left| \frac{\widetilde{g}(\nu)}{\nu} \right|^2,$$

which appears in (eq. 5.19), is called *spectral density* of the stationary increments process X.

Remark 5.8. Observe that the function defined above is not proper a spectral density function, which is defined for stationary processes as the Fourier transform of the autocovariance function.

From the definition above and from (eq. 5.13), the spectral density of a standard fBm is:

$$S_H(\nu) = \left(\frac{|\nu|^{-H+\frac{1}{2}}}{C_2(H)\nu} \right)^2 = \frac{H\Gamma(2H)\sin \pi H}{\pi} |\nu|^{-2H-1}. \tag{5.20}$$

By setting $H = 1/2$, one finds the spectral density of a standard Brownian motion:

$$S_{1/2}(\nu) = \frac{1}{2\pi\nu^2}. \tag{5.21}$$

5.3 Integral representation of fractional Gaussian noise

Let $\{B_H(t)\}_{t\geq 0}$ be a fractional Brownian motion of order H. We consider the following stationary process:

$$z_t = B_H(t+1) - B_H(t), \quad t \geq 0. \tag{5.22}$$

Remark 5.9. Observe that the process z_t, $t \geq 0$, differs from the fGn of Definition 3.15 because the parameter t is taken continuous whereas the time step is fixed to one. We refer to z_t as a sort of "continuous time" fGn.

5.3.1 Time domain representation

Starting from the moving average representation of B_H (eq. 5.1), we obtain the corrispondent representation of z_t:

$$z_t = \frac{1}{C_1(H)} \int_{\mathbb{R}} (t+1-x)_+^{H-\frac{1}{2}} - (t-x)_+^{H-\frac{1}{2}} dB(x), \quad t \geq 0. \tag{5.23}$$

Remark 5.10. By setting $H = 1/2$, one has:

$$z_t = \int_t^{t+1} dB(x) = B(t+1) - B(t).$$

5.3.2 Spectral representation

Starting from the spectral representation of B_H (eq. 5.13) with $E(B_H(1)^2) = \sigma^2$, we obtain the corresponding spectral representation of z_t with variance σ^2:

$$z_t = \frac{\sigma}{C_2(H)} \int_{\mathbb{R}} e^{i\nu t} \frac{e^{i\nu} - 1}{i\nu} \, |\nu|^{1/2-H} \, d\widetilde{B}(\nu), \quad t \geq 0. \tag{5.24}$$

Because z_t is stationary, the autocovariance function $E(z_{t_1} z_{t_2})$ depends only of $|t_2 - t_1|$. One has:

$$\gamma(t) = E(z_0 z_t) = \frac{\sigma^2}{C_2(H)^2} \int_{\mathbb{R}} e^{i\nu t} \frac{\left|e^{i\nu} - 1\right|^2}{|\nu|^{2H+1}} d\nu. \tag{5.25}$$

Thus, the spectral density function of z_t is:

$$f_z(\nu) = \frac{\sigma^2}{C_2(H)^2} \frac{\left|e^{i\nu} - 1\right|^2}{|\nu|^{2H+1}} = \frac{\sigma^2}{C_2(H)^2} \left(4 \sin^2 \frac{\nu}{2}\right) |\nu|^{-2H-1}. \tag{5.26}$$

Remark 5.11. The covariance function can then be rewritten as:

$$\gamma(t) = \frac{8\sigma^2}{C_2(H)^2} \int_0^\infty \cos \nu t \left(\sin^2 \frac{\nu}{2}\right) \nu^{-2H-1} d\nu \tag{5.27}$$

Moreover, because

$$\sigma^2 = \gamma(0) = \int f_z(\nu) d\nu,$$

then, one has

$$\gamma(t) = \sigma^2 \frac{E z_0 z_t}{\sigma^2} = \sigma^2 \frac{\int e^{i\nu t} f_z(\nu) d\nu}{\int f_z(\nu) d\nu}$$

$$\sigma^2 \frac{\int_0^\infty \cos \nu t \left(\sin^2 \frac{\nu}{2}\right) \nu^{-2H-1} d\nu}{\int_0^\infty \left(\sin^2 \frac{\nu}{2}\right) \nu^{-2H-1} d\nu}.$$

5.3.3 Spectral density of fGn

Let us consider now the fGn process $\{Z_k,\ k = 0, 1, 2, \dots\}$ as defined in (Def. 3.15). The covariance is obtained from (eq. 5.25) just evaluating $\gamma(t)$ on $k \in \mathbb{Z}_+$:

$$\gamma(k) = \int_{\mathbb{R}} e^{i\nu k} f_z(\nu) d\nu = \int_{-\pi}^{\pi} e^{i\nu k} \sum_{k\in\mathbb{Z}} f_z(\nu + 2\pi k) d\nu.$$

Using (eq. 5.26) one finds:

$$f_Z(\nu) = \frac{\sigma^2}{C_2(H)^2} \left\{ \frac{\left|e^{i\nu} - 1\right|^2}{|\nu|^{2H+1}} + \sum_{k\neq 0} \frac{\left|e^{i\nu} - 1\right|^2}{|\nu + 2k\pi|^{2H+1}} \right\} \tag{5.28}$$

$$= c_f 2(1 - \cos\nu) \sum_{k\in\mathbb{Z}} |\nu + 2k\pi|^{-2H-1},\ \ \nu \in [-\pi, \pi], \tag{5.29}$$

where:

$$c_f = \frac{\sigma^2}{C_2(H)^2} = \sigma^2 (2\pi)^{-1} \sin(\pi H)\Gamma(2H + 1). \tag{5.30}$$

Remark 5.12. We observe that

$$f_Z(\nu) \sim c_f\, |\nu|^{1-2H},\ \ \nu \to 0. \tag{5.31}$$

Then, $f_Z(\nu)$ diverges as $\nu \to 0$ for any $H > 1/2$.

5.4 Integral representations of linear processes

We introduce now some useful integral representation of stationary linear processes. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Moreover, let $(E, \mathcal{E}, m)$ be a measurable space and $M = \{M(A),\ \ A \in \mathcal{E}_0\}$ a random measure defined on it. We introduce the following common terminology:

i. The measurable space $(E, \mathcal{E})$ is called *frequency space* $\nu \in E =]-\pi, \pi]$.

ii. The measure M is called *spectral random measure* and it is denoted with Z.

iii. The control measure m is indicated with F and it is termed *spectral measure*.

If F is absolutely continuous with respect to the Lebesgue measure of E, then we denote with $f(\nu) = dF(\nu)/d\nu$ the corresponding *spectral density function*.

5.4.1 White noise spectral representation

Let Z be a spectral random measure with spectral measure $F(d\nu) = \frac{\sigma^2}{2\pi} d\nu$. We have the following,

Prop. 5.6. *The random variables*

$$\epsilon_n = \int_{-\pi}^{\pi} e^{in\nu} dZ(\nu),\ \ n \in \mathbb{Z}, \tag{5.32}$$

are independent with zero mean and variance σ^2. Moreover, if $Z = Z_1 + iZ_2$ is such that $Z_1(d\nu) = Z_1(-d\nu)$ and $Z_2(d\nu) = -Z_2(-d\nu)$, then $\{\epsilon_n\}_{n\in\mathbb{Z}}$ defines a discrete time real valued stochastic process.

Proof: using the properties of the stochastic integral, one has:

$$E(\epsilon_n \overline{\epsilon}_m) = \int_{-\pi}^{\pi} e^{i(n-m)\nu} \frac{\sigma^2}{2\pi} d\nu = \begin{cases} \sigma^2, & n = m, \\ 0, & n \neq m. \end{cases}$$

If $Z = Z_1 + iZ_2$ with $Z_1(d\nu) = Z_1(-d\nu)$ and $Z_2(d\nu) = -Z_2(-d\nu)$, then:

$$\overline{\epsilon_n} = \int_{-\pi}^{\pi} e^{-in\nu} \overline{Z(d\nu)} = \int_{-\pi}^{\pi} e^{-in\nu} Z(-d\nu) = \int_{-\pi}^{\pi} e^{in\nu} Z(d\nu) = \epsilon_n. \ \square$$

Equation (5.32) defines the spectral representation of a discrete-time white noise with autocovariance:

$$\gamma(k) = E(\epsilon_{t+k}\epsilon_t) = \int_{-\pi}^{\pi} e^{ik\nu} F(d\nu) = \int_{-\pi}^{\pi} e^{ik\nu} \frac{\sigma^2}{2\pi} d\nu = \begin{cases} \sigma^2, & k = 0, \\ 0, & k \neq 0. \end{cases} \tag{5.33}$$

Remark 5.13. Let B be the lag operator. Then,

$$B\epsilon_t = \epsilon_{t-1} = \int_{-\pi}^{\pi} e^{i(t-1)\nu} Z(d\nu) = \int_{-\pi}^{\pi} e^{it\nu} (e^{-i\nu} Z(d\nu)).$$

Therefore, the action of the backward operator B corresponds to the modification of the spectral random measure by a factor $e^{-i\nu}$.

5.4.2 Spectral representation of a linear process

Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a linear process. Namely,

$$X_t = \sum_{j=-\infty}^{\infty} c_j \epsilon_{t-j} = \sum_{j=-\infty}^{\infty} c_j B^j \epsilon_t, \ \ j \in \mathbb{Z},$$

where $\sum_{j=-\infty}^{\infty} c_j^2 < \infty$ and ϵ_t, $t \in \mathbb{Z}$, is a zero-mean white noise process with variance σ^2. By Remark 5.13 we have the following:

Prop. 5.7. *The process X_t has the following spectral representation:*

$$X_t = \int_{-\pi}^{\pi} e^{it\nu} (\sum_j c_j e^{-ij\nu}) Z(d\nu), \ \ t \in \mathbb{Z}, \tag{5.34}$$

where Z is the random spectral measure of the white noise process.

From the proposition above follows the corollary below:

Cor. 5.1. *The spectral measure and the autocovariance function of a linear process X are:*

$$F_X(d\nu) = f_X(\nu)d\nu = \Big| \sum_j c_j e^{-ij\nu} \Big|^2 \frac{\sigma^2}{2\pi} d\nu, \tag{5.35}$$

$$\gamma_X(k) = E(X_{t+k}X_t) = \int_{-\pi}^{\pi} e^{ik\nu} F_X(d\nu) = \int_{-\pi}^{\pi} e^{ik\nu} \Big| \sum_j c_j e^{-ij\nu} \Big|^2 \frac{\sigma^2}{2\pi} d\nu. \tag{5.36}$$

Example 5.1 (Spectral density of ARMA processes)**.** From the results above, we can write down the spectral density function of an ARMA(p,q) process

$$\Phi_p(B)X_t = \Theta_q(B)\epsilon_t.$$

In fact, from (eq. 5.35) one has:

$$f_X(\nu) = \left|\frac{\Theta_q(e^{-i\nu})}{\Phi_p(e^{-i\nu})}\right|^2 \frac{\sigma^2}{2\pi}. \tag{5.37}$$

5.4.3 Spectral representation of FARIMA processes

Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a FARIMA$(0,d,0)$ with $d < 1/2$ process. Namely,

$$X_t = (1-B)^{-d}\epsilon_d = \left(\sum_{j=0}^{\infty} b_j B^j\right)\epsilon_t,$$

where b_j, $j \geq 0$, is given by (eq. 3.64) and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise with zero mean and variance σ^2. From (eq. 5.34) one finds the spectral representation of X:

$$X_t = \int_{-\pi}^{\pi} e^{it\nu}\left(\sum_j b_j e^{-ij\nu}\right) Z(d\nu) = \int_{-\pi}^{\pi} e^{it\nu}(1-e^{-i\nu})^{-d} Z(d\nu). \tag{5.38}$$

Then, the FARIMA$(0,d,0)$ autocovariance function is:

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\nu}\left|1-e^{-i\nu}\right|^{-2d}\frac{\sigma^2}{2\pi}d\nu. \tag{5.39}$$

5.4.4 Spectral density function of a FARIMA process

Using the equations above, it is easy to write down the expression of the spectral density function of a FARIMA$(0,d,0)$ process:

$$f(\nu) = \frac{\sigma^2}{2\pi}\left|1-e^{-i\nu}\right|^{-2d} = \frac{\sigma^2}{2\pi}(2(1-\cos\nu))^{-d} = \frac{\sigma^2}{2\pi}(4\sin^2\frac{\nu}{2})^{-d}. \tag{5.40}$$

Remark 5.14. We observe that, like the spectral density of the fGn (see eq. 5.31), the spectral density function of the FARIMA process blows up at the origin if $d > 0$:

$$f(\nu) \simeq \frac{\sigma^2}{2\pi}\nu^{-2d}, \;\; \nu \to 0. \tag{5.41}$$

Moreover, the condition $d < 1/2$ corresponds actually to the integrability condition of the spectral density function $f(\nu)$ at the origin.

In the general case, consider a FARIMA(p,d,q). The spectral representation is just:

$$X_t = \int_{-\pi}^{\pi} e^{it\nu}(1-e^{-i\nu})^{-d}\frac{\Theta_q(e^{-i\nu})}{\Phi_p(e^{-i\nu})}Z(d\nu). \tag{5.42}$$

Then, the spectral density function is:

$$f(\nu) = \frac{\sigma^2}{2\pi}(4\sin^2\frac{\nu}{2})^{-d}\left|\frac{\Theta_q(e^{-i\nu})}{\Phi_p(e^{-i\nu})}\right|^2. \tag{5.43}$$

Remark 5.15. The range $0 < d < 1/2$ still corresponds to the blow up of the spectral density at the origin.

Remark 5.16. We can finally evaluate the autocovariance function (eq. 3.63) of a FARIMA(0,*d*,0) process by using (eq. 5.40):

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\nu} f(\nu) d\nu = \frac{\sigma^2}{\pi} \int_0^{\pi} \cos k\nu \left(2 \sin \frac{\nu}{2}\right)^{-2d} d\nu$$

(by eq. 5.45)

$$= \sigma^2 \frac{(-1)^k \gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}, \tag{5.44}$$

where we have used the relation (see [1]):

$$\int_0^{\pi/2} \cos hx (\sin x)^{\nu-1} dx = \frac{\pi \cos(h\pi/2)\Gamma(\nu) 2^{1-\nu}}{\Gamma\left(\frac{\nu+h+1}{2}\right)\Gamma\left(\frac{\nu-h+1}{2}\right)}. \tag{5.45}$$

Using the reflection formula for the Γ-function (eq. 3.65), one finds:

$$\Gamma(1-k-d) = (-1)^k \frac{\Gamma(d)\Gamma(1-d)}{\Gamma(k+d)}.$$

Thus, equation (5.44) becomes

$$\gamma(k) = \sigma^2 \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(d)\Gamma(1-d)},$$

which is (eq. 3.63). Moreover, by setting $k = 0$ one finds (eq. 3.62).

Bibliography

[1] P.J. Brockwell, R.A. Davis, *Time series: theory and methods*, Springer-Verlag, second edition, 1991.

[2] L. Decreusefond and A. S.Üstünel, Fractional Brownian motion: theory and applications, *EASIM Proceedings of Fractional Differential Systems: Models, Methods and Applications* pp. 75-86, Vol 5 (1998). (Electronic book)

[3] L. Feyel and A. de La Pradelle, On fractional Brownian processes, *Potential Analysis* **10** No (1999) 273-288.

[4] B.B. Mandelbrot, Note on the definition and stationarity of fractional Gaussian noise, *Journal of Hydrology* **30** (1976) 407-409.

[5] B.B. Mandelbrot, Gaussian Self-Affinity and Fractals, *Springer-Verlag, New York, 2002*, pp. 654.

[6] B.B. Mandelbrot and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10** (1968) 422-433.

[7] D. Marinucci, P.M. Robinson Alternative forms of fractional Brownian motion *Journal of Statistical Planning and Inference* **80** (1999) 111-122.

[8] M.S. Taqqu, Fractional Brownian motion and long-range dependence, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and Applications of Long-range Dependence*, Birkäuser, Boston - Basel - Berlin, 2003, pp. 5-38.

6 Long-range dependence

We want now to introduce the fundamental concept of long-range dependence (shortly LRD). We have already observed that, typically, stationary Markovian processes have autocovariance function which decrease exponentially to zero as the time lag becomes infinity. This is indeed obvious in the case of Gaussian processes because of equation (3.15). For instance, the AR(1) process (see Example 3.16) and the Ornstein-Ulhenbeck process (Example 3.3).

In contrast, non-Markovian processes, like fractional Gaussian noise and FARIMA processes, have shown an asymptotic power-like behavior of their autocovariance function. In fact, let $B_H = \{B_H(t)\}_{t\geq 0}$ be a fBm, defined on the probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, and let $Z = \{Z_t\}_{t\in\mathbb{Z}}$ be the corresponding fGn. From (eq. 3.43) we observe that the Z autocovariance function $\gamma(k)$ tends to zero, as $k \to \infty$, like a power function. Moreover, when $1/2 < H < 1$, $\gamma(k)$ tends to zero so slowly that it does not result integrable, that is the sum $\sum_k \gamma(k)$ diverges.

We shall say that the fGn $Z = \{Z_t\}_{t\in\mathbb{Z}}$ with $1/2 < H < 1$ exhibits *long-range dependence*. In view of this first rough definition, from equation (3.67) it is clear that also a stationary FARIMA process with $d > 0$ manifests long-range dependence.

More precisely, we give the following definition:

Def. 6.1 (Long-range dependence)**.** Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a stationary stochastic process defined on a certain probability space. We shall say that X exhibits long-range dependence (or strong dependence or long-memory) if:

1. the X autocovariance function tends to zero with a power-like behavior (power-law) at infinity:
$$\gamma(k) \sim k^{\alpha}, \quad k \to \infty.$$
2. $\sum_{k=-\infty}^{\infty} \gamma(k) = \infty.$

Example 6.1. Let us consider a fGn Z in the antipersistent domain. That is, $0 < H < 1/2$. Then, because $\sum_{k=1}^{\infty} k^{2H-2} < \infty$, one has
$$\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty.$$
Moreover, $\sum_{k=-\infty}^{\infty} \gamma(k) = 0$, because of the **telescopic**[1] nature of $\gamma(k)$.

Summarizing, we have that:

Prop. 6.1. *Let $Z = \{Z_t\}_{t\in\mathbb{Z}}$ be a fractional Gaussian noise of order H. Then,*

1. *if $0 < H < 1/2$ (antipersistent domain) the process Z does not exhibits long-range dependence.*
2. *if $1/2 < H < 1$ (persistent domain) the process Z exhibits long-range dependence.*

1) A sum where consecutive terms cancel each other.

Prop. 6.2. *Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a stationary FARIMA$(0,d,0)$ process. Then, X exhibits long-range dependence if and only if $0 < d < 1/2$. In fact, by (eq. 3.67), $\sum_k \gamma(k) = \infty$ if and only if $\sum_{k=1}^{\infty} k^{2d-1} = \infty$, that is $d > 0$. The same result is still true for a general FARIMA process (see Remark 6.2).*

Remark 6.1. General FARIMA(p,d,q) processes provide a powerful modeling instrument. In fact, they make possible to model, not only the long-range correlation structure, but also the short-range correlation structure. In other words, due to the ARMA(p,q) term, using FARIMA processes one can model the whole frequency power spectrum (see below). Therefore, they are in general preferred to the simpler but less flexible fGn.

6.1 Long-range dependence in the frequency domain

Suppose that the autocovariance function $\gamma(k)$ of a stationary process behaves like $|k|^{2H-2}$, up to a multiplicative constant, for $k \to \infty$. Moreover, let $\{f(\nu)\}_{\nu\in[-\pi,\pi]}$ be the spectral density function of $\gamma(k)$. Namely,

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i\nu k} f(\nu) d\nu, \quad k \in \mathbb{Z};$$

$$f(\nu) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\nu k} \gamma(k), \quad \nu \in [-\pi, \pi].$$

Heuristically, one expects that the long-range dependence behavior in the time domain translates on a behavior of the spectral density function around the origin. This is because low frequencies take actually into account of large time lags. Indeed, the following result holds:

Prop. 6.3. *Let $\gamma(k)$ be the autocovariance function of a stationary process and let $f(\nu)$ be its spectral density function. Therefore, from the so called Abelian and Tauberian theorems, under suitable conditions, follow that:*

$$\begin{gathered} \gamma(k) \sim |k|^{2H-2}, \quad k \to \infty \\ \Updownarrow \\ f(\nu) \sim c\nu^{1-2H}, \quad \nu \to 0, \end{gathered} \tag{6.1}$$

for any $0 < H < 1$ and where c is some positive constant.

The proposition above tells us that long-range dependence is expressed in the frequencies domain by the blow-up of the spectral density function $f(\nu)$ as $\nu \to 0$. Moreover, if $\gamma(k)$ behaves as a power-law for large times, the same behavior is associated to the spectral density function for small frequencies.

Remark 6.2. Equations (6.1) are verified by the autocovariance function of the fGn (see eq. 5.31). They also hold for a general FARIMA(p,d,q) process (eq. 5.43) provided that $d = H - 1/2$. Thus, a FARIMA(p,d,q) process exhibits long-memory when $d > 0$ (Rem. 5.16).

6.2 Long-range dependence definitions

Other than Def. 6.1, there exist other different definitions of LRD. Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a discrete time stationary stochastic process. We introduce the concept of slowly varying function:

Def. 6.2 (Slowly varying function)**.** A real valued function $L : \mathbb{R} \to \mathbb{R}$ is called *slowly varying function* in zero (infinity) if it is bounded on any finite interval $I \subseteq \mathbb{R}$ and if, for each $a > 0$, one has

$$\frac{L(ax)}{L(x)} \to 1$$

as x tends to zero (infinity).

Such a functions vary slower than any power function. Logarithms and constants are typical examples of slowly varying functions.

Def. 6.3 (Long-range dependence II)**.** Let $\gamma(k)$ be the autocovariance function of a stationary process X. The following three definitions of long-range dependence are quite common in literature (see [2, 3, 11, 20]):

i) The process X has long-range dependence if there exists $0 < \alpha < 1$ such that:

$$\sum_{k=-n}^{n} \gamma(k) \sim n^{\alpha} L_1(n), \quad n \to \infty, \tag{6.2}$$

where L_1 is a slowly varying function at infinity.

ii) The process X has long-range dependence if there exists $0 < \beta < 1$ such that:

$$\gamma(k) \sim k^{-\beta} L_2(k), \quad k \to \infty, \tag{6.3}$$

where L_2 is a slowly varying function at infinity.

iii) The process X has long-range dependence if there exists $0 < \gamma < 1$ such that:

$$f(\nu) \sim |\nu|^{-\gamma} L_3(|\nu|), \quad \nu \to 0, \tag{6.4}$$

where L_3 is a slowly varying function at zero.

The parameters α, β, γ measure the LRD *intensity*, in the sense that greater the long-range dependence greater the values of α and γ and smaller the values of β.

The function $|x|^p L(x)$, which is the product of a power function with a slowly varying function, is called: *regular variation function*. Such a functions, even if they are a modification of a power function, preserve many of its characteristics, such as the rules of summation and integration. For instance, one can show that for $p > 1$:

$$\int_0^y |x|^p L(x)dx \sim \frac{1}{p} y^{p+1} L(y), \quad y \to \infty.$$

or for $p < -1$

$$\int_y^{\infty} |x|^p L(x)dx \sim \frac{1}{|p|} y^{p+1} L(y), \quad y \to \infty.$$

Similar rules hold for the passage from $\gamma(k)$ to the spectral density $f(\nu)$ and are described by the Abelian theorems. The Tauberian theorems, which typically require more stringent conditions, such as monotony, describe the inverse passage from the frequencies domain to the time domain.

Prop. 6.4. *Let $\gamma(k)$ be ultimately monotone[2] as $k \to \infty$. Then, the definitions (6.2), (6.3) and (6.4) are equivalent with:*

$$\alpha = 1 - \beta, \quad L_1(x) = \frac{2}{1-\beta} L_2(x). \tag{6.5}$$

Moreover, $\gamma = 1 - \beta$ and:

$$L_3(x) = \frac{1}{2\pi}\Gamma(\alpha+1)\sin\frac{\pi(1-\alpha)}{2} L_1(1/x) = \frac{1}{\pi}\Gamma(1-\beta)\sin\frac{\pi\beta}{2} L_2(1/x). \tag{6.6}$$

Example 6.2. Consider a fGn with $1/2 < H < 1$. Then, the three definitions above are equivalent and one has from (eq. 3.43), (eq. 6.5) and (eq. 6.6):

$$\begin{cases} \beta = 2 - 2H, & L_2(x) = \sigma^2 H(2H-1); \\ \alpha = 2H - 1, & L_1(x) = \sigma^2 2H; \\ \gamma = 2H - 1, & L_2(\nu) = \sigma^2 \pi^{-1} H\Gamma(2H)\sin(\pi H). \end{cases} \tag{6.7}$$

6.3 Interpretations of long-range dependence

We have seen that a stationary process exhibits long-range dependence if its autocovariance function tends to zero with a power-law and in such a manner that it does not result integrable (Def. 6.1). Time series with long-memory properties are often observed in nature. After the pioneer works of Mandelbrot, the study of long-range dependence have aroused a growing interest, followed by a rapid development. Because of the large variety of the applications, the corresponding literature splits in a large number of journals related to many different fields of study. For instance: agronomy; astronomy; chemistry; economy; engineering; idrology; geology; telecommunication sciences; mathematics and physics.

Heuristically, a data series possesses long-range dependence if observations far away in the past remain, in a certain sense, significantly correlated each other. Such a physical system is said to have long memory properties. In fact, past events continue to affect future events for large time intervals. Oppositely, we have seen that Markovian stationary time series have an autocorrelation function which decrease exponentially to zero. Naively, this means that the memory of the past is lost exponentially fast. Therefore, in these cases we spoke about short-memory processes.

Suppose we have observed a certain time series $\{x(t),\ t = 1, \ldots, n\}$, for instance possessing long-range dependence. The observations are generated by a certain stochastic process X, termed *generator process*, a priori unknown. Assuming stationarity, one tries to model the behavior of the observed time series through a certain class of parametric models $\{Y_\theta,\ \theta \in \Theta\}$ which have in general no physical meaning. More precisely, suppose that the parametric model depends on a certain parametric vector θ. One tries to obtain an "optimal" estimate of θ starting from the observed dada, with the goal to forecast future observations or obtaining a confidence interval for the mean μ. For this purpose it is not needed to establish a direct link between the estimated parameters $\widehat{\theta}$ and the generating process X.

Although useful for statistic purposes, the above approach does not result satisfactory from a scientific point of view. In fact, a theoretical process, developed in a pure statistic context and lacking any direct connection with the observed phenomenon, does not hardly ever allow to give a physical explanation

2) A function $f : \mathbb{R} \to \mathbb{R}$ is ultimately monotone if there exists $y \in \mathbb{R}$ such that $f(x)$ is monotone for any $x > y$.

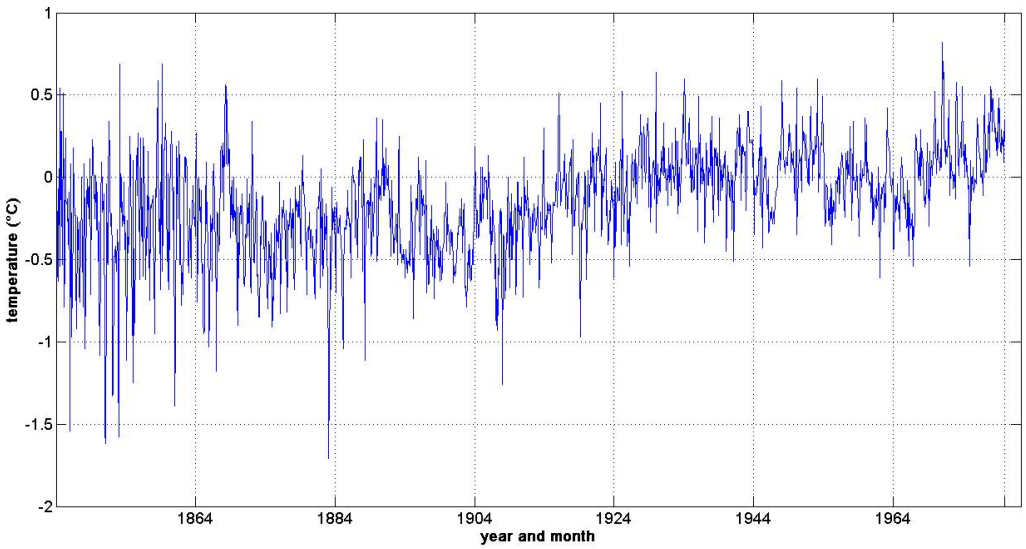

Figure 6.1: Northern hemisphere average monthly temperatures (1854-1989). The plot shows the temperature difference in Celsius degrees with respect to the average temperature of the period 1950-1979.

of the observed dynamical system behavior. On the other hand, one should ever try to understand, at every turn, what is the physical meaning of an observed behavior.

6.3.1 Physical modeling

It is then natural to wonder in which way it is possible to understand LRD from a physical point of view. The cases in which a satisfactory physical model is available are very rare. Thus, stochastic processes can be used in order to model the behavior of a certain time series from a pure statistical point of view. However, it is very important to understand the mechanisms that can explain the presence of long-range correlation in a set of observations.

6.3.1.1 Aggregation of short-memory processes

Very often, observed time series can be interpreted as derived from the aggregation of an high number of individual time series.

Example 6.3. The time series of the average monthly temperature in the northern hemisphere (Figure 6.1) is obtained through the (spatial) aggregation of the time series related to the single data stations.

Let $\{X_t(i)\}_{i\in\mathbb{Z}}$ be a family of short-memory stationary processes, defined in the same probability space. We define:

$$X_t = \sum_{i=1}^{\infty} X_t(i). \tag{6.8}$$

The following result holds:

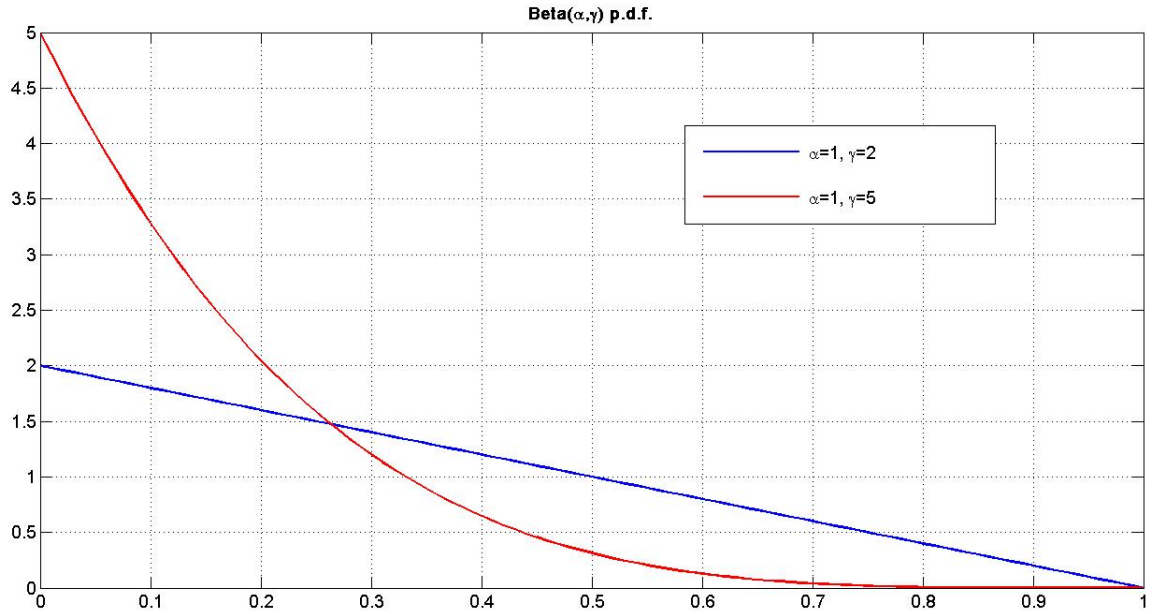

Figure 6.2: **Beta**-distribution probability density function.

Prop. 6.5. *The long-range dependence property can be "artificially" induced by an average over short-memory stationary processes.*

This implies that:

i. Sometimes can be convenient to develop a model which describes the single time series $x_t(i)$ rather than trying to analyze directly the aggregated series.

ii. The aggregation mechanism provides a possible explanation of the long-range dependence generating process in a certain time series.

Proof: suppose that the processes $\{X_t(i)\}_{i\in\mathbb{Z}}$ are Markovian AR(1) independent stationary processes (see [9]). Then,

$$X_t(i) = \alpha_i X_{t-1}(i) + \epsilon_t(i) \quad -1 < \alpha < 1,$$

where $\{\epsilon_t(i)\}_{i\in\mathbb{Z}}$ are independent Gaussian $WN(0,\sigma_i)$. Consider the process

$$X_t^{(N)} = X_t(1) + X_t(2) + \cdots + X_t(N);$$

with spectral density (see eq. 5.37):

$$f^{(N)}(\nu) = \sum_{i=1}^{N} f_i(\nu), \quad f_i(\nu) = \frac{\sigma_i^2}{2\pi}\frac{1}{|1-\alpha_i e^{-i\nu}|^2}.$$

We suppose that $\{\alpha_i\}_{i\in\mathbb{Z}}$ and $\{\sigma_i^2\}_{i\in\mathbb{Z}}$ are realizations of independent random variables α and σ^2, with mean μ_α and μ_{σ^2}, defined on the same probability space $(\Omega,\mathcal{F},P)$. For large N, the spectral density

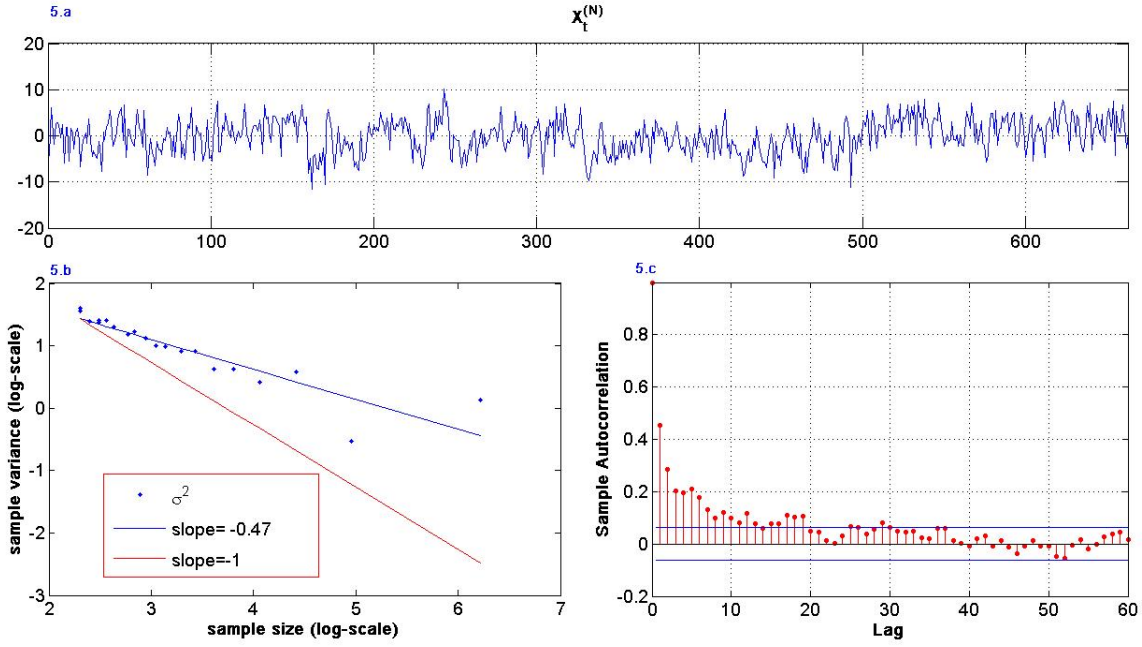

Figure 6.3: Aggregated process $X_t^{(N)}$ (5.a). Variance of the sample mean as a function of the sample size in logarithmic scale. For long-memory series the slope of the line is typically smaller than one (5.b). Sample autocorrelation function (5.c).

function of $X_t^{(N)}$ can be approximated with the expected value:

$$f^{(N)}(\nu) \sim \frac{N}{2\pi} E(\sigma^2 |1 - \alpha e^{-i\nu}|^{-2}) = \frac{N}{2\pi} \mu_{\sigma^2} E(|1 - \alpha e^{-i\nu}|^{-2})$$

$$= \frac{N}{2\pi} \mu_{\sigma^2} \int \frac{1}{|1 - x e^{-i\nu}|^2} dP^{\alpha}(x). \tag{6.9}$$

With a suitable choice of the law of α, for instance a **Beta**(β, γ)-distribution with suitable chosen parameters, one can show that (eq. 6.9) has an integrable pole at the origin. Thus, the aggregated process $X_t^{(N)}$ converges in distribution to a process X_t which manifests long-range dependence. □

Example 6.4. Consider $N = 200$ stationary AR(1) processes. We choose for the r.v. α a **Beta**(β, γ) distribution with parameters $\beta = 1$, $\gamma = 2$ ad for σ^2 a **Beta**(β, γ) distribution with $\beta = 1$, $\gamma = 5$ (Figure (6.2)). In (Figure 6.3) is presented a single realization of length $n = 1000$ of the aggregated process processo $X_t^{(N)}$. The plot of the variance of the sample mean suggests the presence of long-range dependence in the simulated time series (see Section 6.4).

Then, if one observes the presence of long-memory in one aggregated time series (**macroscopic level**), it is possible that the single time series (**microscopic level**) have not long-range correlations. In order to find out the origin of the long-range dependence is necessary, if possible, studying the phenomenon at a microscopic level. However, in many cases, the individual time series can not be observed.

6.3.1.2 Critical phenomena

Long-range correlations arises naturally in the study of phase transition in thermodynamics. For example, in the Ising model, the Hamiltonian is given by:

$$H_\Lambda = -\sum_{i,j\in\Lambda} J_{ij}S_iS_j - h\sum_{i\in\Lambda} S_i,$$

where S_i is the spin in position i of the lattice $\Lambda \subset \mathbb{Z}^d$ and J_{ij} describes the interaction between the spins. Moreover, the right hand term characterizes the interaction with an external field. At the thermodynamic limit, namely $\Lambda \to \mathbb{Z}^d$, one finds that there exists a critical temperature below that (phase transition) the correlation $\rho(i-j) = \text{corr}(S_i, S_j)$ decays to zero very slowly with the space distance (long-range dependence).

Long-range dependence appears naturally also in: turbulence, quantum field theory, $1/f$ noises. In particular, was indeed the study of turbulence which made Kolmogorov define fractional Brownian motion.

6.3.1.3 Partial differential equations

Let Y_ν be a stochastic process, defined in a certain probability space $(\Omega, \mathcal{F}, P)$, such that, for any $\omega \in \Omega$, $Y_\nu(\omega)$ defines a real scalar field defined on $(x,t) \in \mathbb{R}^m \times \mathbb{R}_+$, where t is a time variable.

Suppose that Y_ν is a solution of the stochastic partial differential equation (see eq. 4.4):

$$\partial_t Y_\nu + \alpha Y_\nu = \frac{1}{2}\nabla^2 Y_\nu + \epsilon_\nu, \tag{6.10}$$

where ϵ_ν is a suitable stochastic noise. One can show that when $m > 2$ the solution of (eq. 6.10) is a stationary stochastic process with spatial autocovariance function:

$$\mathbf{cov}(Y_{x,t}, Y_{x',t}) = \gamma(\|x-x'\|) \sim \|x-x'\|^{2-m}, \quad \|x-x'\| \to \infty. \tag{6.11}$$

Imposing different covariance structure for the process ϵ_ν and different boundary conditions for Y_ν, one can obtain many power-laws for the spatial autocorrelation [23].

Therefore, long-memory properties emerge naturally in the stationary solutions of certain stochastic partial differential equations. The reader interested in a discussion concerning how to model LRD through partial differential equation can see Gay et al. [8] .

6.4 Statistics of long-range dependence

Time series with LRD are recognizable thanks to certain quantitative features, for instance the self-similarity parameter H of a fGn, and qualitative features, which are listed below.

The linear plot

Qualitatively, the linear plot of a long-memory time series exhibits the following characteristics:

i. For long time periods the observations tend to stay above or below the mean value.

ii. If one studies small time scale, the series seems to be non-stationary. In fact, one can observes apparent cycles or trends, which disappear if one looks at the series in its completeness.

iii. The series is stationary.

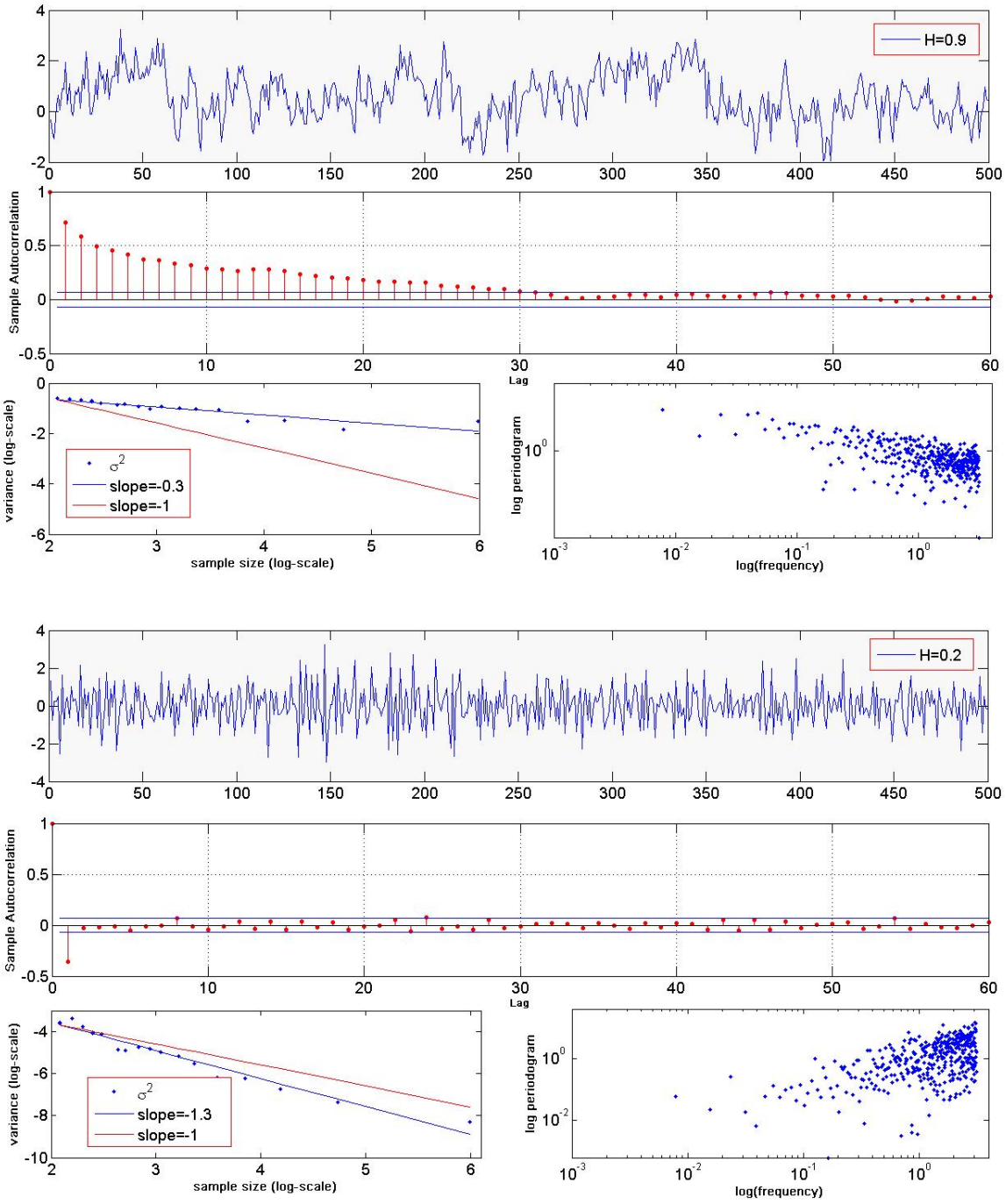

Figure 6.4: Persistent fGn (top figures) $H = 0.9$. Antipersistent fGn (bottom figures) $H = 0.2$.

The variance of the sample mean

The variance of the sample mean seems to converge to zero, as the sample size n increase, slower than the typical behavior n^{-1}. In good approximation, the decay is proportional to $n^{-\alpha}$, with $0 < \alpha < 1$ (see [3]).

The autocorrelation function

The sample autocorrelation function $\widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0)$:

$$\widehat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(x_t - \overline{x}_n)(x_{t+|k|} - \overline{x}_n), \tag{6.12}$$

where $\overline{x}_n$ is the sample mean, goes to zero quite slowly and in good approximation proportional to $k^{-\alpha}$, with $0 < \alpha < 1$. The plot of the autocorrelation function is called *correlogram*.

The periodogram

The periodogram $I(\nu)$ is the sampling analogous of the spectral density function and it is defined as:

$$I(\nu_j) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n}(x_t - \overline{x}_n)e^{it\nu_j}\Big|^2 = \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\widehat{\gamma}(k)e^{ik\nu_j}, \tag{6.13}$$

where $\nu_j = 2\pi j/n$ with $j = 1, \ldots, n^*$ and $n^* = \text{Int}[(n-1)/2]$ are the Fourier frequencies. In logarithmic scale, near the origin, the periodogram is distributed approximately around a line with negative slope (see eq. 6.1). In other words, the periodogram is dominated by low frequencies.

All these features are recognizable in Figure (6.4). Here in the top is shown a long-memory time series. Whereas, in the bottom figure is shown a typical short-memory process. The first figure shows a persistent behavior. In fact, the series is stationary but exhibits local trends and periodicities, while the second series does not. The sample autocorrelation in the top figure decreases to zero quite slowly and, anyhow, very slower than the sample autocorrelation in the bottom. Moreover, the first periodogram is clearly dominated by the low frequencies, while the second periodogram is clearly dominated by high frequencies. Finally, also the slope of the best fit line in the sample mean variance plot indicates the presence of LRD in the data on the top.

The first time series is just a realization of a fractional Gaussian noise in the persistent domain. That is, $H = 0.9$. Whereas, the second time series is a realization of a fGn with $H = 0.2$, that means no long-range dependence (Prop. 6.1).

Remark 6.3. The same analysis of Figure 6.4 could have been obtained with synthetic FARIMA time series with $d = H - 1/2$.

6.4.1 Time series examples

We now study qualitatively some typical example of real time series which manifest long-range dependence.

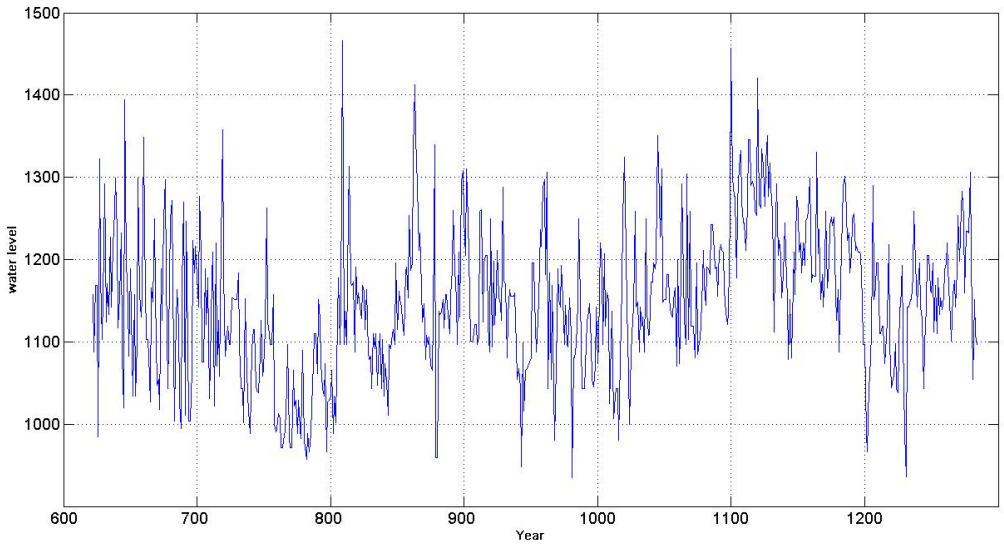

Figure 6.5: Nile river year minima (622-1284)

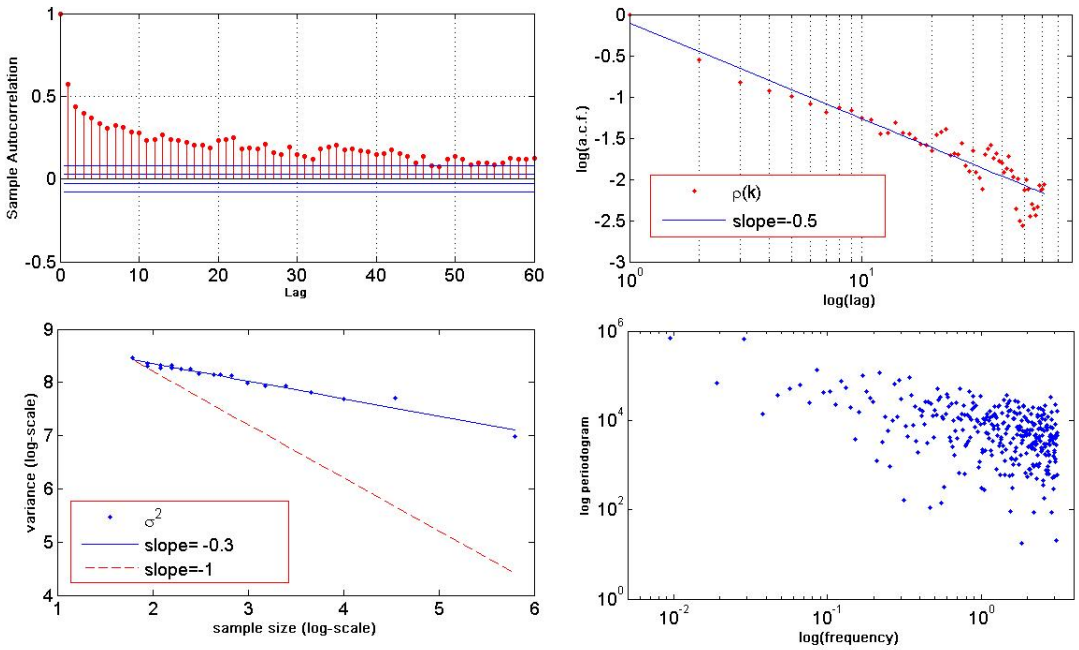

Figure 6.6: Nile river minima: sample autocorrelation function in linear and logarithmic scale; sample mean variance in logarithmic scale; periodogram in logarithmic scale.

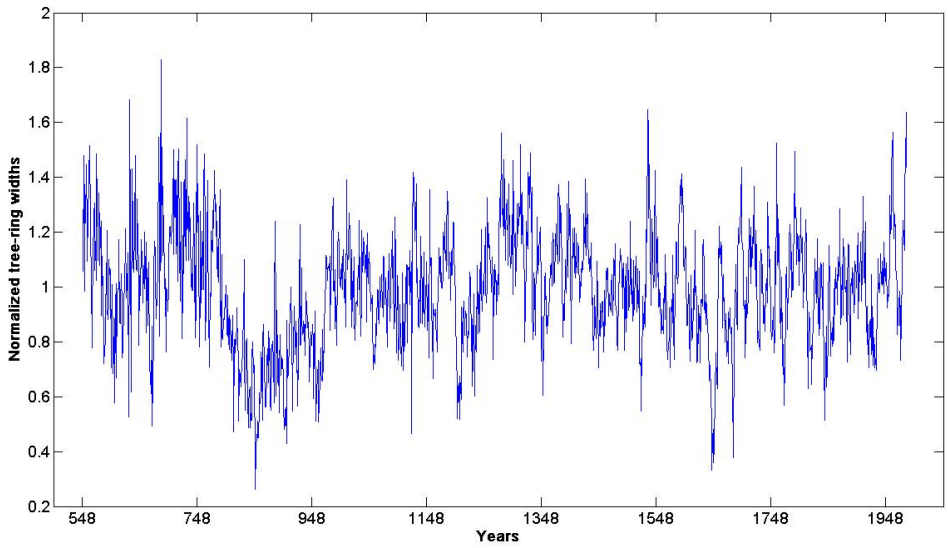

Figure 6.7: Tree rings: "standardized" width rings of a Rocky Mountain Pine (Arizona, 548-1983).

6.4.1.1 Nile river minima

In Figure (6.5) is shown the time series which describes the year minimum levels of the Nile river between the years 622-1284. This data set has a great historical interests. In fact, its analysis contributed to the discovery of the so called **Hurst effect**. Then, motivated from the empirical studies of Hurst, Mandelbrot introduced the fractional Gaussian noise as a standard statistic model for the long-range dependence.

The presence of long-memory in the data is suggested by Figure (6.6), which shows that the variance of the sample mean tends to zero slower than n^{-1}. The best feet line is:

$$r(n) = 9.00 - 0.32 \log n.$$

Further evidence of LRD are given by the plots of the autocorrelation function and by the plot of the periodogram in logarithmic scale, which appears clearly dominated by the low frequencies.

Moreover, the plot of (Figure 6.5) presents all the qualitative characteristics of a typical long-range dependence time series. In fact, it presents locally a non-stationary behavior even if the whole series seems stationary. On the other hand, since from ancient times, the Nile river was known for its long-term behavior. Long period of famine were followed by long period of abundance, both associated to the floods of the river. Mandelbrot called this behavior **Joseph effect**, reminding to the Bible words "*seven years of great abundance in the land of Egypt, but seven years of famine will follow*".

6.4.1.2 Tree rings

Figure (6.7) shows the the plot of the "standardized" rings width of a Rocky Mountains pine in Arizona located at an altitude of 3523 meters. The tree rings analysis provides information about the local cli-

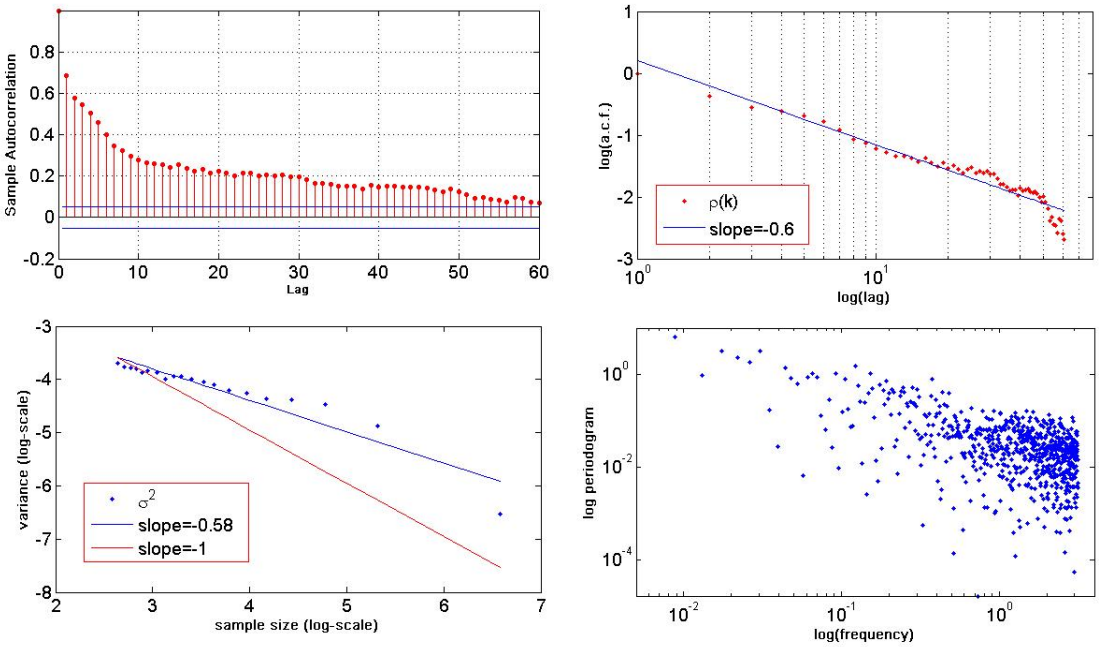

Figure 6.8: Tree rings: autocorrelation function in linear scale and logarithmic scale (top panels); sample mean variance in logarithmic scale and periodogram in logarithmic scale (bottom panels).

mate variations occurred during the plant lifetime. In fact, many trees can live up to a hundreds of years (some times even a thousands of years), for this reason the rings study can give information hardly accessible in different ways. Each ring represents in good approximation one year. Thus, by counting them, it is possible to obtain information about the tree age. Moreover, because the plant growth is greater if the environmental conditions are good, their measurement can be used to obtain many information about the setting, like forest fires, dry seasons, insect attacks or floods. The science which studies climate changes from the analysis of tree rings is called *dendroclimatology*.

The data of the series presented in Figure (6.7) are standardized by a process acted to remove possible non-stationarity components, for example the growing trend due to the normal physiological mutation process suffered by the plant during the obsolescence. For instance, and index equal to 0.75 for a given year indicates a growth below the normal (indicated with 1.00).

The plots in Figure 6.8 show strong evidence of long-range dependence in the data. In fact, the autocorrelation function decreases to zero like a power function with $\alpha \sim 0.59$. The same behavior is highlighted by the sample mean variance and from the graph of the periodogram.

6.4.1.3 Global temperatures

Figure (6.1) shows the monthly average temperature for the northern hemisphere in the years 1854-1989. The data corresponds to the difference between the average monthly temperature and the monthly average on the period 1950-1979. The plot suggests the presence of a linear trend, which could be related to a global worming in the last hundred years. After subtracting such a linear trend (Figura 6.9), except

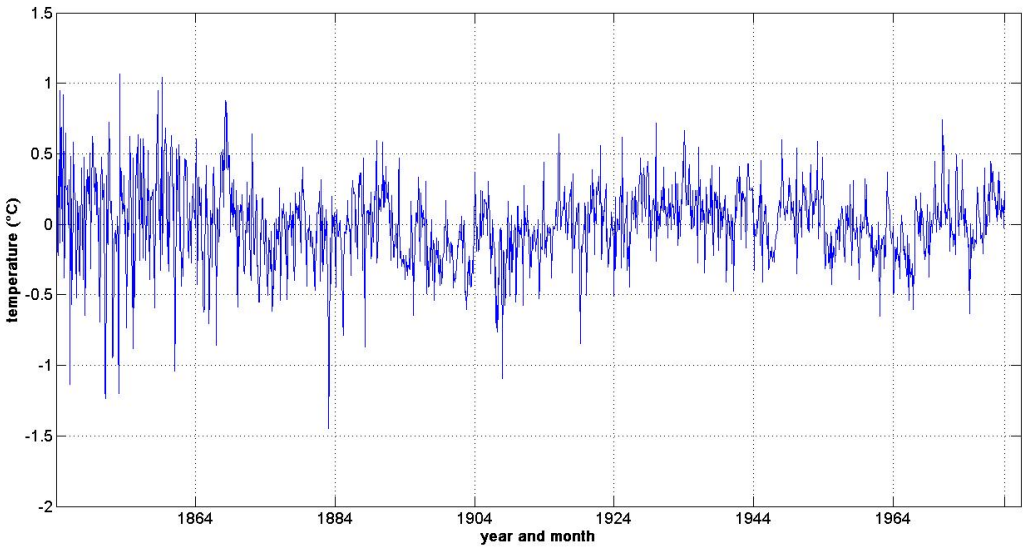

Figure 6.9: Monthly average temperatures in the northern hemisphere after a linear trend subtraction.

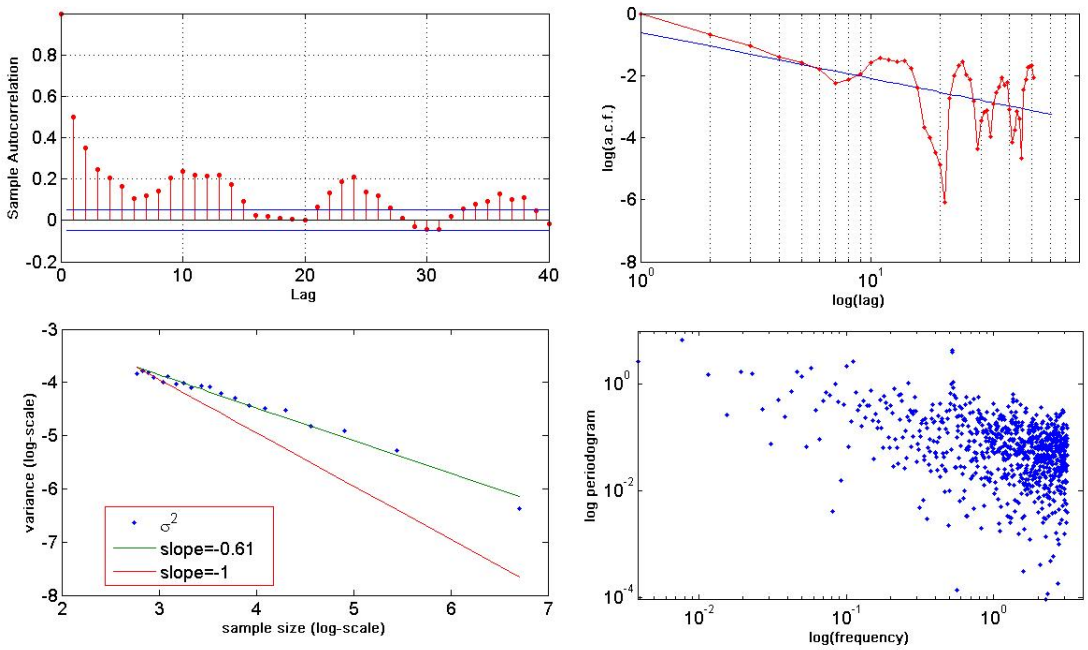

Figure 6.10: Global temperatures: autocorrelation (top panels); sample mean variance an periodogram (bottom panels).

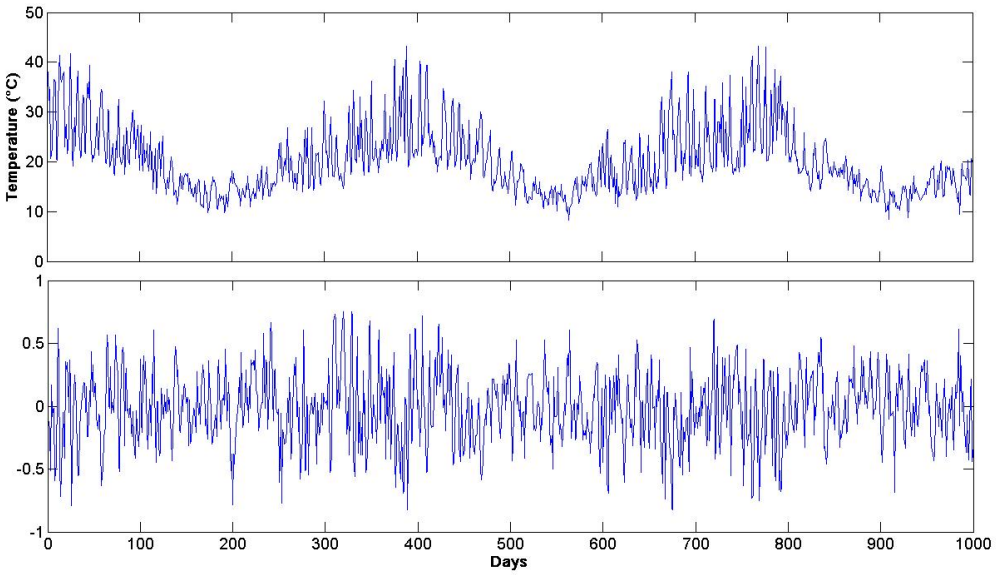

Figure 6.11: Daily maximum temperatures in Melbourn (1981-1990) and deseasonalized data (bottom).

for a quite bigger variance in the beginning, perhaps due to a lack of precision in the measuring instruments, the plot seems by and large to be stationary.

The plots in Figure (6.10) suggest the presence of LRD in the observations. The sample autocorrelation function shows a strong dependence even between observations very far away in time (6.10.a). Moreover, the sample mean variance seems to tend to zero with a typical behavior of a long-memory time series ($\alpha = -0.61$). Finally, the periodogram is dominated by the low frequencies.

We observe that despite to the fact that the series has already been treated in order to remove seasonal components, a sort of seasonality seems to persist in the data.

6.4.1.4 Local temperatures

We close this sequence of real data series providing a typical example of short-range dependence time series.

Figure (6.11) contains the plot of the daily maximum temperatures (over 1000 days) registered in Melbourne between the years 1981-1990 (source Australian Bureau of Meteorology). Moreover, the bottom panel shows the deseasonalized time series.

Figure (6.12) highlights the short-range dependence between the observations. Temperatures above the mean tend, with good probability, to be followed by temperatures above the mean just some day after.

Remark 6.4. The Global temperature time series (Figure 6.9) is actually an aggregation of Local temperature time series (Figure 6.11). Therefore, the long-range dependence in this kind of data could be

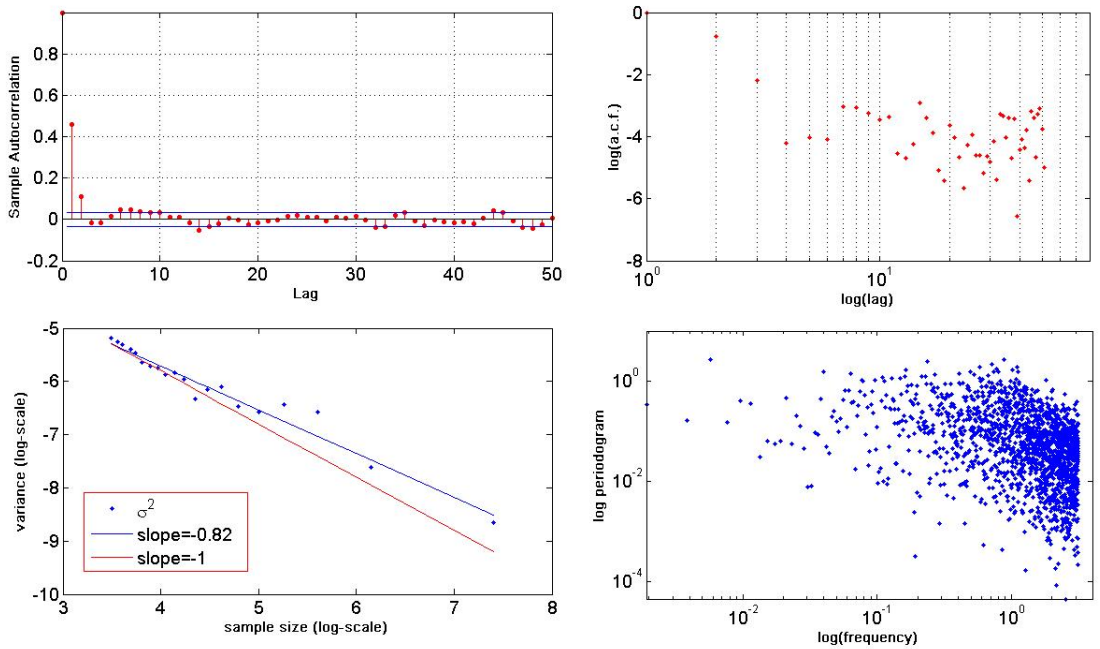

Figure 6.12: Local temperatures: autocorrelation (top); sample mean variance, periodogram (bottom).

explained, from a statistical point of view, through the process of aggregation of many short-memory series corresponding to the local temperatures data.

6.5 Estimation of long-range dependence

We have briefly discussed the qualitative features of long-range dependence time series. We focus now in the problem of the quantitative measure of LRD for a given (possible) long-memory data. We have seen that:

- Fractional Gaussian noise is characterized by the scaling exponent $H \in (0,1)$ of the corresponding self-similar process. The autocovariance function behaves, for large k, like a power function $\gamma(k) \sim c_r k^{2H-2}$. Then, one has long-range dependence if $1/2 < H < 1$;
- For FARIMA(p,d,q) processes, the long-range dependence is measured by the parameter $d \in (-1/2, 1/2)$. In fact, the autocovariance function behaves like $\gamma(k) \sim c_s k^{2d-1}$ when k goes to infinity and LRD occurs when $0 < d < 1/2$.

As already observer (see Rem. 6.2), the two parameters can be formally related by setting $d = H - 1/2$. Then, in both the cases, greater the parameter H greater the long-range dependence in the observations. Therefore, in the conditions of (Prop. 9.1), it is useful to characterize the LRD, (eq. 6.3), (eq. 6.2) and (eq. 6.4), with a unique real parameter $H \in]1/2, 1[$ such that:

$$\beta = 2 - 2H \Rightarrow \alpha = 2H - 1, \;\; \gamma = 2H - 1.$$

Given an observed time series, one has the problem of how to estimate the parameter H, which by definition provides a measure of the long-range dependence in the data.

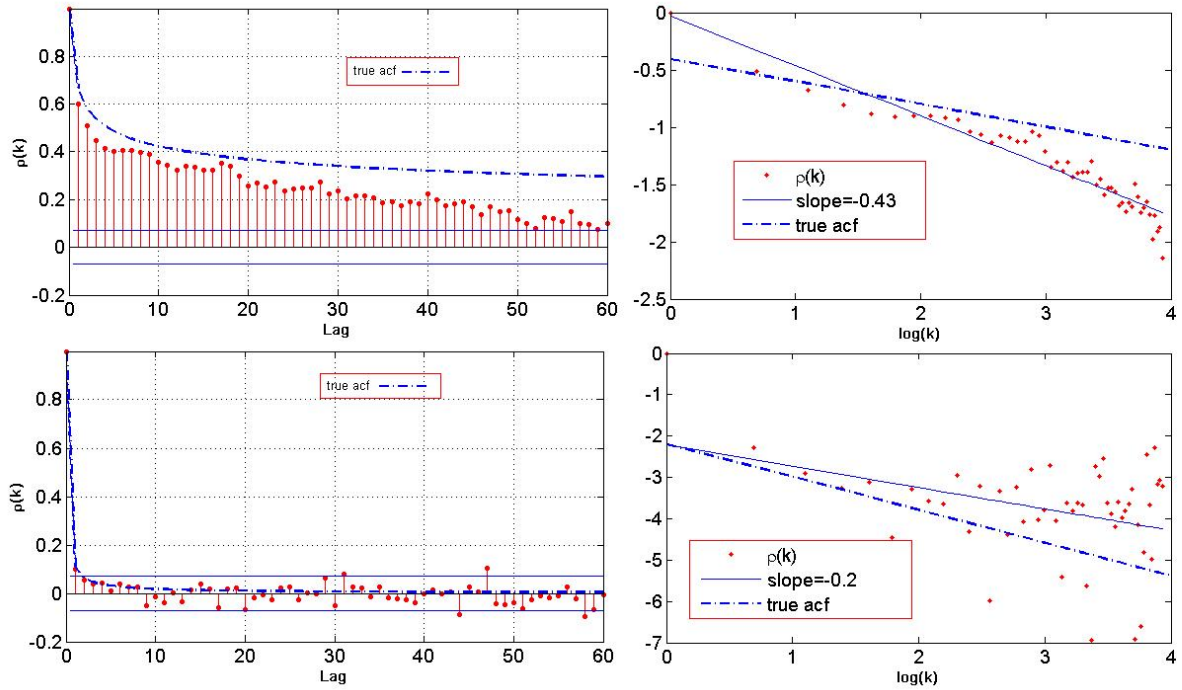

Figure 6.13: FARIMA autocorrelation function. The case $H = 0.9$ (top figures): linear plot (on the left) with the theoretical curve (eq. 3.63) in dashed line; logarithmic scale (on the right) with the best fit line (solid line) and the theoretical curve in dashed line. The same for the case $H = 0.6$ (bottom figures).

6.5.1 Heuristic methods

Many heuristic methods have been purposed. From an historical point of view, it is noteworthy to mention the so called *R/S statistics*, first introduced by Hurst [13], in an hidrological context(see also [3]). Other methods, just introduced in the last section, are:

1. the study of the correlogram;
2. the study of the sample mean variance;
3. the study of the periodogram.

In fact, in all these cases a rough estimate of the H parameter can be obtained evaluating the slope of the best-fit line in the logarithmic scale graphs.

These methods are very useful as a first diagnostic instrument. They give am approximative first idea of the long-range dependence behavior in the observations. They are not efficient and not much flexible and this makes the results interpretation very difficult.

Example 6.5 (The correlogram)**.** Consider the correlogram of two synthetics FARIMA$(0, H - 1/2, 0)$ with $H = 0.9$ and $H = 0.6$ (Figure 6.13). We observe that:

β	H_0	$\widehat{H}$	Confidence
0.1	0.88	0.54	[0.51,0.55]
0.1	0.76	0.53	[0.51,0.55]
0.1	0.82	0.53	[0.51,0.55]
0.5	0.75	0.51	[0.49,0.54]
0.5	0.65	0.51	[0.49,0.53]
0.5	0.66	0.49	[0.47,0.51]
0.9	0.58	0.49	[0.47,0.51]
0.9	0.39	0.49	[0.47,0.51]
0.9	0.43	0.50	[0.48,0.52]
1	0.41	0.50	[0.48,0.52]

Table 6.1: Estimates of the LRD parameter H for a time series of length $n = 10000$, realization of a white noise ϵ_t process plus a trend term $a_t = 2t^{-\beta}$. H_0 is the estimate obtained with the sample mean variance method. Whereas, $\widehat{H}$ is the estimate obtained with a maximum likelihood method calibrating a FARIMA$(0, H - 1/2, 0)$ process using the Whittle approximation. In the last column is given the 95% confidence interval for $\widehat{H}$ (eq. 6.30).

- In the case $H = 0.9$ (Fig. 6.13 top) the LRD behavior is much more evident than in the case $H = 0.6$ (Fig. 6.13 bottom). In this case it is indeed very difficult to say that the autocorrelation function follows an asymptotic behavior k^{2H-2} with $1/2 < H < 1$, rather than an exponential behavior c^k with $0 < c < 1$.

- The absolute values of the correlation coefficients can be anyhow arbitrarily small and can stay within the range $\pm 2/\sqrt{n}$, which represents the null value confident interval.

- The LRD is an asymptotic notion for the correlations. Therefore, one should study the autocorrelation function for large values of the time lag. However, as the lag k increases, the estimates $\widehat{\rho}(k)$ become soon less reliable.

The previous example, together with Figure (6.13), proves that the study of the sample autocorrelation function is more useful in the case of a strong LRD or for very long time series. For relatively short time series, or for a value of H close to $1/2$, it is very difficult to estimate the asymptotic behavior of $\rho(k)$ and then to decide if there is or not a long-memory property in the observed data.

Example 6.6 (Series with trend)**.** Consider a time series realization of the stochastic process:

$$X_t = a_t + \epsilon_t, \quad t \geq 0,$$

where ϵ_t is a Gaussian $WN(0,1)$ and $a_t = 2t^{-\beta}$, with $0 < \beta < 1$, is a trend term which asymptotically disappears. Observing Figures (6.14) and (6.15), the method of the sample mean variance induces wrongly to think that the first data set has long-range dependence, while the second does not.

One can show (see [3]) that the presence of a trend which decrease fairly slowly to zero in a series of observations (for instance when $\beta < 1/2$ in the example before) can manifest itself in certain heuristic methods, like the method of the sample mean variance (previous example), or the R/S analysis, with the presence of long-range dependence. This is not the case of the maximum likelihood methods which will be introduced in the next section (see Table 6.1).

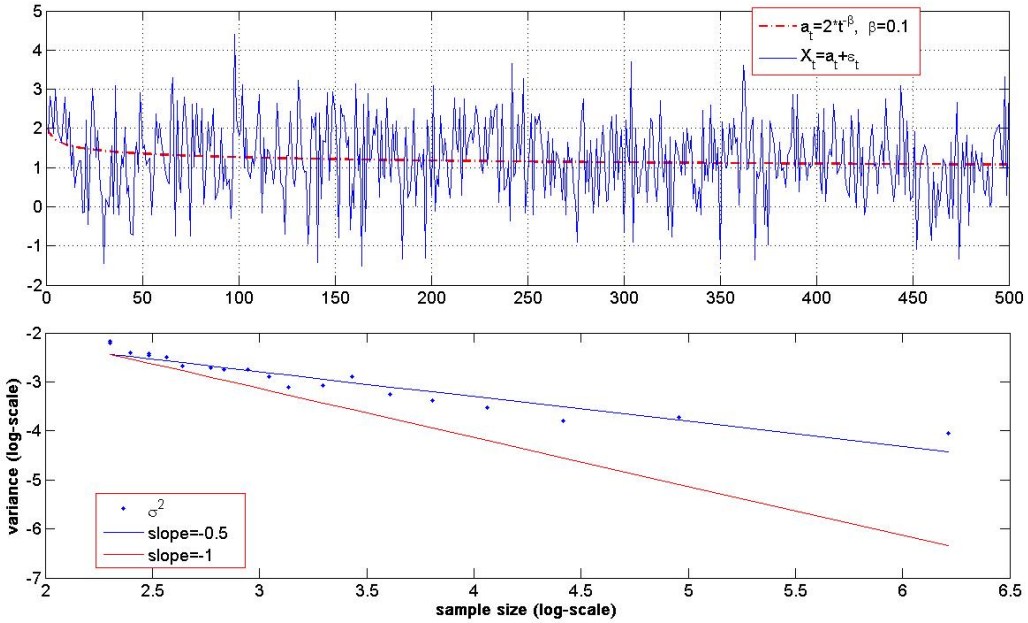

Figure 6.14: White-noise ϵ_t, plus a trend term $a_t = 2t^{-\beta}$ and sample mean variance plot in logarithmic scale in the case $\beta = 0.1$.

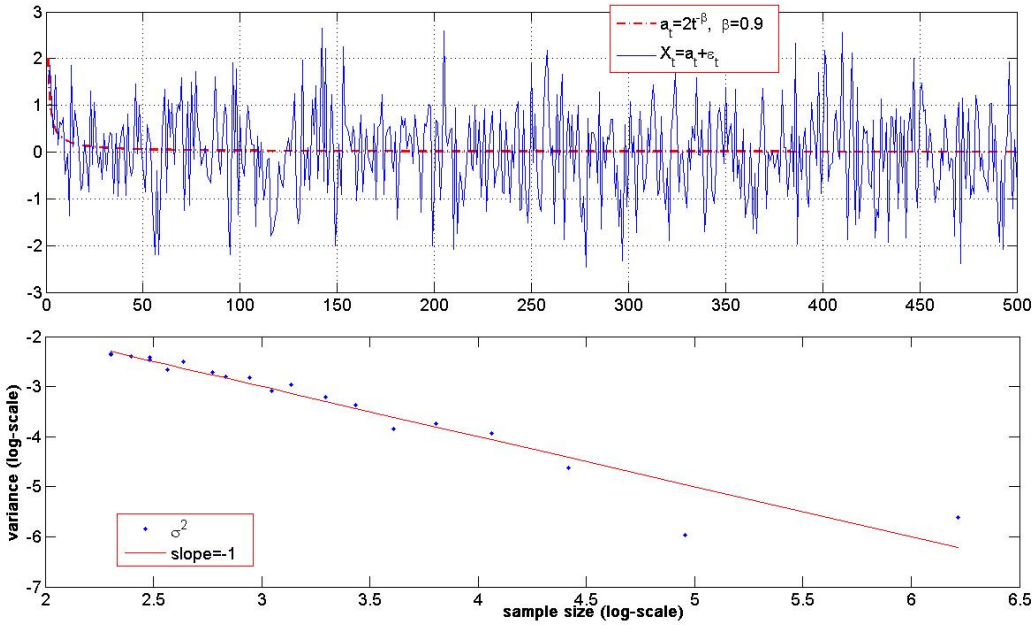

Figure 6.15: White-noise ϵ_t, plus a trend term $a_t = 2t^{-\beta}$ and sample mean variance plot in logarithmic scale in the case $\beta = 0.9$.

6.5.2 Maximum likelihood methods

The heuristic methods, briefly exposed in the previous section, are useful tools for a first survey of LRD in a certain observed data. However, these methods cannot give information about the properties of the short-range dependence structure. For this purpose it is necessary to use methods which consent to model the whole correlaion structure, or equivalently the spectral density function at all the frequencies. A possible approach is to use *parametric models* and the to estimate the parameters, for instance, maximizing the likelihood function.

Parametric methods are preferable even if the objective is just the estimation of the LRD parameter H. In fact, if one is able to build a reasonable parametric model, maximum likelihood methods (MLE methods) are more efficient and more indicated for the *statistic inference*. Howerver, a not so complicated suitable model is often hard to find. Moreover, an elementary model can lead to distorted estimates of H.

For the seek of simplicity, we focus only on Gaussian MLE methods. This not mean that such a methods can be applied only to Gaussian processes. However, if this is not the case, a certain prudence is expected.

6.5.2.1 Characterization of the parametric model

We suppose we want to describe a series of observations $\{x_i\}_{i=1,\dots,n}$, which exhibits long-range dependence, through a certain stochastic model. For this purpose, we introduce a process $X = \{X_t\}_{t\in\mathbb{Z}}$ that verifies the following properties:

Stationarity

Suppose that $X = \{X_t\}_{t\in\mathbb{Z}}$ is a stationary stochastic process with mean μ and variance σ^2, defined on a certain probability space $(\Omega, \mathcal{F}, P)$. We also suppose that X possesses LRD characterized by the parameter H, with $1/2 < H < 1$.

Remark 6.5. In the following we always suppose that μ is known and equal to zero: $\mu = 0$.

Linearity

We suppose that X is a causal linear process. Namely:

$$X_t = \sum_{j=1}^{\infty} b_j X_{t-j} + \epsilon_j \Leftrightarrow X_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \tag{6.14}$$

where ϵ_t is a $\mathrm{WN}(0, \sigma_\epsilon^2)$. The coefficients a_j and b_j verify:

$$\{b_j\}_{j=1,\dots,\infty} \in l^1, \quad \{a_j\}_{j=0,\dots,\infty} \in l^2. \tag{6.15}$$

In fact, the following result holds:

Prop. 6.6. *there exist two positive constants c_a e c_b, such that:*

$$b_k \sim c_b k^{-H-\frac{1}{2}}, \quad a_k \sim c_a k^{H-\frac{3}{2}}, \quad k \to \infty. \tag{6.16}$$

Example 6.7 (FARIMA processes)**.** FARIMA(0,*H-1/2*,0) processes (eq. 3.61) provide an example of causal linear processes which can show long-range dependence. In this case one has:

$$X_t = (1-B)^{-(H-1/2)}\epsilon_t = \sum_{j=0}^{\infty} a_j B^j \epsilon_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j},$$

with (see eq. 3.64)

$$a_j = \frac{\Gamma(j+H-1/2)}{\Gamma(j+1)\Gamma(H-1/2)}, \quad j \in \mathbb{N}.$$

Using Stirling approximation:

$$a_k \sim \frac{k^{H-3/2}}{\Gamma(H-1/2)}, \quad k \to \infty.$$

Thus, $\{a_k\}$ belong to l^2 if $H < 1$. in the same way, if one expresses X in autoregressive from, the coefficients b_k are still given by (eq. 3.64) just by setting $-d \to d$. Then, one has:

$$b_k \sim \frac{k^{-H-1/2}}{\Gamma(1/2-H)}, \quad k \to \infty,$$

which is on l^1 if $H > 1/2$.

Spectral density function

We suppose that the spectral density function f (or equivalently the autocovariance function) of X is characterized ny a finite number of parameters:

$$\theta^0 = (\sigma_0^2, H, \theta_3^0, \theta_4^0, \dots, \theta_M^0). \tag{6.17}$$

This means that f belongs to a parametric class of spectral densities characterized by a parameter M-dimesional vector:

$$f(\nu) \in \{f(\nu;\theta)\}_{\theta\in\Theta}, \quad \Theta \subset \mathbb{R}^M, \quad \nu \in]-\pi, \pi]. \tag{6.18}$$

Given a sequence of observations $\{x_1, x_2, \dots, x_n\}$ of X, the purpose is to estimate the unknown parameter vector $\theta^0 \in \Theta$, and thus the spectral density $f(\nu;\theta^0)$.

The following rsulte will be useful:

Prop. 6.7. *if $f(\nu;\theta^0)$ is the spectral density function of the process X defined in (eq. 6.14), then:*

$$\sigma_\epsilon^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\nu;\theta^0) d\nu\right). \tag{6.19}$$

In particular, it follows that:

$$f(\nu) = \frac{\sigma_\epsilon^2}{2\pi} f_1(\nu) \Rightarrow \int_{-\pi}^{\pi} \log f_1(\nu) d\nu = 0. \tag{6.20}$$

Example 6.8. Suppose for instance that X is a FARIMA$(0, H-1/2, 0)$ process. Then, the spectral density is given by (eq. 5.40), that is

$$f(\nu) = \frac{\sigma_\epsilon^2}{2\pi}\left|1-e^{-i\nu}\right|^{1-2H} = \frac{\sigma_\epsilon^2}{2\pi}(2(1-\cos\nu))^{1/2-H} = \frac{\sigma_\epsilon^2}{2\pi}(4\sin^2\frac{\nu}{2})^{1/2-H}.$$

Then, $f = f(\nu;\theta_1,\theta_2)$, with $\theta_1 = \frac{\sigma_\epsilon^2}{2\pi}$ and $\theta_2 = H$. Furthermore, we are in the case of (eq. 6.20), then

$$\int_{-\pi}^{\pi} \log(2-2\cos\nu) d\nu = 0. \tag{6.21}$$

Normality

Suppose that the process ϵ_t is a Gaussian white noise. This implies that X is a Gaussian process. Then, the finite dimensional distribution are Gaussian with covariance matrix that depends on the vector parameter θ^0. Then, if we consider the random vector $\{X_1, \dots, X_n\}$ and we indicate with $\Sigma(\theta^0)$ the covariance matrix, the joint probability density is:

$$h(x;\theta^0) = \frac{1}{\sqrt{(2\pi)^n |\Sigma(\theta^0)|}} e^{-\frac{1}{2}x'\Sigma^{-1}(\theta^0)x}, \quad x \in \mathbb{R}^n. \tag{6.22}$$

Def. 6.4 (Log-likelihood)**.** We define the *log-likelihood* function as:

$$L_n(x;\theta^0) = \log h(x;\theta^0).$$

Then,

$$L_n(x;\theta^0) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma(\theta^0)| - \frac{1}{2}x'\Sigma^{-1}(\theta^0)x. \tag{6.23}$$

6.5.2.2 Exact MLE estimates

Consider the class $\{h(x;\theta)\}_{\theta\in\Theta}$ of probability distributions in $\mathbb{R}^n$. Suppose we have n observations $x = \{x_i\}_{i=1,\dots,n}$, which correspond to a sigle relaization of the process X up to time $t = n$. The maximum likelihood estimation (MLE estimate) of the parameter vector θ^0 corresponds to the choice of the parameter vector $\widehat{\theta}$ such that the probability (eq. 6.22) to observe $\{x_1, \dots, x_n\}$ is maximum.

Maximizing $h(x;\theta)$ is equivalent to maximizing $L_n(x;\theta)$ with respect to the M-dimensional vector parameter θ. Then, under suitable regularity conditions, the MLE estimate $\widehat{\theta}$ is obtained solving the following system of M equations:

$$L_n'(x;\theta) = 0, \tag{6.24}$$

dove:

$$L_n'(x;\theta)_j = \partial_{\theta_j} L_n(x;\theta)$$

$$= -\frac{1}{2}\partial_{\theta_j}\log|\Sigma(\theta)| - \frac{1}{2}x'[\partial_{\theta_j}\Sigma^{-1}(\theta)]x, \quad j = 1, \dots, M \tag{6.25}$$

Consider now the estimator $\widehat{\theta}$ define by (eq. 6.24). We wonder which is its asymptotic distribution. First of all, we define the $M \times M$ matrix:

$$D_{ij}(\theta^0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \partial_{\theta_i}\log f(\nu;\theta)\partial_{\theta_j}\log f(\nu;\theta)d\nu\big|_{\theta^0}, \tag{6.26}$$

where f is the spectral density function of the process $X(t)$ (eq. 6.18). Then, the following result holds:

Theorem 6.1. *Let X be a stochastic process as defined before. Namely, it is: stationary; Gasussian; causal; linear and with LRD. Consider $\widehat{\theta}$ defined by (eq. 6.24). Then:*

- *$\widehat{\theta}$ is asymptotically correct. That is,*

$$\widehat{\theta} \to \theta^0, \quad \textit{almost surely}. \tag{6.27}$$

- *Moreover,*

$$\sqrt{n}\widehat{\theta} \to_d \zeta, \tag{6.28}$$

where ζ is a Gaussian M-dimesional random vector with mean θ^0 and covariance matrix

$$C = 2D^{-1},$$

with D defined by (eq. 6.26) and the convergence is meant in distributional sense.

Therefore, the maximum likelihood Gaussian estimator has a normal asymptotic distribution and a rate of convergence equal to $1/\sqrt{n}$. This is in contrast with certain estimators, such as the sample mean, that, for long-range dependence processes, have a rate of convergence slower than the corresponding estimators of short-memory processes.

Remark 6.6. The covariance matrix C is diagonal if the functions $f_j(\nu) = \partial_{\theta_j} \log f(\nu)$ are orthogonal, i.e.

$$\langle f_i, f_j \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_i(\nu) f_j(\nu) d\nu = 0, \;\; i \neq j.$$

In this case the MLE estimates of the single parameters are asymptotically independent. Moreover, if $f_j(\nu;\theta)$ does not depend on θ, then the variance of $\widehat{\theta}_j$ is the same for each θ.

Let us consider the following example:

Example 6.9. Suppose we have observed a certain time series $\{x_i\}_{i=1,\dots,n}$ which exhibits LRD. Consider a Gaussian FARIMA$(0, H-1/2, 0)$ process (eq. 3.61) $X(t)$ as a stochastic model for the observed data. The spectral density of $X(t)$ is (see eq. 5.40):

$$f(\nu) = \frac{\sigma_\epsilon^2}{2\pi} \left|1 - e^{-i\nu}\right|^{1-2H} = \frac{\sigma_\epsilon^2}{2\pi} (2(1-\cos\nu))^{1/2-H} = \frac{\sigma^2}{2\pi} (4\sin^2\frac{\nu}{2})^{1/2-H}.$$

We have two parameters (θ_1, θ_2) to estimate. Indeed, $f = f(\nu;\theta_1,\theta_2)$, with $\theta_1 = \frac{\sigma_\epsilon^2}{2\pi}$ and $\theta_2 = H$. Let us study the asymptotic distribution of the maximum likelihood estimator. First, we evaluate the matrix D. One has:

$$\begin{cases} f_1 = \partial_{\theta_1} \log(\theta_1 (2-2\cos\nu)^{1/2-\theta_2}) = 1/\theta_1, \\ f_2 = \partial_{\theta_2} \log(\theta_1 (2-\cos\nu)^{1/2-\theta_2}) = -\log(2-2\cos\nu). \end{cases}$$

$$D = \begin{pmatrix} |f_1|^2 & \langle f_1, f_2 \rangle \\ \langle f_1, f_2 \rangle & |f_2|^2 \end{pmatrix} = \begin{pmatrix} 2\pi/\sigma_\epsilon^4 & 0 \\ 0 & \pi^2/3 \end{pmatrix}.$$

In fact, by (eq. 6.21):

$$\int_{-\pi}^{\pi} \log(2-2\cos\nu) d\nu = 0$$

and one can show

$$\int_{-\pi}^{\pi} (\log(2-2\cos\nu))^2 d\nu = \frac{2}{3}\pi^3.$$

Then, by definition:

$$C = \begin{pmatrix} \sigma_\epsilon^4/\pi & 0 \\ 0 & 6/\pi^2 \end{pmatrix}.$$

Therefore, $\widehat{H}$ and $\widehat{\sigma}_\epsilon^2 = 2\pi\theta_1$ are asymptotically independent with:

$$\mathbf{var}(\widehat{H}) = \frac{6}{\pi^2}, \;\; \mathbf{var}(\widehat{\sigma}_\epsilon^2) = 2\sigma_\epsilon^4. \tag{6.29}$$

Starting from the n observations we have, the 95% confident interval for the MLE estimate of H is:

$$\left[\widehat{H} - \frac{2}{\pi}\sqrt{\frac{6}{n}}, \widehat{H} + \frac{2}{\pi}\sqrt{\frac{6}{n}}\right]. \tag{6.30}$$

6.5.2.3 Whittle approximation

Approximated methods are needed in order to reduce the calculation times in the likelihood estimations. Computational problems may emerge especially for large size time series or for a large number of parameters to estimate .

The two terms depending on θ in (eq. 6.23) are:

$$\log|\Sigma_n(\theta)|, \qquad x'\Sigma_n^{-1}(\theta)x.$$

The index n indicates the size of the covariance matrix. Let us study these two terms distinctly:

1. one can show that:
$$\lim_{n\to\infty}\frac{1}{n}\log|\Sigma_n(\theta)| = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\nu;\theta)d\nu.$$
Then, we approximate $\log|\Sigma_n(\theta)|$ with $\frac{n}{2\pi}\int_{-\pi}^{\pi}\log f(\nu;\theta)d\nu$.

2. The matrix $\Sigma_n^{-1}(\theta)$ can be substituted with a matrix having elements easier to evaluate. Let
$$A_n(\theta) = \alpha(k-j) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\frac{1}{f(\nu;\theta)}e^{i(k-j)\nu}d\nu, \quad k,j = 1,\dots,n. \tag{6.31}$$
Therefore, one can show that, asymptotically, $A_n(\theta) \to \Sigma_n^{-1}(\theta)$.

Combining these two approximations we obtain one approximated version of the log-likelihood function:

$$L_n^*(x;\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\nu;\theta)d\nu - \frac{1}{2}x'A_n(\theta)x. \tag{6.32}$$

Finally, an approximated MLE estimation of θ^0 is obtained mimimazing the likelihood function

$$L_W(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\nu;\theta)d\nu + \frac{1}{n}x'A(\theta)x, \tag{6.33}$$

called *Whittle likelihood function*.

Under suitable regularity conditions, minimizing (eq. 6.33) for a certain sample x_t corresponds to solve the system of M equations:

$$\frac{\partial}{\partial\theta_j}L_W(x;\theta) = 0, \quad j = 1,\dots,M.$$

That is,

$$\frac{1}{2\pi}\frac{\partial}{\partial\theta_j}\int_{-\pi}^{\pi}\log f(\nu;\theta)d\nu + \frac{1}{n}\frac{\partial}{\partial\theta_j}x'A(\theta)x = 0. \tag{6.34}$$

Scaling choice of the parameters

A further simplification of equation (6.34) can be obtained with a special choice of the parameters. In fact, let us indicate with

$$\eta = (\theta_2, \theta_3, \dots, \theta_M), \quad \theta^* = (1, \eta).$$

We then choose the scaling parameter θ_1 such that:

$$f(\nu;\theta) = \theta_1 f(\nu;\theta^*), \quad \int_{-\pi}^{\pi} \log f(\nu;\theta^*) d\nu = 0. \tag{6.35}$$

Then, due to (Prop. 6.7), θ_1 coincides with $2\pi\sigma_\epsilon^2$. with this choice equations (6.34) become:

$$\begin{cases} \dfrac{1}{2\pi\theta_1} - \dfrac{1}{n\theta_1^2} S(\theta^*) = 0, \\ \dfrac{\partial}{\partial\theta_j} S(\theta^*) = 0, \quad j = 2, \dots, M \end{cases} \tag{6.36}$$

because, by definition (eq. 6.31), $A(\theta) = \dfrac{1}{\theta_1} A(\theta^*)$ and we set: $S(\theta) = x' A(\theta) x$.

The estimation procedure

Let $S^*(\eta) = S(\theta^*)$, the estimation procedure is:

- one obtains $\widehat{\eta}$ solving:

$$\partial_{\theta_j} S^*(\eta) = 0, \quad j = 2, \dots, M \tag{6.37}$$

- the one finds $\widehat{\sigma}_\epsilon^2$ by setting:

$$\widehat{\sigma}_\epsilon^2 = \frac{2\pi}{n} S^*(\widehat{\eta}). \tag{6.38}$$

What we can say about the asymptotic distribution of the estimator $\widehat{\theta}$ defined by (eq. 6.34)? The following result holds:

Theorem 6.2. *Let $\widehat{\theta}$ be the estimator defined by (eq. 6.34). Then, under the same conditions of Theorem 6.1, one has:*

- *$\widehat{\theta}$ is asymptotically correct:*

$$\widehat{\theta} \to \theta_0, \quad \textit{quasi sicuramente}$$

- *moreover,*

$$\sqrt{n}\widehat{\theta} \to_d \zeta,$$

where ζ is defined in Theorem 6.1.

Therefore, the Whittle estimator has the same asymptotic distribution of the exact maximum likelihood estimator. Furthermore, it is asymptotically efficient fot Gaussian processes[3].

Remark 6.7. Observe that the choice (eq. 6.35), other than simplifying (eq. 6.34), implies by definition (see eq. 6.26):

$$D_{1j} = D_{j1} = 0, \quad j \neq 1.$$

Thus, $\operatorname{cov}(\theta_1, \theta_j) = 0$ for any $j \neq 1$. The scaling parameter σ_ϵ^2 is then asymptotically independent of the other estimated parameters.

3) A discussion about the efficiency of the estimators for LRD processes can be found in [6].

The Whittle MLE estimation requires the evaluation of the matrix elements (eq. 6.31) for any value of the trial parameter θ. So that, when we deal with very long time series or one has to estimate a large number of parameters, the calculation can be computationally prohibitive.

Given that $1/f$ is regular at the origin, we can approximate (eq. 6.31) with the Riemann sums:

$$\widetilde{\alpha}(k-l) = 2\frac{1}{4\pi^2}\sum_{j=1}^{n^*}\frac{1}{f(\nu_j;\theta)}e^{i(k-l)\nu_j}\frac{2\pi}{n}, \tag{6.39}$$

where $\nu_j = \frac{2\pi}{n}j$, with $j = 1,\ldots,n^*$ and n^* is the integer part of $(n-1)/2$, are the *Fourier frequencies*.

A discrete version of the Whittle estimator

We have the following approximation of the Whittle estimator:

Prop. 6.8. *The following discrete estimator provides an approximation of the Whittle MLE (eq. 6.33):*

$$\widetilde{L}_W(\theta) = 2\frac{1}{2\pi}\left[\sum_{j=1}^{n^*}\log f(\nu_j;\theta)\frac{2\pi}{n} + \sum_{j=1}^{n^*}\frac{I(\nu_j)}{f(\nu_j;\theta)}\frac{2\pi}{n}\right]. \tag{6.40}$$

Proof: remember that the periodogram, set $\overline{x}_n = 0$ for large n (zero mean hypothesis), is given that (eq. 6.13):

$$I(\nu_j) = \frac{1}{2\pi n}\sum_{k,l}x_k e^{i(k-l)\nu_j}x_j.$$

The term $\frac{1}{n}x'Ax$ can be rewritten by using (eq. 6.39) as:

$$\frac{1}{n}\sum_{k,l}x_k\widetilde{A}_{kl}x_l = \frac{2}{n}\frac{1}{4\pi^2}\sum_{j=1}^{n^*}\sum_{k,l}\frac{x_k e^{i(k-l)\nu_j}x_l}{f(\nu_j;\theta)}\frac{2\pi}{n}$$

(using the periodogram formula)

$$= \frac{2}{n}\sum_{j=1}^{n^*}\frac{I(\nu_j)}{f(\nu_j;\theta)}. \quad \square$$

Because the periodogram can be evaluated by using **FFT**, $\widetilde{L}_W$ can be calculated very quickly for any θ. With the parameters choice $\theta = (\theta_1, \eta)$, minimizing (eq. 6.40) equals (see eq. 6.37 and eq. 6.38) to minimize

$$\widetilde{S}(\eta) = 2\sum_{j=1}^{n^*}\frac{I(\nu_j)}{f(\nu_j;\theta^*)}, \tag{6.41}$$

with respect to η and then setting:

$$\sigma_\epsilon^2 = \frac{2\pi}{n}\widetilde{S}(\widehat{\eta}). \tag{6.42}$$

Remark 6.8. By the time the parametric model is chosen, the method described above is equivalent substantially in calibrating the theoretical spectral density function $f(\nu;\theta)$ to the periodogram estimated through the available observations (estimation in the frequency domain).

H	$\widehat{H}$ FARIMA	Confidenza	$\widehat{H}$ fGn	Confidenza
0.6	0.60	[0.55,0.65]	0.54	[0.50,0.60]
0.7	0.73	[0.68,0.78]	0.66	[0.61,0.71]
0.8	0.80	[0.75,0.85]	0.71	[0.67,0.77]
0.9	0.87	[0.82,0.92]	0.78	[0.73,0.83]

Table 6.2: Estimations of the LRD parameter H of $n = 1000$ length time series generated by a FARIMA$(0, H-1/2, 0)$ process, performed calibrating the spectral density function of FARIMA and fGn.

Estimation through FARIMA$(0, H-1/2, 0)$

Let the stochastic model $X(t)$ be a Gaussian FARIMA$(0, H-1/2, 0)$ process (eq. 3.61). The X spectral density function is given by (eq. 5.40), namely:

$$f(\nu;\theta) = \frac{\sigma_\epsilon^2}{2\pi}(2(1-\cos\nu))^{1/2-H} = \frac{\sigma^2}{2\pi}(4\sin^2\frac{\nu}{2})^{1/2-H}.$$

The parameters to be estimated are $\theta_1 = \frac{\sigma_\epsilon^2}{2\pi}$ and $\theta_2 = H$. Observe that (eq. 6.35) holds.

Estimation through fGn

Let the stochastic model $X(t)$ be a fGn process (**Def.** 3.15). The spectral density function is given by (eq. 5.28), that is:

$$f_X(\nu) = \frac{\sigma^2}{\pi}\Gamma(2H+1)\sin(\pi H)(1-\cos\nu)\sum_{k\in\mathbb{Z}}|\nu+2k\pi|^{-2H-1}.$$

We set $\sigma^2 = 1$, and we make the usual choice of the parameters $\theta = (\theta_1, \eta)$. Then, we write:

$$f(\nu;\theta) = \theta_1 f_1(\nu;H) = \theta_1\left[C(H)f_X(\nu;H)\right],$$

where $C(H)$ is chosen such that $\int_{-\pi}^{\pi}\log f_1(\nu;H)d\nu = 0$ holds. Then, one finds:

$$C(H) = \exp\left(-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_X(\nu)d\nu\right). \tag{6.43}$$

Example 6.10. Tables (Tab. 6.2) and (Tab. 6.3) provide the estimates of $\widehat{H}$ performed calibrating FARIMA$(0, H-1/2, 0)$ and fGn models to time series generated by a FARIMA$(0, H-1/2, 0)$ process (Tab. 6.2) and generated by a fractional Gaussian noise (Tab. 6.3).

We apply now the Whittle MLE method in order to estimate the long-range dependence parameter H for some real data introduced before.

Nile-data

Let us consider the times series which describes the Nile river minima (Fig. 6.5). After subtracting the mean value, the H parameter estimates are:

$$\widehat{H} = 0.90 \pm 0.05,$$

H	$\widehat{H}$ FARIMA	Confidence	$\widehat{H}$ fGn	Confidence
0.6	0.63	[0.59,0.68]	0.58	[0.53,0.62]
0.7	0.71	[0.66,0.76]	0.65	[0.60,0.70]
0.8	0.87	[0.82,0.92]	0.78	[0.73,0.83]
0.9	1.0	[0.97,1.07]	0.91	[0.87,0.96]

Table 6.3: Estimations of the LRD parameter H of $n = 1000$ length time series generated by a fGn process, performed calibrating the spectral density function of FARIMA and fGn.

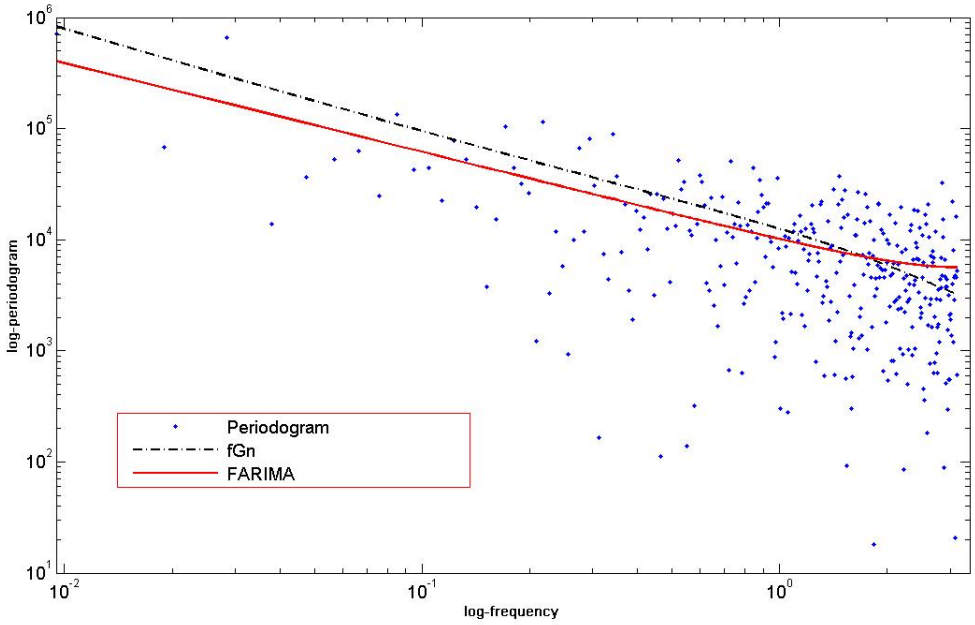

Figure 6.16: Periodogram and calibrated spectral densities in logarithmic scale for the Nile river data.

obtained calibrating a FARIMA$(0, H - 1/2, 0)$ model through the Whitte method. Differently, if one used a fGn model, one would find:

$$\widehat{H} = 0.80 \pm 0.05.$$

Figure (6.16) shows the periodogram and the corresponding spectral densities od the calibrated models.

Tree-rings

Consider the tree rings time series (Fig. 6.7). The H parameter estimates are:

$$\widehat{H} = 0.99 \pm 0.03,$$

for a FARIMA$(0, H - 1/2, 0)$ model and:

$$\widehat{H} = 0.89 \pm 0.03,$$

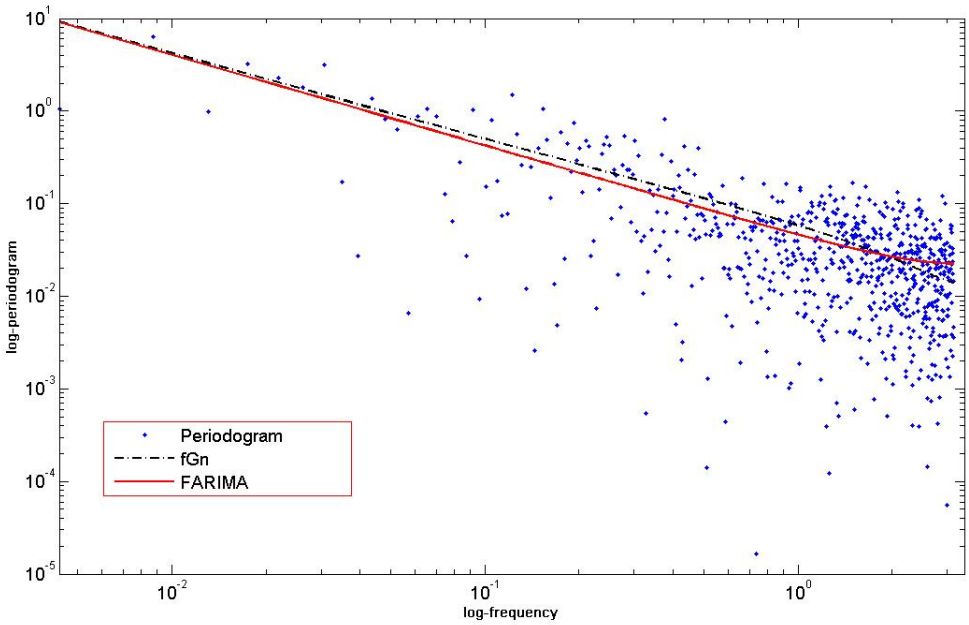

Figure 6.17: Periodogram and calibrated spectral densities in logarithmic scale for the tree rings data.

if a fGn is used. Figure (6.17) shows the behavior of the periodogram and the corresponding theoretical spectral density functions of the calibrated models.

Global temperature-data

Finally, we evaluate H for the global temperature data of the northern hemisphere (Fig. 6.9). One finds:

$$\widehat{H} = 0.87 \pm 0.03,$$

for a FARIMA$(0, H - 1/2, 0)$ model and:

$$\widehat{H} = 0.78 \pm 0.03,$$

for a fGn model (Fig. 6.18).

6.5.3 Robustness

Along this section we have seen that MLE methods are preferable with respect to certain heuristic methods because they provide results which are much more easier to be interpreted. Moreover, such methods turn out to be more robust and efficient. We want to briefly analyze what concerning the robustness of maximum likelihood methods with respect to certain deviation from the original hypothesis.

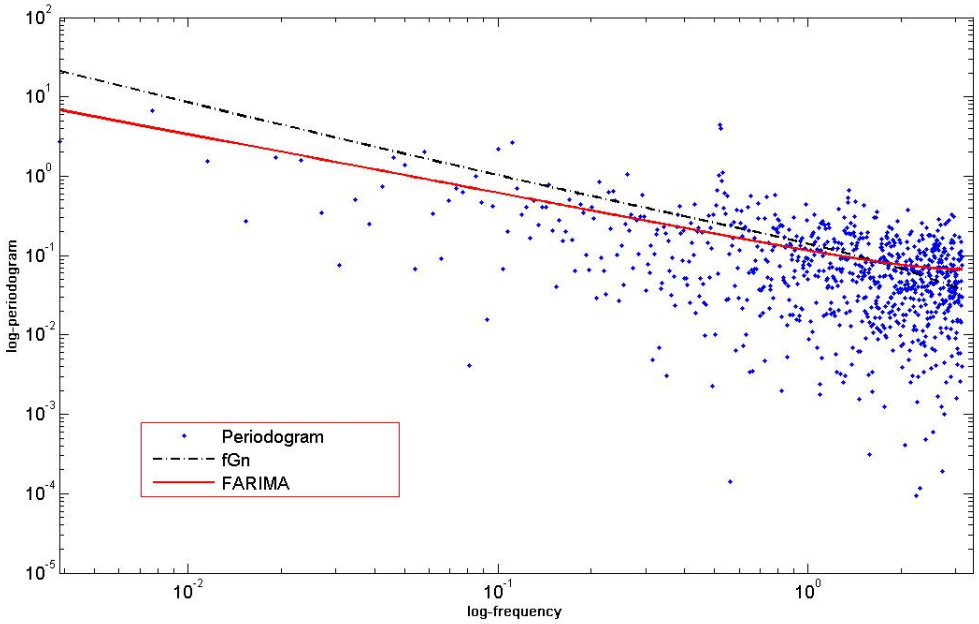

Figure 6.18: Periodogram and calibrated spectral densities in logarithmic scale for the global temperature data.

Deviation from stationarity

In the case of a slight deviation from the stationary hypothesis, for instance introducing an asymptotically nullifying trend, we have seen that the MLE estimates result to be only slightly distorted, especially if compared with the estimates obtained with heuristic methods (see Tab. 6.1).

Deviation from normality

Deviation from normality must be taken into account but do not disturb exceedingly the estimates of H (Figure 6.19 and Table 6.4). Then, if one has a sequence of observations which show a significative deviation from normality, Gaussian stochastic models cannot be of course a satisfactory model for the observe data but can be anyway used for estimations.

Deviation from linearity

Deviation from linearity, that means a wrong model choice for the estimations, can drastically affect the estimates. In this case, non linear models has to be used. As an example of a fairly used class of non-linear models, we cite the so called *fractional exponential* models, shortly FEXP (for more detailed examples and discussion we remind to [3]).

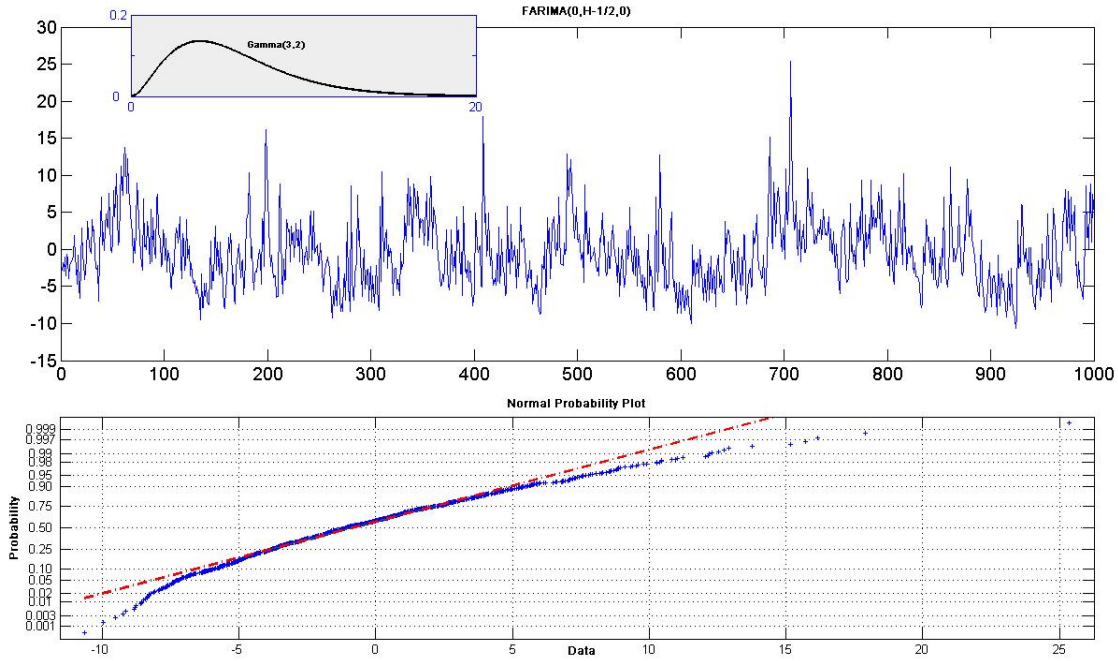

Figure 6.19: Realization of length $n = 1000$ of a FARIMA$(0, H - 1/2, 0)$ process, with $H = 0.9$, obtained through fractional integration of a $WN(0, 12)$ having Gamma(k, θ) distribution with $k = 3$ and $\theta = 2$ (top figure). Normal probability plot (bottom figure) .

N	$\hat{H}$ farima	Confidence	$\hat{H}$ fGn	Confidence
1	0.93	[0.88,0.98]	0.83	[0.78,0.88]
2	0.89	[0.84,0.94]	0.79	[0.74,0.84]
3	0.85	[0.80,0.90]	0.76	[0.70,0.80]
4	0.91	[0.87,0.97]	0.82	[0.77,0.87]
5	0.93	[0.88,0.98]	0.83	[0.79,0.89]
6	0.86	[0.81,0.91]	0.77	[0.72,0.82]
7	0.89	[0.84,0.94]	0.80	[0.76,0.86]
8	0.88	[0.83,0.93]	0.78	[0.73,0.83]
9	0.85	[0.80,0.90]	0.76	[0.70,0.80]
10	0.88	[0.83,0.93]	0.78	[0.73,0.83]
mean	$\overline{H} = 0.89$	$\sigma = 0.03$	$\overline{H} = 0.79$	$\sigma = 0.03$

Table 6.4: Gaussian MLE estimates for the parameter H of 10 time series realization of length$n = 1000$ of a non-Gaussian FARIMA$(0, H - 1/2, 0)$ process (Fig. 6.19) with $H = 0.9$.

N	$\hat{H}$	Confidence
1	0.61	[0.55,0.68]
2	0.62	[0.56,0.69]
3	0.72	[0.66,0.79]
4	0.55	[0.49,0.61]
5	0.61	[0.55,0.68]
6	0.68	[0.62,0.75]
7	0.49	[0.43,0.56]
8	0.64	[0.58,0.71]
9	0.59	[0.53,0.65]
10	0.55	[0.49,0.62]
mean	$\overline{H} = 0.61$	$\sigma = 0.07$

Table 6.5: MLE estimations of the H parameter of a FARIMA$(0, H - 1/2, 0)$ process with $H = 0.9$, contaminated by the presence of occasional outliers with probability $p = 5\%$.

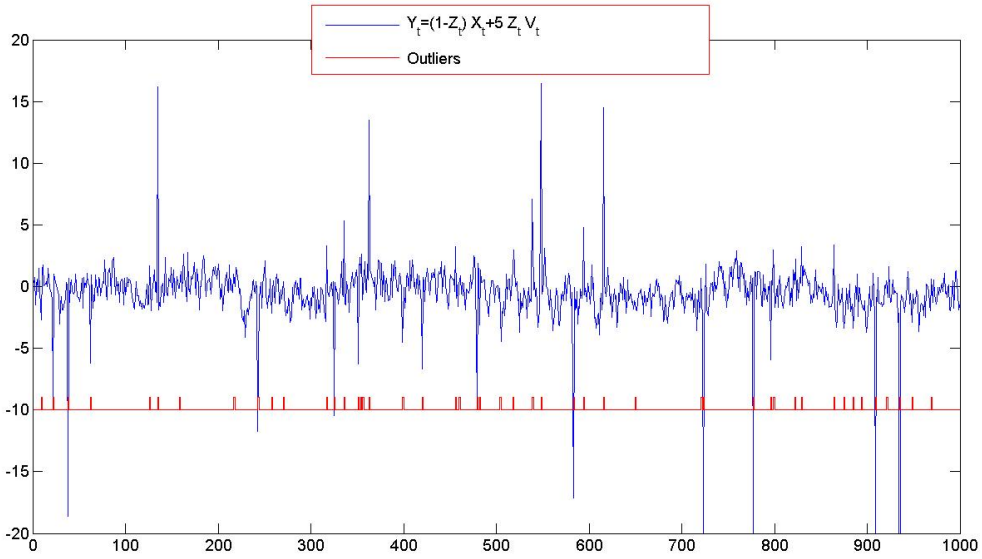

Figure 6.20: Gaussian FARIMA$(0, H - 1/2, 0)$ process with $H = 0.9$, contaminated by the presence of outliers with probability $p = 5\%$ and "amplitude" $c = 5$..

Deviations due to the presence of outliers

What happens to the MLE estimations of H if in the time series are present outlier observations? Suppose that $X(t)$ is a FARIMA$(0, H - 1/2, 0)$ process. Then, let us consider the process:

$$Y_t = (1 - Z_t)X_t + cZ_tV_t, \tag{6.44}$$

where:

- $\{Z_t\}_{t=1,\ldots,n}$ are independent *Bernoulli* variables with $P(Z_t = 1) = p$.

- $\{V_t\}_{t=1,\dots,n}$ is a sequence of **i.i.d.** outcomes extracted form a Student t distribution with two degrees of freedom.

The process Y_t has to be interpreted as the process X_t "contaminated" by a sequence of occasional outliers, which verifies with probability $p = P(Z_t = 1)$. The parameter c governs the "amplitude" of the outliers (Figure 6.20).

Table 6.5 shows that the presence of the **independent** outliers $\{V_t\}$ pushes the MLE estimate of H towards the value $\hat{H} = 0.5$. Thus, instead of estimating the dependence structure of the majority of the data (around 95% in Fig. 6.20), the estimate $\hat{H}$ reflects the independence of the occasional outliers.

For any of the typology of deviations from the original model that one can encounter, there exists different methodologies. However, no one of the known methods are universally applicable and the robustness with respect to a certain characteristic does in general not hold for another [3].

Bibliography

[1] R.T. Baillie and M.L. King (Editors), Fractional Differencing and Long Memory Processes, *Journal of Econometrics* **73** No 1 (1996) 1-324. Annals of Econometrics: Special Issue of J. Econometrics: Editors' Introduction pp. 1-2

[2] R.T. Baillie , Long Memory Processes and fractional integration in econometrics, *Journal of Econometrics* **73** No 1 (1996) 5-59.

[3] J. Beran, *Statistics for Long–Memory Process*, Chapman & Hall, London, 1994.

[4] P.J. Brockwell, R.A. Davis, *Time series: theory and methods*, Springer-Verlag, second edition, 1991.

[5] C. Chatfield, *the Analysis of Time Series, an Introduction* Chapman & Hall, London, 1994.

[6] R. Dahlhaus, Efficient Parameter Estimation for Self-Similar Processes *Annals of Statistics, Vol. 17, No. 4 (Dec., 1989)*, pp. 1749-1766

[7] K. Falconer, *Fractal Geometry* Wiley, Chichester, 1990.

[8] R. Gay, C.C. Heyde, On a class of random field models which allows long range dependence, *Biometrika*, **77** (1990)

[9] C.W.J. Granger, Long memory relationship and the aggregation of dynamic models, *J. Econometrics*, **14** (1980) 227-238.

[10] C.W.J. Granger and Z. Ding (1996): Varieties in long memory models, *J. Econometrics (Annals of Econometrics)* **73** No 1 (1996) 61-78.

[11] C.W.J. Granger and R. Joyeux, An introduction to long memory time series models and fractional differencing, *J. Time Series Analysis* **1** (1980) 15-29.

[12] J.R.M. Hosking, Asymptotic distributions of the sample mean, autocovariance and autocorrelation of long memory time series, *J. Econometrics (Annals of Econometrics)* **73** No 1 (1996) 261-284.

[13] H.E Hurst, Long-term storage capacity of reservoirs, *Trans. Amer. Soc. Civil Eng.*/**116** (1951) 770-808.

[14] B.B. Mandelbrot, A fast fractional Gaussian noise generator, *Water Resources Research* **7** No 3 (1971) 543-553.

[15] B.B. Mandelbrot, Note on the definition and stationarity of fractional Gaussian noise, *Journal of Hydrology* **30** (1976) 407-409.

[16] B.B. Mandelbrot, *Gaussian Self-Affinity and Fractals* (Springer-Verlag, New York, 2002), pp. 654 [Selected Works of Benoit B. Mandelbrot, SELECTA H]

[17] B.B. Mandelbrot, Statistical Methodology for nonperiodic cycles: From the covariance of R/S analysis, *Ann.Econ.Soc.Meas.*, **1,No.3** (259-290)

[18] B.B. Mandelbrot, Limit Theorems of the self-normalized range for weakly and strongly dependent processes *Z. Whar. verw. Beb.*, **31** (271-285)

[19] B.B. Mandelbrot and J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10** (1968) 422-433.

[20] M.S. Taqqu, Fractional Brownian Motion and Long-Range Dependence, to appear in: P. Doukan, G. Oppenheim and M.S. Taqqu (Editors), *Long-range Dependence: Theory and Applications* (Birkäuser, Basel and Boston, 2002).

[21] K.C. Wang and C.W. Lung, Long-time correlation effects and fractal Brownian motion, *Phys. Lett. A* **151** No 3/4 (1990), 119-121.

[22] K.C. Wang, Long-time correlation effects and biased anomalous diffusion, *Phys. Rev. A* **45** No 2 (1992), 833-837.

[23] P.Whittle, Topographic correlation, power-law covariance functions, ad diffusion. *Biometrika, 49* (1962).

7 Introduction to Fractional calculus

Let X_t, $t \in \mathbb{Z}$, be a simple ARIMA$(0,k,0)$, with $k \in \mathbb{Z}_+$, process. Namely, X_t is defined by the following differences equation

$$(1-B)^k X_t = \sum_{j=0}^{k} \binom{k}{j} (-1)^j X_{t-j} = \epsilon_t, \quad t \in \mathbb{Z},$$

where ϵ_t is WN$(0,\sigma^2)$ process. In the left hand side we have a discrete k-th order derivative, so that one can actually say that the k-th discrete derivative of an ARIMA$(0,k,0)$ process is a white noise process.

Observe that the present value X_t depends on the past only up to time $t-k$. One can then heuristically think that an ARIMA$(0,k,0)$ process is, in a certain sense, a "short-memory" process[1]. In order to make long-memory to appear, we let the parameter k to be a real number α.

Let $\alpha > 0$ be given, then the ARIMA process becomes a FARIMA$(0,\alpha,0)$ process (see eq. 3.61). Namely, X_t solves the following equation:

$$(1-B)^\alpha X_t = \sum_{j=0}^{\infty} \binom{\alpha}{j} (-1)^j X_{t-j} = \epsilon_j, \tag{7.1}$$

where:

$$\binom{\alpha}{j} = \frac{\Gamma(1+\alpha)}{\Gamma(j+1)\Gamma(1-j+\alpha)}.$$

The value at the present time t now depends on the whole past history and, according to (Def. 6.1), the process is stationary and actually exhibits LRD for any $0 < \alpha < 1/2$.

The operator $(1-B)^\alpha$ has to be interpreted as a certain type of discrete "time-fractional derivative" of order α. In the same way $(1-B)^{-\alpha}$ would represent a discrete "fractional integral" of order α.

Consider now a smooth real function $f(t)$ defined on a closed interval $I \subset \mathbb{R}$. We introduce a one dimensional lattice $\mathcal{G}_{\delta t}^N = \{t_n = \min(I) + n\delta t, \quad n \in \mathbb{Z}_N\}$, contained on I and such that $\delta t = \frac{d(I)}{(N-1)}$, where $d(I)$ indicates the diameter of the interval. Furthermore, let $f(t_n) = f_n$ indicate the restriction of f to $\mathcal{G}_{\delta t}^N$ evaluated on $t_n \in \mathcal{G}_{\delta t}^N$.

The restriction of the k-th derivative $D_t^k f(t) = f^{(k)}(t)$ to $\mathcal{G}_{\delta t}^N$ could be written as

$$D_t^k f(t_n) = \frac{(1-B)^k}{\delta t^k} f_{n+1}, \quad k \geq 0, \quad t_n \in \mathcal{G}_{\delta t}^N, \tag{7.2}$$

which actually represents an approximation of order $O(\delta t)$ of $f^k(t_n)$. Expanding the binomial in (eq. 7.2), one has:

$$D_t^k f(t_n) = \frac{1}{\delta t^k} \sum_{j=0}^{k_*} (-1)^j \binom{k}{j} f_{n+1-j}, \tag{7.3}$$

where $k_* = \min(k, n+1)$.

1) An ARIMA process is always not stationary, in fact the characteristic polynomial has at least k roots equal to one. Then, we are just naive speaking and not according to the definitions of Section 6 which require stationarity.

Remark 7.1. Observe that in this formulation we must not consider the terms closest to the boundary. The approximation is indeed valid only in the points t_n, with $n \geq k-1$.

Consider $0 < \alpha < 1$, then, as in (eq. 7.1), one is naturally led to take

$$D_t^\alpha f(t_n) = \frac{1}{\delta t^\alpha} \sum_{j=0}^{n+1} (-1)^j \binom{\alpha}{j} f_{n+1-j}, \tag{7.4}$$

and call it (discrete) fractional derivative of order α. Then, taking the limit $\delta t \to 0$, one has the following definition:

Def. 7.1 (Grünwald-Letnikov derivative)**.** Let f be a sufficiently well behaved function, for instance one can take f absolutely continuous on I, then the *Grünwald-Letnikov* (GL) fractional derivative of order $0 < \alpha < 1$ is defined as:

$$D_t^\alpha f(t) = \lim_{\delta t \to 0} \frac{1}{\delta t^\alpha} \sum_{j=0}^{[t/\delta t]} (-1)^j \binom{\alpha}{j} f(t - j\delta t). \tag{7.5}$$

Remark 7.2. Observe that the GL fractional derivative is *non-local* in the sense that as $\delta t \to 0$ the sum in (eq. 7.5) involves an infinite number of elements, namely the whole "history" of the function up to "time" t.

It is possible to show [19] that the limit in (eq. 7.5) converges to the integral equation:

$$\mathcal{D}_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t f(s)(t-s)^{-\alpha} ds, \tag{7.6}$$

for any $t \in (0, b)$, $b > 0$, and $0 < \alpha < 1$. This is the so called *Riemann-Liouville* (RL) fractional derivative of order α.

Remark 7.3. As a consequence, equation (7.4) defines indeed a discrete approximation of the RL fractional derivative and can be used for numerical purposes.

As it turns out to be, the RL fractional derivative is the left inverse of the so called (left-side) *Riemann-Liouville* fractional integral of order $0 < \alpha < 1$:

$$\mathcal{I}_t^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s) ds, \quad t \in [a, b]. \tag{7.7}$$

The above examples suggest that the so called *fractional calculus*, that is the theory of fractional integrals and derivatives, seems naturally to be very appropriate for studying and modeling systems which exhibit memory effects. For instance, FARIMA models can be seen as a discrete fractional integral of a white noise. In the same way, as we will see, also fractional Brownian motion can be seen as a fractional integral of a Gaussian white noise (giving reason of the name fractional Brownian motion).

Remark 7.4. We observe that (eq. 7.7) is a convolution type integral, that is

$$\mathcal{I}_t^\alpha f(t) = \int_0^t K(t-s) f(s) ds, \tag{7.8}$$

where

$$K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)} \tag{7.9}$$

is a memory kernel.

Let us consider a time homogeneous stochastic diffusion $Q(t)$, $t \geq 0$, with drift $a(x)$ and diffusion term $\sigma(x)$. Its marginal density function (see Remark 4.14) satisfies the Fokker-Planck equation (eq. 3.33), which in integral form reads:

$$u(x,t) = u_0(x) + \int_0^t \mathcal{P}_x u(x,s)ds, \tag{7.10}$$

where $u(x,0) = u_0(x)$ is the initial density, and $\mathcal{P}_x$ is the time homogeneous Fokker-Planck operator

$$\mathcal{P}_x = \partial_{xx}\sigma(x) - \partial_x a(x) + [\partial_x \sigma(x) - a(x)]\,\partial_x + \frac{1}{2}\sigma^2(x)\partial_{xx}. \tag{7.11}$$

We say that (eq. 7.10) is a "Markovian equation" in the sense that it comes from a Markovian diffusion. Clearly, it is also a local equation in the sense that can be written in differential form (see eq. 3.33). However, also non-Markovian processes can have marginal density function which satisfies local equations. For instance, the standard fBm marginal density function satisfies the "stretched" diffusion equation:

$$\partial_t u(t,x) = \frac{1}{2}\alpha t^{\alpha-1}\partial_{xx}u(x,t), \quad 0 < 2\alpha = H < 1. \tag{7.12}$$

Let us now introduce a memory kernel in (eq. 7.10), namely

$$u(x,t) = u_0(x) + \int_0^t K(t-s)\mathcal{P}_x u(x,s)ds. \tag{7.13}$$

These Volterra type equations, clearly no longer local in time, will be studied in Chapter 9. They are often encountered when studying physical phenomena exhibiting memory effects and are usually related to relaxation and diffusion problems in complex systems (see Srokowsky [22] for examples). Such equations are often argued by phenomenological considerations and, some times, can be more or less rigorously derived starting from a microscopic description [2,23,5,22].

If the memory kernel is taken in a suitable way, (eq. 7.13) has fundamental solution that can be interpreted as a probability density function (see Section 9.2.1 of Chapter 9). If this is the case, one can ever find a number of stochastic processes which have marginal density function that evolves in time according to (eq. 7.13). These processes are obviously not Markovian. Therefore, there is actually a whole equivalence class of such processes with the same marginal density function but which differ in the sense of finite-dimensional distributions. We will back later on this topic. At this point we just observe that a suitable choice of the memory kernel $K(t)$ would be just the power kernel (eq. 7.9) with $0 < \alpha < 1$. This choice leads to the so called *time-fractional Fokker-Plack equation* of order α:

$$u(x,t) = u_0(x) + \int_0^t \frac{1}{\Gamma(\alpha)}(t-s)^{\alpha-1}\mathcal{P}_x u(x,s)ds = u_0(x) + \mathcal{I}_t^\alpha \mathcal{P}_x u(x,t). \tag{7.14}$$

This case is noteworthy because, as we will see more in detail later, it corresponds to the only suitable kernel choice which can provide self-similar processes (see Theorem 9.1).

In this chapter we will introduce the basic notions of fractional calculus, providing useful examples and applications, both from a deterministic and from a stochastic point of view.

7.1 Fractional integrals and derivatives

One of the common way to motivate the definition of fractional integrals is to start with an n-tuple iterated integral and show that it can be expressed as a single integral involving the parameter n. The

fractional integral of order $\alpha > 0$ is then defined just replacing the integer n by the positive real α.

Let $a < b$ be two real numbers and ϕ a function defined on $I = [a, b]$. Then, we have the following result:

Prop. 7.1. *By induction it is easy to show that, for any integer $n \geq 1$, a multiple integral of ϕ can be expressed as:*

$$\int_a^{t_n} \cdots \left\{ \int_a^{t_2} \left\{ \int_a^{t_1} \phi(s)ds \right\} dt_1 \right\} \cdots dt_{n-1} = \frac{1}{(n-1)!} \int_a^{t_n} (t_n - s)^{n-1} \phi(s)ds. \tag{7.15}$$

By replacing the integer n by a positive real number α, we obtain the following definition:

Def. 7.2 (Riemann-Liouville fractional integral)**.** Let $\phi \in L^1([a, b])$ and $\alpha > 0$. Then, for any $t \in (a, b)$, the integrals

$$\mathcal{I}_{a+}^{\alpha} \phi(t) = \frac{1}{\Gamma(\alpha)} \int_a^t (t - s)^{\alpha-1} \phi(s)ds = \frac{1}{\Gamma(\alpha)} \int_a^b (t - s)_+^{\alpha-1} \phi(s)ds, \tag{7.16}$$

$$\mathcal{I}_{b-}^{\alpha} \phi(t) = \frac{1}{\Gamma(\alpha)} \int_t^b (s - t)^{\alpha-1} \phi(s)ds = \frac{1}{\Gamma(\alpha)} \int_a^b (s - t)_+^{\alpha-1} \phi(s)ds, \tag{7.17}$$

are called left-side and right-side *Riemann-Liouville fractional integrals* of order $\alpha > 0$.

Remark 7.5. The fractional integrals $\mathcal{I}_{a+}^{\alpha} \phi(t)$ and $\mathcal{I}_{b-}^{\alpha} \phi(t)$ are well defined for any $\phi \in L^p([a, b])$, $p \geq 1$. In fact, suppose $\phi \in L^1([a, b])$, then, for any $t \in (a, b)$,

$$\Gamma(\alpha) \int_a^b |\mathcal{I}_{a+}^{\alpha} \phi(t)|dt \leq \Gamma(\alpha) \int_a^b \mathcal{I}_{a+}^{\alpha} |\phi(t)|dt = \int_a^b \int_a^b |\phi(s)|(t - s)_+^{\alpha-1} dsdt$$

(by changing the integration order)

$$= \alpha^{-1} \int_a^b |\phi(t)|(b - s)^{\alpha} ds \leq \alpha^{-1} (b - a)^{\alpha} \int_a^b |\phi(s)|ds < \infty,$$

and the same for the right-side integral.

Example 7.1. Evaluate the α-th fractional integral of $\phi(t) = (t - a)^{-\alpha}$, $0 < \alpha < 1$. One has,

$$\frac{1}{\Gamma(\alpha)} \int_a^t (t - s)^{\alpha-1} (s - a)^{-\alpha} ds$$

(after the change of variables $s = a + (t - a)z$)

$$= \frac{1}{\Gamma(\alpha)} \int_0^1 (1 - z)^{\alpha-1} z^{-\alpha} dz = \frac{1}{\Gamma(\alpha)} \mathcal{B}(\alpha, 1 - \alpha) = \Gamma(1 - \alpha),$$

where:

$$\mathcal{B}(x, y) = \int_0^1 s^{x-1} (1 - s)^{y-1} ds = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}, \quad x, y > 0.$$

is the Beta function. Thus,

$$\mathcal{I}_{a+}^{\alpha} (t - a)^{-\alpha} = \Gamma(1 - \alpha). \tag{7.18}$$

Example 7.2. In the same way, one can easily show that, for any $\gamma > -1$ and $\alpha > 0$:

$$\mathcal{I}_{a+}^{\alpha} (t - a)^{\gamma} = \frac{\Gamma(\gamma + 1)}{\Gamma(\alpha + \gamma + 1)} (t - a)^{\gamma+\alpha}, \quad t \in (a, b). \tag{7.19}$$

Remark 7.6. By convention is set: $\mathcal{I}^0 \phi(t) = \phi(t)$.

Properties of fractional integrals

Let $\alpha > 0$, then the fractional integrals, both the left-side and right-side, $\mathcal{I}^\alpha$ have the following properties (see [20]):

1. Reflection property: let Q be the reflection operator: $Q\phi(t) = \phi(a - b - t)$, $t \in [a, b]$. Then,
$$Q\mathcal{I}^\alpha_{a+} = \mathcal{I}^\alpha_{b-} Q. \tag{7.20}$$
2. Semigroup property: for any $\phi \in L^1([a, b])$
$$\mathcal{I}^\alpha \mathcal{I}^\beta \phi(t) = \mathcal{I}^{\alpha+\beta}\phi(t), \quad \alpha, \beta > 0. \tag{7.21}$$
3. $\mathcal{I}^\alpha \phi = 0$ implies that $\phi = 0$ almost everywhere.
4. Fractional integration by parts formula: let $\phi \in L^p([a, b])$ and $\psi \in L^q([a, b])$ either with $\alpha \geq 1$, $p = q = 1$, or with $0 < \alpha < 1$, $\frac{1}{p} + \frac{1}{q} \leq 1 + \alpha$, $p, q > 1$. Then,
$$\int_a^b \phi(s)\mathcal{I}^\alpha_{a+}\psi(s)ds = \int_a^b \mathcal{I}^\alpha_{b-}\phi(s)\psi(s)ds. \tag{7.22}$$

Remark 7.7. Observe that the second property implies that if $f = \mathcal{I}^\alpha_{a+}\phi_1$ and $f = \mathcal{I}^\alpha_{a+}\phi_2$, then $\phi_1 = \phi_2$ a.e., that is, the inverse of the fractional integral is a.e. unique. This property is used in the definition of fractional derivatives.

In fact, as in classical calculus, fractional derivatives are defined as the left inverse of fractional integrals. In other words, because of (eq. 7.21), we have to extend in some way the Definition (7.2) to $\alpha < 0$.

Remark 7.8. Observe that we can not simply substitute $\alpha < 0$ in Definition (7.2) since the integrals diverge. In fact, as in the case of ordinary calculus, the definition of derivative involves more restrictions than integrals.

Def. 7.3 (Rimeann-Liouville fractional derivatives)**.** Let $0 < \alpha < 1$. Then, for any $t \in (a, b)$, the integrals
$$\mathcal{D}^\alpha_{a+} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t - s)^{-\alpha} f(s)ds, \tag{7.23}$$
$$\mathcal{D}^\alpha_{b-} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_t^b (t - s)^{-\alpha} f(s)ds, \tag{7.24}$$
are called left-side and right-side *Riemann-Liouville* fractional derivatives of order $0 < \alpha < 1$.

Fractional derivatives of order $0 < \alpha < 1$ are well defined if, for example, f is differentiable. In fact, if this is the case, one has[2]:
$$\begin{aligned}
\mathcal{D}^\alpha_{a+} f(t) &= \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t - s)^{-\alpha} \left(f(a) + \int_a^s f'(u)du \right) ds \\
&= \frac{1}{(1-\alpha)\Gamma(1-\alpha)} \frac{d}{dt} \left((t - a)^{1-\alpha} + \int_a^u (t - u)^{1-\alpha} f'(u)du \right) \\
&= \frac{1}{\Gamma(1-\alpha)} \left(f(a)(t - a)^{-\alpha} + \int_a^t f'(u)(t - u)^{-\alpha}du \right).
\end{aligned}$$

2) The case of te right-side derivative is the same.

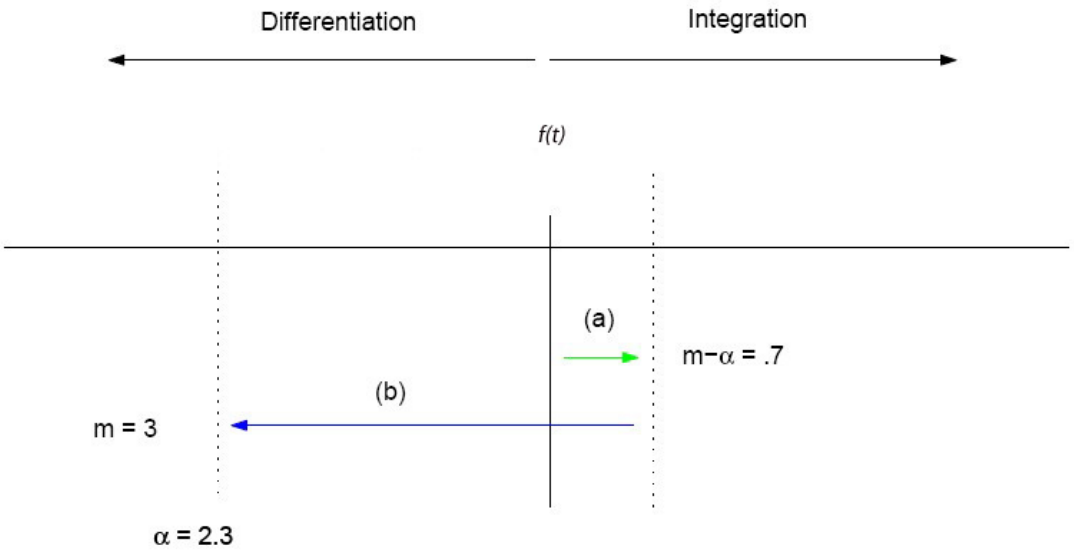

Figure 7.1: Graphical representation of the Riemann-Liouville fractional derivative of order $\alpha = 2.3$. First one integrates by order $m - \alpha = 0.7$ (a), then one differentiates by order $m = 3$ (b).

This is not surprising because we expect that if a function is differentiable then it also should be fractional differentiable of order $0 < \alpha < 1$.

In the following we just consider the left-side derivative. The corresponding properties of the right-side derivative can be obtained using (eq. 7.20), which is still valid for $-1 < \alpha < 0$.

Example 7.3. Let $f(t) = (t-a)^{\alpha-1}$, $t \in (a,b)$ and $0 < \alpha < 1$. Then,

$$\mathcal{D}^{\alpha}_{a+} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^{-\alpha} (s-a)^{\alpha-1} ds = \Gamma(\alpha) \frac{d}{du} 1 = 0. \tag{7.25}$$

Example 7.4. Let $C \in \mathbb{R}$, then:

$$\mathcal{D}^{\alpha}_{a+} C = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^{-\alpha} C ds = \frac{C}{\Gamma(1-\alpha)} (t-a)^{-\alpha}. \tag{7.26}$$

The Riemann-Liouville fractional derivative of a constant in not zero.

Properties of fractional derivatives

Let $0 < \alpha < 1$. The fractional derivative $\mathcal{D}^{\alpha}_{a+}$ has the following properties (see [20]):

1. For any $\phi \in L^1([a,b])$, we have that:
$$\mathcal{D}^{\alpha}_{a+} \mathcal{I}^{\alpha}_{a+} \phi = \phi. \tag{7.27}$$
2. For any $f = \mathcal{I}^{\alpha}_{a+} \phi$, we have that:
$$\mathcal{I}^{\alpha}_{a+} \mathcal{D}^{\alpha}_{a+} f = f. \tag{7.28}$$

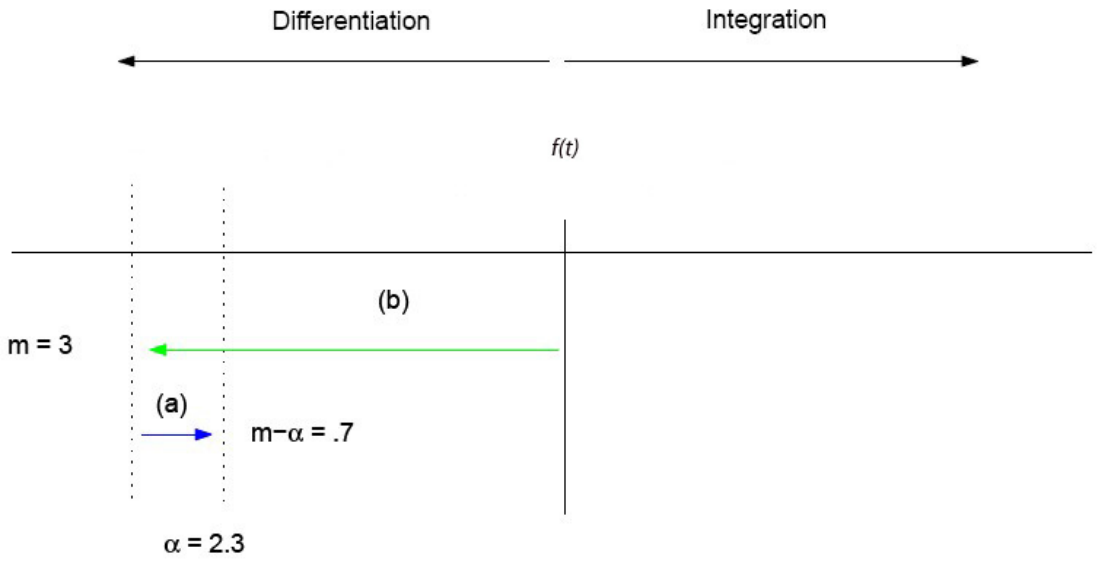

Figure 7.2: Graphical representation of the Caputo fractional derivative of order $\alpha = 2.3$. In this case, one first differentiates by order $m = 3$ (b), then one integrates by order $m - \alpha = 0.7$ (a).

3. The latter can be generalized. In fact, if the function $\mathcal{I}_{a+}^{1-\alpha} f$ is absolutely continuous on $[a, b]$, then:

$$\mathcal{I}_{a+}^{\alpha} \mathcal{D}_{a+}^{\alpha} f(t) = f(t) - \frac{\mathcal{I}_{a+}^{1-\alpha} f(a)}{\Gamma(\alpha)} (t-a)^{\alpha-1}, \quad t \in (a, b), \tag{7.29}$$

where $\mathcal{I}_{a+}^{1-\alpha} f(a) = \lim_{s \to a+} (\mathcal{I}_{a+}^{1-\alpha} f)(s)$, which is in general non-zero.

7.1.1 Two forms of fractional derivatives

It is possible to define fractional derivative operators for $\alpha \geq 1$ as well. The idea is to use usual integer order derivative operators. Consider an integer m such that $m - 1 < \alpha \leq m$. The first step is to integrate f by order $m - \alpha$ and then differentiate by m to obtain a resultant differentiation of order α (left-hand definition, see Figure 7.1). That is,

$$\mathcal{D}_{a+}^{\alpha} f(t) = \begin{cases} \dfrac{1}{\Gamma(m-\alpha)} \dfrac{d^m}{dt^m} \displaystyle\int_a^t (t-s)^{m-\alpha-1} f(s) ds, & m-1 < \alpha < m, \\ \dfrac{d^m}{dt^m} f(t), & \alpha = m. \end{cases} \tag{7.30}$$

Remark 7.9. Heuristically, the fractional derivative of order α is well defined if $f \in C^m([a, b])$, where $m - 1 < \alpha \leq m$.

There is another way to define fractional derivatives, which is schematized in Figure (7.2). One just has to invert the integration and derivation operations of before (right-hand definition). Then, one can define:

Def. 7.4 (Caputo-Dzherbashyan derivative)**.** Let $\alpha > 0$, then for any $t \in (a, b)$

$$ {}_*\mathcal{D}^{\alpha}_{a+} f(t) = \begin{cases} \dfrac{1}{\Gamma(m-\alpha)} \displaystyle\int_a^t (t-s)^{m-\alpha-1} \frac{d^m}{ds^m} f(s) ds, & m-1 < \alpha < m, \\ \dfrac{d^m}{dt^m} f(t), & \alpha = m. \end{cases} \tag{7.31} $$

is called (left-side) Caputo-Dzherbashyan (CD) derivative of order $\alpha > 0$.

Remark 7.10. Let $0 < \alpha < 1$ and $a = 0$. In this case we shall write $\mathcal{I}^{\alpha}_{0+} = \mathcal{I}^{\alpha}_t$. Consider a well defined function f, for instance take $f \in C^1(\mathbb{R}_+)$. Then, one has:

$$ \mathcal{I}^{\alpha}_t \, {}_*\mathcal{D}^{\alpha}_t f(t) = \mathcal{I}^{\alpha}_t \mathcal{I}^{1-\alpha}_t \frac{d}{dt} f(t) = \mathcal{I}_t \frac{d}{dt} f(t) = f(t) - f(0_+). \tag{7.32} $$

Therefore, (by using eq. 7.26) we have the following relationships between the RL and CD derivatives of order $0 < \alpha < 1$:

$$ {}_*\mathcal{D}^{\alpha}_t f(t) = \mathcal{D}^{\alpha}_t (f(t) - f(0_+)) = \mathcal{D}^{\alpha}_t f(t) - \frac{f(0_+)}{\Gamma(1-\alpha)} t^{-\alpha}. \tag{7.33} $$

Remark 7.11. The above equation (eq. 7.33), allows us to write down immediately the Grünwald-Letnikov approximation for the CD derivative (see eq. 7.5):

$$ {}_*D^{\alpha}_t f(t_n) = \frac{1}{\delta t^{\alpha}} \sum_{j=0}^{n+1} (-1)^j \binom{\alpha}{j} (f_{n+1-j} - f_0), \quad 0 < \alpha < 1. \tag{7.34} $$

Provided that the order is not integer, the two fractional derivatives differ because the derivative of order m does not generally commute with the fractional integral. We point out that the *Caputo* fractional derivative satisfies the relevant property of being zero when applied to a constant (see eq. 7.26 and eq. 7.33). More in general, it is zero when it is applied to any power function of non-negative integer degree less than m, where $m - 1 < \alpha \leq m$.

A generalization of (eq. 7.33) can be found in [10]

$$ {}_*\mathcal{D}^{\alpha}_t f(t) = \mathcal{D}^{\alpha}_t \left(f(t) - \sum_{k=0}^{m-1} f^{(k)}(0_+) \frac{t^k}{k!} \right) = \mathcal{D}^{\alpha}_t f(t) - \sum_{k=0}^{m-1} \frac{f^{(k)}(0_+) t^{k-\alpha}}{\Gamma(k-\alpha+1)}, \tag{7.35} $$

where $\alpha > 0$ and $m - 1 < \alpha < m$.

Remark 7.12. The CD fractional derivative represents a sort of regularization (in the origin) of the RL fractional derivative. Moreover, in order for the CD derivative to exist, all the limiting values $f^{(k)}(0_+)$ are required to be finite for any $k \leq m - 1$. Then, because the derivative of order m is required to exist and is subjected to some regularity conditions, the Caputo fractional derivative is in this sense more restrictive than the Riemann-Liouville derivative.

The RL fractional derivative $\mathcal{D}^{\alpha}_t$ is continuous with respect to α, at any positive integer, whereas ${}_*\mathcal{D}^{\alpha}_t$ is left-continuous. In fact, one can show that,

$$ \begin{cases} \lim\limits_{\alpha \to m-1_+} {}_*\mathcal{D}^{\alpha}_t f(t) = f^{(m-1)}(t) - f^{(m-1)}(0_+), \\ \lim\limits_{\alpha \to m_-} {}_*\mathcal{D}^{\alpha}_t f(t) = f^{(m)}(t). \end{cases} $$

The Caputo fractional derivative turns out to be very useful in treating initial-value problems for physical and engineering applications. In fact, in this case the initial conditions can be expressed in terms of integer-order derivatives. On the other hands, in the case of RL derivatives the required initial conditions are themselves of non-integer orders. This can be easily seen using the Laplace transformation.

Remark 7.13. Let $\alpha > 0$. We indicate with $\mathcal{L}\{f(t), s\} = \widetilde{f}(s)$ the Laplace transform of the function f with respect to t evaluated on $s \geq 0$, namely:

$$\widetilde{f}(s) = \int_0^\infty e^{-ts} f(t) dt, \quad s \geq 0. \tag{7.36}$$

Then, using the Laplace convolution theorem, one easily finds:

$$\mathcal{L}\{\mathcal{I}_t^\alpha f(t), s\} = s^{-\alpha} \widetilde{f}(s). \tag{7.37}$$

Keeping the remark above in mind, for any $\alpha > 0$ and $m - 1 < \alpha < m$, one has:

$$\mathcal{L}\{\mathcal{D}_t^\alpha f(t), s\} = \mathcal{L}\{\mathcal{D}_t^m \mathcal{I}_t^{m-\alpha} f(t), s\}$$

(by setting $g(t) = \mathcal{I}_t^{m-\alpha} f(t)$)

$$= \mathcal{L}\{\mathcal{D}_t^m g(t), s\} = s^m \widetilde{g}(s) - \sum_{k=0}^{m-1} s^{m-1-k} g^{(k)}(0_+)$$

$$= s^m \mathcal{L}\{\mathcal{I}^{m-\alpha} f(t), s\} - \sum_{k=0}^{m-1} s^k g^{(m-1-k)}(0_+)$$

and, by using (eq. 7.37), one finally finds:

$$\mathcal{L}\{\mathcal{D}_t^\alpha f(t), s\} = s^\alpha \widetilde{f}(s) - \sum_{k=0}^{m-1} s^k \mathcal{D}_t^{\alpha-1-k} f(0_+). \tag{7.38}$$

Thus, the needed initial conditions are, for each integer k from zero to $n-1$, fractional order derivatives of $f(t)$.

Let us consider the CD derivative. With an analogue calculation, one finds:

$$\mathcal{L}\{{}_*\mathcal{D}_t^\alpha f(t), s\} = s^\alpha \widetilde{f}(s) - \sum_{k=0}^{m-1} s^{\alpha-k-1} f^{(k)}(0_+). \tag{7.39}$$

In this case, the order α does not appear in the derivatives of $f(t)$, but rather in the preceding multiplier $s^{\alpha-k-1}$. So that, quite conveniently, integer order derivatives of $f(t)$ are used as initial conditions.

7.1.2 Fractional integrals and derivatives on the real line

Fractional integral and derivatives can be defined also in the real line:

Def. 7.5 (Fractional integrals on the real line)**.** Let $\alpha > 0$. The integrals,

$$\mathcal{I}_+^\alpha \phi(t) = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^t (t-s)^{\alpha-1} \phi(s) ds = \frac{1}{\Gamma(\alpha)} \int_{\mathbb{R}} (t-s)_+^{\alpha-1} \phi(s) ds, \quad t \in \mathbb{R}. \tag{7.40}$$

$$\mathcal{I}_-^\alpha \phi(t) = \frac{1}{\Gamma(\alpha)} \int_t^\infty (s-t)^{\alpha-1} \phi(s) ds = \frac{1}{\Gamma(\alpha)} \int_{\mathbb{R}} (s-t)_+^{\alpha-1} \phi(s) ds, \quad t \in \mathbb{R}, \tag{7.41}$$

are called *fractional integrals* of order $\alpha > 0$ on the real line.

Remark 7.14. We observe that $\mathcal{I}^{\alpha}\phi$ is well defined for instance if $0 < \alpha < 1$ and $\phi \in L^p(\mathbb{R})$ with $1 \leq p < 1/\alpha$ (see [18]).

In the same way, one can generalize (left-hand) fractional derivatives to the the real line. We concentrate on the case $0 < \alpha < 1$.

Def. 7.6 (Fractional derivatives on the real line)**.** Let $0 < \alpha < 1$. For any $t \in \mathbb{R}$, the integrals,

$$\mathcal{D}_{+}^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_{-\infty}^{t} (t-s)^{-\alpha} f(s) ds, \tag{7.42}$$

$$\mathcal{D}_{-}^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_{t}^{\infty} (t-s)^{-\alpha} f(s) ds, \tag{7.43}$$

are called *fractional derivatives* of order $0 < \alpha < 1$ on the real line.

Example 7.5. Let $a < b$, it is easy to show that, for any $\alpha > 0$ and $t \in \mathbb{R}$,

$$\mathcal{I}_{\pm}^{\alpha} 1_{[a,b)}(t) = \frac{1}{\Gamma(1+\alpha)} \left[(b-t)_{\mp}^{\alpha} - (a-t)_{\mp}^{\alpha} \right], \tag{7.44}$$

where we remember that $x_{-} = -\min(x,0) = \max(-x,0) = (-x)_{+}$. Moreover, for any $0 < \alpha < 1$, we have that:

$$\mathcal{D}_{\pm}^{\alpha} 1_{[a,b)}(t) = \frac{1}{\Gamma(1-\alpha)} \left[(b-t)_{\mp}^{-\alpha} - (a-t)_{\mp}^{-\alpha} \right]. \tag{7.45}$$

The fractional integrals $\mathcal{I}_{\pm}^{\alpha}$ have properties analogous to those of fractional integrals on an interval. For instance if $Qf(t) = f(-t)$ is the reflection operator on $\mathbb{R}$, one has $Q\mathcal{I}_{\pm}^{\alpha} f(t) = \mathcal{I}_{\mp}^{\alpha} Qf$. Also, the semigroup property (eq. 7.21)and the integration by parts formula (eq. 7.22) still hold.

Remark 7.15. Equation (7.27) is valid for functions $\phi \in L^1(\mathbb{R})$. This condition is not strong enough in many context.

The above remark suggests that one could need another more flexible definition for fractional derivative on the real line. Let f be a "sufficiently good" function, and $0 < \alpha < 1$, one has:

$$\mathcal{D}_{+}^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_{\mathbb{R}} (t-s)_{+}^{-\alpha} f(s) ds = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_{0}^{\infty} u^{-\alpha} f(t-u) du$$

$$\frac{1}{\Gamma(1-\alpha)} \int_{0}^{\infty} u^{-\alpha} \frac{d}{dt} f(t-u) du = \frac{\alpha}{\Gamma(1-\alpha)} \int_{0}^{\infty} \frac{d}{dt} f(t-u) \int_{u}^{\infty} \frac{1}{z^{\alpha+1}} dz du$$

(changing the integration order)

$$= \frac{\alpha}{\Gamma(1-\alpha)} \int_{0}^{\infty} \frac{1}{z^{\alpha+1}} \int_{0}^{z} \frac{d}{dt} f(t-u) du dz,$$

that is:

$$\mathcal{D}_{+}^{\alpha} f(t) = \frac{\alpha}{\Gamma(1-\alpha)} \int_{0}^{\infty} \frac{f(t) - f(t-z)}{z^{\alpha+1}} dz, \quad 0 < \alpha < 1. \tag{7.46}$$

In the same way, one finds:

$$\mathcal{D}_{-}^{\alpha} f(t) = \frac{\alpha}{\Gamma(1-\alpha)} \int_{0}^{\infty} \frac{f(t) - f(t+z)}{z^{\alpha+1}} dz, \quad 0 < \alpha < 1. \tag{7.47}$$

Def. 7.7 (Marchaud fractional derivatives)**.** Let $0 < \alpha < 1$. The operators:

$$\mathbf{D}^{\alpha}_{\pm} f(t) = \lim_{\epsilon \to 0} \mathbf{D}^{\alpha}_{\pm,\epsilon} f(t), \tag{7.48}$$

where:

$$\mathbf{D}^{\alpha}_{\pm,\epsilon} f(t) = \frac{\alpha}{\Gamma(1-\alpha)} \int_{\epsilon}^{\infty} \frac{f(t) - f(t \mp z)}{z^{\alpha+1}} dz, \tag{7.49}$$

are called *Marchaud fractional derivatives* of order $0 < \alpha < 1$ on the real line. The operators $\mathbf{D}^{\alpha}_{\pm,\epsilon}$ are called *truncated* Marchaud fractional derivatives.

Remark 7.16. The functions $\mathcal{D}^{\alpha}_{\pm} f(t)$ and $\mathbf{D}^{\alpha}_{\pm} f(t)$ are identical for many good enough function (see example 7.6). However, the operators $\mathcal{D}^{\alpha}_{\pm}$ and $\mathbf{D}^{\alpha}_{\pm}$ have different domains. For instance, we may take f to be constant, or even growing slowly enough at infinity, that $\mathbf{D}^{\alpha}_{\pm} f(t)$ exists, but $\mathcal{D}^{\alpha}_{\pm} f(t)$ is not defined since constants are not $L^1(\mathbb{R})$ functions.

We have the following result (see [20]).

Theorem 7.1. *Let $0 < \alpha < 1$ and let $\phi \in L^p(\mathbb{R})$ with $1 \leq p < 1/\alpha$. Then,*

$$\mathbf{D}^{\alpha}_{\pm} \mathcal{I}^{\alpha}_{\pm} \phi = \phi. \tag{7.50}$$

Example 7.6. Let $a < b$ and let $0 < \alpha < 1$. Evaluate $\mathbf{D}^{\alpha}_{\pm} 1_{[a,b)}(t)$. Inspired by (eq. 7.45), suppose that:

$$\mathbf{D}^{\alpha}_{\pm} 1_{[a,b)}(t) = \frac{1}{\Gamma(1-\alpha)} \left[(b-t)^{-\alpha}_{\mp} - (a-t)^{-\alpha}_{\mp} \right]. \tag{7.51}$$

Because of the theorem above, it is sufficient to show that:

$$\mathcal{I}^{\alpha}_{\pm} \left[\frac{1}{\Gamma(1-\alpha)} \left[(b-s)^{-\alpha}_{\mp} - (a-s)^{-\alpha}_{\mp} \right] \right] (t) = 1_{[a,b)}(t).$$

Let us check out the case $\mathcal{I}^{\alpha}_{+}$. One has to evaluate:

$$\mathcal{I}(t) = \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)} \int_{\mathbb{R}} (t-s)^{\alpha-1}_{+} \left[(s-b)^{-\alpha}_{+} - (s-a)^{-\alpha}_{+} \right] ds,$$

which is clearly equal to zero if $t < a$. If $a \leq t \leq b$, one has:

$$\mathcal{I}(t) = \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)} \int_{t}^{a} (t-s)^{\alpha-1} (s-a)^{-\alpha} ds,$$

$$= \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)} \int_{0}^{1} z^{\alpha-1} (1-z)^{-\alpha} dz = 1.$$

Finally if $t > b$, one has:

$$\mathcal{I}(t) = \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)} \left[\int_{b}^{t} (t-s)^{\alpha-1} (s-b)^{-\alpha} ds - \int_{a}^{t} (t-s)^{\alpha-1} (s-a)^{-\alpha} ds \right] = 0.$$

Remark 7.17. The definitions of fractional derivatives on the real line can be extended to the case $\alpha \geq 1$, see [20] for details.

It is interesting to study the behavior of fractional integrals with respect to Fourier transformations.

Remark 7.18. Let us indicate with $\mathcal{F}\{\phi(y), x\} = \widehat{\phi}(x)$ the Fourier transform of the function ϕ with respect to y evaluated on $x \in \mathbb{R}$. That is,

$$\widehat{\phi}(x) = \int_{\mathbb{R}} e^{ixy}\phi(y)dy. \tag{7.52}$$

Consider the fractional integral $\mathcal{I}_+^\alpha$, with $0 < \alpha < 1/2$. Then,

$$\mathcal{I}_+^\alpha \phi(t) = \frac{1}{\Gamma(\alpha)} \int_{\mathbb{R}} (t-s)_+^{\alpha-1}\phi(s)ds, \tag{7.53}$$

is well defined (Remark 7.14) for any $\phi \in L^2(\mathbb{R})$. Then, using the Fourier convolution theorem, one has:

$$\widehat{\mathcal{I}_+^\alpha \phi}(x) = \widehat{\phi}(x)\widehat{\psi}_\alpha(x),$$

where $\psi_\alpha(y) = y_+^{\alpha-1}/\Gamma(\alpha)$. Observe that, in distributional sense, one has:

$$\widehat{\psi}_\alpha(x) = |x|^{-\alpha} e^{\,\mathrm{sign}(x)\frac{\alpha\pi i}{2}} = (-ix)^{-\alpha}.$$

In fact, heuristically:

$$\widehat{\psi}(x)_\alpha = \frac{1}{\Gamma(\alpha)} \int_{\mathbb{R}} e^{ixy} y_+^{\alpha-1} dy = \frac{|x|^{-\alpha}}{\Gamma(\alpha)} \left[1_{\{x>0\}} \int_{\mathbb{R}} e^{iz} z_+^{\alpha-1} dz + 1_{\{x<0\}} \int_{\mathbb{R}} e^{-iz} z_+^{\alpha-1} dz \right].$$

One has similar results if using $\mathcal{I}_-^\alpha$. For these reasons, we will write:

$$\widehat{\mathcal{I}_\pm^\alpha \phi}(x) = \widehat{\phi}(x)/(\mp ix)^\alpha. \tag{7.54}$$

7.2 Applications

In this section we present several noteworthy applications of fractional calculus.

7.2.1 Fractional relaxation of single order

The classical phenomenon of relaxation in its simplest form is known to be governed by a first order linear ordinary differential equation, possibly non-homogeneous, that hereafter we recall with the corresponding solution. We denote by $t \geq 0$ the time variable. Moreover, let $u = u(t)$ be the field variable and let $\mathcal{D}_t^1$ indicate the first-order time derivative. Then, the (homogeneous) *relaxation* differential equation reads:

$$\mathcal{D}_t^1 u(t) = -\lambda u(t), \quad t \geq 0, \tag{7.55}$$

where λ is a positive constant denoting the inverse of some characteristic time. The solution of (eq. 7.55), with initial condition $u(0^+) = 1$, is called the *fundamental solution* and reads:

$$u(t) = e^{-\lambda t}, \quad t \geq 0. \tag{7.56}$$

We want to generalize the standard relaxation problem in order to introduce some memory effect in the system. From a fractional calculus point of view, two different formulations are possible. One uses RL fractional derivatives and the other uses CD derivatives (defined in an interval). Then, let $0 < \beta \leq 1$, for any $t \geq 0$, one could write:

$$\mathcal{D}_t^1 u(t) = -\lambda \mathcal{D}_t^{1-\beta} u(t), \quad 0 < \beta \leq 1, \tag{7.57}$$

and

$$ {}_*\mathcal{D}_t^\beta u(t) = -\lambda u(t), \quad 0 < \beta \leq 1, \tag{7.58}$$

Remark 7.19. In the case of the fractional relaxation, the relaxation parameter λ has dimension $[t]^{-\beta}$.

These two formulation are equivalent. In fact, (eq. 7.58) reduces to (eq. 7.57) simply by multiplying on the left by $\mathcal{D}_t^{1-\beta}$.

Remark 7.20. Observe that (eq. 7.55) can be written,

$$u(t) = u_0 - \lambda \int_0^t u(s)ds, \quad t \geq 0. \tag{7.59}$$

Then, the fractional relaxation is obtained just by inserting a power-like memory kernel, that is considering:

$$u(t) = u_0 - \lambda \mathcal{I}_t^{\beta} u(t), \quad 0 < \beta < 1. \tag{7.60}$$

This formulation is equivalent to the previous two. In fact, the RL fractional relaxation is obtained just differentiating both sides of (eq. 7.60). Whereas, (eq, 7.60) is obtained by (eq. 7.58) integrating both sides by order β (see eq. 7.32).

In order to study the fundamental solution, we evaluate the Laplace transform of equation (7.58) with the initial condition $u(0_+) = 1$. One finds (see eq. 7.39):

$$\widetilde{u}(s) = \frac{s^{\beta-1}}{s^{\beta} + \lambda}. \tag{7.61}$$

7.2.2 The Mittag-Leffler function

As quite common in fractional calculus, the so called *Mittag-Leffler* function plays a fundamental role [3, **?**, 10, 19]. It is defined by the following series representation, which is valid in the whole complex plane:

$$E_{\alpha}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + 1)}, \quad \alpha > 0, \quad z \in \mathbb{C}. \tag{7.62}$$

$E_{\alpha}(z)$ is an *entire function* of order $\rho = 1/\alpha$ of the first type.

Remark 7.21. The Mittag-Leffler function provides a simple generalization of the exponential function, which is indeed obtained when $\alpha = 1$.

Example 7.7. Other than the elementary exponential case $\alpha = 1$, one has the following noteworthy cases: for any $x \in \mathbb{R}$,

$$\begin{cases} E_2(x^2) = \cosh x, \\ E_2(-x^2) = \cos x, \\ E(1/2)(\pm|x|^{1/2}) = e^{|x|}\left[1 + \mathrm{erf}(\pm|x|^{1/2})\right] = e^{|}x|\mathrm{erfc}(\mp|x|^{1/2}), \end{cases} \tag{7.63}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy, \quad \mathrm{erfc}(x) = 1 - \mathrm{erf}(x). \tag{7.64}$$

A further generalization can be obtained by the introduction of a second complex valued parameter β:

$$E_{\alpha,\beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)}, \quad \alpha > 0, \quad \beta \in \mathbb{C}, \quad z \in \mathbb{C}. \tag{7.65}$$

Example 7.8. One can find that, for any $x > 0$:

$$\begin{cases} E_{1,2}(x) = \dfrac{e^x - 1}{x}, \\ E_{2,2}(x) = \dfrac{\sinh\sqrt{x}}{\sqrt{x}}. \end{cases} \tag{7.66}$$

The following results hold:

Prop. 7.2. *Let $E_{\alpha,\beta}(z)$ be the generalized Mittag-Leffler function (eq. 7.65). Then, one has the following useful relations:*

- *Let $z \in \mathbb{C}$ then,*
$$E_{\alpha,\beta}(z) = \frac{1}{\Gamma(\beta)} + zE_{\alpha,\alpha+\beta}(z), \tag{7.67}$$
- *Moreover,*
$$E_{\alpha,\beta}(z) = \beta E_{\alpha,\beta+1}(z) + \alpha z \frac{d}{dz} E_{\alpha,\beta+1}(z). \tag{7.68}$$
- *For any $x, q > 0$ and $0 < \alpha < 1$,*
$$qx^{\alpha-1}E_{\alpha,\alpha}(-qx^\alpha) = -\frac{d}{dx}E_\alpha(-qx^\alpha). \tag{7.69}$$
- *Finally, for any $x > 0$ and $p \in \mathbb{N}$*
$$\frac{d^p}{dx^p}\left[x^{\beta-1}E_{\alpha,\beta}(x^\alpha)\right] = x^{\beta-p-1}E_{\alpha,\beta-p}(x^\alpha). \tag{7.70}$$

The Mittag-Leffler functions are connected with Laplace integrals by

$$\int_0^\infty e^{-u}E_\alpha(u^\alpha z)du = \frac{1}{1-z} = \int_0^\infty e^{-u}u^{\beta-1}E_{\alpha,\beta}(u^\alpha z)du, \quad \alpha, \beta > 0, \tag{7.71}$$

which can be found easily by inserting the series representation (eq. 7.62) and evaluating explicitly the integral. Then, by using the above equation, one can evaluate Laplace transforms involving the Mittag-Leffler function. For instance, we have:

$$\mathcal{L}\{E_\alpha(-\lambda t^\alpha), s\} = \frac{s^{\alpha-1}}{s^\alpha + \lambda}, \quad s, \lambda > 0. \tag{7.72}$$

Moreover,

$$\mathcal{L}\{t^{\beta-1}E_{\alpha,\beta}(-\lambda t^\alpha), s\} = \frac{s^{\alpha-\beta}}{s^\alpha + \lambda}, \quad s, \lambda > 0. \tag{7.73}$$

Remark 7.22. Using Tauberian theorems and (eq. 7.72), one can easily obtain the asymptotic behavior of $E_\alpha(-\lambda t^\alpha)$ for $t \gg 1$:

$$E_\alpha(-\lambda t^\alpha) \sim \frac{t^{-\alpha}}{\lambda\Gamma(1-\alpha)}, \quad t \to \infty. \tag{7.74}$$

Prop. 7.3. *Looking at (eq. 7.61) and (eq. 7.72), it is evident that the fundamental solution of equation (7.58) is:*

$$u(t) = E_\beta(-\lambda t^\beta), \quad t \geq 0, \quad 0 < \beta \leq 1. \tag{7.75}$$

Remark 7.23. We remark that such $u(t)$ is the fundamental solution of equation (7.57) and equation (7.60) as well.

In Figure 7.3 is shown the behavior of the fundamental solution (eq. 7.75) for different values of the parameter: $\beta_1 = [1/4, 1/2, 3/4, 1]$, where we set $\lambda = 1$. In the top figure, the time interval is $[0, 10]$ (linear scales), whereas in the bottom plate , the time interval is $[10^1, 10^7]$ (logarithmic scales). Moreover, in the bottom figure, we presented the asymptotic values for large times (dotted lines) in order to visualize the power-law decay expressed by $t^{-\beta_1}/\Gamma(1-\beta_1)$ for the cases $0 < \beta_1 < 1$.

In both the plates we have shown (dashed line) the singular solution for the limiting case $\beta_1 = 0$, stretching the definition of the Mittag-Leffler function to $E_0(z) = 1/(1-z)$, the geometric series,

$$u(t) = \begin{cases} E_0(0) = 1, & t = 0, \\ E_0(-t^0) = E_0(-1) = 1/2, & t > 0. \end{cases}$$

Remark 7.24. The Mittag-Leffler function is actually a generalization of the exponential function. Now, we see in which way this function is associated to fractional calculus, namely to generalizations of standard differential equations to fractional differential equation.

7.2.3 Fractional relaxation of distributed order

The simple fractional relaxation equations (eq. 7.57), (eq. 7.58) and (eq. 7.60) can be generalized by using the notion *fractional derivative of distributed order*. Then, we consider the so called *fractional relaxation equation of distributed order* in the two alternative forms involving the RL and the CD derivatives. Namely, we write:

$$\mathcal{D}_t^1 u(t) = -\lambda \int_0^1 p(\beta) \mathcal{D}_t^{1-\beta} u(t) d\beta \tag{7.76}$$

and

$$\int_0^1 p(\beta) \, {}_*\mathcal{D}_t^{\beta} v(t) \, d\beta = -\lambda v(t), \tag{7.77}$$

with the initial condition $u(0^+) = v(0^+) = 1$. We suppose that,

$$p(\beta) \geq 0, \quad \int_0^1 p(\beta) d\beta = c > 0. \tag{7.78}$$

We set $c = 1$ in order to have the integral normalized. Moreover, we conveniently require that:

- one has
$$\lim_{\beta \to 0} P(\beta) = \lim_{\beta \to 0} \int_0^{\beta} p(\beta') d\beta' = 0. \tag{7.79}$$
- $P(\beta)$ is right-continuous in the origin.
- $P(\beta)$ has at most finitely many (upwards) jump points in the half-open interval $0 < \beta \leq 1$.

Remark 7.25. These jump points allow delta contributions in $p(\beta)$, which are relevant for discrete order distributions.

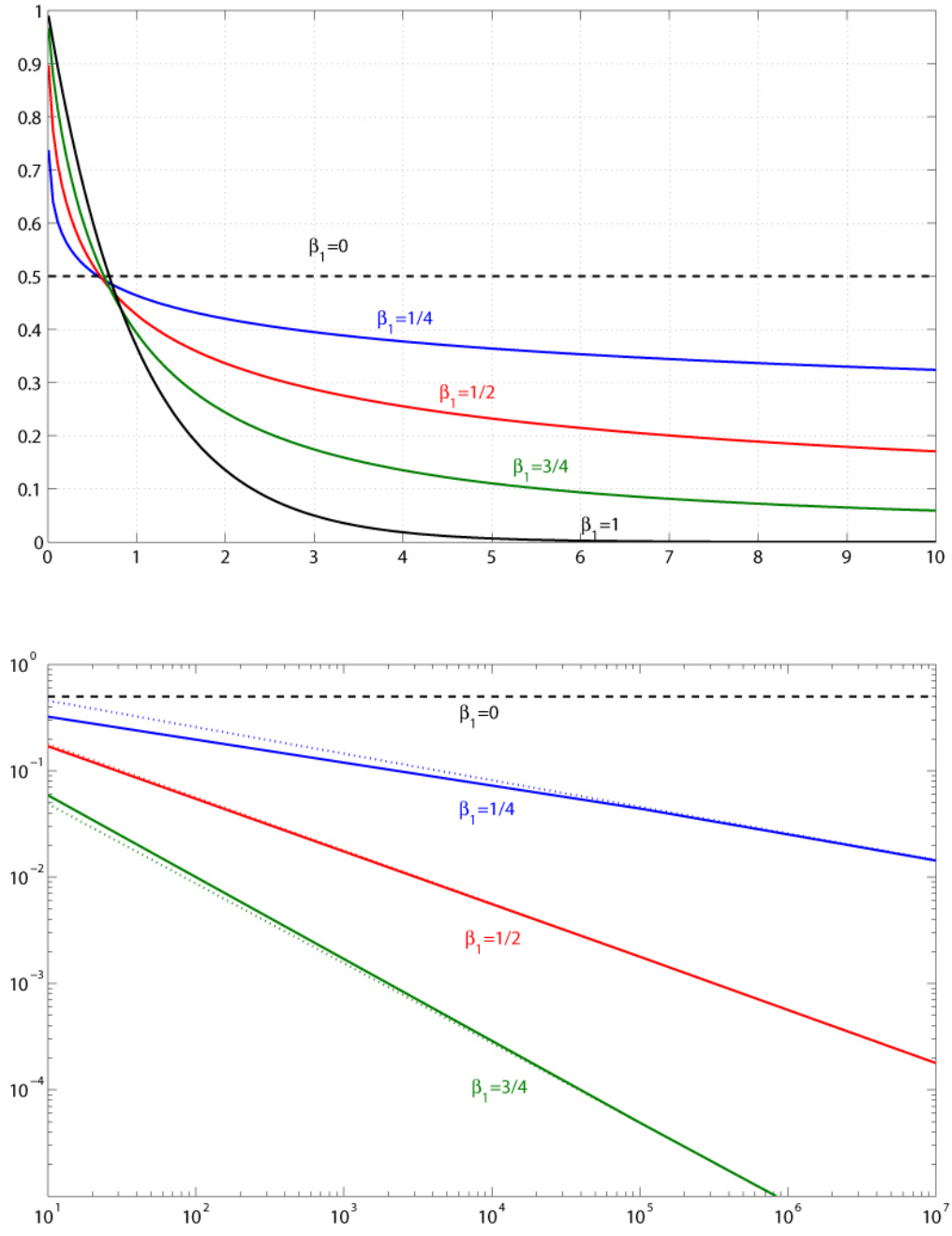

Figure 7.3: Fundamental solutions of the fractional relaxation of order $\beta_1 = [1/4, 1/2, 3/4, 1]$. Linear scale (top figure) and in logarithmic scale (bottom figure).

In the case of distributed order fractional derivatives, there is no reason such that the two formulations provide the same solutions. In fact, the RL formulation (eq. 7.76) turns out to be equivalent to the equation

$$u(t) = u_0 - \lambda \int_0^1 \mathcal{I}_t^{\beta} u(t) d\beta, \tag{7.80}$$

which however cannot be anymore obtained from (eq. 7.77). For these reasons, from now on, we study the two fractional equations (7.76) and (7.77) distinguishing between their fundamental solutions.

The present analysis is based on the application of Laplace transformation. We focus on some special cases depending on the choice of the order density $p(\beta)$, $0 < \beta < 1$, providing many different plots related to the corresponding fundamental solutions.

Let us introduce the auxiliary functions:

$$A(s) = s \int_0^1 p(\beta)\, s^{-\beta}\, d\beta,$$

and

$$B(s) = \int_0^1 p(\beta)\, s^{\beta}\, d\beta.$$

Then, we apply Laplace transform to (eq. 7.76) and (eq. 7.77). By using the rules (7.38 and 7.39), one easily finds the Laplace transform of the corresponding fundamental solutions:

$$\widetilde{u}(s) = \frac{1}{s + \lambda A(s)}, \tag{7.81}$$

and

$$\widetilde{v}(s) = \frac{B(s) s^{-1}}{\lambda + B(s)}. \tag{7.82}$$

Remark 7.26. We note that in the case $p(\beta) = \delta(\beta - \beta_1)$ we have: $A(s) = s^{1-\beta_1}$, $B(s) = s^{\beta_1}$. Then, Eqs. (7.76) and (7.77) provide the same result (eq. 7.61) of the single order fractional relaxation.

By inverting the Laplace transforms in (eq. 7.81) and (eq. 7.82), one obtains the fundamental solutions of the RL and CD fractional relaxation of distributed order.

Let us start with the RL case. In virtue of Laplace inversion Titchmarsh's theorem, we get:

$$u(t) = -\frac{1}{\pi} \int_0^{\infty} e^{-rt} \mathrm{Im} \left\{ \widetilde{u}\left(r e^{i\pi} \right) \right\} dr,$$

which requires the evaluation of $-\mathrm{Im}\left\{1/[s + \lambda A(s)]\right\}$ along the ray $s = r\, e^{i\pi}$, $r > 0$, (the branch cut of the function $s^{-\beta}$). We write

$$A\left(r\, e^{i\pi} \right) = \rho\, \cos(\pi\gamma) + i\rho \sin(\pi\gamma),$$

where

$$\begin{cases} \rho = \rho(r) = |A\left(r e^{i\pi} \right)|, \\ \gamma = \gamma(r) = \dfrac{1}{\pi} \arg\left[A\left(r e^{i\pi} \right) \right] \end{cases}$$

Then, after simple calculations, we get:

$$u(t) = \int_0^{\infty} e^{-rt} H(r; \lambda) dr, \tag{7.83}$$

with

$$H(r;\lambda) = \frac{1}{\pi}\frac{\lambda\,\rho\,\sin(\pi\gamma)}{r^2 - 2\lambda\, r\,\rho\,\cos(\pi\gamma) + \lambda^2\rho^2} \geq 0. \tag{7.84}$$

In the same way, in the CD case, we obtain:

$$v(t) = -\frac{1}{\pi}\int_0^\infty e^{-rt}\mathrm{Im}\left\{\widetilde{v}\left(re^{i\pi}\right)\right\} dr,$$

which requires the evaluation of $-\mathrm{Im}\left\{B(s)/[s(\lambda + B(s))]\right\}$ along the ray $s = re^{i\pi}$, $r > 0$ (the branch cut of the function s^β). We write

$$B\left(re^{i\pi}\right) = \rho_* \cos(\pi\gamma_*) + i\rho_* \sin(\pi\gamma_*),$$

where

$$\begin{cases} \rho_* = \rho_*(r) = |B\left(re^{i\pi}\right)| \\ \gamma_* = \gamma_*(r) = \dfrac{1}{\pi}\arg\left[B\left(re^{i\pi}\right)\right]. \end{cases}$$

After simple calculations, we get

$$v(t) = \int_0^\infty e^{-rt}K(r;\lambda)d, \tag{7.85}$$

where now,

$$K(r;\lambda) = \frac{1}{\pi r}\frac{\lambda\rho_*\sin(\pi\gamma_*)}{\lambda^2 + 2\lambda\,\rho_*\,\cos(\pi\gamma_*) + \rho_*^2} \geq 0. \tag{7.86}$$

Remark 7.27. We note that, since $H(r;\lambda)$ and $K(r;\lambda)$ are non-negative functions of r for any $\lambda \in \mathbb{R}^+$, then the fundamental solutions $u(t)$ and $v(t)$ are *completely monotone*[3]. Therefore, the fundamental solutions of the fractional relaxation of distributed order, in both the cases, maintain the fundamental property of being complete monotone, as in the case of the standard relaxation. However, the basic solutions exhibit a quite different behavior with a slower decay to zero for large times.

The integral expressions (7.83) and (7.85) provide a sort of "spectral representation" of the fundamental solutions, which can be actually used in order to evaluate them numerically.

Furthermore, it is quite instructive to compute the asymptotic expressions of $u(t)$ and $v(t)$ as $t \to 0$ and $t \to \infty$. In fact, they provide the analytical (even if approximated) representations for sufficiently short and long time respectively. Moreover, they are useful in order to check the numerical evaluation. In order to derive these asymptotic representations, we shall apply Tauberian theory of Laplace transforms. According to this, the asymptotic behavior of a function $f(t)$ as $t \to \infty$ and $t \to 0$ is (formally) obtained from the asymptotic behavior of its Laplace transform $\widetilde{f}(s)$ for $s \to 0^+$ and for $s \to +\infty$, respectively. For this purpose, we note the asymptotic representations, from (7.81):

$$\widetilde{u}(s) \sim \begin{cases} \dfrac{1}{\lambda\, A(s)}, & s \to 0^+ \ \text{ being } \ A(s)/s \gg \lambda \\ \dfrac{1}{s}\left[1 - \lambda\dfrac{A(s)}{s}\right], & s \to +\infty \ \text{ being } \ A(s)/s \ll 1/\lambda, \end{cases} \tag{7.87}$$

3) A non-negative function $\varphi(t)$, possessing derivative of any order and such that $(-1)^k\varphi^{(k)}(t) \geq 0$, $t > 0$, for any $k \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$.

and from (7.82):

$$\widetilde{v}(s) \sim \begin{cases} \dfrac{1}{\lambda}\dfrac{B(s)}{s}, & s \to 0^+ \ \text{ being } \ B(s) \ll \lambda \\ \dfrac{1}{s}\left[1 - \dfrac{\lambda}{B(s)}\right], & s \to +\infty \ \text{ being } \ B(s) \gg \lambda. \end{cases} \tag{7.88}$$

Finding explicit solution formulas is not possible for the relaxation equations (7.76) and (7.77), then we shall concentrate to some typical choices of $p(\beta)$. For these choices we present the numerical evaluation of the Titchmarsh integral formulas (eq. 7.83 and 7.85).

The numerical results have been checked by verifying the asymptotic matching with the asymptotic expression of $u(t)$ and $v(t)$, obtained through Tauberian theorem.

Example 7.9 (Double-order fractional relaxation)**.** Let us consider:

$$p(\beta) = p_1\delta(\beta - \beta_1) + p_2\delta(\beta - \beta_2)\,, \quad 0 < \beta_1 < \beta_2 \leq 1\,, \tag{7.89}$$

where the constants p_1 and p_2 are both positive, such that $p_1 + p_2 = 1$. Then, for the RL case we have:

$$A(s) = p_1\, s^{1-\beta_1} + p_2\, s^{1-\beta_2}. \tag{7.90}$$

So that, (eq. 7.81) becomes:

$$\widetilde{u}(s) = \frac{1}{s[1 + \lambda(p_1\, s^{-\beta_1} + p_2\, s^{-\beta_2})]} \tag{7.91}$$

Similarly, for the CD case, we have

$$B(s) = p_1\, s^{\beta_1} + p_2\, s^{\beta_2}, \tag{7.92}$$

and (eq. 7.81) turns out to be:

$$\widetilde{v}(s) = \frac{p_1\, s^{\beta_1} + p_2\, s^{\beta_2}}{s[\lambda + p_1\, s^{\beta_1} + p_2\, s^{\beta_2}]}. \tag{7.93}$$

The asymptotic expressions can be easily evaluated. In the RL case, if $\beta_2 < 1$,

$$\widetilde{u}(s) \sim \begin{cases} \dfrac{1}{\lambda p_2} s^{\beta_2 - 1}, \ \ s \to 0_+, \\ \dfrac{1}{s}\left(1 - \lambda p_1 s^{-\beta_1}\right), \ \ s \to \infty. \end{cases} \tag{7.94}$$

Therefore,

$$u(t) \sim \begin{cases} \dfrac{1}{\lambda p_2}\dfrac{t^{-\beta_2}}{\Gamma(1-\beta_2)}, \ \ t \to \infty, \\ 1 - \lambda p_1 \dfrac{t^{\beta_1}}{\Gamma(1+\beta_1)}, \ \ t \to 0_+. \end{cases} \tag{7.95}$$

Observe that (eq. 7.94), and (eq. 7.94) as well, loses its meaning if $\beta_2 = 1$. In this case, we need a different approach. We consider the expression for $s \to 0$ provided by (eq. 7.87) as it stands. Namely,

$$\widetilde{u}(s) \sim \frac{1}{\lambda\,[p_1 s^{1-\beta_1} + p_2]} = \frac{1}{\lambda\, p_1}\,\frac{1}{s^{1-\beta_1} + p_2/(\lambda\, p_1)} \tag{7.96}$$

In virtue of (eq. 7.73)

$$\mathcal{L}\{t^{\nu-1}E_{\mu,\nu}\left(-qt^{\mu}\right), s\} = \frac{s^{\mu-\nu}}{s^{\mu} + q},$$

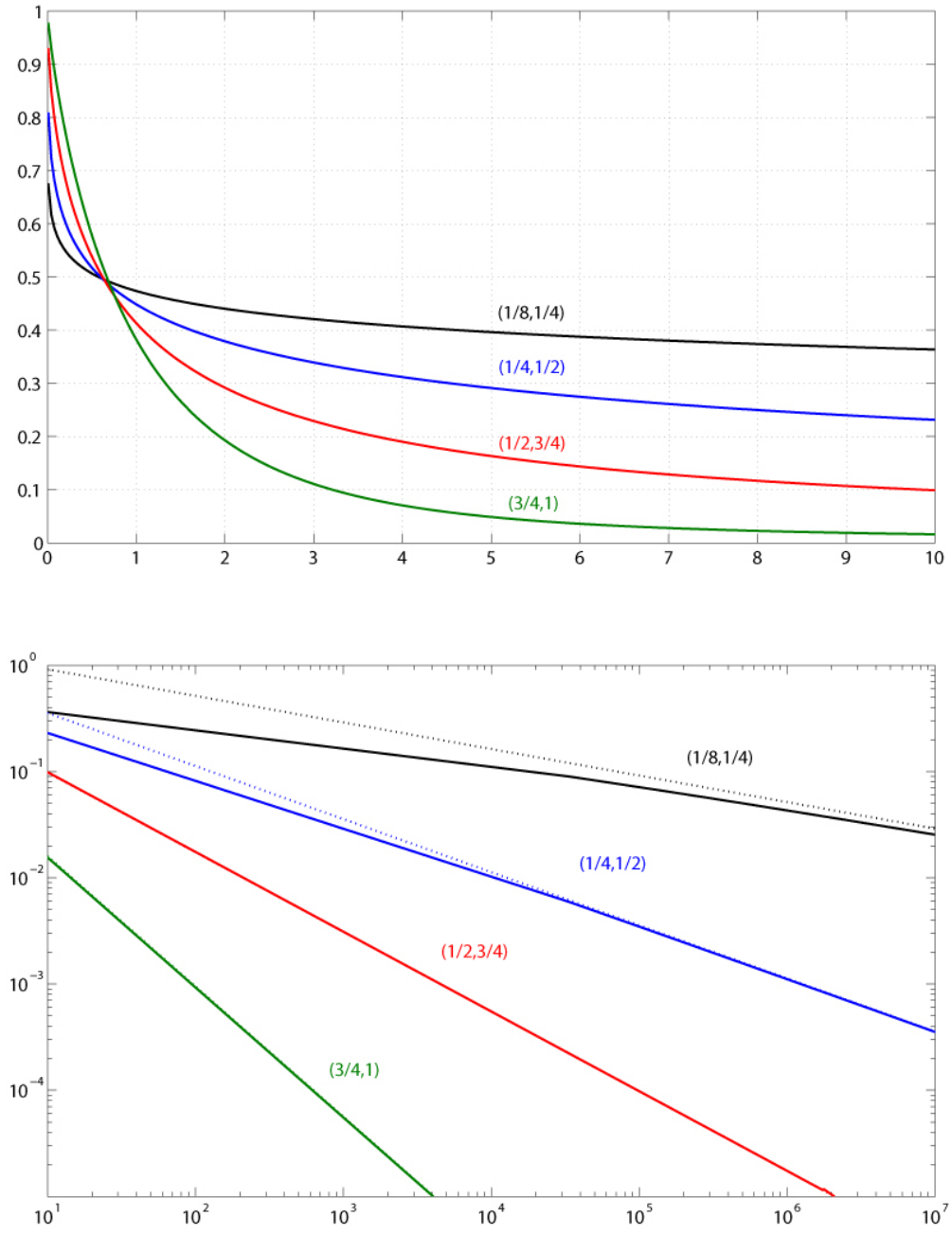

Figure 7.4: Fundamental solutions of the RL-fractional relaxation of double order in some $\{\beta_1, \beta_2\}$ combinations: $\{1/8, 1/4\}$; $\{1/4, 1/2\}$; $\{1/2, 3/4\}$; $\{3/4, 1\}$. Top: linear scales, Bottom: logarithmic scales.

we get for large t, by setting $q = p_2/(\lambda\, p_1)$ and $\mu = \nu = 1 - \beta_1$, and using (eq. 7.69):

$$u(t) \sim \frac{1}{\lambda p_1} t^{-\beta_1} E_{1-\beta_1,1-\beta_1}\left(-qt^{1-\beta_1}\right) = -\frac{1}{\lambda p_1}\frac{d}{dt} E_{1-\beta_1}\left(-qt^{1-\beta_1}\right). \tag{7.97}$$

Moreover, taking into account the asymptotic behavior of the Mittag-Leffler function, we finally get

$$u(t) \sim \lambda\, \frac{p_1}{p_2}\, \frac{1-\beta_1}{\Gamma(\beta_1)}\, t^{-(2-\beta_1)} \tag{7.98}$$

Similarly, in the CD case, one gets:

$$\widetilde{v}(s) \sim \begin{cases} \dfrac{p_1}{\lambda} s^{\beta_1 - 1}, \;\; s \to 0_+, \\ \dfrac{1}{s}\left(1 - \dfrac{\lambda}{p_2} s^{-\beta_2}\right), \;\; s \to \infty. \end{cases} \tag{7.99}$$

So that,

$$v(t) \sim \begin{cases} \dfrac{p_1}{\lambda}\dfrac{t^{-\beta_1}}{\Gamma(1-\beta_1)}, \;\; t \to \infty, \\ 1 - \dfrac{\lambda}{p_2}\dfrac{t^{\beta_2}}{\Gamma(1+\beta_2)}, \;\; t \to 0_+. \end{cases} \tag{7.100}$$

In the figures, we exhibit many fundamental solution plots for both the cases of fractional relaxation and for different choices of the parameter orders $\{\beta_1, \beta_2\}$. We have chosen $p_1 = p_2 = 1/2$ and, as usual $\lambda = 1$. From the plots, the reader is expected to verify the role of the different orders for small and large times, according to the corresponding asymptotic expressions (eq. 7.94, 7.95, 7.99, 7.100).

Example 7.10 (Uniformly distributed fractional relaxation)**.** Consider the uniformly distributed case. Namely,

$$p(\beta) = 1, \;\; 0 < \beta < 1.$$

In the RL case, we have

$$A(s) = s \int_0^1 s^{-\beta} d\beta = \frac{s-1}{\log s}.$$

Hence,

$$\widetilde{u}(s) = \frac{\log s}{s\log s + \lambda\,(s-1)}. \tag{7.101}$$

Also, in the CD case:

$$B(s) = \int_0^1 s^{\beta}\, d\beta = \frac{s-1}{\log s}.$$

Thus,

$$\widetilde{v}(s) = \frac{1}{s}\frac{s-1}{\lambda\log s + s - 1} = \frac{1}{s} - \frac{1}{s}\frac{\lambda \log s}{\lambda \log s + s - 1}. \tag{7.102}$$

Remark 7.28. We note that for this special order distribution, we have $A(s) = B(s)$. However, the corresponding fundamental solutions are quite different, as we see from their Laplace transforms.

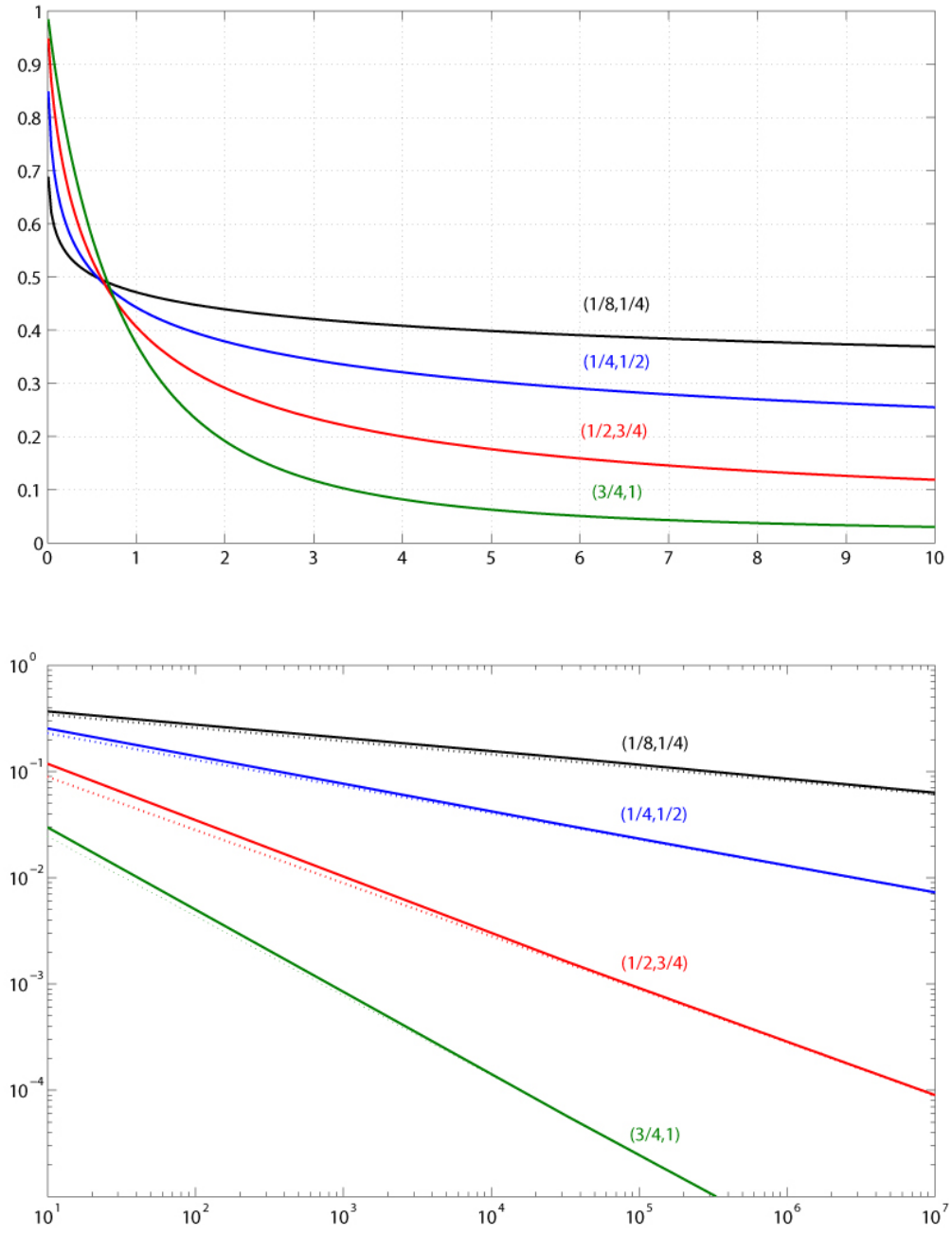

Figure 7.5: Fundamental solutions of the CD-fractional relaxation of double order in some $\{\beta_1, \beta_2\}$ combinations: $\{1/8, 1/4\}$; $\{1/4, 1/2\}$; $\{1/2, 3/4\}$; $\{3/4, 1\}$. Top: linear scales. Bottom: logarithmic scales.

Then, invoking the Tauberian theory for *regular variation functions* (see for instance Prop. 9.1), we have the following asymptotic expressions for the RL and CD cases.

$$\widetilde{u}(s) \sim \begin{cases} \dfrac{\log s}{\lambda(s-1)}, & s \to 0_+, \\ \dfrac{1}{s}\left(1 - \lambda\dfrac{s-1}{s\log s}\right), & s \to \infty. \end{cases} \tag{7.103}$$

Then, we find:

$$u(t) \sim \begin{cases} \dfrac{1}{\lambda}e^t\mathcal{E}_1(t) \sim \dfrac{1}{\lambda t}, & t \to \infty, \\ 1 - \dfrac{\lambda}{\log(1/t)}, & t \to 0_+, \end{cases} \tag{7.104}$$

where

$$\mathcal{E}_1(t) = \int_t^\infty \frac{e^{-u}}{u}du \tag{7.105}$$

denotes the *exponential integral* [1]. In the CD case, one finds:

$$\widetilde{v}(s) \sim \begin{cases} \dfrac{1}{\lambda s\log(1/s)}, & s \to 0_+, \\ \dfrac{1}{s} - \dfrac{\lambda\log s}{s^2}, & s \to \infty. \end{cases} \tag{7.106}$$

Hence, we find:

$$v(t) \sim \begin{cases} \dfrac{1}{\lambda\log t}, & t \to \infty, \\ 1 + \lambda t\log t, & t \to 0_+. \end{cases} \tag{7.107}$$

In Figure (7.6) are shown the plots of the fundamental solutions for both the cases of uniformly distributed fractional relaxation, adopting as previously, in the top plate, linear scales ($0 \le t \le 10$), and in the bottom plate, logarithmic scales ($10^1 \le t \le 10^7$).

For comparison, in the top plate the plots for single orders $\beta_1 = [0, 1/2, 1]$ are displayed. We note that, for $1 < t < 10$, the RL and CD plots are close to those corresponding to $\beta_1 = 1/2$ from above and from below, respectively. In the bottom plate (where the β_1 plot is not visible because of its faster exponential decay) we have added, in dotted lines, the asymptotic solutions for large times.

Remark 7.29. Observe that the CD plot is decaying much slower than any power law. Whereas, the RL plot is decaying like t^{-1}. This means that for large times these plots are the border lines for all the plots corresponding to single order relaxation with $\beta_1 \in (0, 1)$.

7.2.4 Fractional drift equation

Consider the non-random process $l(t) = t$, $t \ge 0$, which depicts a non-random linear time evolution and let $f_l(\tau, t)$ denote its density function at time t. Therefore one has $f_l(\tau, t) = \delta(\tau - t)$. It is natural to interpret $f_l(\tau, t)$ as the fundamental solution[4] of the standard forward drift equation:

$$\partial_t u(\tau, t) = -\partial_\tau u(\tau, t), \quad \tau, t \ge 0, \tag{7.108}$$

4) Here *fundamental solution* means a solution $u(\tau, t)$ with "deterministic" initial value, namely $u(\tau, 0) = \delta(\tau)$.

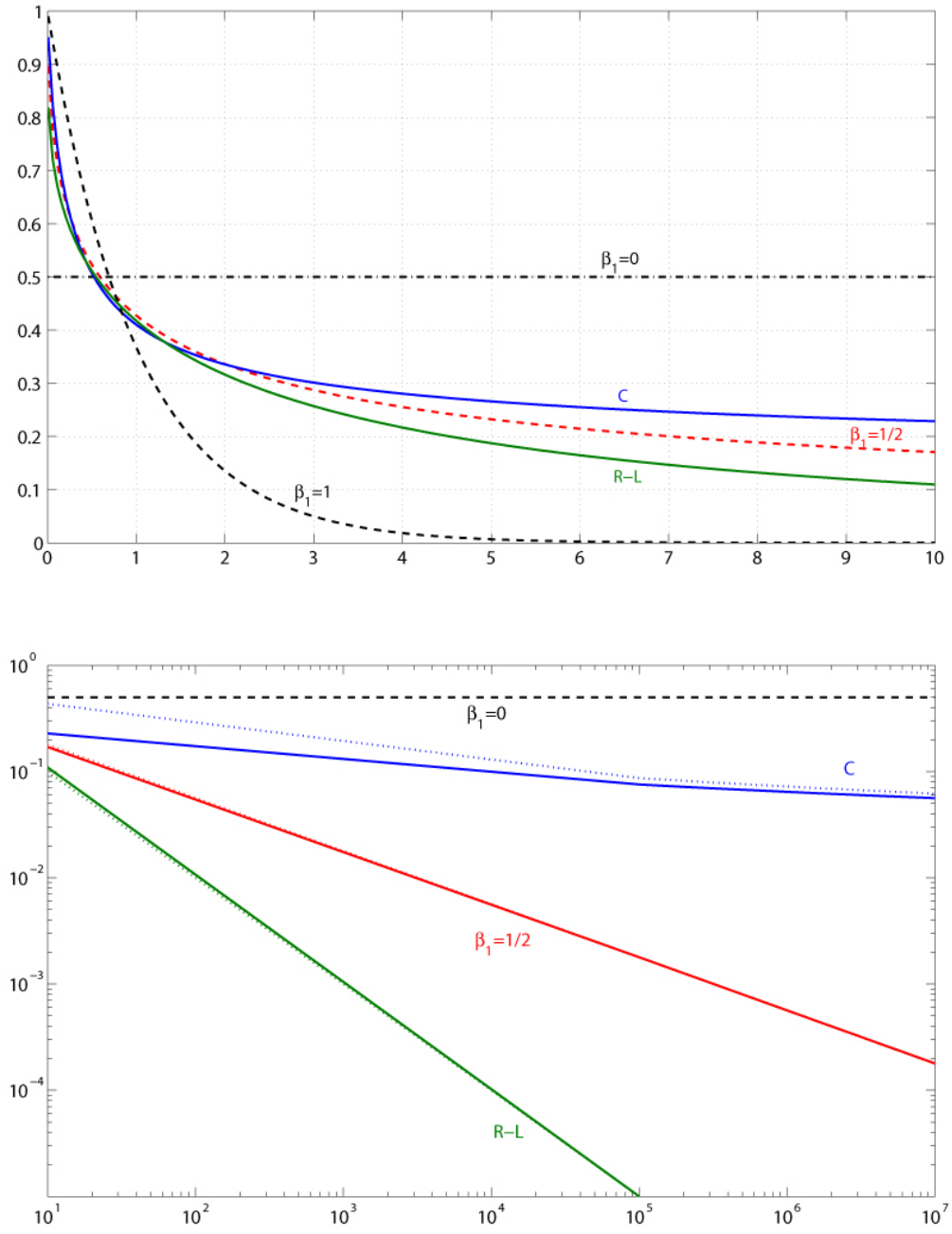

Figure 7.6: Fundamental solutions for the RL and CD uniformly distributed fractional relaxation compared with a single order case. Top: linear scales. Bottom: logarithmic scales.

which in integral form reads:

$$u(\tau,t) = u_0(\tau) - \int_0^t \partial_\tau u(\tau,s)ds, \quad u_0(\tau) = u(\tau,0). \tag{7.109}$$

The general solutions are waves of the form $u(\tau,t) = u_0(\tau - t)$ and thus, when $u_0(\tau) = \delta(\tau)$, the solution of (eq. 7.108) is indeed $u(\tau,t) = \delta(\tau - t)$. Observe that the variable $\tau \geq 0$ plays the role of a space variable.

If we introduce a power-like memory kernel, we obtain the *fractional Forward drift equation* of order $0 < \beta \leq 1$, namely:

$$u(\tau,t) = u_0(\tau) - \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} \partial_\tau u(\tau,s)ds, \quad \tau, t \geq 0, \tag{7.110}$$

Remark 7.30. We observe that (eq. 7.110) can be written:

$$u(\tau,t) = u_0(\tau) - \mathcal{I}_t^\beta \partial_\tau u(\tau,t), \tag{7.111}$$

which differentiating both sides with respect to t gives:

$$\frac{\partial}{\partial t} u(\tau,t) = -\mathcal{D}_t^{1-\beta} \frac{\partial}{\partial \tau} u(\tau,t). \tag{7.112}$$

Then, the latter provides an equivalent formulation of (eq. 7.110) provided that $u(\tau,0) = u_0(\tau)$. Another equivalent formulation is given by:

$${}_*\mathcal{D}_t^\beta u(\tau,t) = -\frac{\partial}{\partial \tau} u(\tau,t), \tag{7.113}$$

which indeed reduces to (eq. 7.110) integrating both sides by order β.

We indicate with $h(\tau,t)$ its fundamental solution. It is easy to show that:

$$\mathcal{L}\{h(\tau,t);t,s\} = s^{\beta-1}\exp\Big(-\tau s^\beta\Big), \quad \tau, s \geq 0, \tag{7.114}$$

and zero for $\tau < 0$. Indeed, taking the Laplace transform with respect to the variable t in (eq. 7.110) one finds (see eq. 7.37):

$$\partial_\tau \widetilde{u}(\tau,s) = u_0(\tau)s^{\beta-1} - \widetilde{u}(\tau,s)s^\beta \tag{7.115}$$

and (eq. 7.114) is a solution, in the distributional sense, when $u_0(\tau) = \delta(\tau)$.

Remark 7.31. Actually the general solution of (eq. 9.27) with $u(\tau) = \delta(\tau)$ is:

$$\varphi(\tau,s) = \theta(\tau)s^{\beta-1}\exp\Big(-\tau s^\beta\Big) + C\exp\Big(-\tau s^\beta\Big), \quad \tau \in \mathbb{R},$$

where C is a real constant and where

$$\theta(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0 \end{cases} \tag{7.116}$$

is the Heaviside's step function. Since we require $\varphi(\tau,t) = 0$ for $\tau < 0$, we get $C = 0$ i.e. (eq. 7.114).

We circumvent the direct inversion of the Laplace transform in (eq. 7.114) by passing through the intermediate step of the Mellin transform. Then, we indicate with

$$\mathcal{M}\{\varphi(x);x,u\} = \int_0^\infty \varphi(x)x^{u-1}dx, \tag{7.117}$$

the Mellin transform of a function $\varphi(x)$, $x \geq 0$, with respect to x evaluated in $u \geq 0$. Therefore,

$$\mathcal{M}\{h(\tau,t);t,u\} = \frac{1}{\Gamma(1-u)}\int_0^\infty ds s^{-u}\mathcal{L}\{h(\tau,t);t,s\}$$

$$= \frac{1}{\Gamma(1-u)}\int_0^\infty s^{\beta-1-u}e^{-\tau s^\beta}ds$$

(with the change of variables $\tau s^\beta = z$)

$$= \frac{\tau^{\beta/u-1}}{\beta\Gamma(1-u)}\int_0^\infty z^{-u/\beta}e^{-z}dz = \frac{\tau^{\beta/u-1}}{\beta}\frac{\Gamma(1-u/\beta)}{\Gamma(1-u)}.$$

Inverting the Mellin transform we get:

$$h(\tau,t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{\tau^{u/\beta-1}}{\beta}\frac{\Gamma(1-u/\beta)}{\Gamma(1-u)}t^{-u}du$$

where $0 < \gamma < 1$. This is the same as:

$$\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\tau^{z-1}\frac{\Gamma(1-z)}{\Gamma(1-\beta z)}t^{-\beta z}dz = \frac{1}{\tau}\int_{\gamma-i\infty}^{\gamma+i\infty}\left(\tau t^{-\beta}\right)^z\frac{\Gamma(1-z)}{\Gamma(1-\beta z)}dz = \frac{1}{\tau}H_{1,1}^{1,0}\left(\tau t^{-\beta}\middle|\begin{array}{c}(1,\beta)\\(1,1)\end{array}\right).$$

Here, the Fox H-function

$$H_{p,q}^{m,n}(z) = H_{p,q}^{m,n}\left(z\middle|\begin{array}{c}(a_i,\alpha_i)_{i=1,\ldots,p}\\(b_j,\beta_j)_{j=1,\ldots,q}\end{array}\right),$$

is characterized by its Mellin transform:

$$\mathcal{M}\{H_{p,q}^{m,n}(z);z,u\} = \frac{A(u)B(u)}{C(u)D(u)}, \tag{7.118}$$

with

$$A(u) = \prod_{i=1}^{m}\Gamma(b_j+\beta_j u), \qquad B(u) = \prod_{j=1}^{n}\Gamma(1-a_j-\alpha_j u),$$
$$C(u) = \prod_{i=m+1}^{q}\Gamma(1-b_j-\beta_j u), \quad D(u) = \prod_{j=n+1}^{p}\Gamma(a_j+\alpha_j u),$$

where: $1 \leq m \leq q$, $0 \leq n \leq p$, $\alpha_j, \beta_j > 0$ and $a_j, b_j \in \mathbb{C}$ (see [21,13] for more details). The Fox H-function satisfies the property:

$$z^\sigma H_{p,q}^{m,n}\left(z\middle|\begin{array}{c}(a_i,\alpha_i)_{i=1,\ldots,p}\\(b_j,\beta_j)_{j=1,\ldots,q}\end{array}\right) = H_{p,q}^{m,n}\left(z\middle|\begin{array}{c}(a_i+\sigma\alpha_i,\alpha_i)_{i=1,\ldots,p}\\(b_j+\sigma\beta_j,\beta_j)_{j=1,\ldots,q}\end{array}\right). \tag{7.119}$$

We have:

$$h(\tau,t) = \frac{1}{\tau}H_{1,1}^{1,0}\left(\tau t^{-\beta}\middle|\begin{array}{c}(1,\beta)\\(1,1)\end{array}\right) = t^{-\beta}(\tau t^{-\beta})^{-1}H_{1,1}^{1,0}\left(\tau t^{-\beta}\middle|\begin{array}{c}(1,\beta)\\(1,1)\end{array}\right) \tag{7.120}$$

Using (eq. 7.119) we find

$$h(\tau,t) = t^{-\beta} H_{1,1}^{1,0}\left(\tau t^{-\beta}\Big| \begin{matrix}(1-\beta,\beta)\\(0,1)\end{matrix}\right) := t^{-\beta} M_\beta(\tau t^{-\beta}) := \mathcal{M}_\beta(\tau,t). \tag{7.121}$$

It is possible to show that the function $M_\beta(r)$ is defined for $0<\beta<1$ by the power series [11]:

$$\begin{aligned} M_\beta(r) &= \sum_{k=0}^{\infty} \frac{(-r)^k}{k!\Gamma\left[-\beta k+(1-\beta)\right]} \\ &= \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(-r)^k}{k!}\Gamma\left[(\beta(k+1))\right]\sin\left[\pi\beta(k+1)\right], \quad r\geq 0. \end{aligned} \tag{7.122}$$

The above series defines a transcendental function (entire of order $1/(1-\beta)$) and satisfies the following properties:

1. $\mathcal{M}_\beta(\tau,t)$, $0<\beta\leq 1$, defines a probability density function in $\tau\geq 0$ for any $t\geq 0$ (see Remark 9.4).
2. The Laplace transform of $\mathcal{M}_\beta(\tau,t)$ with respect to t is:
$$\mathcal{L}\{\mathcal{M}_\beta(\tau,t);t,s\} = s^{\beta-1}e^{-\tau s^\beta}, \quad \tau,s\geq 0. \tag{7.123}$$
3. While, the Laplace transform with respect to τ is:
$$\mathcal{L}\{\mathcal{M}_\beta(\tau,t);\tau,u\} = E_\beta(-ut^\beta), \quad t,u\geq 0, \tag{7.124}$$
where $E_\beta(x)$ is the Mittag-Leffler function defined in (eq. 7.62).
4. Equation (7.123) suggests that in the singular limit $\beta\to 1$ one has:
$$\mathcal{M}_1(\tau,t) = \delta(\tau-t), \quad \tau,t\geq 0. \tag{7.125}$$
5. If $\beta=1/2$:
$$\mathcal{M}_{1/2}(\tau,t) = \frac{1}{\sqrt{\pi t}}\exp(-\tau^2/4t), \quad \tau,t\geq 0. \tag{7.126}$$
6. For any $\eta,\beta\in(0,1)$:
$$\mathcal{M}_\nu(x,t) = \int_0^\infty \mathcal{M}_\eta(x,\tau)\mathcal{M}_\beta(\tau,t)d\tau, \quad \nu=\eta\beta \quad x\geq 0. \tag{7.127}$$

The first one will be shown later. Properties 2-5 follow directly from the definition. In order to see (eq. 7.124) one can just take the Laplace transform of (eq. 7.123) with respect to τ

$$\mathcal{L}\{s^{\beta-1}e^{-\tau s^\beta};\tau,u\} = \frac{s^{\beta-1}}{s^\beta+u},$$

and then inverting with respect to s by using (eq. 7.72). We are now interested in showing (eq. 7.127). We have:

$$X(x,t) = \int_0^\infty \mathcal{M}_\eta(x,\tau)\mathcal{M}_\beta(\tau,t)d\tau = \int_0^\infty \tau^{-\eta}M_\eta(x\tau^{-\eta})t^{-\beta}M_\beta(\tau t^{-\beta})d\tau$$

(with the change of variables $y = \tau^\eta$)

$$= \int_0^\infty \frac{1}{y} M_\eta\left(\frac{x}{y}\right) \frac{t^{-\beta} y^{\frac{1}{\eta}-1}}{\eta} M_\beta(y^{\frac{1}{\eta}} t^{-\beta}) dy.$$

Therefore:

$$X(x,t) = (M_\eta \star Y_t)(x) = \int_0^\infty \frac{1}{y} M_\eta(x/y) Y_t(y) dy, \tag{7.128}$$

where $(a \star b)(x)$ indicates the Mellin convolution and

$$Y_t(x) = \frac{t^{-\beta} x^{\frac{1}{\eta}-1}}{\eta} M_\beta(x^{\frac{1}{\eta}} t^{-\beta}). \tag{7.129}$$

Using the Mellin convolution theorem we get:

$$\mathcal{M}\{X(x,t), x, u\} = \mathcal{M}\{M_\eta(x), x, u\} \mathcal{M}\{Y_t(x), x, u\}.$$

Using:

$$M_\eta(x) = H^{1,0}_{1,1}\left(\tau \middle| \begin{matrix} (1-\eta, \eta) \\ (0,1) \end{matrix}\right),$$

and (eq. 7.118), we get:

$$\mathcal{M}\{M_\eta(x), x, u\} = \frac{\Gamma(u)}{\Gamma(1-\eta+\eta u)}.$$

Moreover:

$$\mathcal{M}\{Y_t(x), x, u\} = \int_0^\infty \frac{t^{-\beta} x^{\frac{1}{\eta}-1}}{\eta} M_\beta(x^{\frac{1}{\eta}} t^{-\beta}) x^{u-1} dx$$

(with the change of variables $z = x^{\frac{1}{\eta}} t^{-\beta}$)

$$= t^{\eta\beta(u-1)} \int_0^\infty z^{\eta(u-1)} M_\beta(z) dz = t^{\eta\beta(u-1)} \mathcal{M}\{M_\beta(x); x, 1-\eta+\eta u\}$$

$$= t^{\eta\beta(u-1)} \frac{\Gamma(1-\eta+\eta u)}{\Gamma(1-\beta+\beta(1-\eta+\eta u))} = t^{\eta\beta(u-1)} \frac{\Gamma(1-\eta+\eta u)}{\Gamma(1-\beta\eta+\beta\eta u))}.$$

Finally:

$$\mathcal{M}\{X(x,t), x, u\} = t^{\eta\beta(u-1)} \frac{\Gamma(u)}{\Gamma(1-\beta\eta+\beta\eta u))}.$$

Inverting the Mellin transform and using (eq. 7.118) one finds:

$$X(x,t) = t^{-\eta\beta} H^{1,0}_{1,1}\left(x t^{-\eta\beta} \middle| \begin{matrix} (1-\eta\beta, \eta\beta) \\ (0,1) \end{matrix}\right),$$

that is:

$$X(x,t) = t^{-\eta\beta} M_{\eta\beta}(x t^{-\eta\beta}) = \mathcal{M}_{\eta\beta}(x,t).$$

Remark 7.32. Because $M_\beta(x)$, $x \geq 0$, is a probability density function, there are a number of non-Markovian stochastic processes which have $\mathcal{M}_\beta(\tau, t)$ as marginal density function. We will meet many of them later.

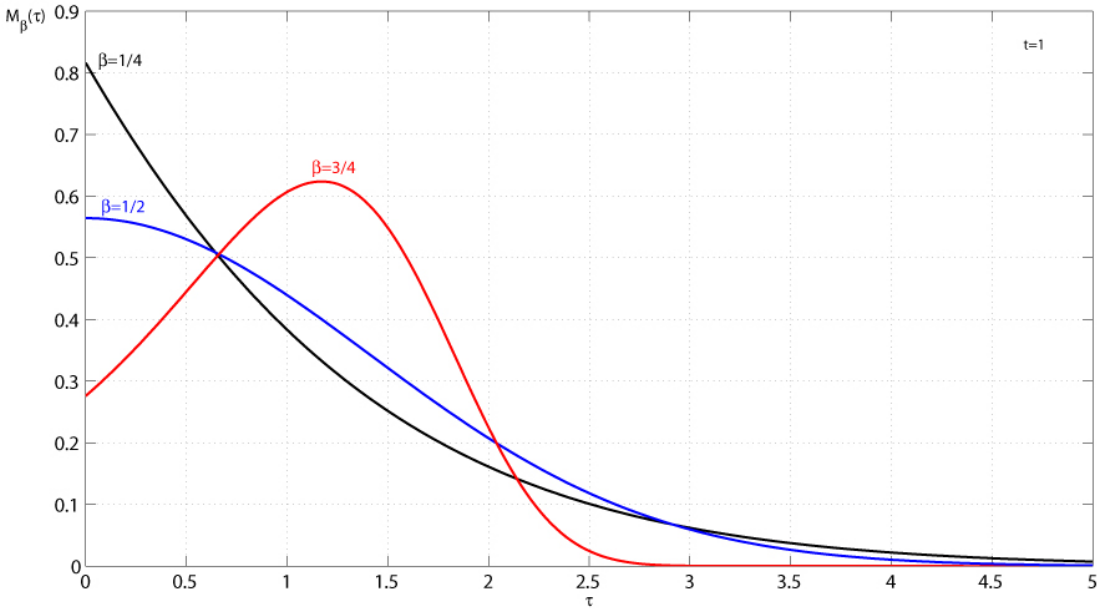

Figure 7.7: Plot of the density function $h(\tau,t) = t^{-\beta}M(\tau t^{-\beta})$ at time $t = 1$, for different values of the parameter $\beta = [1/4, 1/2, 3/4]$.

7.2.5 Time-fractional diffusion of single order

The main physical purpose for adopting and investigating diffusion equations of fractional order, is to describe phenomena of *anomalous diffusion*, which are usually met in transport processes through complex and/or disordered systems, including fractal media. The *standard diffusion equation* in re-scaled non-dimensional variables is known to be

$$\frac{\partial}{\partial t}\,u(x,t) = \frac{\partial^2}{\partial x^2}\,u(x,t)\,, \quad x \in \mathbb{R}\,, \quad t \geq 0\,, \tag{7.130}$$

where $u(x,t)$ is the field variable. We assume that $u(x,t)$ is subjected to the initial condition $u(x,0^+) = u_0(x)$ where $u_0(x)$ denotes a given ordinary or generalized function defined on $\mathbb{R}$, that we assume to be Fourier transformable in ordinary or generalized sense, respectively.

Remark 7.33. We assume to work in a suitable space of generalized functions where it is possible to deal freely with delta functions, integral transforms of Fourier, Laplace and Mellin type, and fractional integrals and derivatives.

The fundamental solution of the standard diffusion equation (eq. 9.1) is the Gaussian density

$$G(x,t) = \frac{1}{2\sqrt{\pi}}\,t^{-1/2}\,\mathrm{e}^{-x^2/(4t)}, \quad t \geq 0, \tag{7.131}$$

with time growing variance $\sigma^2(t) = 2t$ consistently with a law of *normal diffusion*[5].

5) The asymptotic behavior of the variance as $t \to \infty$ is relevant to distinguish between *normal diffusion*, which occurs when $\sigma^2(t)/t \to c$, $c > 0$, and anomalous diffusion, usually divided in *sub-diffusion*, $\sigma^2(t)/t \to 0$ and *super-diffusion*, $\sigma^2(t)/t \to +\infty$.

Remark 7.34. We note the *scaling property* of the fundamental solution,

$$G(x,t) = t^{-1/2}\, U(xt^{-1/2}), \quad \text{with} \quad U(x) := u(x,1). \tag{7.132}$$

The function $U(x)$ depending on the single variable x turns out to be an even function $U(x) = U(|x|)$ and is called the *reduced Green function*. The variable $X := x/t^{1/2}$ acts as the similarity variable.

As usual we rewrite Equation (9.1) in integral form:

$$u(x,t) = u_0(x) + \int_0^t \left[\frac{\partial^2}{\partial x^2}\, u(x,s)\right] ds. \tag{7.133}$$

The above Cauchy problem can be generalized in order to obtain the so-called *time-fractional diffusion equation*. As in the case of fractional relaxation, two distinct (but mathematically equivalent) choices are possible. They can both be obtained by inserting a power-like memory kernel in (eq. 7.133). If one let β denote a real number such that $0 < \beta < 1$, one gets the *time-fractional diffusion equation* of order β.

$$u(x,t) = u_0(x) + \mathcal{I}_t^{\beta}\left[\frac{\partial^2}{\partial x^2}\, u(x,t)\right]. \tag{7.134}$$

Then, the Riemann-Liouville time-fractional diffusion equation is obtained differentiating with respect to t both sides of (eq. 7.134) and taking account of the initial condition:

$$\frac{\partial}{\partial t} u(x,t) = \mathcal{D}_t^{1-\beta}\, \frac{\partial^2}{\partial x^2}\, u(x,t), \quad x \in \mathbb{R},\ t \geq 0,; \quad u(x,0^+) = u_0(x). \tag{7.135}$$

Furthermore, the Caputo-Dzherbashyan time-fractional diffusion can be formally obtained generalizing the time-derivative in (eq. 9.1) with a CD derivative of order β, namely:

$${}_*\mathcal{D}_t^{\beta}\, u(x,t) = \frac{\partial^2}{\partial x^2}\, u(x,t), \quad x \in \mathbb{R},\ t \geq 0,; \quad u(x,0^+) = u_0(x), \tag{7.136}$$

which reduces to (eq. 7.134) integrating both sides through a time-fractional integration of order β. In all the cases, when $\beta = 1$ we recover the standard diffusion equation.

Remark 7.35. We observe that (eq. 7.135) can be put in a *conservative form* as a *continuity equation*. Namely,

$$\frac{\partial}{\partial t}\, u(x,t) + \frac{\partial}{\partial x}\, F\left[u(x,t)\right] = 0, \tag{7.137}$$

where F is the *flux* given by

$$F\left[u(x,t)\right] = -\frac{\partial}{\partial x}\left[\mathcal{D}_t^{1-\beta}\, u(x,t)\right] = -\frac{\partial}{\partial x}\left\{\frac{1}{\Gamma(\beta)}\frac{\partial}{\partial t}\left[\int_0^t (t-s)^{\beta-1} u(x,s)ds\right]\right\}. \tag{7.138}$$

For $\beta = 1$, we recover in the limit the standard *Fick law*

$$F\left[u(x,t)\right] = -\frac{\partial}{\partial x}\, u(x,t),$$

which leads to the standard diffusion equation by using the continuity equation (7.137). We also note that (eq. 7.138) can be interpreted as a generalized Fick law[6] where memory effects are taken into account througha time-fractional derivative of order $1-\beta$ in the *Riemann-Liouville* sense.

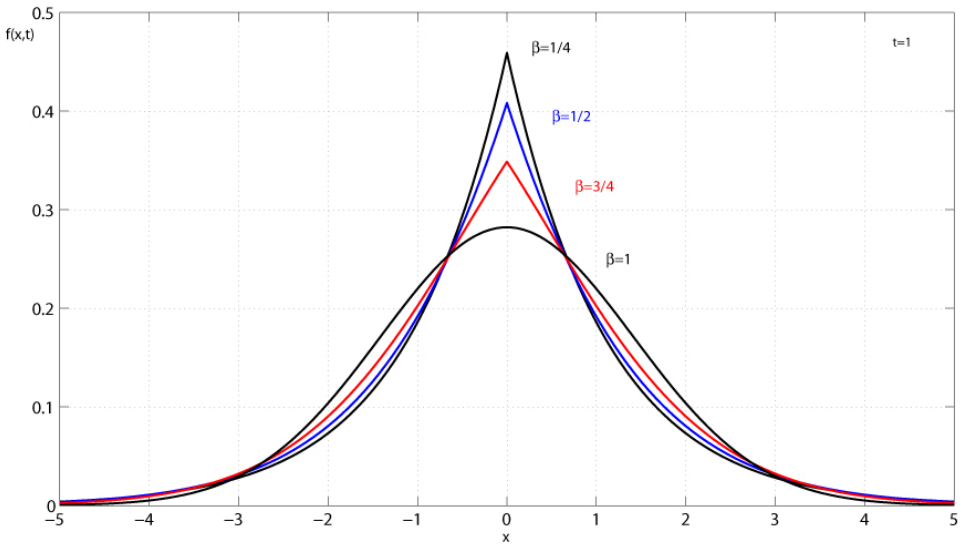

Figure 7.8: Plot of the fundamental solution $u(x,t)$ given by (eq. 7.140) at time $t = 1$, for different values of the parameter $\beta = [1/4, 1/2, 3/4, 1]$. For $\beta = 1$ one recovers the standard Gaussian density $G(x,t)$.

As we will see in Chapter 9 (Theorem 9.2 with $g(t) = t$ and $K = t^{\beta-1}/\Gamma(\beta)$), the fundamental solution of (eq. 7.134) turns out to be:

$$u(x,t) = \int_0^\infty G(x,\tau)h(\tau,t)d\tau, \tag{7.139}$$

where $G(x,t)$ is the Gaussian density and $h(\tau,t)$ is the fundamental solution of the time-fractional forward drift equation (eq. 7.110). Now, by using (eq. 7.121) and (eq. 7.126), one has:

$$u(x,t) = \frac{1}{2}\int_0^\infty \mathcal{M}_{1/2}(|x|,\tau)\mathcal{M}_\beta(\tau,t)d\tau,$$

which by (eq. 7.127) gives:

$$u(x,t) = \frac{1}{2}\mathcal{M}_{\beta/2}(|x|,t) = \frac{t^{-\beta/2}}{2}M_{\beta/2}(|x|t^{-\beta/2}). \tag{7.140}$$

Remark 7.36. Because $h(\tau,t) = \mathcal{M}_\beta(\tau,t)$, $0 < \beta \leq 1$, is a probability density function in $\tau \geq 0$, for any $t \geq 0$, then from (eq. 7.139) follows that $u(x,t)$, $t \geq 0$, is a probability density in $x \in \mathbb{R}$.

Of particular interest is the variance time evolution. In fact, by (eq. 7.139), one has:

$$\sigma^2(t) = 2\int_0^\infty \tau\mathcal{M}_\beta(\tau,t)d\tau. \tag{7.141}$$

6) We recall that the Fick law is essentially a phenomenological law, which represents the simplest relationship between the flux F and the gradient of the concentration u. If u is a temperature, F is the heat-flux, so we speak of the Fourier law. In both cases the law can be replaced by a more suitable phenomenological relationship which may account for possible non-local, non-linear and memory effects, without violating the conservation law expressed by the continuity equation.

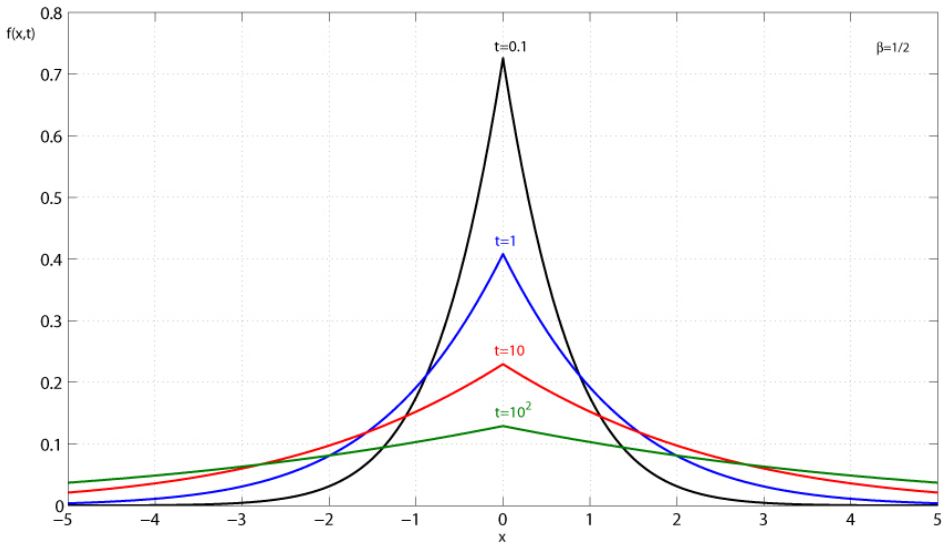

Figure 7.9: Plot of the density function $u(x,t)$ for fixed $\beta = 1/2$, at different times $t = [0.1, 1, 10, 10^2]$.

We observe that:

$$\int_0^\infty \tau s^{\beta-1} e^{-\tau s^\beta} d\tau = s^{-\beta-1},$$

then, by using (eq. 7.123), and inverting the Laplace transform (see also eq. (9.60) with $m = 1$), one finds

$$\sigma^2(t) = 2\frac{t^\beta}{\Gamma(\beta+1)}, \quad 0 < \beta \leq 1. \tag{7.142}$$

So that, we have a sub-linear growth in time, consistent with an anomalous process of *slow diffusion*.

In order to better visualize the decay of the tails of $u(x,t)$, in Figure (7.10) we used a semi-logarithmic scales. In this case the decay-plot of the queues is ranging from a straight line ($\beta = 0$) to a parabolic line ($\beta = 1$).

Remark 7.37. From a stochastic point of view, the fundamental solution (eq. 7.140) can be seen as the marginal density function of a self-similar stochastic process with $H = \beta/2$. Clearly, there exist many different stochastic processes which serve as model for the time-fractional diffusion. Here, we just observe that, in view of the subordination equation (eq. 7.139), a natural choice is the process $X = B(l(t))$, $t \geq 0$, where $l(t)$, $t \geq 0$, is a non-Markovian random time, with marginal density function satisfying the time-fractional forward drift equation (eq. 7.110), independent of the Brownian motion $B(t)$. We will back later on this topic.

7.2.6 Time-fractional diffusion of distributed order

As in th case of fractional relaxation, the time-fractional diffusion equations (eq. 7.135) and (eq. 7.136) can be generalized by using the notion of time-fractional derivative of distributed order. Then, let us introduce the normalized density function $p(\beta)$, $0 < \beta \leq 1$, which, as in (eq. 7.78), represents the weight

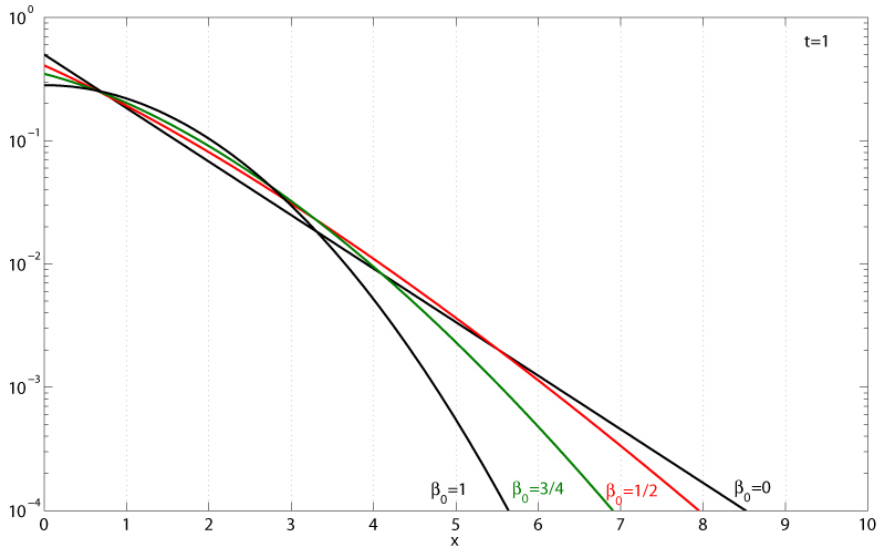

Figure 7.10: Plots (in linear-logarithmic scales) of $\frac{1}{2}M_{\beta_0/2}(x))$ versus x (in the interval $0 \leq x \leq 10$), for $\beta_0 = [0, 1/2, 3/4, 1]$.

distribution of the fractional derivatives orders. We suppose $p(\beta)$ verifies the same regularity properties of (eq. 7.78). Thus, we can write the two alternative forms of fractional diffusion of distributed order:

$$\frac{\partial}{\partial t}\, u(x,t) = \int_0^1 p(\beta) \mathcal{D}_t^{1-\beta} \left[\frac{\partial^2}{\partial x^2}\, u(x,t) \right] d\beta\,, \quad x \in \mathbb{R},\ t \geq 0\,, \tag{7.143}$$

and

$$\int_0^1 p(\beta) \left[{}_*\mathcal{D}_t^{\beta}\, v(x,t) \right] d\beta = \frac{\partial^2}{\partial x^2}\, v(x,t)\,, \quad x \in \mathbb{R},\ t \geq 0\,. \tag{7.144}$$

Remark 7.38. As in the case of fractional relaxation of distributed order, the two forms above are no-longer equivalent. We observe that only the RL time-fractional diffusion of distributed order with initial condition $u_0(x)$ is equivalent to the integral equation:

$$u(x,t) = u_0(x) + \int_0^1 p(\beta) \mathcal{I}_t^{\beta} \partial_{xx} u(x,t) d\beta. \tag{7.145}$$

We are interested in finding out a general representation of the fundamental solution corresponding to a generic choice of $p(\beta)$.

Remark 7.39. Let us remark that the flux formula (7.138) can be generalized in the case of the RL fractional diffusion equation of distributed order. In fact,

$$F\left[u(x,t)\right] = -\frac{\partial}{\partial x} \left\{ \int_0^1 p(\beta) \left[{}_tD^{1-\beta}\, u(x,t) \right] d\beta \right\}. \tag{7.146}$$

In order to find a suitable representation of the fundamental solutions, we apply the Laplace transform with respect to $t \geq 0$, and Fourier transform with respect to $x \in \mathbb{R}$, to the equations (7.143) and (7.144).

After introducing the auxiliary functions,

$$A(s) = s \int_0^1 p(\beta)\, s^{-\beta}\, d\beta \tag{7.147}$$

$$B(s) = \int_0^1 p(\beta)\, s^{\beta}\, d\beta\,, \tag{7.148}$$

we get:

$$\widehat{\widetilde{u}}(\kappa, s) = \frac{1}{s + \kappa^2 A(s)}\,, \tag{7.149}$$

and

$$\widehat{\widetilde{v}}(\kappa, s) = \frac{B(s)/s}{\kappa^2 + B(s)}\,. \tag{7.150}$$

Remark 7.40. We note that in the particular case $p(\beta) = \delta(\beta - \beta_0)$, one has $A(s) = s^{1-\beta_0}$, and $B(s) = s^{\beta_0}$. Then, one gets the same result of the time-fractional diffusion of a single order $\beta = \beta_0$.

We are inverting first the Fourier transform. We observe that[7]

$$\mathcal{F}\left\{\frac{c}{2\sqrt{d}} e^{-|x|\sqrt{d}}, x, k\right\} = \frac{c}{d + \kappa^2}, \quad d > 0. \tag{7.151}$$

Then, by comparing the latter with (eq. 7.149) and (eq. 7.150), we recognize the same Fourier transform couple with:

- In the RL case:
$$\begin{cases} c = c(s) := 1/A(s), \\ d = d(s) := s/A(s). \end{cases}$$
- In the CD case:
$$\begin{cases} c = c(s) := B(s)/s, \\ d = d(s) := B(s). \end{cases}$$

Then, one obtains:

$$\widetilde{u}(x, s) = \frac{1}{2\sqrt{sA(s)}} \exp\left\{-|x|\sqrt{s/A(s)}\right\}. \tag{7.152}$$

$$\widetilde{v}(x, s) = \frac{\sqrt{B(s)}}{2s} \exp\left\{-|x|\sqrt{B(s)}\right\}. \tag{7.153}$$

Following a standard procedure in complex analysis, the Laplace inversion requires the integration along the borders of the negative real semi-axis in the s-complex cut plain. In fact, this semi-axis, defined by $s = re^{\pm i\pi}$, $r > 0$, turns out to be the common branch-cut of the functions $s^{1-\beta}$, which appears in $A(s)$ and s^{β} which appears in $B(s)$. Then, in virtue of the Titchmarsh theorem on Laplace inversion, we get the representations in terms of real integrals of Laplace type.

$$u(x, t) = -\frac{1}{\pi} \int_0^{\infty} \mathrm{e}^{-rt}\, \mathrm{Im}\left\{\widetilde{u}\left(x, re^{i\pi}\right)\right\} dr\,, \tag{7.154}$$

7) The result is just a straightforward exercise in complex analysis based on residue theorem and Jordan's lemma.

which requires the evaluation of $A(s)$ along the ray $s = r\,\mathrm{e}^{i\pi}$ with $r > 0$. At this purpose, we write

$$A\left(r\,\mathrm{e}^{i\pi}\right) = \rho\,\cos(\pi\gamma) + i\rho\sin(\pi\gamma)\,, \tag{4.12a}$$

where

$$\begin{cases} \rho = \rho(r) = \left|A\left(r\,\mathrm{e}^{i\pi}\right)\right|\,, \\ \gamma = \gamma(r) = \frac{1}{\pi}\arg\left[A\left(r\,\mathrm{e}^{i\pi}\right)\right]\,. \end{cases} \tag{7.155}$$

In the same way, in the CD case, one gets:

$$v(x,t) = -\frac{1}{\pi}\int_0^\infty \mathrm{e}^{-rt}\,\mathrm{Im}\left\{\widetilde{u}_*\left(x, r\mathrm{e}^{i\pi}\right)\right\}\,dr, \tag{7.156}$$

which requires the evaluation of $B(s)$ along the ray $s = r\,\mathrm{e}^{i\pi}$ with $r > 0$. Again, we write

$$B\left(r\,\mathrm{e}^{i\pi}\right) = \rho_*\,\cos(\pi\gamma_*) + i\rho_*\sin(\pi\gamma_*)\,, \tag{7.157}$$

where

$$\begin{cases} \rho_* = \rho_*(r) = \left|B\left(r\,\mathrm{e}^{i\pi}\right)\right|\,, \\ \gamma_* = \gamma_*(r) = \frac{1}{\pi}\arg\left[B\left(r\,\mathrm{e}^{i\pi}\right)\right]\,. \end{cases} \tag{7.158}$$

Finally, we have found an integral representation of the fundamental solutions:

$$u(x,t) = \int_0^\infty \mathrm{e}^{-rt}\,P(x,r)\,dr\,, \quad P(x,r) = -\frac{1}{\pi}\,\mathrm{Im}\left\{\widetilde{u}\left(x, r\mathrm{e}^{i\pi}\right)\right\}\,, \tag{7.159}$$

and

$$v(x,t) = \int_0^\infty \mathrm{e}^{-rt}\,P_*(x,r)\,dr\,, \quad P_*(x,r) = -\frac{1}{\pi}\,\mathrm{Im}\left\{\widetilde{v}\left(x, r\mathrm{e}^{i\pi}\right)\right\}\,, \tag{7.160}$$

where the functions $P(x,r)$ and $P_*(x,r)$ must be derived by using (eq. 7.152) and (eq. 7.153), respectively.

Remark 7.41. We recognize that the expressions of P and P_* are related to each other by the transformation

$$\rho_*(r) \Longleftrightarrow r/\rho(r)\,, \quad \gamma_*(r) \Longleftrightarrow 1 - \gamma(r)\,. \tag{7.161}$$

By the remark above, it is sufficient to provide the explicit expression in the CD case:

$$\begin{aligned} P_*(x,r) &= \frac{1}{2\pi r}\,\mathrm{Im}\left\{\rho_*^{1/2}\,\mathrm{e}^{i\pi\gamma_*/2}\exp\left[-\mathrm{e}^{i\pi\gamma_*/2}\rho_*^{1/2}x\right]\right\} \\ &= \frac{1}{2\pi r}\,\rho_*^{1/2}\,\mathrm{e}^{-\rho_*^{1/2}x\cos(\pi\gamma_*/2)}\,\sin\left[\pi\gamma_*/2 - \rho_*^{1/2}x\sin(\pi\gamma_*/2)\right]\,. \end{aligned} \tag{7.162}$$

Then, the corresponding expression of $P(x;r)$ is obtained by applying the transformation (7.161).

Remark 7.42. The expressions (7.159) an (7.160) can be evaluated numerically in order to get plots of the fundamental solutions.

Remark 7.43. It is possible to show (see Example 9.6) that the fundamental solutions are probability densities. However, in contrast with the simple single order equation, the self-similarity is no-longer.

We are now interested in the study of their variance. Indeed, the variance time behavior is fundamental in the classification of the type of diffusion. One has:

$$\text{RL}\ :\ \sigma^2(t) = -\frac{\partial^2}{\partial\kappa^2}\,\widehat{u}(\kappa=0,t)\,;\quad \text{CD}\ :\ \sigma^2_*(t) = -\frac{\partial^2}{\partial\kappa^2}\,\widehat{v}(\kappa=0,t)\,. \tag{7.163}$$

As a consequence, we just have to invert the Laplace transforms in (eq. 7.149) and (eq. 7.150). We have, for $\kappa \to 0$,

$$\widehat{\widetilde{u}}(\kappa,s) = \frac{1}{s}\left(1-\kappa^2\,\frac{A(s)}{s}+\dots\right) \Leftrightarrow \widetilde{\sigma^2}(s) = -\frac{\partial^2}{\partial\kappa^2}\widehat{\widetilde{u}}(\kappa=0,s) = \frac{2A(s)}{s^2}\,, \tag{7.164}$$

and

$$\widehat{\widetilde{u}}_*(\kappa,s) = \frac{1}{s}\left(1-\kappa^2\,\frac{1}{B(s)}+\dots\right) \Leftrightarrow \widetilde{\sigma^2}(s) = -\frac{\partial^2}{\partial\kappa^2}\widehat{\widetilde{u}}(\kappa=0,s) = \frac{2}{s\,B(s)}. \tag{7.165}$$

We now study the examples of *double order* and *uniformly distributed order* fractional diffusion.

Example 7.11 (Double order fractional diffusion)**.** Let $p_1+p_2=1$ two positive numbers, and

$$p(\beta) = p_1\delta(\beta-\beta_1)+p_2\delta(\beta-\beta_2)\,,\quad 0<\beta_1<\beta_2\le 1. \tag{7.166}$$

Therefore,

- in the RL case, one has:
$$A(s) = p_1\,s^{1-\beta_1}+p_2\,s^{1-\beta_2}$$
which, inserted in (eq. 7.149), gives
$$\widehat{\widetilde{u}}(\kappa,s) = \frac{1}{s[1+\kappa^2(p_1\,s^{-\beta_1}+p_2\,s^{-\beta_2})]} \tag{7.167}$$
- In the CD case, one has:
$$B(s) = p_1\,s^{\beta_1}+p_2\,s^{\beta_2}\,,$$
which, inserted in (eq. 7.150), gives:
$$\widehat{\widetilde{u}}_*(\kappa,s) = \frac{p_1\,s^{\beta_1}+p_2\,s^{\beta_2}}{s[\kappa^2+p_1\,s^{\beta_1}+p_2\,s^{\beta_2}]} \tag{7.168}$$

We leave as algebraic exercise the evaluation of the functions $P(x,r)$ an $P_*(x,r)$, which have to be used in the numerical computation of the solutions. We focus now on the study of the variance. We start from the corresponding Laplace transforms (eq. 7.164) and (eq. 7.165):

$$\begin{cases} \widetilde{\sigma^2}(s) = 2\,p_1\,s^{-(1+\beta_1)}+2\,p_2\,s^{-(1+\beta_2)}, \\ \widetilde{\sigma^2}_*(s) = \dfrac{2}{p_1\,s^{(1+\beta_1)}+p_2\,s^{(1+\beta_2)}}. \end{cases}$$

In the RL case one easily finds:

$$\sigma^2(t) = 2\,p_1\,\frac{t^{\beta_1}}{\Gamma(\beta_1+1)}+2\,p_2\,\frac{t^{\beta_2}}{\Gamma(\beta_2+1)} \sim \begin{cases} 2p_1\dfrac{t^{\beta_1}}{\Gamma(1+\beta_1)}, & t\to 0_+, \\ 2p_2\dfrac{t^{\beta_2}}{\Gamma(1+\beta_2)}, & t\to+\infty. \end{cases} \tag{7.169}$$

Whereas, in the CD case, by using (eq. 7.73), one finds:

$$\sigma_*^2(t) = \frac{2}{p_2}\, t^{\beta_2}\, E_{\beta_2-\beta_1,\beta_2+1}\left(-\frac{p_1}{p_2}\, t^{\beta_2-\beta_1}\right) \sim \begin{cases} \dfrac{2}{p_2}\,\dfrac{t^{\beta_2}}{\Gamma(1+\beta_2)}, & t \to 0_+ \\ \dfrac{2}{p_1}\,\dfrac{t^{\beta_1}}{\Gamma(1+\beta_1)}, & t \to +\infty\,. \end{cases} \tag{7.170}$$

Therefore, we see that: in the RL case we have an explicit combination of two power-laws. The smallest exponent (β_1) dominates at small times, whereas the largest exponent (β_2) dominates at large times. In the CD case, we have a generalized Mittag-Leffler function which gives only asymptotically a combination of power-laws. Precisely, we get an opposite behavior than before: the largest exponent (β_2) dominates at small times and the smallest exponent (β_1) dominates at large times.

Remark 7.44. One could have obtained the same results directly from the Laplace transforms by applying the Tauberian theory.

In Figure (7.11), we chose $\{\beta_1 = 1/4, \beta_2 = 1\}$ in order to better contrast the different evolution of the fundamental solution for the RL and the CD forms. We exhibit the plots of the corresponding solution versus x (in the interval $|x| \leq 5$), at different times, selected as $t = 0.1$, $t = 1$ and $t = 10$. In this limited spatial range we can note how the time evolution of the probability density functions depends on the different time-asymptotic behavior of the variance.

Example 7.12 (Uniformly distributed fractional diffusion)**.** Let $p(\beta) = 1$ for any $0 < \beta \leq 1$. With this choice,

- in the RL case, one has:

$$A(s) = s\int_0^1 s^{-\beta}\, d\beta = \frac{s-1}{\log s},$$

  which, inserted in (eq. 7.149), gives

$$\widehat{\widetilde{u}}(\kappa, s) = \frac{\log s}{s\log s + \kappa^2\,(s-1)}\,. \tag{7.171}$$

- In the CD case, one has:

$$B(s) = \int_0^1 s^{\beta}\, d\beta = \frac{s-1}{\log s}\,,$$

  which, inserted in (eq. 7.150), gives:

$$\widehat{\widetilde{u}}_*(\kappa, s) = \frac{1}{s}\frac{s-1}{\kappa^2 \log s + s - 1} = \frac{1}{s} - \frac{1}{s}\frac{\kappa^2 \log s}{\kappa^2 \log s + s - 1}\,. \tag{7.172}$$

Also here, we leave as algebraic exercise the evaluation of the functions $P(x, r)$ an $P_*(x, r)$, and focus now on the study of the variance. In the RL case:

$$\widetilde{\sigma^2}(s) = 2\left[\frac{1}{s\log s} - \frac{1}{s^2 \log s}\right]. \tag{7.173}$$

Then, by inversion (see [16, 17]),

$$\sigma^2(t) = 2\,[\nu(t,0) - \nu(t,1)] \sim \begin{cases} 2/\log(1/t), & \to 0, \\ 2t/\log t, & t \to \infty, \end{cases} \tag{7.174}$$

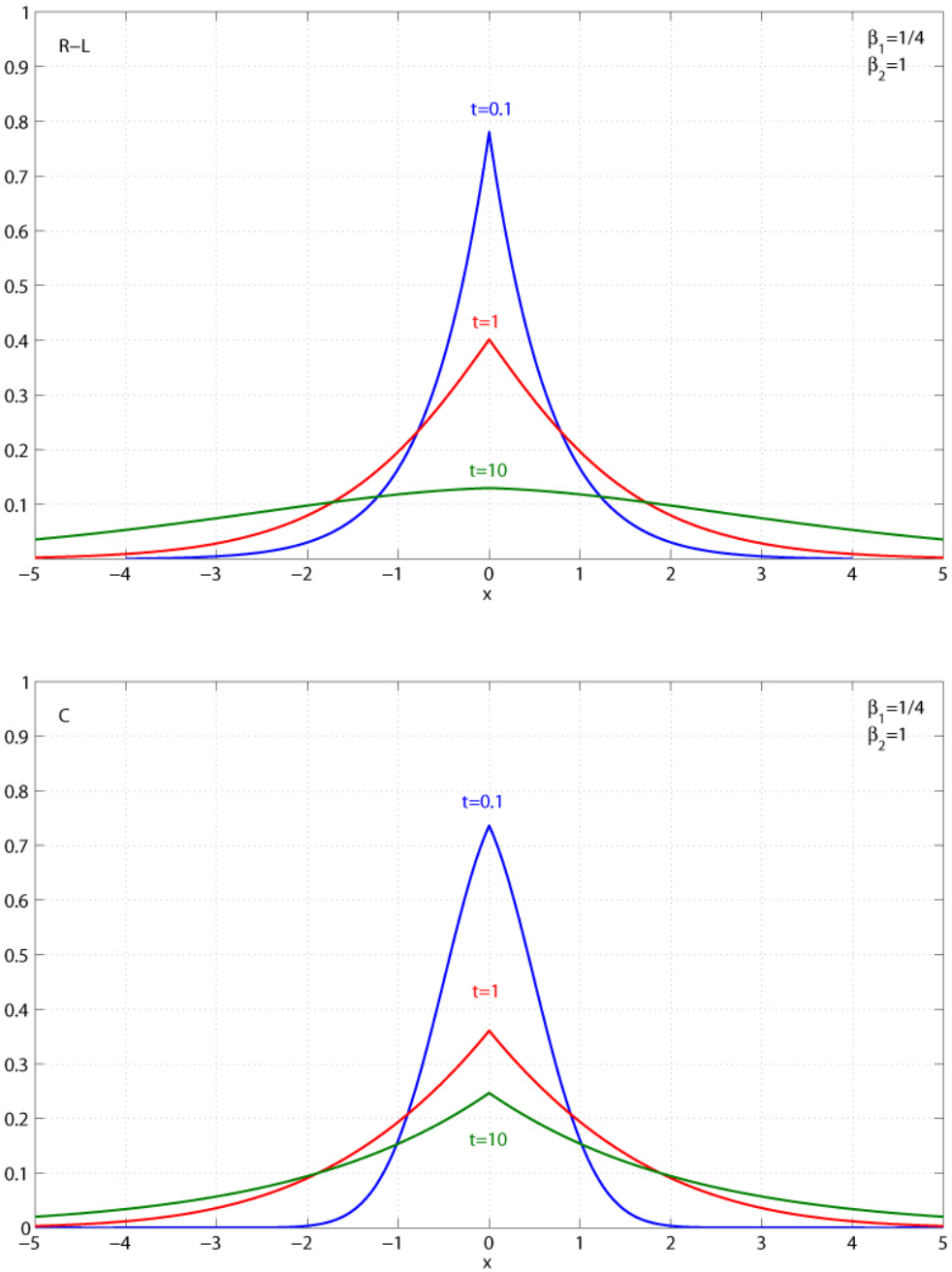

Figure 7.11: Plots of the fundamental solution versus x (in the interval $|x| \leq 5$), for the case $\{\beta_1 = 1/4, \beta_2 = 1\}$ at times $t = [0.1, 1, 10]$. Top: RL case; bottom: CD case.)

where

$$\nu(t,a) = \int_0^\infty \frac{t^{a+\tau}}{\Gamma(a+\tau+1)}\, d\tau\,, \quad a > -1.$$

In the CD case, we have:

$$\widetilde{\sigma_*^2}(s) = \frac{2}{s}\,\frac{\log s}{s-1}. \tag{7.175}$$

Which, by inversion (see [16, 17]), gives:

$$\sigma_*^2(t) = 2\left[\log t + \gamma + \mathrm{e}^t\,\mathcal{E}_1(t)\right] \sim \begin{cases} 2t\log(1/t), & t \to 0, \\ 2\log(t), & t \to \infty, \end{cases} \tag{7.176}$$

where $\mathcal{E}_1(t)$ denotes the exponential integral function (eq. 7.105) and $\gamma = 0.57721...$ is the so-called Euler-Mascheroni constant.

We find instructive to compare in Figure (7.6) the solutions corresponding to RL and CD cases with the solutions of the fractional diffusion of a single order $\beta_0 = [1/4, 3/4, 1]$ at fixed times $t = 1, 10$.

Then, in Figures (7.13) and (7.14), we compare the variance at moderate times $0 \le t \le 10$, using linear scales, and at large times $10^1 \le t \le 10^7$, using logarithmic scales, respectively. Here we have inserted the plot for $\beta_0 = 1/2$. In order to interpret these asymptotic behaviors, we observe that, due to the logarithmic constituents in the RL case, the variance grows slightly slower than linearly for $t \to \infty$, but extremely slowly near $t = 0$. In the CD case the variance exhibits a slightly super-linear growth near $t = 0$, but an extremely slow growth for $t \to \infty$.

7.2.7 Fractional representation of fractional Brownian motion

The name fractional Brownian motion suggests that it can be in some way regarded as a fractional integral of Brownian motion. Here, we want to show that it is best viewed as a fractional integral of a Gaussian white noise.

Let $B_H(t)$, $t \in \mathbb{R}$, be a (two-sided) fractional Brownian motion of index $0 < H < 1$ (see Section 3.7). In the context of fractional integration, it is convenient to use another parameterization of fBm: we set $\kappa = H - 1/2$, and we indicate with $B^\kappa(t)$ a fBm of order κ. Clearly, the range $0 < H < 1$ corresponds to the range $-1/2 < \kappa < 1/2$.

Remark 7.45. In the statistics of long range dependence (see Chapter 6), the parameter κ has been indicated with $d = H - 1/2$ and it represents just the order of the fractional derivative appearing in the definition of a FARIMA process (eq. 3.61). However, d can be confusing in this context.

We start with the moving average representation of standard fractional Brownian motion (eq. 5.1), namely:

$$B^\kappa(t) \stackrel{d}{=} \frac{1}{C_1(\kappa)} \int_{\mathbb{R}} \left((t-s)_+^\kappa - (-s)_+^k\right) dB^0(s), \quad t \in \mathbb{R}, \tag{7.177}$$

where

$$C_1(\kappa)^2 = \int_0^\infty \left((1+s)^\kappa - s^\kappa\right)^2 ds + \frac{1}{2\kappa+1}. \tag{7.178}$$

We have the following:

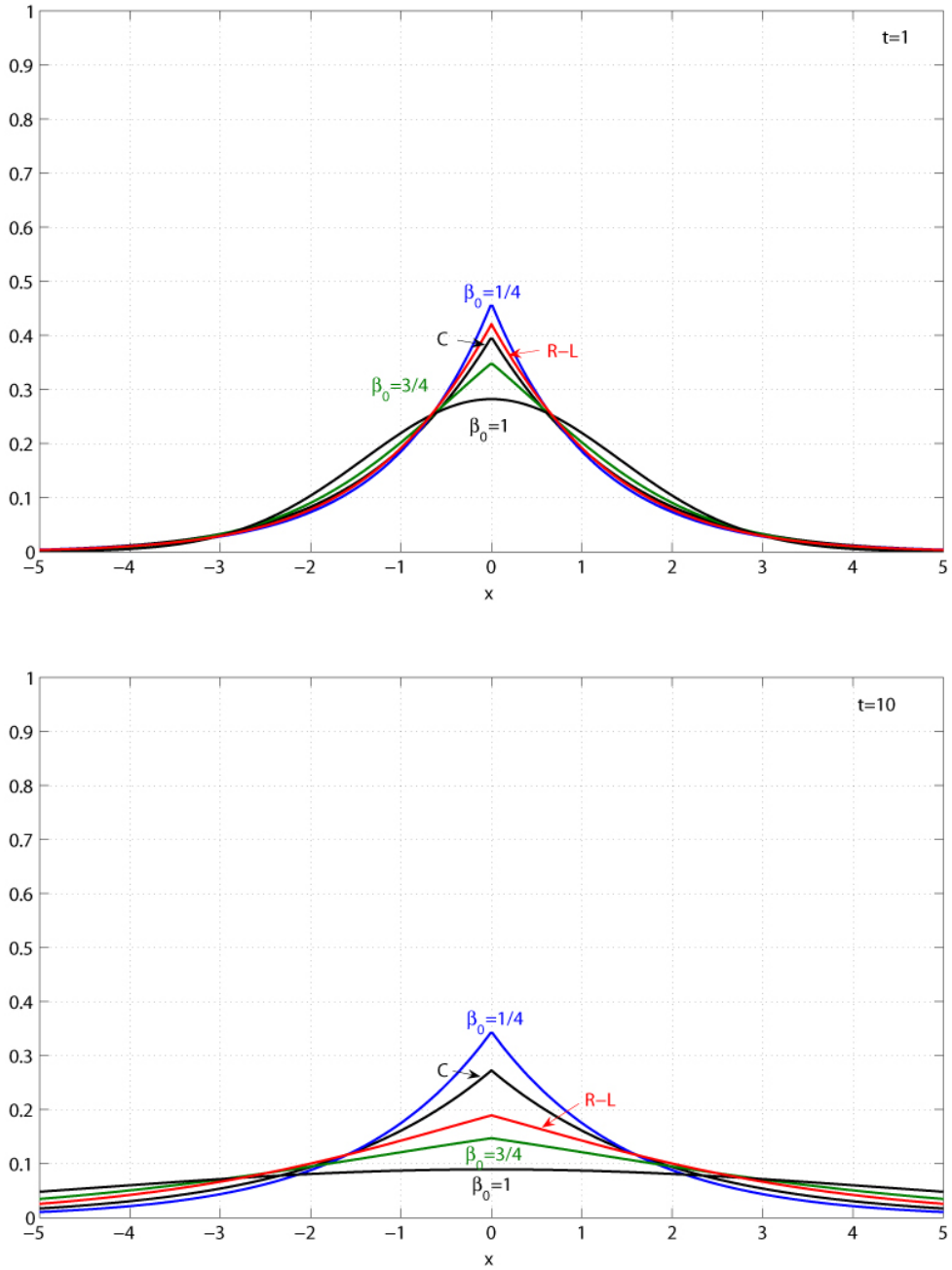

Figure 7.12: Plots of the fundamental solution versus x (in the interval $|x| \leq 5$), for the uniform order distribution in RL and CD forms compared with the solutions of some cases of single order at $t = 1$ (top) and $t = 10$ (bottom).

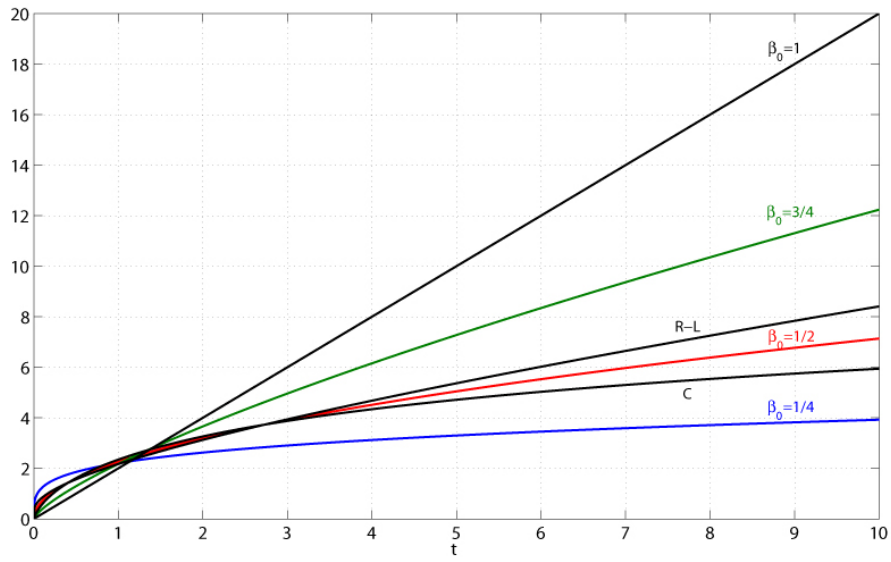

Figure 7.13: Plots of the variance versus t in the interval $0 \leq t \leq 10$ (linear scales), for the uniform order distribution.

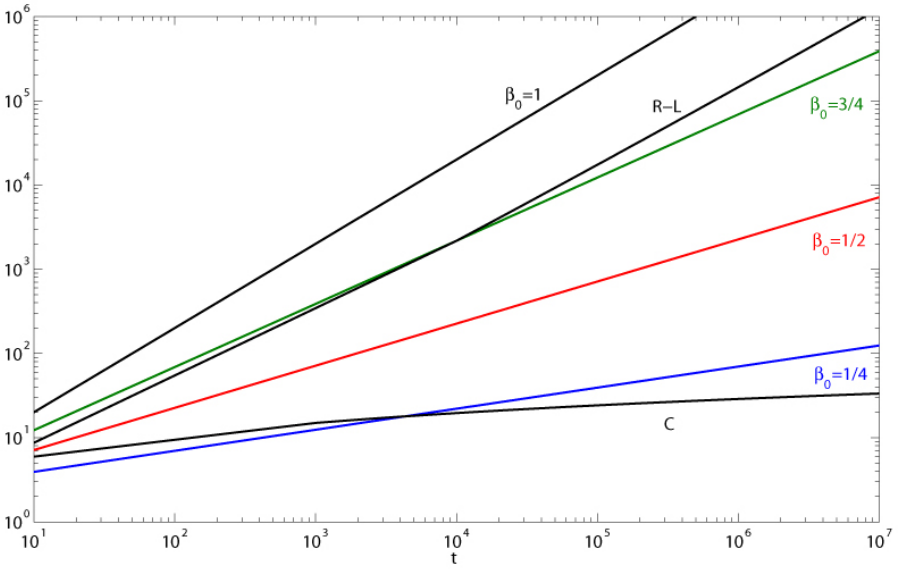

Figure 7.14: Plots of the variance versus t in the interval $10^1 \leq t \leq 10^7$ (logarithmic scales) for the uniform order distribution.

Prop. 7.4. *Let $-1/2 < \kappa < 1/2$ and suppose that B^k is a standard fBm. Then,*

$$B^\kappa(t) \overset{d}{=} \frac{\Gamma(\kappa+1)}{C_1(\kappa)} \int_{\mathbb{R}} \mathcal{I}_-^\kappa 1_{[0,t)}(s) dB^0(s), \quad t \in \mathbb{R}. \tag{7.179}$$

In fact, by using (eq. 7.44) for $0 < \kappa < 1/2$ and (eq. 7.51) for $-1/2 < \kappa < 0$, by meaning $\mathcal{I}_\pm^{-\alpha} := \mathbf{D}_\pm^\alpha$, $\alpha > 0$, one has:

$$\Gamma(\kappa+1)\mathcal{I}_-^\kappa 1_{[0,t)}(s) = (t-s)_+^\kappa - (-s)_+^\kappa,$$

for any $t \in \mathbb{R}$ and $-1/2 < \kappa < 1/2$.

Remark 7.46. One can observe that the spectral representation (eq. 5.13) can be directly obtained form (eq. 7.179) by using (Prop. 5.3), namely:

$$B^\kappa(t) \overset{d}{=} \frac{\Gamma(\kappa+1)}{\sqrt{2\pi}C_1(\kappa)} \int_{\mathbb{R}} \widehat{\mathcal{I}_-^\kappa 1}_{[0,t)}(\nu) d\widetilde{B}(\nu), \tag{7.180}$$

where $d\widetilde{B}(\nu) = dB_1(\nu) + i dB_2(\nu)$ is a complex valued Gaussian measure such that $dB_1(-\nu) = dB_1(\nu)$ and $dB_2(-\nu) = -dB_2(\nu)$.

In fact, the factor $\sqrt{\pi}$ comes after comparing the definitions (eq. 5.7) and (eq. 7.52), and gives (eq. 5.5). Moreover, by using (eq. 7.54), on finds

$$\widehat{\mathcal{I}_-^\kappa 1}_{[0,t)}(\nu) = \frac{\widehat{1}_{[0,t)}(\nu)}{(i\nu)^\kappa} = \frac{e^{i\nu t} - 1}{i\nu} |x|^{-\kappa} e^{-\mathrm{sign}(x)\frac{i\kappa\pi}{2}},$$

which actually gives the expected result (see Remark 5.5).

Heuristically, the representation (7.179) says that the fractional Gaussian noise is the κ-fractional integral of the white noise $\dot{B}^0$. That is, formally speaking:

$$\dot{B}^\kappa(t) = \mathcal{I}_+^\kappa \dot{B}^0(t). \tag{7.181}$$

In fact, we have naively that:

$$B^\kappa(t) = \int_{\mathbb{R}} 1_{[0,t)}(s) \dot{B}^\kappa(s) ds = \int_{\mathbb{R}} \mathcal{I}_-^\kappa 1_{[0,t)}(u) \dot{B}^0(u) du = \int_{\mathbb{R}} 1_{[0,t)}(u) \mathcal{I}_+^\kappa \dot{B}^0(u) du,$$

where we have used the integration by parts formula (eq. 7.22).

Remark 7.47. It is also possible to relate fractional Brownian motion to fractional integrals defined on an interval (see [18]).

Bibliography

[1] Abramowitz, M and Stegun, I.A., 1965, *Handbook of Mathematical Functions*, Dover, New York.

[2] M. Dentz et al., Time behavior of solute transport in heterogeneous media: transition from anomalous to normal transport, *Adv. Water Resources* **27** (2004), 155–173.

[3] Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F.G., 1955, *Higher Transcendental Functions*,Vol. 3, McGraw-Hill, New York.

[4] C. Fox, The *G* and *H*-Functions as Symmetrical Fourier Kernels. *Trans. Amer. Math. Soc. 98*, 395-429, 1961.

[5] V. M. Kenkre, R. S. Knox, Generalized-master-equation theory of excitation transfer, *Physical Review B* **9** (1974) 5279-5290.

[6] Kilbas, A.A., Srivastava, H.M. and Trujillo, J.J., 2006, *Theory and Applications of Fractional Differential Equations*, North-Holland, Amsterdam.

[7] R. Gorenflo, Yu. Luchko and F. Mainardi, Analytical properties and applications of the Wright function, *Fractional Calculus and Applied Analysis* **2** (1999) 383-414.

[8] R. Gorenflo and F. Mainardi, Fractional diffusion processes: Probability distributions and continuous time random walk, in: G. Rangarajan and M. Ding (Editors), *Processes with Long Range Correlations*, Springer Verlag, Berlin, 2003, pp. 148-161. [Lecture Notes in Physics, No 621]

[9] R. Gorenflo, and F. Mainardi, 1997, "Fractional calculus: integral and differential equations of fractional order", in *Fractals and Fractional Calculus in Continuum Mechanics*, A. Carpinteri and F. Mainardi, eds., Springer Verlag, Wien, pp. 223–276. [Reprinted in `http://www.fracalmo.org`]

[10] R. Gorenflo, F. Mainardi: Fractional calculus: integral and differential equations of fractional order. In: *Fractals and Fractional Calculus in Continuum Mechanics*, ed by A. Carpinteri, F. Mainardi (Springer Verlag, Wien and New York 1997) pp. 223–276 [Reprinted in `http://www.fracalmo.org`]

[11] F. Mainardi, Fractional calculus: some basic problems in continuum ad statistical mechanics, in: A. Carpinteri and F. Mainardi (Editors), *Fractals and Fractional Calculus in Continuum Mechanics*, Springer Verlag, Wien and New York, 1997, pp. 291-348. [Reprinted in `http://www.fracalmo.org`]

[12] F. Mainardi, Yu. Luchko and G. Pagnini, The fundamental solution of the space-time fractional diffusion equation, *Fractional Calculus and Applied Analysis* **4**(2001) 153-192. [E-print `http://arxiv.org/abs/cond-mat/0702419`]

[13] F. Mainardi, G. Pagnini and R. K. Saxena, Fox H functions in fractional diffusion, *J. Comput. Appl. Math. vol. 178, 2005*, Elsevier Science Publishers B. V. pp. 321-331.

[14] F. Mainardi, G. Pagnini and R. Gorenflo, Mellin transform and subordination laws in fractional diffusion processes, *Fractional Calculus and Applied Analysis* **6** (2003) 441-459.

[15] F. Mainardi, A. Mura, R. Gorenflo and M. Stojanovic, Two forms of fractional relaxation of distributed order, *Journal of Vibration and Control*, **13**, No. 9-10, pp.1249-1268 (2007).

[16] F. Mainardi, A. Mura, G. Pagnini and R. Gorenflo, *Sub-diffusion equations of fractional order and their fundamental solution*, in K. Tas, J. A. Tenreiro-Machado and D. Baleanu (Editors): Mathematical Methods in Engineering, Springer-Verlag, Dordrecht 2007, pp. 23-55. ISBN 978-1-4020-2 (HB).

[17] F. Mainardi, A. Mura, G. Pagnini and R. Gorenflo, Time-fractional diffusion of distributed order, *Journal of Vibration and Control*, to appear (2007).

[18] V. Pipiras, M.S. Taqqu, Fractional Calculus and its connections to fractional Brownian motion, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and Applications of Long-range Dependence*, Birkäuser, Boston - Basel - Berlin, 2003, pp. 165-201.

[19] I. Podlubny, *Fractional Differential Equations*, Academic Press, San Diego-Boston-New York-London-Tokyo-Toronto, 1999.

[20] S. G. Samko, A. A. Kilbas, O. I. Marichev, *Fractional Integrals and Derivatives*, Gordon and Breach Science Publishers, 1993.

[21] W.R. Schneider, Fractional diffusion, in: R. Lima, L. Streit and D. Vilela Mendes (Editors), *Dynamics and Stochastic Processes, Theory and Applications*, Springer Verlag, Heidelberg, 1990, pp. 276-286. [Lecture Notes in Physics No 355]

[22] T. Srokowski, Non-Markovian Lévy diffusion in non homogeneous media, *Physical Review E* **75** (2007) 051105/1-8. [E-print `http://arxiv.org/abs/cond-mat/0611056`]

[23] R. Zwanzig, Memory effects in irreversible thermodynamics, *Physical Review* **124** (1961) 983-992.

8 Grey Brownian motion

The grey noise theory introduced by Schneider (see [17, 18]) leads naturally to a class of self-similar stochastic processes $\{B_\beta(t),\ 0 < \beta \leq 1\}$. These processes, called *grey Brownian motion* (gBm), provide stochastic models for the slow-anomalous diffusion described by the time fractional diffusion equation (eq. 7.134). This means that the marginal density function of the grey Brownian motion is the fundamental solution (eq. 7.140) of the time fractional diffusion equation.

In this chapter we introduce an extended class of stochastic processes $\{B_{\alpha,\beta}(t)\}$, with $0 < \alpha < 2,\ \ 0 < \beta \leq 1$, which is called *"generalized" grey Brownian motion* (ggBm) (see Mura [12, 13]). This class includes non-Markovian stochastic models either for slow and fast-anomalous diffusion. After having presented and motivated the mathematical construction, we will show that this class is made up of H-sssi processes and contain either Gaussian and non-Gaussian processes (like fractional Brownian motion and grey Brownian motion). The time evolution of the $B_{\alpha,\beta}(t)$ marginal density function is described by partial integro-differential equations of fractional type, which means featuring a power-like memory kernel, indeed including the time-fractional diffusion equations.

8.1 White noise

The ggBm is defined through the explicit construction of the underline probability space. The same construction could be used in order to define Brownian motion, which is actually a particular case of grey Brownian motion. So that, we start describing this construction in the case of Bm. At this purpose it is crucial to better clarify the concept of white noie.

We have already encountered the notion of white noise several times. We stated that it could have be seen as a sort of Bm time derivative (see Remark 3.24 and considerations here below), but, being Brownian motion trajectory almost surely not differentiable, this intuition remained obscure. Then, we hinted that a useful and actually rigorous definition must have be given in terms of generalized processes (see the comments below eq. 4.4). We try to better clarify these concepts in the following.

We have seen that a *Gaussian white noise* $\{z(t)\}_{t\in\mathbb{R}}$ is heuristically defined as a zero-mean **stationary** process such that $z(t)$'s are independent with mean 0 and variance ∞, in the sense that:

$$E(z_t z_s) = \gamma(t-s) = \delta(t-s), \quad t, s \in \mathbb{R}, \tag{8.1}$$

where $\delta(x)$ is the Dirac delta function. The name white noise comes from the fact that, by (eq. 8.1), it has a constant nonzero spectral density:

$$f_z(\nu) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\nu t}\gamma(t)dt = \frac{1}{2\pi},$$

so that its average power is uniformly distributed in frequency, which is a characteristic of white light. The equation (8.1) suggests that Gaussian white noise is an unusual stochastic process.

In fact, let $\{W_t\}_{t\in\mathbb{R}}$ be a standard Wiener process and for fixed $h > 0$ define a new process:

$$\{z_h(t)\}_{t\in\mathbb{R}} = \left\{\frac{W(t+h) - W(t)}{h}\right\}_{t\in\mathbb{R}}. \tag{8.2}$$

The standard Wiener process W_t has a spectral representation (eq. 5.17):

$$W_t = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{e^{it\nu} - 1}{i\nu} d\widetilde{W}(\nu), \tag{8.3}$$

where, as usual, $d\widetilde{W}(\nu)$ is a complex Gaussian measure such that:

$$d\widetilde{W}(\nu) = dW_1(\nu) + i dW_2(\nu),$$

where W_1 and W_2 are independent standard Wiener processes with:

$$dW_1(\nu) = dW_1(-\nu), \quad dW_2(\nu) = -dW_2(-\nu).$$

Starting from (eq. 8.3) and using the definition (eq. 8.2), we find the corresponding spectral representation for the process $z_h(t)$:

$$z_h(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\nu t} \frac{e^{ih\nu} - 1}{i\nu h} d\widetilde{W}(\nu).$$

Because the $\{z_h(t)\}$'s stationarity, we could write:

$$\gamma(t) = E\left(z_h(t) z_h(0)\right) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\nu t} \frac{|e^{ih\nu} - 1|^2}{\nu^2 h^2} d\nu,$$

therefore the spectral density is (compare with eq. 5.26 with $H = 1/2$ and $\sigma = 1$):

$$f_{z_h}(\nu) = \frac{1}{2\pi} \frac{|e^{ih\nu} - 1|^2}{\nu^2 h^2} = \frac{1}{2\pi} \frac{4\sin^2(h\nu/2)}{\nu^2 h^2}; \tag{8.4}$$

this density converges to $1/2\pi$ for all $\nu \neq 0$ as h converges to 0, which suggests that the process z_h converges in some sense to a Gaussian white noise that could than be seen as the derivative $\dot{W}$ of a Wiener process. However the sample paths of a Wiener process are, almost surely, not differentiable anywhere. Thus a Gaussian white noise process cannot be a stochastic process in the usual sense, but, as suggested by (8.1), it must be interpreted in the sense of generalized functions like the Dirac delta function.

Remark 8.1. White noise cannot be realized physically, but can be approximated to any desired degree of accuracy by conventional stochastic processes with broad banded spectra such as (8.2).

8.1.1 White noise as a generalized stochastic process

As suggested by our previous considerations, the white noise should be thought as a sort of generalized process. We know that a generalized function is a function $f(\xi)$ depending linearly on test functions ξ. In the same way a *generalized stochastic process* $X(\xi)$ shall be a function depending linearly on test functions ξ such that for each ξ, $X(\xi)$ is a random variable.

Therefore, we define:

Def. 8.1 (White noise)**.** a white noise is a generalized stochastic process X such that for each test function ξ:

- $X(\xi)$ is Gaussian;
- $E(X(\xi)) = 0$,

- $\sigma^2(X(\xi)) = \int_{\mathbb{R}} \xi(t)^2 dt$.

We observe now that if it were possible to write:

$$X(\xi) = \int_{\mathbb{R}} \xi(t) z(t) dt, \tag{8.5}$$

we would have:

$$E(z(t)z(s)) = \delta(t-s),$$

that is (eq. 8.1). Indeed, for each pair of test functions ξ and η:

$$E(X(\xi+\eta)^2) = \int_{\mathbb{R}} \left[\xi^2(t) + \eta^2(t) + 2\xi(t)\eta(t)\right] dt$$

(by linearity)

$$= E(X(\xi)^2) + E(X(\eta)^2) + 2E(X(\xi)X(\eta)),$$

therefore, using (8.5), we get the result:

$$\int_{\mathbb{R}} \xi(t)\eta(t)dt = E(X(\xi)X(\eta)) = \int_{\mathbb{R}^2} \xi(t)\eta(s)E(z(t)z(s))dtds.$$

Does white noise as such a generalized stochastic process $X(\xi)$ exists? The answer is affirmative. In fact, it could ever be defined by making use of stochastic integrals. Indeed, let $\{W(t)\}_{t\in\mathbb{R}}$ be a standard Brownian motion, then

$$X(\xi) = \int_{\mathbb{R}} \xi(t) dW(t)$$

defines a white noise $\dot{W}$ as a generalized process. We write formally:

$$X(\xi) = \int_{\mathbb{R}} \xi(t)\dot{W}(t)dt = \int_{\mathbb{R}} \xi(t) dW(t). \tag{8.6}$$

The equation (8.6) suggests that the space of test function for X be the space of the integrands for the Ito integral (for example $L^2(\mathbb{R})$).

Example 8.1. Starting from the above definition, we could define objects like $\ddot{W}$, regarded as a generalized stochastic process defined by:

$$Y(\xi) = \int_{\mathbb{R}} \xi(t)\ddot{W}(t)dt = -\int_{\mathbb{R}} \xi'(t) dW(t). \tag{8.7}$$

Remark 8.2. It's important to observe that when white noise is defined as a generalized stochastic process, it is regarded as a whole $\dot{W}$ and, for each $t \in \mathbb{R}$, the quantity $\dot{W}(t)$ has no meaning. In other words, we cannot find the quantity $\dot{W}(t)$ alone, but always forming a stochastic differential in the combination

$$\dot{W}(t)dt = dW(t).$$

Remark 8.3. However, it is possible to show that the quantity $\dot{W}(t)$ can become meaningful as a generalized function defined on an **infinite dimensional** space (T. Hida's theory). Such a theory, called *white noise distribution theory*, is indeed a generalization of the Schwartz distribution theory to infinite dimensional spaces. We remind the interested reader to the reach literature covering this interesting topic (see for instance [5,9]).

8.1.2 Characteristic functional and Minlos theorem

We begin introducing some basic concepts and facts. Let X be a vector space over a $\mathbb{K}$-field and let $\{||\cdot||_p,\ p \in I\}$ be a countable family of Hilbert seminorms defined on it. We remember that:

Def. 8.2 (Seminorms)**.** A seminorm on X is a map $||\cdot||_p : X \to \mathbb{R}$ such that

1. $||\cdot||_p$ is positive or positive semidefinite: $||x||_p \geq 0$, $x \in X$;
2. $||\cdot||_p$ is positive homogeneous or positive scalable: $||\lambda x||_p = |\lambda|||x||_p$, with $\lambda \in \mathbb{K}$ and $x \in X$;
3. $||\cdot||_p$ is subadditive: $||x+y||_p \leq ||x||_p + ||y||_p$, $x, y \in X$

Notation 8.1. In the following, we are going to indicate with X_p the completion of X with respect to the seminorm $||\cdot||_p$.

Def. 8.3 (Hilbert seminorms)**.** A seminorm $||\cdot||_p$ is an **Hilbert seminorm** if X_p is an Hilbert space, or equivalently if $||\cdot||_p$ comes from a sesquilinear positive semidefinite form on X.

Remark 8.4. If $||\cdot||_p$ satisfies positive definiteness, which means that $||x||_p = 0$ if and only if $x = 0$, then $||\cdot||_p$ is a norm.

Def. 8.4 (Topological vector space)**.** The space X along with the family of seminorms $\{||\cdot||_p,\ p \in I\}$, is said to be a *locally convex space* (or simple a *topological vector space*), if it carries, as natural topology, the *initial topology*[1] of the seminorms and the vector space operations (addition and scalar multiplication).

Let X be a topological vector space with the topology given by a family $\{||\cdot||_p,\ p \in I\}$ of Hilbert norms. For any $u, v \in X$, we define:

$$d(u,v) = \sum_{n=1}^{\infty} 2^{-p_n} \frac{||u-v||_{p_n}}{1+||u-v||_{p_n}}.$$

Then, d is a metric on X and it is Cauchy if and only if X is Cauchy in each norm $||\cdot||_p$.

Def. 8.5 (Countably-Hilbert space)**.** A topological vector space X, with the topology generated by a family of Hilbert norms $\{||\cdot||_p\}_p$, is called a *countably-Hilbert space* if it is complete with respect to its topology or, equivalently, with respect to the corresponding metric.

For instance, if one assumes that the family $\{||\cdot||_p\}_p$ is increasing. Namely, for any $x \in X$:

$$||x||_{p_1} \leq ||x||_{p_2} \leq \cdots \leq ||x||_{p_n},\ \ p_1 < p_2 < \cdots < p_n.$$

Then, the following inclusions are continuous:

$$X \subset \cdots \subset X_{p_{n+1}} \subset X_{p_n} \subset \cdots \subset X_{p_1},$$

and the following result holds:

Prop. 8.1. *X is complete if and only if $X = \bigcap_{n=1}^{\infty} X_{p_n}$.*

Conversely, suppose that $\{(Y_p,\ ||\cdot||_p),\ p \in I\}$ is a countable sequence of Hilbert spaces such that $Y_{p_{n+1}}$ is continuously embedded in Y_{p_n} for each $p_n \in I$. We set:

$$Y = \bigcap_{n=1}^{\infty} Y_{p_n}.$$

1) The coarsest topology on X which makes those functions continuous.

Def. 8.6 (Projective limit topology)**.** We define the *projective limit topology* on Y as the coarsest topology such that for each $n \geq 1$ the inclusion map from Y to Y_{p_n} is continuous. The topological vector space Y is said the *projective limit* of $\{Y_{p_n},\ n \geq 1\}$.

Remark 8.5. Obviously, Y is a countably-Hilbert space with the norms $\{||\cdot||_p,\ p \in I\}$.

In the same way, let $\{(W_n, ||\cdot||_n),\ n \geq 1\}$ be a sequence of Hilbert spaces such that W_n is continuously embedded in W_{n+1} for each n. We set

$$W = \bigcup_{n=1}^{\infty} W_n.$$

We define:

Def. 8.7 (Inductive limit topology)**.** We define the *inductive limit topology* on W as the finest locally convex topology such that for each $n \geq 1$ the inclusion map from W_n to W is continuous. The topological vector space W is said the *inductive limit* of $\{W_n\}_n$.

We consider now dual spaces. Let X be a countably-Hilbert space with an increasing sequence $\{||\cdot||_p,\ p \in I\}$ of norms. Then, the dual space X' of X is given by:

$$X' = \bigcup_{n=1}^{\infty} X'_{p_n},$$

and we have the inclusions:

$$X'_{p_1} \subset \cdots \subset X'_{p_n} \subset X'_{p_{n+1}} \subset \cdots \subset X'. \tag{8.8}$$

Let $\langle \cdot, \cdot \rangle$ denote the natural bilinear pairing between X and X'. There are several topologies on the dual space X'.

A. *Weak topology*: the weak topology on X' is defined to be the coarsest topology on X' such that the functional $\langle \cdot, x \rangle$ is continuous for any $x \in X$. Equivalently this topology has a base of neighborhoods of zero given by sets of form:

$$N(x_1, \ldots, x_k; \epsilon) = \{\xi \in X'; |\langle \xi, x_j \rangle| < \epsilon,\ 1 \leq j \leq k\}, \tag{8.9}$$

where $x_1, \ldots, x_k \in X$ and $\epsilon > 0$. Obviously the inclusion map from X'_p into X' is continuous and X'_p is dense in X' relative to the weak topology.

B. *Strong topology*: the strong topology of X' is defined to be the topology with a base of neighborhoods of zero given by:

$$N(D, \epsilon) = \{\xi \in X';\ \sup_{x \in D} |\langle \xi, x \rangle| < \epsilon\},$$

where D is any bounded[2] subset of X and $\epsilon > 0$. The strong topology is finer than the weak topology. Moreover, the inclusion map between X'_p and X' is continuous and X'_p is dense in X'.

C. *Inductive topology*: if X is a contably-Hilbert space, its dual space X', in view of (8.8), can be regarded as the inductive limit of $\{X'_n,\ n \geq 1\}$. Thus X' has the inductive limit topology. It turns out that this inductive topology is the same as the strong topology on X'.

2) $D \subset X$ is bounded if for any neighborhood $N \subset X$ of zero there is a positive number $c > 0$ such that $D \subset cN$.

Remark 8.6. It is possible to show that the σ-fields on X' generated by the three topologies (weak, strong and inductive limit) are all the same if X is a so called *Nuclear space*.

Def. 8.8 (Nuclear space)**.** A topological vector space X, with the topology defined by a family of Hilbert norms, is said a *nuclear space* if for any Hilbert norm $||\cdot||_p$ there exists a larger norm $||\cdot||_q$ such that the inclusion map $X_q \hookrightarrow X_p$ is an Hilbert-Schmidt operator[3].

Nuclear spaces have many of the good properties of the finite dimensional Euclidean spaces $\mathbb{R}^d$. For example, a subset of a nuclear space is compact if and only if is bounded and closed. Moreover, spaces whose elements are "smooth" is some sense tend to be nuclear spaces.

In the following example we see how nuclear spaces could be constructed naturally starting from an Hilbert space and an operator (see Kuo [9]).

Example 8.2. Let H be an Hilbert space and A an operator defined on it. Suppose that there exists an orthonormal bases $\{h_n,\ n = 1, 2, \dots\}$ satisfying the following properties:

1. They are eigenvectors of A; i.e. for any $n > 0$: $Ah_n = \lambda_n h_n,\ \ \lambda_n \in \mathbb{R}$.
2. $\{\lambda_n\}_{n>0}$ is a non-decreasing sequence such that: $1 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$
3. There exists a positive integer a such that: $\sum_{n=1}^{\infty} \lambda_n^{-a} < \infty$.

For any non-negative rational number $p \in \mathbb{Q}_+$, we define a sequence of norms $\{||\cdot||_p,\ p \in \mathbb{Q}_+\}$ such that: $||\xi||_p = ||A^p \xi||,\ \ \xi \in H$. That is:

$$||\xi||_p = \left(\sum_{n=1}^{\infty} \lambda_n^{2p} (\xi, h_n)^2 \right)^{1/2}, \tag{8.10}$$

where $(\cdot, \cdot)$ indicates the H inner product.

Remark 8.7. For any $p \in \mathbb{Q}_+$, the norm $||\cdot||_p$ is an Hilbert-norm. Indeed, it comes from the scalar product:

$$(\xi, \eta)_p = \sum_{n=1}^{\infty} \lambda_n^{2p} (\xi, h_n)(\eta, h_n). \tag{8.11}$$

For any $p \in \mathbb{Q}_+$ we define: $X_p = \{\xi \in H;\ ||\xi||_p < \infty\}$. In view of the above remark, X_p is an Hilbert space. Moreover, it is easy to see that for any $p \geq q \geq 0$:

$$X_p \subset X_q. \tag{8.12}$$

We have the following proposition:

Prop. 8.2. *For any $p \in \mathbb{Q}_+$, the inclusion map $X_{p+a/2} \hookrightarrow X_p$ is an Hilbert-Schmidt operator.*

Proof: we set $h_n^p = \dfrac{1}{\lambda_n^p} h_n$. The collection $\{h_n^p,\ n = 1, 2, \dots\}$ is an orthonormal bases of X_p. In fact, for any positive integers n and m:

$$(h_n^p, h_m^p)_p = \sum_{k=1}^{\infty} \lambda_k^p (h_n^p, h_k)(h_m^p, h_k) = \sum_{k=1}^{\infty} \frac{\lambda_k^{2p}}{\lambda_n^p \lambda_m^p} \delta_{nk} \delta_{mk} = \delta_{nm}.$$

3) An Hilbert-Schmidt operator is a bounded operator A, defined on an Hilbert space H, such that there exists an orthonormal basis $\{e_i\}_{i \in I}$ of H with the property $\sum_{i \in I} ||Ae_i||^2 < \infty$.

For each $\xi \in X_{p+a/2}$, we indicate with $i(\xi) = \xi \in X_p$ the inclusion map. Therefore, for any $n > 0$:

$$i(h_n^{p+a/2}) = h_n^{p+a/2} = \frac{1}{\lambda_n^{p+a/2}} \lambda_n^p h_n^p = \lambda^{-a/2} h_n^p,$$

and thus by hypothesis

$$\sum_{n=1}^{\infty} ||i(h_n^{p+a/2})||_p^2 = \sum_{n=1}^{\infty} \lambda_n^{-a} < \infty. \ \square$$

Consider the vector space $X = \bigcap_{p \in \mathbb{Q}_+} X_p$. In view of the above proposition X along with the family of Hilbert-norms $\{||\cdot||_p\,,\ p \in \mathbb{Q}_+\}$ is a nuclear space.

Let X be a vector space.

Def. 8.9 (Characteristic functional)**.** A continuous map $\Phi : X \to \mathbb{C}$ is called a characteristic functional on X if it's normalized:

$$\Phi(0) = 1,$$

and positive defined:

$$\sum_{i,j=1}^{m} \bar{c}_i \Phi(\xi_i - \xi_j) c_j \geq 0, \quad m \in \mathbb{Z}, \ \ \{c_i\}_{i=1,\dots,m} \in \mathbb{C}, \quad \{\xi_i\}_{i=1,\dots,m} \in X.$$

Example 8.3. For instance, consider the so called *generating functional* of a probability measure μ defined on $\mathbb{R}^n$, that is the Fourier transform:

$$\chi_\mu(\xi) = \int_{\mathbb{R}^n} e^{i(x,\xi)} d\mu(x), \quad \xi \in \mathbb{R}^n, \tag{8.13}$$

where (ξ, x) indicates the Euclidean scalar product. Then, χ_μ is a characteristic functional. In fact, it is continuous and normalized. Moreover, if we define

$$f(x) = \sum_{i=1}^{m} c_i e^{i(\xi_i, x)},$$

then we have:

$$\sum_{i,j=1}^{m} \bar{c}_i \chi(\xi_i - \xi_j) c_j = \int_{\mathbb{R}^n} |f(x)|^2 d\mu(x) = ||f||_\mu^2 \geq 0,$$

where $||\cdot||_\mu$ is the $L^2(\mathbb{R}^n, \mu)$ norm.

The finite dimensional Bochner theorem states that the converse is also true.

Theorem 8.1 (Bochner's theorem)**.** *For any characteristic functional χ on $\mathbb{R}^n$ there exists a* **unique** *probability measure μ on $\mathbb{R}^n$ such that χ is its generating functional. Namely,*

$$\chi(\xi) = \int_{\mathbb{R}^n} e^{i(x,\xi)} d\mu(x), \ \ \xi \in \mathbb{R}^n.$$

Example 8.4 (Standard Gaussian measure)**.** Consider the Euclidean $\mathbb{R}^n$ and define the function:

$$\chi(\xi) = e^{-\frac{1}{2}||\xi||^2}, \ \ \xi \in \mathbb{R}^n.$$

Obviously, $\chi : \mathbb{R}^n \to \mathbb{R}_+$ defines a characteristic functional. In fact, it is normalized, continuous and positive defined; to see the latter observe that:

$$\exp\big(-\frac{1}{2}||y_1 - y_2||^2\big) \geq \exp\big(-||y_1||^2 - ||y_2||^2\big);$$

then, for any $\{c_i\}_i \in \mathbb{C}$ and any $\{\xi_i\}_i \in \mathbb{R}^n$:

$$\sum_{i,j} \bar{c}_i \exp(-\frac{1}{2}||\xi_i - \xi_j||^2) c_j \geq \sum_i \bar{c}_i \exp(-||\xi_i||^2) \sum_j c_j \exp(-||\xi_j||^2) \geq 0.$$

By Bochner's theorem there is a unique probability measure μ_n on $(\mathbb{R}^n, \mathcal{B})$ such that:

$$\int_{\mathbb{R}^n} d\mu_n(x) e^{i(x,\xi)} = e^{-\frac{1}{2}||\xi||^2}.$$

This measure is called *Standard Gaussian measure*.

It's important to observe that such a measure does not exist in an infinite dimensional Hilbert space. In order to see this, let H be a real separable topological (infinite dimensional) Hilbert space and suppose that a Standard Gaussian measure μ exists. Let $\{h_n;\ n = 1, 2, \dots\}$ be an orthonormal basis for H. Then:

$$\int_H d\mu(x) e^{i(x,h_n)} = e^{-\frac{1}{2}}.$$

Note that for every $x \in H$ we must have $(x, h_n) \to 0$ as $n \to \infty$. Hence, using the Lebesgue dominated convergence theorem, we get a contradiction $1 = e^{-1/2}$. However, if the space is nuclear, there is an analogue result due to Minlos.

Let then X be a topological vector space. In the characterization of typical configurations of measures on infinite dimensional spaces the so called Minlos theorem plays a very important role. This theorem is an *infinite dimensional* generalization of the Bochner theorem and indeed requires the space X to be a nuclear space.

Theorem 8.2 (Minlos theorem)**.** *Let X be a nuclear space. For any characteristic functional Φ defined on X there exists a unique probability measure μ defined on the measurable space $(X', \mathcal{B})$, where $\mathcal{B}$ is regarded as the Borel σ-algebra generated by the weak topology on X', such that:*

$$\int_{X'} e^{i\langle\omega,\xi\rangle} d\mu(\omega) = \Phi(\xi), \quad \xi \in X. \tag{8.14}$$

Remark 8.8. We have considered the dual space X' equipped with the weak topology just for the seek of simplicity. In fact, because X is nuclear, the Borel σ-algebra generated by the three topology weak, strong and inductive, is the same (see Remark 8.6).

Characteristic functional on Hilbert spaces can be defined starting from completely monotonic functions[4]. In fact we have the following proposition:

Prop. 8.3. *Let F be a completely monotonic function defined on the positive real line. Therefore, there exists a unique characteristic functional Φ, defined on a real separable Hilbert space H, such that:*

$$\Phi(\xi) = F(||\xi||^2), \quad \xi \in H.$$

This is obvious because completely monotonic functions are associated to non-negative measure defined on the positive real line (see Feller [2]). The converse is also true (see Schneider [17,18]).

4) We remember that a function $F(t)$ is completely monotone if it is non-negative and possesses derivatives of any order such that: $(-1)^k \frac{d^k}{dt^k} F(t) \geq 0, \quad t > 0, \ k \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$.

8.1.3 White noise space

Consider the Schwartz space $\mathcal{S}(\mathbb{R})$. We remember that a complex valued function ξ on $\mathbb{R}^n$ is called *rapidly decreasing* if it is smooth and tends to 0 with all its derivatives faster than any power function. More precisely:

Def. 8.10 (Schwartz space)**.** The space $\mathcal{S}(\mathbb{R}^n)$ is the space of all the functions $\xi \in C^\infty(\mathbb{R}^n)$, such that for any multi-indices $\mathbf{j} = (j_1, j_2, \dots, j_n)$ and $\mathbf{k} = (k_1, k_2, \dots, k_n)$:

$$\sup_{x \in \mathbb{R}^n} |x^{\mathbf{j}} D^{\mathbf{k}} \xi(x)| < \infty. \tag{8.15}$$

Remark 8.9. Clearly a C_0^∞ class function is also rapidly decreasing; therefore:

$$C_0^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n).$$

Moreover, one can prove that:

$$\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n).$$

In fact, if $\xi \in \mathcal{S}$, for any $N \in \mathbb{Z}$ and multi-index $\mathbf{k}$, exists a constant $C_{N,\mathbf{k}}$ such that:

$$(1 + |x|^2)^N |D^{\mathbf{k}} \xi(x)| \leq C_{N,\mathbf{k}}.$$

Therefore,

$$|D^{\mathbf{k}} \xi(x)| \leq \frac{C_{N,\mathbf{k}}}{(1 + |x|^2)^N},$$

this implies that $\xi \in L^1$, with all its derivatives, if we chose $2N > n$. In the same way, we can see that $\xi \in L^p$ for any $1 \leq p \leq \infty$. Because the C_0^∞ functions are dense in L^p (with the L^p topology), we have that $\mathcal{S}$ is also dense in L^p.

We can define a convergence notion in the space $\mathcal{S}$:

Def. 8.11. A sequence of functions $\xi_i \in \mathcal{S}$, $i \in \mathbb{N}$, is said to be convergent to $\xi \in \mathcal{S}$ in $\mathcal{S}$, if and only if for any multi-indices $\mathbf{j} = (j_1, j_2, \dots, j_n)$ and $\mathbf{k} = (k_1, k_2, \dots, k_n)$:

$$\sup_{x \in \mathbb{R}^n} |x^{\mathbf{j}} D^{\mathbf{k}} \xi_i(x) - x^{\mathbf{j}} D^{\mathbf{k}} \xi(x)| \xrightarrow[i \to \infty]{} 0, \tag{8.16}$$

and we write:

$$\xi_i \xrightarrow[i \to \infty]{\mathcal{S}} \xi, \qquad \mathcal{S}\text{-}\lim_{i \to \infty} \xi_i = \xi.$$

The space of *tempered distributions* $\mathcal{S}'$ is defined as follows:

Def. 8.12. The set of linear and continuous functionals defined on $\mathcal{S}(\mathbb{R}^n)$ is called tempered distributions space $\mathcal{S}'(\mathbb{R}^n)$.

Remark 8.10. Then, a linear functional

$$T : \mathcal{S} \to \mathbb{C}$$

defines a tempered distribution if:

$$\xi_i \xrightarrow[i \to \infty]{\mathcal{S}} 0 \Longrightarrow \langle T, \xi_i \rangle \xrightarrow[i \to \infty]{} 0.$$

On the space $\mathcal{S}$ of test functions, we have continuous operators such as differential operators, Laplacian operators, translation operators, scaling operator, multiplication, convolution, Fourier transform. These operators could be extended by continuity to continuous operators in $\mathcal{S}'$ using the translation invariance of the Lebesgue measure. For example we could define elementary operations like the multiplication for a smooth function f and derivation:

- $\langle fT, \xi \rangle = \langle T, f\xi \rangle, \quad f \in C^\infty, \ \xi \in \mathcal{S}$
- $\langle D^{\mathbf{k}} T, \xi \rangle = (-1)^{|\mathbf{k}|} \langle T, D^{\mathbf{k}} \xi \rangle, \quad \xi \in \mathcal{S}, \ \mathbf{k} = (k_1, k_2, \ldots, k_n)$

The Fourier transform of a tempered distribution is also defined by continuity:

$$\langle \mathcal{F}T, \xi \rangle = \langle T, \mathcal{F}\xi \rangle, \quad \xi \in \mathcal{S}.$$

Moreover, identifying L^2 with $L^{2'}$, we have the following triple for the Schwartz distribution theory in $\mathbb{R}^n$:

$$\mathcal{S}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n), \tag{8.17}$$

and $\mathcal{S}$ is dense in L^2. Such a triple is often called a **Gel'fand triple**.

Let us consider the Schwartz space $\mathcal{S}(\mathbb{R})$ equipped with the usual scalar product:

$$(\xi, \eta) = \int_{\mathbb{R}} dt \xi(t) \eta(t), \quad \xi, \eta \in \mathcal{S}(\mathbb{R}). \tag{8.18}$$

We indicate the completion of $\mathcal{S}(\mathbb{R})$ with respect to (eq. 8.18) with $\mathcal{S}_0(\mathbb{R})$, which of course, because denseness, is equal to $L^2(\mathbb{R})$. We consider the orthonormal system $\{h_n\}_{n \geq 0}$ of the Hermite functions:

$$h_n(x) = \frac{1}{\sqrt{(2^n n! \sqrt{\pi})}} H_n(x) e^{-x^2/2}, \tag{8.19}$$

where $H_n(x) = (-1)^n e^{x^2} (d/dx)^n e^{-x^2}$ are the Hermite polynomials of degree n. Let A be the "harmonic oscillator" operator:

$$A = -\frac{d^2}{dx^2} + x^2 + 1; \tag{8.20}$$

A is densely defined on $\mathcal{S}_0(\mathbb{R})$ and the Hermite functions are eigenfunctions of A:

$$Ah_n = \lambda_n h_n = (2n + 2) h_n, \quad n = 0, 1, \ldots .$$

We observe that $1 \leq \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_n$ and $\sum_n \lambda_n^{-2} < \infty$. We are in the condition of Example (8.2) with $a = 2$. Therefore, for any non-negative integer p, we can define:

$$||\xi||_p = ||A^p \xi|| = \left(\sum_{n=0}^{\infty} (2n + 2)^{2p} (\xi, h_n)^2 \right)^{1/2},$$

where $|| \cdot ||$ indicates the L^2-norm. The Schwartz space $\mathcal{S}(\mathbb{R})$ could be then "reconstructed" as the *projective limit* of the Hilbert spaces $\mathcal{S}_p(\mathbb{R}) = \{\xi \in L^2(\mathbb{R}); \ ||\xi||_p < \infty\}$. That is:

$$\mathcal{S}(\mathbb{R}) = \bigcap_{p \geq 0} \mathcal{S}_p(\mathbb{R}). \tag{8.21}$$

Therefore, the topological Schwartz space, with the topology defined by the $||\cdot||_p$ norms, is a nuclear space. Since $\mathcal{S}(\mathbb{R})$ is a nuclear space, we can apply the Minlos theorem in order to define probability measures on its dual space $\mathcal{S}'(\mathbb{R})$.

Consider the positive function $F(t) = e^{-t}$, $t \geq 0$. It is obvious that F is a completely monotone function. Therefore, by (Prop. 9.1), the functional $\Phi(\xi) = F(||\xi||^2)$, $\xi \in L^2(\mathbb{R})$, defines a characteristic functional on $\mathcal{S}(\mathbb{R})$. By Minlos theorem, there exists a unique probability measure μ, defined on $(\mathcal{S}'(\mathbb{R}), \mathcal{B})$, such that:

$$\int_{\mathcal{S}'(\mathbb{R})} e^{i\langle\omega,\xi\rangle} d\mu(\omega) = e^{-||\xi||^2}, \quad \xi \in \mathcal{S}(\mathbb{R}). \tag{8.22}$$

The probability space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu)$ is called *white noise space* and the measure μ is called *white noise measure*, or standard Gaussian measure (compare with Example 8.4), on $\mathcal{S}'(\mathbb{R})$.

To understand the reason why it is called white noise space, consider the generalized stochastic process X, defined on the white noise space, such that for each test function $\varphi \in \mathcal{S}(\mathbb{R})$:

$$X(\varphi)(\cdot) = \langle \cdot, \varphi \rangle. \tag{8.23}$$

Clearly, for any $\varphi \in \mathcal{S}(\mathbb{R})$, $X(\varphi)$ is a Gaussian random variable with zero mean and variance $E(X(\varphi)^2) = 2||\varphi||^2$. Moreover, by linearity, for any $\varphi, \phi \in \mathcal{S}(\mathbb{R})$:

$$E(X(\varphi)X(\phi)) = 2(\varphi, \phi). \tag{8.24}$$

We refer to the generalized process X as the *canonical noise* of $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu)$. In view of the above properties, the process X is a white noise (Def. 8.1), and this motivates the name "white noise space" for the probability space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu)$.

Remark 8.11. Observe the choice we made for the white noise variance parameterization. Namely, the factor 2 in the right side of (eq. 8.22). This will be clear in the following.

We have the following:

Prop. 8.4. *For any $h \in L^2(\mathbb{R})$, $X(h)$ is defined almost everywhere on $\mathcal{S}'(\mathbb{R})$. Moreover, it is Gaussian with zero mean and variance $2||h||^2$.*

Proof: we indicate with $(L^2) = L^2(\mathcal{S}'(\mathbb{R}), \mu)$. Clearly, for any $\xi \in \mathcal{S}(\mathbb{R})$, we have that $X(\xi) \in (L^2)$ and:

$$||X(\xi)||^2_{(L^2)} = E(X(\xi)^2) = 2||\xi||^2_{L^2}. \tag{8.25}$$

For each $h \in L^2(\mathbb{R})$, there exists a sequence $\{\xi_n\}_{n\in\mathbb{N}}$ of $\mathcal{S}(\mathbb{R})$-elements which converges to h in the $L^2(\mathbb{R})$-norm. Therefore, from (eq. 8.25), the sequence $\{X(\xi_n)\}_{n\in\mathbb{N}}$ is Cauchy in (L^2) and converges to a limit function $X(h)$, defined on $\mathcal{S}'(\mathbb{R})$. □

The latter proposition states that for every sequence $\{f_t\}_{t\in\mathbb{R}}$ of $L^2(\mathbb{R})$-functions, depending continuously on a real parameter $t \in \mathbb{R}$, there exists a Gaussian stochastic process

$$\{Y(t)\}_{t\in\mathbb{R}} = \{X(f_t)\}_{t\in\mathbb{R}}, \tag{8.26}$$

defined on the probability space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu)$, which has zero mean, variance $E(Y_t)^2 = 2||f_t||^2$ and covariance $E(Y(t_1)Y(t_2)) = 2(f_{t_1}, f_{t_2})$.

Remark 8.12. We have already observed that, if $W(x)$, $x \in \mathbb{R}$, is a Wiener process defined on the probability space $(\Omega, \mathcal{F}, P)$, then the functional

$$X(\varphi) = \int \varphi(x) dW(x), \quad \varphi \in L^2(\mathbb{R}), \tag{8.27}$$

is a white noise on the space $(\Omega, \mathcal{F}, P)$. Therefore, if we indicate with $1_{[0,t)}(x)$, $t \geq 0$, the indicator function of the interval $[0, t)$, the process

$$X(1_{[0,t)}) = \int_0^t dW(x) = W(t), \quad t \geq 0, \tag{8.28}$$

is a one-sided Brownian motion.

Example 8.5 (Brownian motion)**.** Let X be a white noise defined canonically on the white noise space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu)$. Looking at (eq. 8.28), it is natural to state that the stochastic process

$$\{B(t)\}_{t\geq 0} = \{X(1_{[0,t)})\}_{t\geq 0} \tag{8.29}$$

is "standard" Brownian motion[5]. Indeed, the process $\{X(1_{[0,t)})\}_{t\geq 0}$, is Gaussian with covariance:

$$E\left[X(1_{[0,t)})X(1_{[0,s)})\right] = 2(1_{[0,t)}, 1_{[0,s)}) = 2\min(t,s), \quad t,s \geq 0.$$

Example 8.6 (Fractional Brownian motion)**.** The stochastic process:

$$\{B_{\alpha/2}(t)\}_{t\geq 0} = \{X(f_{\alpha,t})\}_{t\geq 0}, \quad 0 < \alpha < 2, \tag{8.30}$$

where

$$f_{\alpha,t}(x) = \frac{1}{C_1(\alpha)}\left((t-x)_+^{\frac{\alpha-1}{2}} - (-x)_+^{\frac{\alpha-1}{2}}\right), \quad x_+ = \max(x, 0), \tag{8.31}$$

and

$$C_1(\alpha) = \frac{\Gamma(\frac{\alpha+1}{2})}{\left(\Gamma(\alpha+1)\sin\frac{\pi\alpha}{2}\right)^{1/2}}, \tag{8.32}$$

is a "standard" fractional Brownian motion of order $H = \alpha/2$ (compare with eq. 5.1).

8.2 Grey noises

We have seen that white noise is a generalized stochastic process X defined canonically on the white noise space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu)$, with space of test functions $L^2(\mathbb{R})$. We have remarked that the white noise could also be defined starting from stochastic integrals with respect to the Brownian motion. In this case the space of test function turns out to be the space of integrands of the stochastic integral. Then, the Brownian motion $B(t)$ could be obtained from the white noise by setting $B(t) = X(1_{[0,t)})$. We want to generalize the previous construction in order to define a general class of H-sssi processes which includes, Brownian motion, fractional Brownian motion and more general processes.

5) With the word "standard" Brownian motion here we mean that $E(B(1)^2) = 2$.

Consider a one-sided fractional Brownian motion $\{B_{\alpha/2}(t)\}_{t\geq 0}$ with self-similarity parameter $H = \alpha/2$ and $0 < \alpha < 2$, defined on a certain probability space $(\Omega, \mathcal{F}, P)$. The fractional Brownian motion has a spectral representation (eq. 5.13):

$$B_{\alpha/2}(t) = \sqrt{C(\alpha)} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{e^{itx} - 1}{ix} |x|^{\frac{1-\alpha}{2}} d\widetilde{B}(x), \quad t \geq 0, \tag{8.33}$$

where as usual $d\widetilde{B}(x)$ is a complex Gaussian measure such that $d\widetilde{B}(x) = dB_1(x) + idB_2(x)$ with $dB_1(x) = dB_1(-x)$, $dB_2(x) = -dB_2(-x)$ and where B_1 and B_2 are independent Brownian motions. Moreover, by (eq. 5.14), one has:

$$C(\alpha) = \Gamma(\alpha + 1) \sin \frac{\pi\alpha}{2}. \tag{8.34}$$

We observe that

$$\frac{1}{\sqrt{2\pi}} \frac{e^{itx} - 1}{ix} = \widetilde{1}_{[0,t)}(x), \tag{8.35}$$

where we have indicated with $\widetilde{f}(x)$ the Fourier transform of the function f evaluated on $x \in \mathbb{R}$:

$$\widetilde{f}(x) = \mathcal{F}\{f(y), x\} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixy} f(y) dy. \tag{8.36}$$

In view of (eq. 8.35), we have:

$$B_{\alpha/2}(t) = \sqrt{C(\alpha)} \int_{\mathbb{R}} \widetilde{1}_{[0,t)}(x) |x|^{\frac{1-\alpha}{2}} d\widetilde{B}(x). \tag{8.37}$$

Therefore, if one defines a generalized stochastic process X such that, for a suitable choice of a test function φ,

$$X_\alpha(\varphi) = \sqrt{C(\alpha)} \int_{\mathbb{R}} \widetilde{\varphi}(x) |x|^{\frac{1-\alpha}{2}} d\widetilde{B}(x), \tag{8.38}$$

one can write:

$$B_{\alpha/2}(t) = X_\alpha(1_{[0,t)}), \quad t \geq 0. \tag{8.39}$$

Remark 8.13. The space of test function can be chosen as the space

$$\widetilde{\Lambda}_\alpha = \{f \in L^2(\mathbb{R}); \ ||f||_\alpha^2 = C(\alpha) \int_{\mathbb{R}} |\widetilde{f}(x)|^2 |x|^{1-\alpha} dx < \infty\}, \tag{8.40}$$

which coincides with a space of deterministic integrands for fractional Brownian motion (see Pipiras and Taqqu [14,15]).

Consider again the Schwartz space $\mathcal{S}(\mathbb{R})$ equipped now with the scalar product:

$$(\xi, \eta)_\alpha = C(\alpha) \int_{\mathbb{R}} \overline{\widetilde{\xi}(x)} \widetilde{\eta}(x) |x|^{1-\alpha} dx, \quad \xi, \eta \in \mathcal{S}(\mathbb{R}), \ 0 < \alpha < 2, \tag{8.41}$$

where $C(\alpha)$ is given by (eq. 8.34). This scalar product generate the α-norm in (eq. 8.40). We indicate with $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$ the completion of $\mathcal{S}(\mathbb{R})$ with respect to (eq. 8.41).

Remark 8.14. If we set $\alpha = 1$ in (eq. 8.41), we have $C(1) = 1$ and:

$$(\xi, \eta)_1 = \int_{\mathbb{R}} \overline{\widetilde{\xi}(x)} \widetilde{\eta}(x) dx = \int_{\mathbb{R}} \xi(y) \eta(y) dy, \tag{8.42}$$

so that, we recover the $L^2(\mathbb{R})$-inner product. Moreover, $\mathcal{S}_0^{(1)}(\mathbb{R}) = \mathcal{S}_0(\mathbb{R}) = L^2(\mathbb{R})$.

Starting from the Hilbert space $(\mathcal{S}_0^{(\alpha)}(\mathbb{R}), ||\cdot||_\alpha)$, it is possible to reproduce the construction of Example (8.2). Then, the space $\mathcal{S}(\mathbb{R})$ turns out to be a nuclear space with respect to the topology generated by the α-norm $||\cdot||_\alpha$ and an operator $A^{(\alpha)}$.

We need to find an orthonormal bases for the space $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$. For this purpose, we introduce the following definition:

Def. 8.13 (Generalized Laguerre polynomials)**.** The generalized *Laguerre polynomials* are defined, for any non-negative integer n, by:

$$L_n^\gamma(x) = \frac{x^{-\gamma}e^x}{\Gamma(n+1)}\frac{d^n}{dx^n}\left(e^{-x}x^{n+\gamma}\right), \quad \gamma > -1, \quad x \geq 0. \tag{8.43}$$

They are orthogonal respect to the weighting function $x^\gamma e^{-x}$:

$$\int_0^\infty x^\gamma e^{-x} L_n^\gamma(x) L_m^\gamma(x) dx = \frac{\Gamma(n+\gamma+1)}{\Gamma(n+1)}\delta_{nm}. \tag{8.44}$$

We have the following important result:

Prop. 8.5. *The generalized Laguerre polynomials satisfies the Laguerre equation:*

$$\left(x\frac{d^2}{dx^2} + (\gamma+1-x)\frac{d}{dx}\right) L_n^\gamma(x) = -nL^\gamma(x). \tag{8.45}$$

Define now a set of $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$ functions $\{h_n^\alpha\}_{n\in\mathbb{Z}_+}$ in terms of their Fourier transform by:

$$\begin{cases} \widetilde{h}_{2n}^\alpha(x) = a_{n,\alpha}e^{-x^2/2}L_n^{-\alpha/2}(x^2), & n \in \mathbb{Z}_+; \\ \widetilde{h}_{2n+1}^\alpha(x) = b_{n,\alpha}e^{-x^2/2}xL_n^{1-\alpha/2}(x^2), & n \in \mathbb{Z}_+. \end{cases} \tag{8.46}$$

Then, one has:

Prop. 8.6. *The set of functions $\{h_n^\alpha\}_{n\in\mathbb{Z}_+}$ is an orthonormal basis for $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$ with the choice:*

$$\begin{aligned} a_{\alpha,n} &= \left(\frac{\Gamma(n+1)}{C_\alpha\Gamma(n+1-\alpha/2)}\right)^{1/2}, \\ b_{\alpha,n} &= \left(\frac{\Gamma(n+1)}{C_\alpha\Gamma(n+2-\alpha/2)}\right)^{1/2}. \end{aligned} \tag{8.47}$$

Proof: clearly, by symmetry, for any $m, n \in \mathbb{Z}_+$:

$$(h_{2n}^\alpha, h_{2m+1}^\alpha)_\alpha = 0.$$

Now, by definition:

$$(h_{2n}^\alpha, h_{2m}^\alpha)_\alpha = 2a_{\alpha,n}a_{\alpha,m}C_\alpha \int_0^\infty dx x^{1-\alpha}e^{-x^2}L_n^{-\alpha/2}(x^2)L_m^{-\alpha/2}(x^2)$$

(with the variable change $x^2 = y$, and using (8.44))

$$= a_{\alpha,n}a_{\alpha,m}C_\alpha \int_0^\infty dy y^{-\alpha/2}e^{-y}L_n^{-\alpha/2}(y)L_m^{-\alpha/2}(y) = a_{\alpha,n}^2 C_\alpha \frac{\Gamma(n+1-\alpha/2)}{\Gamma(n+1)}\delta_{nm}$$

and the same for $b_{\alpha,n}$. □

Remark 8.15. Observe that as $\alpha \to 1$ we are able to recover the white noise space frame. In fact, $\mathcal{S}_0^\alpha(\mathbb{R})$ becomes $\mathcal{S}_0(\mathbb{R})$ (Remark 8.14) and the orthonormal base $\{h_n^\alpha\}_{n\in\mathbb{Z}_+}$ turns into the Hermite bases[6] (eq. 8.19) of $\mathcal{S}_0(\mathbb{R})$. In fact, we recall the well known relationships between Laguerre and Hermite polynomials:

$$\begin{cases} H_{2n}(x) = (-1)^n 2^{2n} n! L_n^{-1/2}(x^2) \\ H_{2n+1}(x) = (-1)^n 2^{2n+1} n! x L_n^{1/2}(x^2). \end{cases} \tag{8.48}$$

At this point, we need to find the analogue of the harmonic oscillator operator A (eq. 8.20). Starting from (eq. 8.45), we obtain that $L_n^{-\alpha/2}(x^2)$ satisfies the following equation:

$$\left[\frac{d^2}{dx^2} + \left(\frac{1-\alpha-2x^2}{x}\right)\frac{d}{dx}\right] L_n^{-\alpha/2}(x^2) = -4n L_n^{-\alpha/2}(x^2). \tag{8.49}$$

Moreover, if we indicate with $\mathcal{L}_n^{1-\alpha/2}(x^2) = xL_n^{1-\alpha/2}(x^2)$, we also have:

$$\left[\frac{d^2}{dx^2} + \left(\frac{1-\alpha-2x^2}{x}\right)\frac{d}{dx} + \left(\frac{\alpha-1}{x^2}\right)\right] \mathcal{L}_n^{1-\alpha/2}(x^2) = -(4n+2)\mathcal{L}_n^{1-\alpha/2}(x^2). \tag{8.50}$$

Remark 8.16. Again, we observe that as $\alpha = 1$ the equations (eq. 8.49) and (eq. 8.50), in view of (eq. 8.48), give the Hermite's equation

$$\left[\frac{d^2}{dx^2} - 2x\frac{d}{dx}\right] H_n(x) = -2nH_n(x) \tag{8.51}$$

for the correspondents Hermite's polynomials. In the first case, we obtain the equation for $H_{2n}(x)$, whereas in the second case, we obtain the equation for $H_{2n+1}(x)$.

Using the above equations (eq. 8.49) and (eq. 8.50), we can show the following

Prop. 8.7. *The functions $\{\widetilde{h}_n^\alpha\}_{n\in\mathbb{Z}_+}$ satisfies:*

$$\begin{cases} \left[-\frac{d^2}{dx^2} + \frac{\alpha-1}{x}\frac{d}{dx} + x^2\right] \widetilde{h}_{2n}^\alpha(x) := A_{\mathbf{even}}^\alpha \widetilde{h}_{2n}^\alpha = (4n+2-\alpha)\widetilde{h}_{2n}^\alpha(x); \\ \left[-\frac{d^2}{dx^2} + \frac{\alpha-1}{x}\frac{d}{dx} + x^2 - \frac{\alpha-1}{x^2}\right] \widetilde{h}_{2n+1}^\alpha(x) := A_{\mathbf{odd}}^\alpha \widetilde{h}_{2n+1}^\alpha = (4n+4-\alpha)\widetilde{h}_{2n+1}^\alpha(x). \end{cases} \tag{8.52}$$

Proof: we multiply equation (eq. 8.49) from the left by $a_{n,\alpha}e^{-x^2/2}$, then we find:

$$e^{-x^2/2}\left[\frac{d^2}{dx^2} + \left(\frac{1-\alpha-2x^2}{x}\right)\frac{d}{dx}\right] e^{x^2/2}\widetilde{h}_{2n}^\alpha(x) = -4n\widetilde{h}_{2n}^\alpha(x);$$

therefore calculating the derivatives we find the first of (eq. 8.52). The same for the second one, starting from (eq. 8.50). □

At this point we have the two non hermitian operators:

$$\begin{cases} A_{\mathbf{even}}^\alpha = \left[-\frac{d^2}{dx^2} + \frac{\alpha-1}{x}\frac{d}{dx} + x^2\right], \\ A_{\mathbf{odd}}^\alpha = \left[-\frac{d^2}{dx^2} + \frac{\alpha-1}{x}\frac{d}{dx} + x^2 - \frac{\alpha-1}{x^2}\right], \end{cases} \tag{8.53}$$

6) Observe that the Hermite functions are invariants under Fourier transform.

defined on $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$ and two orthogonal sets of eigenfunctions $\{h_{2n}^\alpha\}_{n\in\mathbb{Z}_+}$ and $\{h_{2n+1}^\alpha\}_{n\in\mathbb{Z}_+}$ such that:

$$\begin{cases} \mathcal{F}^{-1}A^\alpha_{\mathbf{even}}\mathcal{F}h^\alpha_{2n} := \widetilde{A}^\alpha_{\mathbf{even}}h^\alpha_{2n} = (4n+2-\alpha)h^\alpha_{2n}\,, \;\; n\in\mathbb{Z}_+ \\ \mathcal{F}^{-1}A^\alpha_{\mathbf{odd}}\mathcal{F}h^\alpha_{2n+1} := \widetilde{A}^\alpha_{\mathbf{odd}}h^\alpha_{2n} = (4n+4-\alpha)h^\alpha_{2n+1}\,, \;\; n\in\mathbb{Z}_+\,. \end{cases} \tag{8.54}$$

where $\mathcal{F}:\mathcal{S}_0^{(\alpha)}\to\mathcal{S}_0^{(\alpha)}$ is the Fourier transform. We write:

$$\mathcal{S}_0^{(\alpha)}(\mathbb{R}) = L\oplus L^\perp, \tag{8.55}$$

where L is the subspace generated by $\{h_{2n}^\alpha\}_{n\in\mathbb{Z}_+}$ and $L^\perp$ is its orthogonal subspace, **i.e.** the space generated by $\{h_{2n+1}^\alpha\}_{n\in\mathbb{Z}_+}$. Therefore, we can define the operator

$$A^{(\alpha)} = \overline{A}^\alpha_{\mathbf{even}}\oplus\overline{A}^\alpha_{\mathbf{odd}} + \mathbb{I}, \tag{8.56}$$

where $\overline{A}^\alpha_{\mathbf{even}}$ is the restriction of $\widetilde{A}^\alpha_{\mathbf{even}}$ to L, $\overline{A}^\alpha_{\mathbf{odd}}$ is the restriction of $\widetilde{A}^\alpha_{\mathbf{odd}}$ to $L^\perp$ and $\mathbb{I}$ is the identity operator. Then, we have found an operator $A^{(\alpha)}$, defined on $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$, which has the orthonormal bases $\{h_n^\alpha\}_{n\in\mathbb{Z}_+}$ as eigenfunctions, with eigenvalues $\{2n+2-\alpha+1\}_{n\in\mathbb{Z}_+}$. Namely,

$$A^{(\alpha)}h_n^\alpha = (2n+2-\alpha+1)h_n^\alpha,\;\; n=0,1,\ldots,\;\; 0<\alpha<2. \tag{8.57}$$

Remark 8.17. It's clear that when $\alpha=1$ the operator $A^{(1)}$ results to be defined over $\mathcal{S}_0(\mathbb{R})$ and it is equal to the Harmonic oscillator operator (eq. 8.20).

8.2.1 Generalized grey noise space

By Proposition (9.1), starting from a completely monotonic function F, we can define characteristic functionals on $\mathcal{S}(\mathbb{R})$ by setting $\Phi(\xi) = F(||\xi||^2_\alpha)$. Then, we can use Minlos theorem in order to define probability measures on $\mathcal{S}'(\mathbb{R})$. Because we are looking for stochastic processes in some way related to fractional calculus, we consider the Mittag-Leffler function of order $\beta>0$:

$$E_\beta(x) = \sum_{n=0}^{\infty}\frac{x^n}{\Gamma(\beta n+1)},\;\; x\in\mathbb{R}. \tag{8.58}$$

In fact, we have seen that such a function plays a fundamental role in fractional calculus. For our convenience we introduce:

$$F_\beta(x) = E_\beta(-x),\;\; x\in\mathbb{R}. \tag{8.59}$$

We have the following result:

Prop. 8.8. *The function $F_\beta(t)$ is completely monotonic on $\mathbb{R}_+$ for $0<\beta\le 1$.*

For $\beta=1$ this is obvious as:

$$F_1(t) = e^{-t}. \tag{8.60}$$

We do not prove this statement for $0<\beta<1$ (see Miller et al. [11] for details), but we just say that this implies by Bernstein theorem [2] that F_β is the Laplace transform of a probability measure on $\mathbb{R}_+$ (see Lemma 9.1). In fact, one finds

$$F_\beta(t) = \frac{1}{\beta}\int_0^\infty dx e^{-tx}x^{-(1+\beta^{-1})}f_\beta(x^{-\beta^{-1}}),$$

where f_β denotes the one-sided stable Lévy probability density.

Therefore, the functional $\Phi_{\alpha,\beta}(\xi) = F_\beta(||\xi||_\alpha^2)$, $\xi \in \mathcal{S}_0^{(\alpha)}(\mathbb{R})$, defines a characteristic functional on $\mathcal{S}(\mathbb{R})$. By Minlos theorem, there exists a unique probability measure $\mu_{\alpha,\beta}$, defined on $(\mathcal{S}'(\mathbb{R}), \mathcal{B})$, such that:

$$\int_{\mathcal{S}'(\mathbb{R})} e^{i\langle\omega,\xi\rangle} d\mu_{\alpha,\beta}(\omega) = F_\beta(||\xi||_\alpha^2), \quad \xi \in \mathcal{S}(\mathbb{R}). \tag{8.61}$$

When $\alpha = \beta$ and $0 < \beta \leq 1$, the probability space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_{\beta,\beta})$ is called *grey noise space* and the measure $\mu_{\beta,\beta}$ is called *grey noise measure* (see Schneider [17, 18]). We focus on the more general case $0 < \alpha < 2$ and we call the space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_{\alpha,\beta})$ "generalized" grey noise space and $\mu_{\alpha,\beta}$ "generalized" grey noise measure[7].

Def. 8.14. The generalized stochastic process $X_{\alpha,\beta}$, defined canonically on the "generalized" grey noise space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_{\alpha,\beta})$, is called "generalized" grey noise. Therefore, for each test function $\varphi \in \mathcal{S}(\mathbb{R})$:

$$X_{\alpha,\beta}(\varphi)(\cdot) = \langle \cdot, \varphi \rangle. \tag{8.62}$$

Remark 8.18. By the definition of "generalized" grey noise measure (eq. 8.61), for any $\varphi \in \mathcal{S}(\mathbb{R})$, we have:

$$E(e^{iyX_{\alpha,\beta}(\varphi)}) = E_\beta(-y^2||\varphi||_\alpha^2), \quad y \in \mathbb{R}. \tag{8.63}$$

Using (eq. 8.63) and (eq. 8.58) it easy to show that the "generalized" grey noise has moments of any order:

$$\begin{cases} E(X_{\alpha,\beta}(\xi)^{2n+1}) = 0, \\ E(X_{\alpha,\beta}(\xi)^{2n}) = \dfrac{2n!}{\Gamma(\beta n + 1)} ||\xi||_\alpha^{2n}, \end{cases} \tag{8.64}$$

for any integer $n \geq 0$ and $\xi \in \mathcal{S}(\mathbb{R})$. In fact, for any $y \in \mathbb{R}$

$$E(e^{iyX_{\alpha,\beta}(\xi)}) = \sum_{n=0}^{\infty} \frac{i^n y^n E(X_{\alpha,\beta}(\xi)^n)}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n} ||\xi||_\alpha^{2n}}{\Gamma(\beta n + 1)}.$$

Remark 8.19. As in the white noise case, it is possible to extend the space of test functions to the whole $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$. In fact, for any $\xi \in \mathcal{S}(\mathbb{R})$ we have $X_{\alpha,\beta}(\xi) \in (L^2) = L^2(\mathcal{S}'(\mathbb{R}), \mu_{\alpha,\beta})$. Thus, for any $h \in \mathcal{S}_0^{(\alpha)}(\mathbb{R})$, the function $X_{\alpha,\beta}(h)$ is defined as a limit of a sequence $X_{\alpha,\beta}(\xi_n)$, where $\{\xi_n\}$ belong to $\mathcal{S}(\mathbb{R})$.

Therefore, we have the following:

Prop. 8.9. *For any $h \in \mathcal{S}_0^{(\alpha)}(\mathbb{R})$, $X_{\alpha,\beta}(h)$ is defined almost everywhere on $\mathcal{S}'(\mathbb{R})$ and belongs to (L^2).*

Summarizing: the "generalized" grey noise is defined canonically on the grey noise space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_{\alpha,\beta})$ with the following properties:

1. For any $h \in \mathcal{S}_0^{(\alpha)}(\mathbb{R})$, $X_{\alpha,\beta}(h)$ is well defined and belong to (L^2).
2. $E(e^{iyX_{\alpha,\beta}(h)}) = E_\beta(-y^2||h||_\alpha^2)$ for any $y \in \mathbb{R}$.
3. $E(X_{\alpha,\beta}(h)) = 0$ and $E(X_{\alpha,\beta}(h)^2) = \dfrac{2}{\Gamma(\beta+1)} ||h||_\alpha^2$.

7) The word "generalized" is written in double quotes because it could be confused with the word *generalized* stochastic processes. Its meaning here is not mathematical, but just a generalization of a previous concept.

4. For any h and g which belong to $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$, by linearity, it is easy to show that:

$$E\left(X_{\alpha,\beta}(h)X_{\alpha,\beta}(g)\right) = \frac{1}{\Gamma(\beta+1)}\left[(h,g)_\alpha + \overline{(h,g)}_\alpha\right]. \tag{8.65}$$

8.2.2 Generalized grey Brownian motion

If we put $\beta = 1$ in (eq. 8.63), the measure $\mu_{\alpha,1} := \mu_\alpha$ is a Gaussian measure and $X_{\alpha,1} := X_\alpha$ is a Gaussian noise. In fact, for any $h \in \mathcal{S}_0^{(\alpha)}(\mathbb{R})$, the random variable $X_\alpha(h)$ is Gaussian with zero mean and variance $E(X_\alpha(h)^2) = 2||h||_\alpha^2$ (see eq. 8.63). Moreover, for any sequence $\{f_t\}_{t\in\mathbb{R}}$ of $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$-functions, depending continuously on a real parameter $t \in \mathbb{R}$, the stochastic process $Y(t) = X_\alpha(f_t)$ is Gaussian with auto-covariance given by (eq. 8.65) with $\beta = 1$,

$$E(Y(t)Y(s)) = E(X_\alpha)(f_t)X_\alpha(f_s) = (f_t, f_s)_\alpha + \overline{(f_t, f_s)}_\alpha. \tag{8.66}$$

Finally, when $\alpha = 1$, X_α reduces to a "standard" white noise.

Example 8.7 (Fractional Brownian motion)**.** For any $t \geq 0$ the function $1_{[0,t)}$ belongs to $\mathcal{S}_0^{(\alpha)}(\mathbb{R})$. In fact, one can show that $||1_{[0,t)}||_\alpha^2 < \infty$ when $0 < \alpha < 2$ and

$$||1_{[0,t)}||_\alpha^2 = \frac{C(\alpha)}{2\pi}\int_{\mathbb{R}} dx \frac{2}{|x|^{1+\alpha}}(1-\cos tx) = t^\alpha. \tag{8.67}$$

In fact, because

$$\widetilde{1}_{[0,t)}(x) = \frac{1}{\sqrt{2\pi}}\frac{e^{ixt}-1}{ix}, \tag{8.68}$$

by the definition of the $||\cdot||_\alpha$-norm (eq. 8.41), one indeed finds

$$C(\alpha)\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{2}{|x|^{1+\alpha}}(1-\cos tx)dx = C(\alpha)\frac{t^\alpha}{2\pi}\int_{-\infty}^{+\infty}\frac{2}{|x|^{1+\alpha}}(1-\cos x)dx, \tag{8.69}$$

which is equal, in distributional sense, to:

$$C(\alpha)\frac{t^\alpha}{\pi}\int_{-\infty}^{+\infty}\mathcal{F}\left(\frac{1}{|x|^{\alpha+1}}\right)(\omega)\mathcal{F}\left(1-\cos x\right)(\omega)d\omega.$$

We have:

$$\mathcal{F}\left(1-\cos x\right)(\omega) = \sqrt{2\pi}\left(\delta(\omega) - \frac{1}{2}(\delta(\omega-1)+\delta(\omega+1))\right).$$

Moreover (see Gel'fand [3]):

$$\mathcal{F}\left(\frac{1}{|x|^{\alpha+1}}\right)(\omega) = -\sqrt{\frac{\pi}{2}}\frac{|\omega|^\alpha}{\Gamma(\alpha+1)\sin\frac{\pi\alpha}{2}}.$$

Thus, by looking (eq. 8.34), one finds (eq. 8.67).

Therefore, for any $0 < \alpha < 2$, we can define the process:

$$B_{\alpha/2}(t) = X_\alpha(1_{[0,t)}), \quad t \geq 0. \tag{8.70}$$

The process $B_{\alpha/2}(t)$ is a "standard" fractional Brownian motion with parameter $H = \alpha/2$. Indeed, it is Gaussian with variance $E(B_{\alpha/2}(t)^2) = 2||1_{[0,t)}||_\alpha^2 = 2t^\alpha$ and auto-covariance:

$$E(B_{\alpha/2}(t)B_{\alpha/2}(s)) = (1_{[0,t)}, 1_{[0,s)})_\alpha + \overline{(1_{[0,t)}, 1_{[0,s)})}_\alpha$$

$$= \frac{C(\alpha)}{2\pi}\int_{\mathbb{R}} dx \frac{2}{|x|^{\alpha+1}}\left(1 - \cos tx + 1 - \cos sx - 1 + \cos(t-s)x\right)$$

$$= t^\alpha + s^\alpha - |t-s|^\alpha = \gamma_\alpha(t,s), \quad t,s \geq 0,$$

which is the fractional Brownian motion auto-covariance (eq. 3.36 with $\sigma^2 = 2$ and $H = \alpha/2$).

In view of the above example, X_α could be regarded as a *fractional Gaussian noise* defined on the space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_\alpha)$.

Example 8.8 (Deconvolution of Brownian motion)**.** The stochastic process

$$\{B(t)\}_{t\geq 0} = \{X_\alpha(g_{\alpha,t})\}_{t\geq 0}, \tag{8.71}$$

where, for each $t \geq 0$, the function $g_{\alpha,t}$ is defined by:

$$\widetilde{g}_{\alpha,t}(x) = \frac{1}{\sqrt{C(\alpha)}}\widetilde{1}_{[0,t)}(x)(ix)^{\frac{\alpha-1}{2}}, \tag{8.72}$$

is a "standard" Brownian motion. Indeed, it is Gaussian, with zero mean, variance

$$E(B(t)^2) = 2\int_{\mathbb{R}} |x|^{1-\alpha}|\widetilde{1}_{[0,t)}(x)|^2|x|^{\alpha-1}dx = 2\int_{\mathbb{R}} |\widetilde{1}_{[0,t)}(x)|^2 dx = 2t, \tag{8.73}$$

and autocovariance:

$$E(B(t)B(s)) = \int_{\mathbb{R}} \left(\overline{\widetilde{1}_{[0,t)}}(x)\widetilde{1}_{[0,s)}(x) + \overline{\widetilde{1}_{[0,s)}}(x)\widetilde{1}_{[0,t)}(x)\right) dx = 2\min(t,s). \tag{8.74}$$

Remark 8.20. The representation of Brownian motion in terms of the fractional Gaussian noise (eq. 8.71) corresponds to a particular case of the so called *deconvolution formula*, which expresses the Brownian motion as a stochastic integral with respect to a fractional Brownian motion of order $H = \alpha/2$ (see [16]). More generally, we can represent a fractional Brownian motion $B_{\gamma/2}(t)$ of order $H = \gamma/2$, $0 < \gamma < 2$ in terms of a fractional Gaussian noise of order α, which corresponds to a representation of $B_{\gamma/2}$ in terms of a stochastic integral of a fractional Brownian motion $B_{\alpha/2}$ of order $H = \alpha/2$, $0 < \alpha < 2$ (see example below).

Example 8.9 (Deconvolution of fractional Brownian motion)**.** The stochastic process,

$$\{B_{\gamma/2}(t)\}_{t\geq 0} = \{X_\alpha(g_{\alpha,\gamma,t})\}_{t\geq 0}, \tag{8.75}$$

where:

$$\widetilde{g}_{\alpha,\gamma,t}(x) = \sqrt{\frac{C(\gamma)}{C(\alpha)}}\widetilde{1}_{[0,t)}(x)(ix)^{\frac{\alpha-\gamma}{2}}, \quad 0 < \gamma < 2, \tag{8.76}$$

is a "standard" fractional Brownian motion of order $H = \gamma/2$.

We consider now the general case $0 < \alpha < 2$, $0 < \beta \leq 1$.

Def. 8.15. The stochastic process

$$\{B_{\alpha,\beta}(t)\}_{t\geq 0} = \{X_{\alpha,\beta}(1_{[0,t)})\}_{t\geq 0}, \tag{8.77}$$

is called "generalized" (standard) grey Brownian motion.

The "generalized" grey Brownian motion $B_{\alpha,\beta}$ has the following properties which come directly from the grey noise properties and (eq. 8.67):

1. $B_{\alpha,\beta}(0) = 0$ almost surely. Moreover, for each $t \geq 0$, $E(B_{\alpha,\beta}(t)) = 0$ and

$$E(B_{\alpha,\beta}(t)^2) = \frac{2}{\Gamma(\beta+1)} t^{\alpha}. \tag{8.78}$$

2. The auto-covariance function is:

$$E(B_{\alpha,\beta}(t)B_{\alpha,\beta}(s)) = \gamma_{\alpha,\beta}(t,s) = \frac{1}{\Gamma(\beta+1)} \left(t^{\alpha} + s^{\alpha} - |t-s|^{\alpha}\right). \tag{8.79}$$

3. For any $t, s \geq 0$, the characteristic function of the increments is:

$$E\left(e^{iy(B_{\alpha,\beta}(t)-B_{\alpha,\beta}(s))}\right) = E_{\beta}(-y^2|t-s|^{\alpha}), \quad y \in \mathbb{R}. \tag{8.80}$$

The third property follows from the linearity of the grey noise definition. In fact, suppose $0 \leq s < t$, we have $y(B_{\alpha,\beta}(t) - B_{\alpha,\beta}(s)) = yX_{\alpha,\beta}(1_{[0,t)} - 1_{[0,s)}) = X_{\alpha,\beta}(y1_{[s,t)})$, and $||y1_{[s,t)}||_{\alpha}^2 = y^2(t-s)^{\alpha}$. All these properties are enclosed in the following:

Prop. 8.10. *For any $0 < \alpha < 2$ and $0 < \beta \leq 1$, the process $B_{\alpha,\beta}(t)$, $t \geq 0$, is a self-similar with stationary increments process (H-sssi), with $H = \alpha/2$.*

Proof: This result is actually a consequence of the linearity of the noise definition. Given a sequence of real numbers $\{\theta_j\}_{j=1,\ldots,n}$, we have to show that for any $0 < t_1 < t_2 < \cdots < t_n$ and $a > 0$:

$$E\left(\exp(i\sum_j \theta_j B_{\alpha,\beta}(at_j))\right) = E\left(\exp(i\sum_j \theta_j a^{\frac{\alpha}{2}} B_{\alpha,\beta}(t_j))\right).$$

The linearity of the grey noise definition allows to write the above equality as:

$$E\left[\exp\left(iX_{\alpha,\beta}(\sum_j \theta_j 1_{[0,at_j)})\right)\right] = E\left[\exp\left(iX_{\alpha,\beta}(a^{\frac{\alpha}{2}}\sum_j \theta_j 1_{[0,t_j)})\right)\right].$$

Using (eq. 8.63) we have

$$F_{\beta}\left(||\sum_j \theta_j 1_{[0,at_j)}||_{\alpha}^2\right) = F_{\beta}\left(||a^{\frac{\alpha}{2}}\sum_j \theta_j 1_{[0,t_j)}||_{\alpha}^2\right)$$

which, because the complete monotonicity, reduces to

$$||\sum_j \theta_j 1_{[0,at_j)}||_{\alpha}^2 = a^{\alpha}||\sum_j \theta_j 1_{[0,t_j)}||_{\alpha}^2.$$

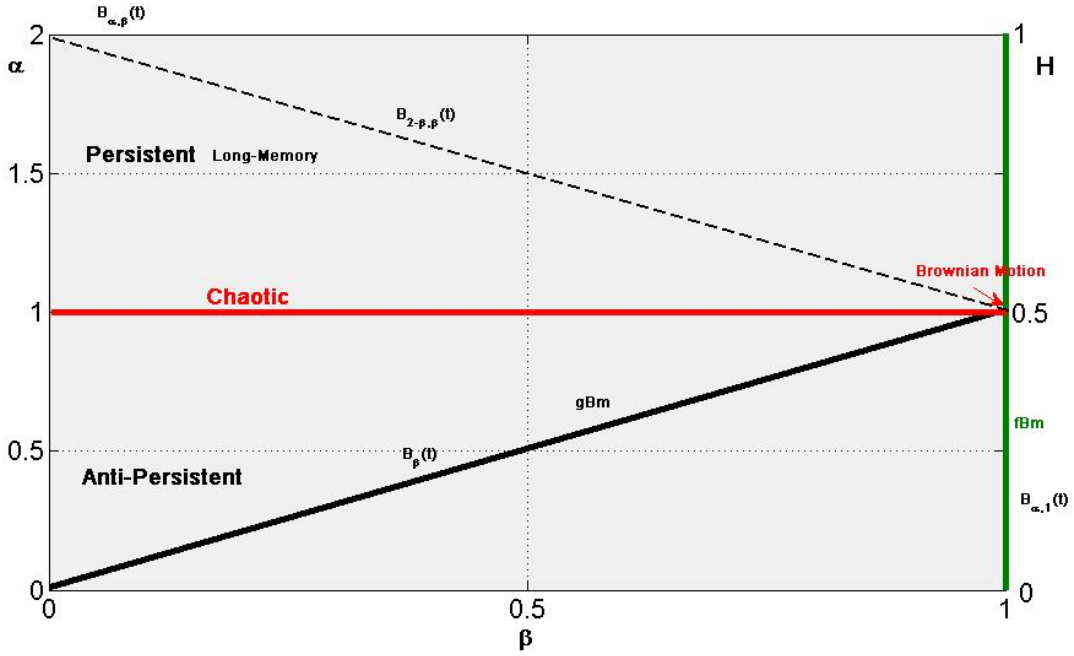

Figure 8.1: Parametric class of generalized grey Brownian motion. The upper diagonal line indicates the "conjugated" process of grey Brownian motion.

In view of the definition (eq. 8.41) and (eq. 8.35), the above equality is checked after a simple change of variable in the integration. In the same way we can prove the stationarity of the increments. We have to show that for any $h \in \mathbb{R}$:

$$E\left[\exp\left(i\sum_j \theta_j(B_{\alpha,\beta}(t_j+h)-B_{\alpha,\beta}(h))\right)\right] = E\left[\exp\left(i\sum_j \theta_j(B_{\alpha,\beta}(t_j)\right)\right].$$

We use the linearity property to write:

$$E\left[\exp\left(iX_{\alpha,\beta}(\sum_j \theta_j 1_{[h,t_j+h)})\right)\right] = E\left[\exp\left(iX_{\alpha,\beta}(\sum_j \theta_j 1_{[0,t_j)})\right)\right].$$

By using the definition and the complete monotonicity, we have:

$$||\sum_j \theta_j 1_{[h,t_j+h)}||_\alpha^2 = ||\sum_j \theta_j 1_{[0,t_j)}||_\alpha^2$$

which is true because:

$$\widetilde{1}_{[h,t_j+h)}(x) = \frac{1}{\sqrt{2\pi}}\frac{e^{ixh}}{ix}\left(e^{ixt_j}-1\right). \quad \square$$

Remark 8.21. In view of Proposition 8.10, $\{B_{\alpha,\beta}(t)\}$ forms a class of H-sssi stochastic processes indexed by two parameters $0 < \alpha < 2$ and $0 < \beta \leq 1$. This class includes fractional Brownian motion ($\beta = 1$),

grey Brownian motion ($\alpha = \beta$) and Brownian motion ($\alpha = \beta = 1$). In Figure 8.1 we present a diagram which allows us to identify the elements of the class. The long-range dependence domain corresponds to the region $1 < \alpha < 2$. The horizontal line represents the processes with purely random increments, that is, processes which possess uncorrelated increments. The fractional Brownian motion is identified by the vertical line ($\beta = 1$). The lower diagonal line represents the grey Brownian motion.

8.2.3 The ggBm master equation

The following proposition characterizes the marginal density function of the process $\{B_{\alpha,\beta}(t),\ t \geq 0\}$:

Prop. 8.11. *The marginal probability density function $f_{\alpha,\beta}(x,t)$ of the process $\{B_{\alpha,\beta}(t),\ t \geq 0\}$ is the fundamental solution of the "stretched" time-fractional diffusion equation:*

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)} \int_0^t \frac{\alpha}{\beta} s^{\alpha/\beta - 1} \left(t^{\frac{\alpha}{\beta}} - s^{\frac{\alpha}{\beta}} \right)^{\beta - 1} \frac{\partial^2}{\partial x^2} u(x,s) ds, \quad t \geq 0. \tag{8.81}$$

Proof: (eq. 8.80) (with $s = 0$) states that $\widetilde{f}_{\alpha,\beta}(y,t) = E_\beta(-y^2 t^\alpha)$. Using (eq. 8.58), we can show that the Mittag-Leffler function satisfies

$$E_\beta(-y^2 (t^{\frac{\alpha}{\beta}})^\beta) = 1 - \frac{y^2}{\Gamma(\beta)} \int_0^{t^{\frac{\alpha}{\beta}}} ds' (t^{\frac{\alpha}{\beta}} - s')^{\beta - 1} E_\beta(-y^2 s'^\beta)$$

$$= 1 - \frac{y^2}{\Gamma(\beta)} \int_0^t \frac{\alpha}{\beta} s^{\alpha/\beta - 1} (t^{\frac{\alpha}{\beta}} - s^{\frac{\alpha}{\beta}})^{\beta - 1} E_\beta(-y^2 s^\alpha) ds,$$

where we have used the change of variables $s' = s^{\alpha/\beta}$. Thus, $f_{\alpha,\beta}(x,t)$ solves (eq. 8.81) with initial condition $u_0(x) = f_{\alpha,\beta}(x,0) = \delta(x)$. □

Remark 8.22. Observe that, if $f(x,t)$ solves the time fractional diffusion equation of order β (eq. 7.134), namely:

$$f(x,t) = f_0(x) + \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta - 1} \partial_{xx} f(x,s) ds,$$

then, the function $u(x,t) = f(x, t^{\alpha/\beta})$, solves (eq. 8.81), with the same initial condition. Therefore, the fundamental solution of (eq. 8.81) is just (see eq. 7.140):

$$u(x,t) = \mathcal{M}_\beta(x, t^{\alpha/\beta}) = \frac{t^{-\alpha/2}}{2} M_{\beta/2}(|x| t^{-\alpha/2}). \tag{8.82}$$

We refer to (eq. 8.81) as the master equation of the marginal density function of the "generalized" grey Brownian motion. Therefore, the diagram in Figure (8.1) can be also read in terms of partial integro-differential equation of fractional type. When $\alpha = \beta$ and $0 < \beta \leq 1$, we recover the time-fractional diffusion equation (eq. 7.134) of order β (lower diagonal line). When $\beta = 1$ and $0 < \alpha < 2$, we have the equation of the fractional Brownian motion marginal density (eq. 7.12), that is the equation of a stretched Gaussian density (vertical line). Finally, when $\alpha = \beta = 1$ we find the standard diffusion equation.

From Proposition (8.11), it follows that the parametric class $\{B_{\alpha,\beta}(t)\}$ provides stochastic models for anomalous diffusions described by (eq. 8.81). Looking at (eq. 8.78) and (eq. 8.79), which describe the variance and the covariance function respectively, it follows that:

When $0 < \alpha < 1$, the diffusion is slow. The increments of the process $B_{\alpha,\beta}(t)$ turn out to be negatively correlated. This implies that the trajectories are very "zigzaging" (antipersistent). The increments form a stationary process which does not exhibit long-range dependence.

When $\alpha = 1$, the diffusion is normal. The increments of the process are uncorrelated. The trajectories are said to be "chaotic".

When $1 < \alpha < 2$, the diffusion is fast. The increments of the process $B_{\alpha,\beta}(t)$ are positively correlated. So that, the trajectories are more regular (persistent). In this case the increments exhibits long-range dependence (see Prop. 6.1).

The stochastic processes considered so far, governed by the master equation (8.81), are of course Non-Markovian. We have already observed (see Remark 7.37) that non-Markovian equations like (eq. 8.81) are often associated to subordinated stochastic processes $D(t) = B(l(t))$, where the parent Markov process $B(t)$ is a "standard" Brownian motion and the random time process $l(t)$ is a self-similar of order $H = \beta$ non-negative non-decreasing non-Markovian process. For example, in Kolsrud [8] the random time $l(t)$ is taken to be related to the *local time* of a $d = 2(1-\beta)$-dimesional fractional Bessel process, while in Meerschaert et al. [10] (see also Gorenflo et al. [4] and Stanislavsky [20]), in the context of Continuous Time Random Walk, it is taken as the *inverse process* of the totally skewed strictly β-stable process (see Chapter 1, Rem. 1.13).

Remark 8.23. Heuristically, our stochastic process $\{B_{\alpha,\beta}(t),\ t \geq 0\}$ cannot be a subordinated process (for example if $\beta = 1$ it reduces to a fractional Brownian motion). Therefore, here we provided an example of a class of stochastic models associated to time-fractional diffusion equations like (eq. 8.81), which are not subordinated processes.

It is important to remark that, starting from a master equation which describes the dynamic evolution of a probability density function $f(x,t)$, it is always possible to define an equivalence class of stochastic processes with the same marginal density function $f(x,t)$. All these processes provide suitable stochastic models for the starting equation. In this paper we focused on a subclass $\{B_{\alpha,\beta}(t),\ t \geq 0\}$ associated to the non-Markovian equation (eq. 8.81). This subclass is made up of processes with *stationary increments*. In this case, the memory effects are enclosed in the typical dependence structure of a H-sssi process (eq. 8.79), while, for instance in the case of a subordinated process, these are due to the subordination process and to the non-Markovian property of the *random time process* itself.

Remark 8.24. It is also interesting to observe that the "generalized" grey Brownian motion turns out to be a direct generalization of a Gaussian process. Indeed, it includes the fractional Brownian motion as particular case when $\beta = 1$. Moreover, for any sequence of real numbers $\{\theta_i\}_{i=1,\dots,n}$, if one considers the collection $\{B_{\alpha,\beta}(t_1),\dots,B_{\alpha,\beta}(t_n)\}$ with $0 < t_1 < t_2 < \cdots < t_n$, it is easy to show that:

$$E\left(\exp\Big(i\sum_{j=1}^{n}\theta_j B_{\alpha,\beta}(t_j)\Big)\right) = E\left(\exp\left(iX_{\alpha,\beta}\Big(\sum_{j=1}^{n}\theta_j 1_{[0,t_j)}\Big)\right)\right)$$

$$= E_\beta\left(-\Big|\Big|\sum_j \theta_j 1_{[0,t_j)}\Big|\Big|_\alpha^2\right) = E_\beta\left(-\Gamma(\beta+1)\frac{1}{2}\sum_{i,j}\theta_i\theta_j\gamma_{\alpha,\beta}(t_i,t_j)\right), \tag{8.83}$$

where $\gamma_{\alpha,\beta}$ is the the autocovariance matrix (eq. 8.79). It is clear that, fixed β, the "generalized" grey Brownian motion is defined only by its covariance structure. In other words, $B_{\alpha,\beta}(t)$ provides an example of a stochastic process characterized only by the first and second moments, which is a property of Gaussian processes. We will use this remark in the next section.

8.3 Characterization of the ggBm

We have seen that the generalized grey Brownian motion (ggBm), is made up off self-similar with stationary increments processes (Prop. 8.10) and depends on two real parameters $\alpha \in (0,2)$ and $\beta \in (0,1]$.

It includes fractional Brownian motion when $\alpha \in (0,2)$ and $\beta = 1$, and time-fractional diffusion stochastic processes when $\alpha = \beta \in (0,1)$. The latters have marginal probability density function governed by time-fractional diffusion equations of order β (see equations 7.134, 7.135 and 7.136).

The ggBm is defined through the explicit construction of the underline probability space. However, we are now going to show that it is possible to define it in an unspecified probability space. For this purpose, we write down explicitly all the finite dimensional probability density functions. Moreover, we shall provide different ggBm characterizations.

The role of the M-Wright function (eq. 7.122), which is related to the fundamental solution of the time-fractional diffusion equation (eq. 7.140), emerges as a natural generalization of the Gaussian distribution (eq. 3.3).

Furthermore, we will show that ggBm can be represented in terms of the product of a random variable and an independent fractional Brownian motion. This representation highlights the H-**sssi** nature of the ggBm and provides a way to study and simulate the trajectories.

For this purpose, we developed a random walk model based on the Grunwald-Letnikov finite difference approximation (eq. 7.4) of the time fractional drift equation (eq. 7.110).

Then, we want to characterize the ggBm through its finite dimensional structure. From (eq. 8.83), we know that all the ggBm finite dimensional probability density functions are characterized only by its autocovriance matrix $\gamma_{\alpha,\beta}$ (eq. 8.79). The following proposition holds:

Prop. 8.12. *Let $B_{\alpha,\beta}$ be a ggBm, then for any collection $\{B_{\alpha,\beta}(t_1), \dots, B_{\alpha,\beta}(t_n)\}$, the joint probability density function is given by:*

$$f_{\alpha,\beta}(x_1, x_2, \dots, x_n; \gamma_{\alpha,\beta}) = \frac{(2\pi)^{-\frac{n-1}{2}}}{\sqrt{2\Gamma(1+\beta)^n \det \gamma_{\alpha,\beta}}} \int_0^\infty \frac{1}{\tau^{n/2}} M_{1/2}\left(\frac{\xi}{\tau^{1/2}}\right) M_\beta(\tau) d\tau, \tag{8.84}$$

with:

$$\xi = \left(2\Gamma(1+\beta)^{-1} \sum_{i,j=1}^{n} x_i {\gamma_{\alpha,\beta}}^{-1}(t_i, t_j) x_j\right)^{1/2}, \quad \gamma_{\alpha,\beta}(t_i, t_j) = \frac{1}{\Gamma(1+\beta)}(t_i^\alpha + t_j^\alpha - |t_i - t_j|^\alpha), \quad i, j = 1, \dots, n.$$

Proof: in order to show (eq. 8.84), we calculate its n-dimensional Fourier transform and we find that it is equal to (eq. 8.83). We have

$$\int_{\mathbb{R}^n} \exp\left(i \sum_{j=1}^{n} \theta_j x_j\right) f_{\alpha,\beta}(x_1, \dots, x_n; \gamma_{\alpha,\beta}) d^n x$$

$$= \frac{(2\pi)^{-\frac{n-1}{2}}}{\sqrt{2\Gamma(1+\beta)^n \det \gamma_{\alpha,\beta}}} \int_0^\infty \frac{1}{\tau^{n/2}} M_\beta(\tau) \int_{\mathbb{R}^n} \exp\left(i \sum_{j=1}^{n} \theta_j x_j\right) M_{1/2}\left(\frac{\xi}{\tau^{1/2}}\right) d^n x d\tau.$$

We remember that $M_{1/2}(r) = \frac{1}{\sqrt{\pi}} e^{-r^2/4}$, thus we get:

$$\int_0^\infty \frac{1}{\tau^{n/2}} M_\beta(\tau) \int_{\mathbb{R}^n} \exp\left(i \sum_{j=1}^{n} \theta_j x_j\right) \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\Gamma(1+\beta)^n \det \gamma_{\alpha,\beta}}} \exp\left(-\Gamma(1+\beta)^{-1} \sum_{i,j=1}^{n} x_i \gamma_{\alpha,\beta}^{-1}(t_i, t_j) x_j / 2\tau\right) d^n x d\tau. \tag{8.85}$$

We make the change of variables $\mathbf{x} = \Gamma(1+\beta)^{1/2}\tau^{1/2}\mathbf{y}$, with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and we get

$$\int_0^\infty M_\beta(\tau) \int_{\mathbb{R}^n} \exp\left(i\Gamma(1+\beta)^{1/2}\tau^{1/2}\sum_{j=1}^n \theta_j y_j\right) \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det \gamma_{\alpha,\beta}}} \exp\left(-\sum_{i,j=1}^n y_i \gamma_{\alpha,\beta}^{-1}(t_i,t_j)y_j/2\right) d^n y d\tau$$

$$= \int_0^\infty M_\beta(\tau) \exp\left(-\Gamma(1+\beta)\tau \sum_{i,j=1}^n \theta_i\theta_j\gamma_\alpha(t_i,t_j)/2\right) d\tau = \int_0^\infty e^{-\tau s} M_\beta(\tau) d\tau = E_\beta(-s),$$

where $s = \Gamma(1+\beta)\sum_{i,j=1}^n \theta_i\theta_j\gamma_{\alpha,\beta}(t_i,t_j)/2$ and where we have used (eq. 7.124). □

Using the Kolmogorov extension theorem (see Theorem 3.1), the above proposition allows us to define the ggBm in an unspecified probability space. In fact, given a probability space $(\Omega, \mathcal{F}, P)$, the following proposition characterizes the ggBm:

Prop. 8.13. *Let $X(t)$, $t \geq 0$, be a stochastic process, defined in a certain probability space $(\Omega, \mathcal{F}, P)$, such that*

i) $X(t)$ has covariance matrix indicated by $\gamma_{\alpha,\beta}$ and finite-dimensional distributions defined by (eq. 8.84).

ii) $EX^2(t) = \frac{2}{\Gamma(1+\beta)}t^\alpha$ for $0 < \beta \leq 1$ and $0 < \alpha < 2$.

iii) $X(t)$ has stationary increments,

then $X(t)$, $t \geq 0$, is a generalized grey Brownian motion.

In fact condition ii) together with condition iii) imply that $\gamma_{\alpha,\beta}$ must be the ggBm autocovariance matrix (eq. 8.79).

Remark 8.25. Using (eq. 7.127), for $n = 1$, (eq. 8.84) reduces to:

$$f_{\alpha,\beta}(x,t) = \frac{1}{\sqrt{4t^\alpha}} \int_0^\infty \mathcal{M}_{1/2}\left(|x|t^{-\alpha/2}, \tau\right) \mathcal{M}_\beta(\tau, 1) d\tau = \frac{1}{2}t^{-\alpha/2} M_{\beta/2}(|x|t^{-\alpha/2}). \tag{8.86}$$

This means that the ggBm marginal density function is indeed the fundamental solution (eq. 8.82) of (eq. 8.81).

Remark 8.26. Moreover, because:

$$M_1(\tau) = \delta(\tau - 1),$$

for $\beta = 1$, putting $\gamma_{\alpha,1} \equiv \gamma_\alpha$, we have that (eq. 8.84) reduces to the Gaussian distribution of the fractional Brownian motion. That is,

$$f_{\alpha,1}(x_1, x_2, \ldots, x_n; \gamma_{\alpha,1}) = \frac{(2\pi)^{-\frac{n-1}{2}}}{\sqrt{2\det\gamma_\alpha}} M_{1/2}\left(\left(2\sum_{i,j=1}^n x_i \gamma_\alpha^{-1}(t_i,t_j)x_j\right)^{1/2}\right).$$

We have the following corollary.

Cor. 8.1. *Let $X(t)$, $t \geq 0$, be a stochastic process defined in a certain probability space $(\Omega, \mathcal{F}, P)$. Let $H = \alpha/2$ with $0 < \alpha < 2$ and suppose that $EX(1)^2 = 2/\Gamma(1+\beta)$. The following statements are equivalent:*

*i) X is H-**sssi** with finite-dimensional distribution defined by (eq. 8.84);*

ii) X is a generalized grey Brownian motion with scaling exponent $\alpha/2$ and "fractional order" parameter β;

iii) X has zero mean, covariance function $\gamma_{\alpha,\beta}(t,s)$, $t, s \geq 0$, defined by (eq. 8.79) and finite dimensional distribution defined by (eq. 8.84).

8.3.1 Representation of ggBm

Up to now, we have seen that the ggBm $B_{\alpha,\beta}(t)$, $t \geq 0$, is an H-**sssi** process, which generalizes Gaussian processes (it is indeed Gaussian when $\beta = 1$) and is defined only by its autocovariance structure. These properties make us think that $B_{\alpha,\beta}(t)$ may be equivalent to a process $\Lambda_\beta X_\alpha(t)$, $t \geq 0$, where $X_\alpha(t)$ is a Gaussian process and Λ_β is a suitable chosen independent random variable. Indeed, the following proposition holds:

Prop. 8.14. *Let $B_{\alpha,\beta}(t)$, $t \geq 0$, be a ggBm, then:*

$$B_{\alpha,\beta}(t) \stackrel{d}{=} \sqrt{L_\beta} X_\alpha(t), \quad t \geq 0, \quad 0 < \beta \leq 1, \quad 0 < \alpha < 2, \tag{8.87}$$

where $\stackrel{d}{=}$ denotes as usual the equality of the finite dimensional distribution, $X_\alpha(t)$ is a standard fBm and L_β is an independent non-negative random variable with probability density function $M_\beta(\tau)$, $\tau \geq 0$.

In fact, after some manipulation, (eq. 8.85) can be written:

$$\int_{\mathbb{R}^n} \int_0^\infty \exp\left(i \sum_{j=1}^n \theta_j y x_j \right) 2y M_\beta(y^2) \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det \gamma_\alpha}} \exp\left(-\sum_{i,j=1}^n x_i \gamma_\alpha^{-1}(t_i, t_j) x_j / 2 \right) dy d^n x.$$

$$= E\left(\exp(i \sum_{j=1}^n \theta_j \sqrt{L_\beta} X_\alpha(t_j)) \right).$$

Remark 8.27. Proposition (8.14) highlights the H-**sssi** nature of the ggBm. Moreover, for $\beta = 1$ from (eq. 7.125) follows that $L_1 = 1$ a.s., thus we recover the fractional Brownian motion of order $H = \alpha/2$.

The representation (8.87) is particularly interesting. In fact, a number of question, in particularly those related to the distribution properties of $B_{\alpha,\beta}(t)$, can be reduced to question concerning the fBm $X_\alpha(t)$, which are easier since $X_\alpha(t)$ is a Gaussian process. For instance, the Hölder continuity of the $B_{\alpha,\beta}(t)$ trajectories follows immediately from those of $X_\alpha(t)$:

$$E(|X_\alpha(t) - X_\alpha(s)|^p) = c_p |t - s|^{p\alpha/2}.$$

Moreover, this factorization is indeed suitable for path-simulation (see next section).

Remark 8.28. From (eq. 8.87), it is clear that Brownian motion, that is $B_{1,1}(t)$, $t \geq 0$, is the only one process of the ggBm class which has independent increments.

8.4 ggBm path simulation

In the previous section we have shown that the ggBm could be represented by a process:

$$B_{\alpha,\beta}(t) = \sqrt{L_\beta} X_\alpha(t), \quad t \geq 0,$$

where L_β is suitable chosen random variable independent of $X_\alpha(t)$. Clearly, in order to simulate ggBm trajectories, we first need a method to generate the random variable L_β.

For this purpose, we consider the time-fractional forward drift equation (eq. 7.110), which we rewrite here for the seek of clearness:

$$u(x,t) = u_0(x) - \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} \frac{\partial}{\partial x} u(x,s) ds, \quad x \in \mathbb{R}, \quad t \geq 0, \quad 0 < \beta \leq 1. \tag{8.88}$$

Remark 8.29. We remember that the term drift equation refers to the fact that when $\beta = 1$ (eq. 8.88) turns out to be the simple one dimensional (forward) drift equation $\partial_t u(x,t) = -\partial_x u(x,t)$ (eq. 9.4), whose fundamental solution is $\delta(x-t)$.

We observe that we could have taken $x \geq 0$ in (eq. 8.88) as in (eq. 7.110). In fact, the fundamental solution of (eq. 8.88), is just (eq. 7.121):

$$u(x,t) = \mathcal{M}_\beta(x,t), \quad x,t \geq 0, \tag{8.89}$$

and always zero for $x < 0$. This function can be interpreted as the marginal density function of a non-negative self-similar stochastic process with scaling parameter $H = \beta$.

Remark 8.30. When $\beta = 1$ we recover $\mathcal{M}_1(x,t) = \delta(x-t)$ (see (eq. 7.125)).

We write (eq. 8.88) in terms of the Caputo-Dzherbashyan fractional derivative of order β (see eq. 7.113). Then, equation (8.88) is equivalent to:

$$\begin{cases} {}_*\mathcal{D}_t^\beta u(x,t) = -\partial_x u(x,t), \\ u(x,0) = u_0(x) = \delta(x), \end{cases} \tag{8.90}$$

with $x \in \mathbb{R}$, $t \geq 0$ and $0 < \beta \leq 1$. Using a random walk model, one can simulate a discrete time random process $X(t)$, $t \geq 0$, governed by the time-fractional forward drift equation (eq. 8.90) (see [6,7]). In this way, for each running up to time $t = 1$, the random variable $X(1)$ has the required distribution $u(x,1) = M_\beta(x)$. The random walk construction follows two steps:

- using the Grünwald-Letnikov discretization of Caputo-Dzherbashyan derivative (eq. 7.34),
- interpretation of the corresponding finite difference scheme as a random walk scheme.

8.4.1 Finite difference schemes

We begin defining the finite difference model. First, the Cauchy problem (eq. 8.90) has to be written in a finite domain:

$$\begin{cases} {}_*\mathcal{D}_t^\beta u(x,t) = -\dfrac{\partial}{\partial x} u(x,t), \quad (x,t) \in \Omega = [-a,a] \times [0,1], \ a > 0, \\ u(x,0) = u_0(x) = \delta(x), \\ u(-a,t) = \Phi_1(t), \quad u(a,t) = \Phi_2(t), \quad t > 0. \end{cases} \tag{8.91}$$

Let N, M be positive integers. Then, we introduce a bi-dimensional lattice

$$\mathcal{G}_{\delta x,\delta t}^{2M,N} = \{(j\delta x, n\delta t), (j,n) \in \mathbb{Z}_{2M} \times \mathbb{Z}_N\},$$

contained on Ω, with $\delta x = \frac{2a}{2M-1}$ and $\delta t = \frac{1}{N-1}$. The lattice elements are indicated with

$$(x_j, t_n) = (j\delta x, n\delta t), \quad j = 0,1,\ldots,2M-1 \ , n = 0,1,\ldots N-1.$$

Let $u : \Omega \to \mathbb{R}$ be a function defined on Ω. We indicate with $u_j^n = u(x_j,t_n)$ the restriction of u to $\mathcal{G}_{\delta x,\delta t}^{2M,N}$ evaluated in (x_j,t_n).

The time-fractional forward drift equation is then replaced by the finite difference equation:

$$ {}_*D_t^\beta u_j^n = -\frac{u_j^n - u_{j-1}^n}{\delta x}, \tag{8.92} $$

where

$$ {}_*D_t^\beta u_j^n = \sum_{k=0}^{n+1} (-1)^k \binom{\beta}{k} \frac{u_j^{n+1-k} - u_j^0}{\delta t^\beta}, \tag{8.93} $$

with $u_j^0 = u_0(j\delta x)$, is the (forward) Grünwald-Letnikov scheme for the Caputo-Dzherbashyan derivative (eq. 7.34). Using the "empty sum" convention:

$$ \sum_{k=p}^{q} \cdot = 0, \quad \text{if } q < p, $$

for any $n \geq 0$, we obtain the explicit equation:

$$ u_j^{n+1} = u_j^0 \sum_{k=0}^{n} (-1)^k \binom{\beta}{k} + \sum_{k=1}^{n} (-1)^{k+1} \binom{\beta}{k} u_j^{n+1-k} + \mu(u_{j-1}^n - u_j^n), \tag{8.94} $$

where $\mu = \delta t^\beta / \delta x$. Equation (8.94) can be written in the following noteworthy form:

$$ u_j^{n+1} = b_n u_j^0 + \sum_{k=1}^{n} c_k u_j^{n+1-k} + \mu(u_{j-1}^n - u_j^n), \tag{8.95} $$

where we have defined:

$$ \begin{cases} c_k = (-1)^{k+1} \binom{\beta}{k}, \quad k \geq 1, \\ b_n = \sum_{k=0}^{n} (-1)^k \binom{\beta}{k}, \quad n \geq 0. \end{cases} \tag{8.96} $$

More precisely, the explicit scheme reads

$$ \begin{cases} u_0^n = \Phi_1(t_n), \quad u_{2M-1}^n = \Phi_2(t_n), \quad n > 0, \\ u_j^1 = (1 - \mu + \mu L) u_j^0, \quad 0 < j < 2M - 1 \\ u_j^{n+1} = (c_1 - \mu + \mu L) u_j^n + c_2 u_j^{n-1} + \cdots + c_n u_j^1 + b_n u_j^0, \quad n > 1, \quad 0 < j < 2M - 1, \end{cases} \tag{8.97} $$

where L is the "lowering" operator: $Lf_j = f_{j-1}$.

Remark 8.31 (Stability)**.** We observe that the explicit scheme is stable if $\mu \leq \beta$, i.e.

$$ \delta x \geq \delta t^\beta / \beta. \tag{8.98} $$

Indeed, this assures that the scheme preserves non-negativity.

Remark 8.32. When $\beta = 1$ all the coefficients c_k and b_n vanish except $b_0 = c_1 = 1$. So that, we recover the finite difference approximation of the simple (forward) drift equation.

It is not difficult to write an implicit scheme. In fact, we just have to use backward approximations for the time-fractional derivative. Therefore, for any $n \geq 0$, we obtain:

$$u_j^{n+1} + \mu(u_j^{n+1} - u_{j-1}^{n+1}) = b_n u_j^0 + \sum_{k=1}^{n} c_k u_j^{n+1-k}. \tag{8.99}$$

Namely,

$$\begin{cases} u_0^n = \Phi_1(t_n), \;\; u_{2M-1}^n = \Phi_2(t_n), \;\; n > 0 \\ (1 + \mu - \mu L) u_j^1 = u_j^0, \quad 0 < j < 2M - 1, \\ (1 + \mu - \mu L) u_j^{n+1} = c_1 u_j^n + c_2 u_j^{n-1} + \cdots + c_n u_j^1 + b_n u_j^0, \;\; n > 0, \quad 0 < j < 2M - 1. \end{cases} \tag{8.100}$$

The above equation can be rewritten in matrix notation:

$$\begin{cases} \Lambda_{ij} u_j^1 = u_i^0 + \psi_i^1, \\ \Lambda_{ij} u_j^{n+1} = c_1 u_i^n + c_2 u_i^{n-1} + \cdots + c_n u_i^1 + b_n u_i^0 + \psi_i^{n+1}, \;\; n \geq 1. \end{cases} \tag{8.101}$$

Λ is the following $2M \times 2M$ matrix, divided in four $M \times M$ blocks:

$$\Lambda = \left(\begin{array}{c|c} \Lambda_1 & 0 \\ \hline \Lambda_2 & A \end{array}\right) = \left(\begin{array}{ccccc|ccccc} 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\ -\mu & 1+\mu & 0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & -\mu & 1+\mu & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\mu & 1+\mu & 0 & 0 & \cdots & 0 & 0 \\ \hline 0 & 0 & 0 & \cdots & -\mu & 1+\mu & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\mu & 1+\mu & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & -\mu & 1+\mu & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -\mu & 1 \end{array}\right) \tag{8.102}$$

Moreover, for any $n \geq 0$, ψ^n is a suitable vector which takes into account of the boundary terms.

Remark 8.33. Because Λ is lower diagonal, it is easy to show that

$$\Lambda^{-1} = \left(\begin{array}{c|c} \Lambda_1^{-1} & 0 \\ \hline -A^{-1}\Lambda_2\Lambda_1^{-1} & A^{-1} \end{array}\right). \tag{8.103}$$

8.4.2 Random walk models

We observe that, for $0 < \beta < 1$:

$$\sum_{k=1}^{\infty} c_k = 1, \;\; 1 > \beta = c_1 > c_2 > \cdots \to 0. \tag{8.104}$$

In fact,

$$\sum_{k=1}^{\infty} c_k = 1 - \sum_{k=0}^{\infty} (-1)^k \binom{\beta}{k} = 1 - \lim_{x \to 1_-} \sum_{k=0}^{\infty} (-1)^k \binom{\beta}{k} x^k = 1 - \lim_{x \to 1_-} (1-x)^\beta = 1.$$

Moreover,

$$\begin{cases} b_0 = 1 = \sum_{k=1}^{\infty} c_k, \;\; b_m = 1 - \sum_{k=1}^{m} c_k = \sum_{k=m+1}^{\infty} c_k, \\ 1 = b_0 > b_1 > b_2 > \cdots \to 0. \end{cases} \tag{8.105}$$

Thus, the coefficients c_k and b_n form a sequence of positive numbers, which do not exceed unity and decrease strictly monotonically to zero.

8.4.2.1 Explicit random walk

In order to build a random walk model, we consider for first the explicit scheme. Omitting the boundary terms, we have:

$$\begin{cases} u_j^1 = (1-\mu)u_j^0 + \mu u_{j-1}^0, \\ u_j^{n+1} = (c_1 - \mu)u_j^n + \mu u_{j-1}^n + c_2 u_j^{n-1} + \cdots + c_n u_j^1 + b_n u_j^0, \;\; n \geq 1. \end{cases}$$

We consider a walker which starts in zero at time zero, namely $x(t=0) = 0$. We interpret the u_j^n as the probability of sojourn in $x_j = j\delta x$ at time $t_n = n\delta t$. Then, we indicate with $x(t_n)$ the position of the particle at time t_n.

At time $t_1 = \delta t$ the walker could be at the position $x(1) = x_1$ with probability μ (that is the probability to come from one space-step behind) or in $x(1) = x_0 = 0$ with probability $1 - \mu$ (that is the probability to stay in the starting position).

The position at time t_{n+1} is determined as follows: from (eq. 8.104) and (eq. 8.105), it is clear that the parameters $c_1, c_2, \ldots c_n, b_n$ can be interpreted as probabilities. Then, we define a partition of events $\{E_{c_1}, E_{c_2}, \ldots, E_{c_n}, E_{b_n}\}$, with $P(E_{c_k}) = c_k$ and $P(E_{b_n}) = b_n$ with $n \geq 1$ and such that:

- $E_{c_1} = \Big\{$*the particle starts in the previous position* $x(t_n)$ *and jumps in* $x(t_n) + \delta x$ *with probability* μ *or stays in* $x(t_n)$ *with probability* $1 - \mu\Big\}$.
- $E_{c_k} = \Big\{$*the particle backs to the position* $x(t_{n+1-k})\Big\}$.
- $E_{b_n} = \Big\{$*the particle backs to the initial position* $x(t_0)\Big\}$.

8.4.2.2 Implicit random walk

Consider the implicit case:

$$\begin{cases} u^1 = \Lambda^{-1} u^0, \\ u^{n+1} = \Lambda^{-1}\left[c_1 u^n + c_2 u^{n-1} + \cdots + c_n u^1 + b_n u^0\right], \end{cases}$$

where Λ is given by eq. (8.103). The probability interpretation of the parameters c_k and b_n is still valid. In this case, however, we must use the *transpose matrix* P of A^{-1} to define the transition probabilities. Indeed, the $M \times M$ matrix A^{-1} propagates in the positive semi-axis. Moreover, from (eq. 9.6), it can be shown that P defines a transition matrix (all the elements are positive numbers less than one and all the rows sum to one, see example below).

Example 8.10. For instance, in the four dimensional case the matrix P would be:

$$P = \begin{pmatrix} 1/(1+\mu) & \mu/(1+\mu)^2 & \mu^2/(1+\mu)^3 & \mu^3/(1+\mu)^3 \\ 0 & 1/(1+\mu) & \mu/(1+\mu)^2 & \mu^2/(1+\mu)^2 \\ 0 & 0 & 1/(1+\mu) & \mu/(1+\mu) \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In the implicit case the random walk is defined as follows: let the particle start in zero. At the first step it could jump up to $M-1$ steps ahead with probabilities defined by the first P row. Then, we have the following partition of events:

- $E_{c_1} = \Big\{$*the particle starts in the previous position* $x(t_n)$ *for instance* x_j*. Then, it could jump up to* $M-j-1$ *steps ahead with probabilities defined by the* $j+1$*-th* P *row*$\Big\}$;
- $E_{c_k} = \Big\{$*the particle backs to the position* $x(t_{n+1-k})$ *for instance* x_j*. Then, it could jump up to* $M-j-1$ *steps ahead with probabilities defined by the* $j+1$*-th* P *row*$\Big\}$;
- $E_{b_n} = \Big\{$*the particle backs to the initial position* $x(t_0)$*. Then, it could jump up to* $M-1$ *steps ahead with probabilities defined by the first* P *row*$\Big\}$.

Remark 8.34. The implicit method is slower than the explicit one. However, we observe that, because of the stability constraint (eq. 8.98), the implicit scheme is advisable for small β. Indeed, we have $\delta x \sim \delta t^{\beta}$, $0 < \beta < 1$ and we are forced to raise the time steps a lot, in order to improve "spatial" resolution.

8.4.3 ggBm trajectories

Along this section, we have provided a method to generate the random variable L_β and the simulation results are shown in Figures 7.2-7.4. In particular

- Figure 8.2 shows a random-walk simulation with $\beta = 0.4$. In this case we used an implicit random-walk scheme. Moreover, we compared the histogram evaluated over $N = 10000$ simulations and the density function $M_\beta(x)$, $x \geq 0$.
- Figure 8.3 shows a random walk when $\beta = 0.5$. Because $M_{1/2}(x) = \frac{1}{\sqrt{\pi}} e^{-x^2/4}$, $x \geq 0$, in this case $L_\beta \stackrel{d}{=} |Z|$ where Z is a Gaussian random variable.
- Figure 8.4 shows a random walk with $\beta = 0.8$. We used an explicit random-walk scheme. Then, we compared the histogram with the corresponding probability density function $M_\beta(x)$, $x \geq 0$.

In all the studied cases, we have found a good agreement between the histograms and the theoretical density functions.

At this point, in order to obtain examples of the $B_{\alpha,\beta}(t) = \sqrt{L_\beta} X_\alpha(t)$ trajectories, we just have to simulate the fractional Brownian motion $X_\alpha(t)$. For this purpose we have used an exact Cholesky method (see [1]).

Path simulations of $B_{\beta,\beta}(t)$ (shortly B_β) and $B_{2-\beta,\beta}(t)$, with $\beta = 1/2$ are shown in Figures 7.5-7.10. The first process provides an example of stochastic model for slow-diffusion (short-memory), the second provides a stochastic model for fast-diffusion (long-memory).

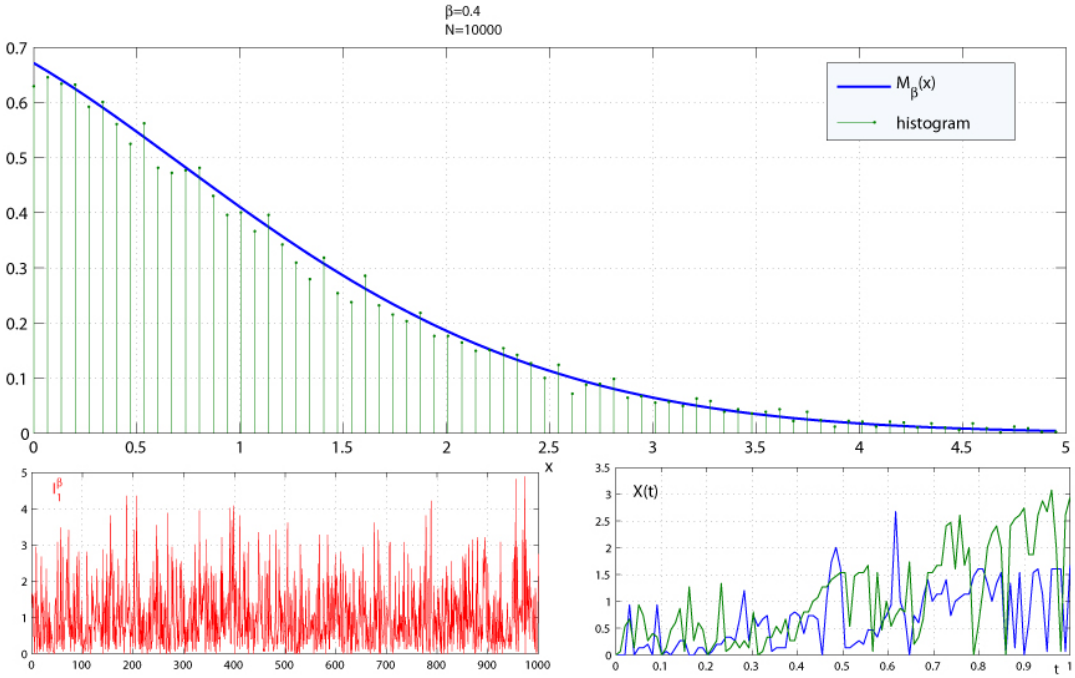

Figure 8.2: In the top panel, the histogram of L_β, which is calculated from a sample of $N = 10000$ outcomes, is obtained with an implicit random-walk scheme and it is compared with the exact PDF $M_\beta(x)$, $x \geq 0$, with $\beta = 0.4$. In the bottom panels, the random variable L_β (left) and two trajectory examples (right) are shown.

- Figure 8.5 shows some typical paths. In the bottom panel, we present the corresponding increment process. Namely,

$$Z_\beta(t_k) = B_\beta(t_k) - B_\beta(t_{k-1}), \quad t_k = k\delta t, \; k = 1, 2, \ldots, M-1.$$

- Figure 8.6 shows the agreement between simulations and the theoretical densities at times $t = 1$ and $t = 2$.
- Figure 8.7 presents the plot of the sample variance in logarithmic scale. Moreover, we evaluated a linear fitting, which shows a good agreement with the theoretical result.
- Figure 8.8 shows some typical paths for the long-memory process $B_{2-\beta,\beta}(t)$.
- Figure 8.9 collects the histograms in the case $\alpha = 2 - \beta$ at time $t = 1$ and time $t = 2$. The super-diffusive behavior (that is the rapid increasing of the variance in time) is highlighted.
- Figure 8.10 presents the plot of the sample variance in logarithmic scale. Even in this case, we evaluated a linear fitting.

8.5 Concluding remarks

The marginal probability density function (eq. 8.86) of the generalized grey Brownian motion $B_{\alpha,\beta}(t)$, $t \geq 0$, evolves in time according to a "stretched" time-fractional diffusion equation of order β (see

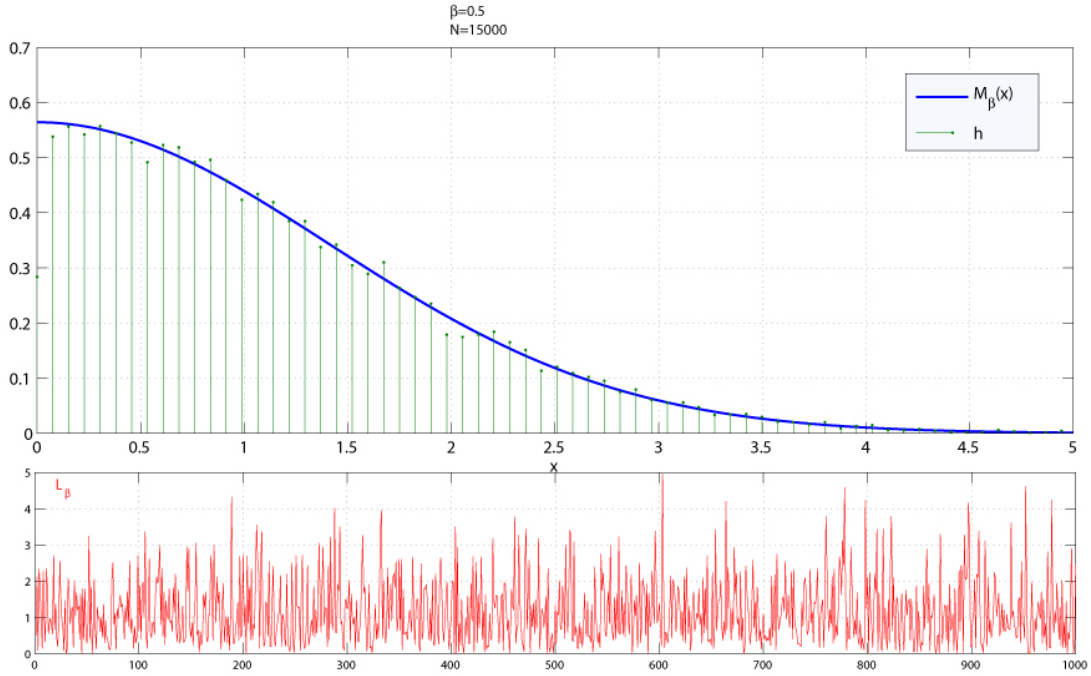

Figure 8.3: In the case $\beta = 0.5$, the histogram (top panel) is calculated from a sample of $N = 15000$ outcomes, which are obtained simulating independent Gaussian random variables. The corresponding L_β is shown in the bottom.

eq. 8.81). Therefore, the ggBm serves as stochastic model for the anomalous diffusion described by these class of fractional equations.

The ggBm is defined canonically (see eq. 8.77) in the so called grey noise space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_{\alpha,\beta})$, where the grey noise measure satisfies (eq. 8.61). However, the ggBm is an H-**sssi** process of order $H = \alpha$ and Proposition (8.13) provides a characterization of $B_{\alpha,\beta}(t)$ notwithstanding the underline probability space.

There are many other processes which serve as stochastic models for a given master equation. In fact, given a master equation for a PDF $f(x,t)$, it is always possible to define an equivalence class of stochastic processes with the same marginal density function $f(x,t)$. The ggBm defines a subclass $\{B_{\alpha,\beta}(t),\ t \geq 0\}$ associated to the non-Markovian equation (8.81). In this case, the memory effects are enclosed in the typical dependence structure of a H-**sssi** process. While, for instance in the case of a subordinated process, these are due to the memory properties of the *random time process*. The latter are preferable because provide a ready-made physical interpretation (see Remark 9.3). However, $B_{\alpha,\beta}(t)$ is interesting because of the stationarity of its increments.

Proposition (8.14) provides an enlighten representation of ggBm. Thus, the generalized grey Brownian motion turns out to be merely a *fractional Brownian motion with stochastic variance*. That is,

$$B_{\alpha,\beta}(t) = \Lambda_\beta X_\alpha(t), \quad t \geq 0,$$

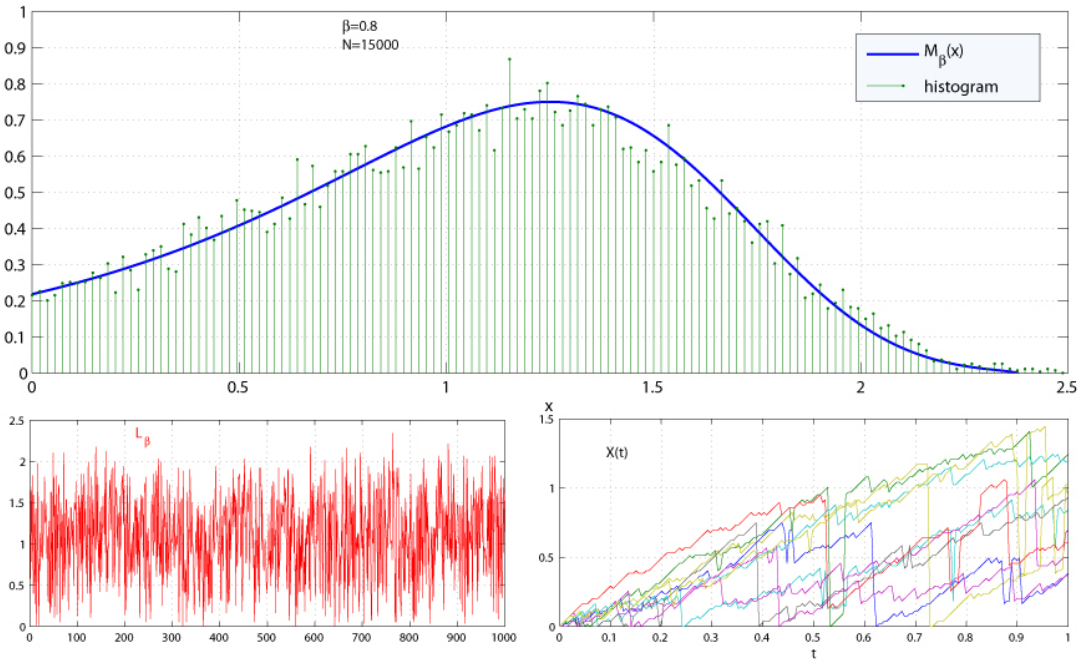

Figure 8.4: In the top panel, the histogram of L_β, which is calculated from a sample of $N = 15000$ outcomes, it is obtained with an explicit random-walk scheme and it is compared with the exact PDF $M_\beta(x)$, $x \geq 0$, with $\beta = 0.8$. In the bottom panels, the random variable L_β (left) and many trajectory examples (right) are shown.

where Λ_β is a suitable independent random variable (see eq. 8.87).

As a final remark, we observe that such a process is not ergodic. In fact, heuristically, the ergodicity of the system is broken by the multiplication with the random variable Λ_β. This appears also from the simulated trajectories. Indeed, it is impossible with a single realization of the system $B_{\alpha,\beta}(t,\omega)$, $\omega \in \Omega$, to distinguish a ggBm from a fBm with variance $2\Lambda_\beta^2(\omega)t^\alpha$, where $\Lambda_\beta(\omega)$ indicates a single realization of the random variable Λ_β.

Remark 8.35. It is interesting to wonder if the generalized grey Brownian motion is the only one stationary increment process which serves as model for time-fractional diffusion equations like (eq. 8.81).

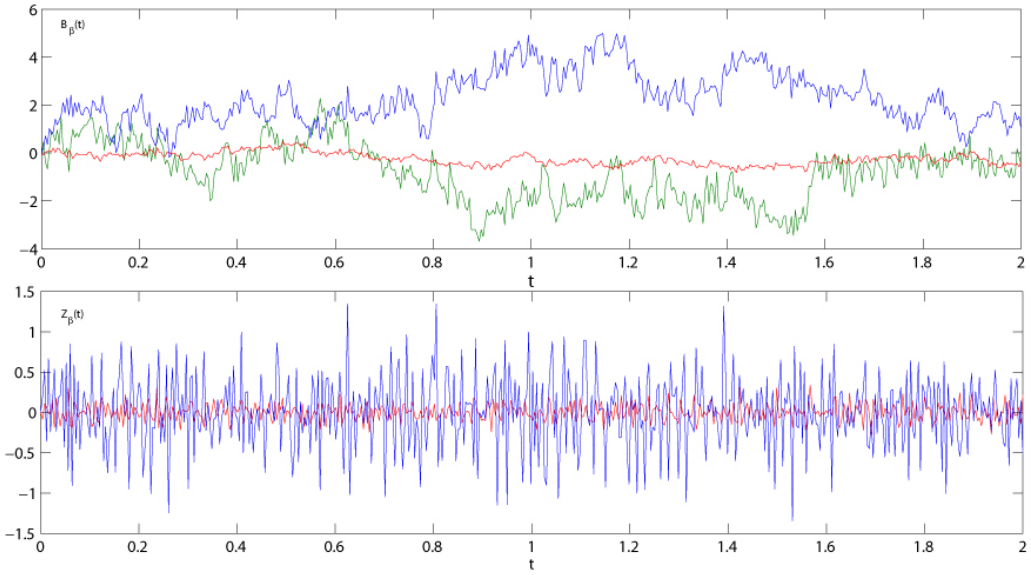

Figure 8.5: $B_\beta(t)$ trajectories in the case $\beta = 0.5$ (to panel) for $0 \leq t \leq 2$. The time series of the corresponding stationary noise $Z_\beta(t)$ is presented in the bottom panel.

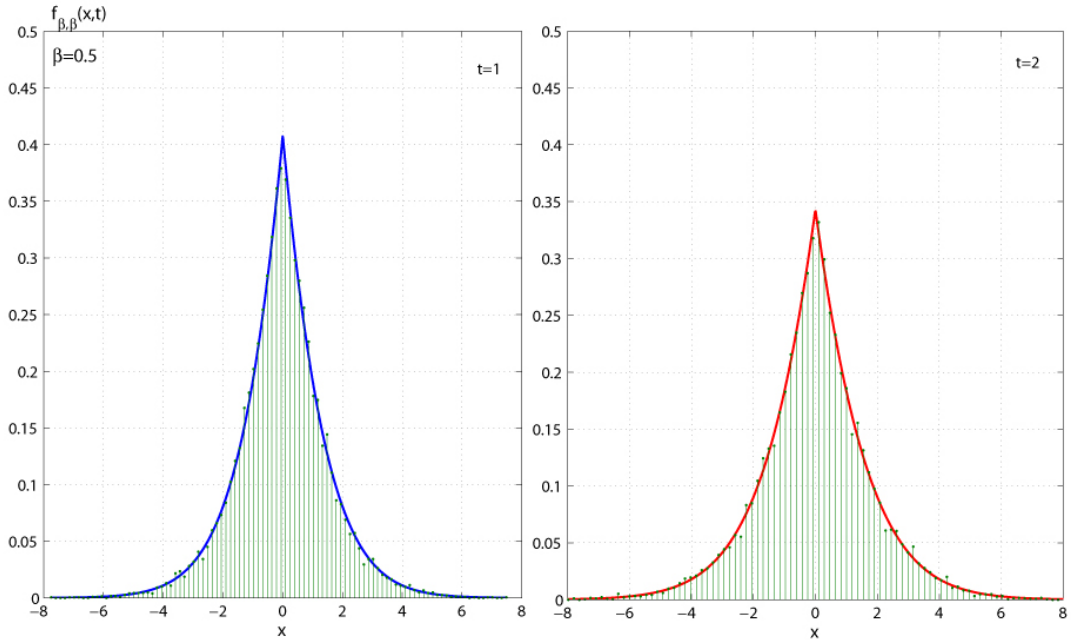

Figure 8.6: Histograms of $N = 15000$ simulations with $\beta = 0.5$ and exact marginal density at $t = 1$ and $t = 2$.

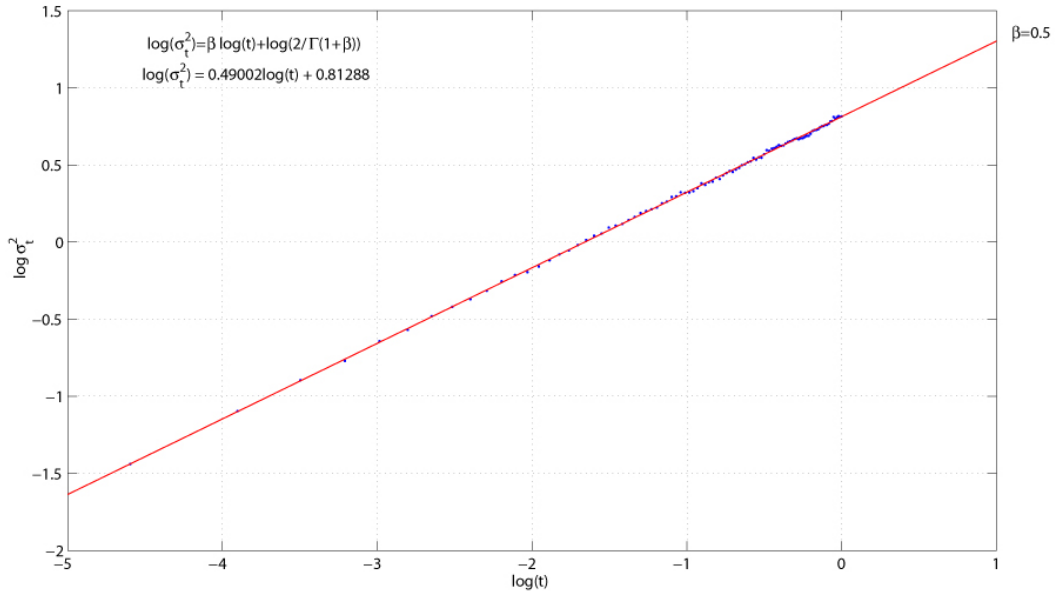

Figure 8.7: Sample variance in logarithmic scale and linear fitting ($N = 10^4$).

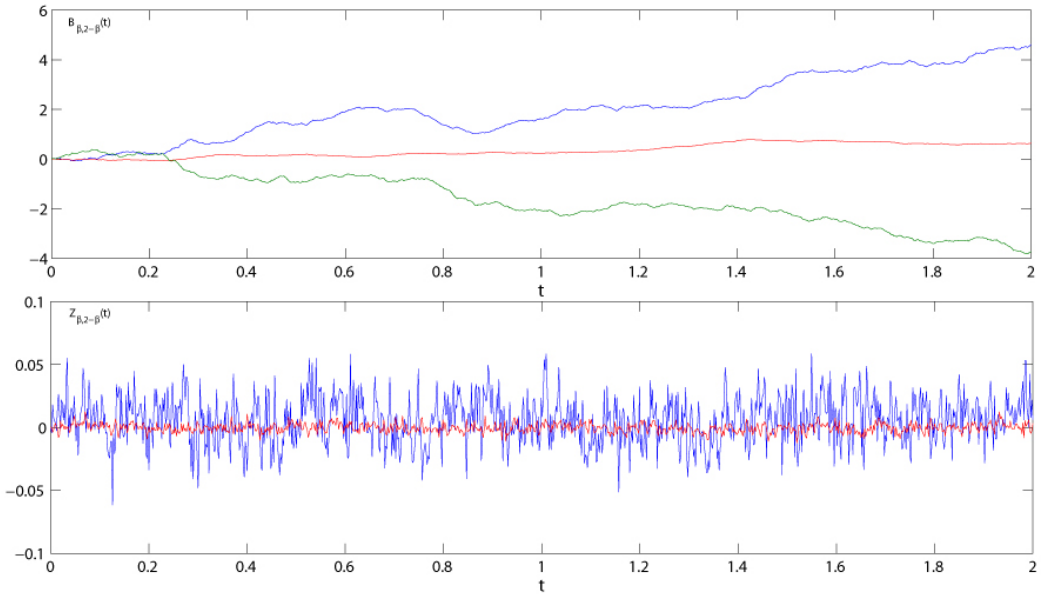

Figure 8.8: $B_{2-\beta,\beta}(t)$ trajectories in the case $\beta = 0.5$ (top panel) for $0 \leq t \leq 2$. The corresponding stationary noise time series is presented in the bottom panel.

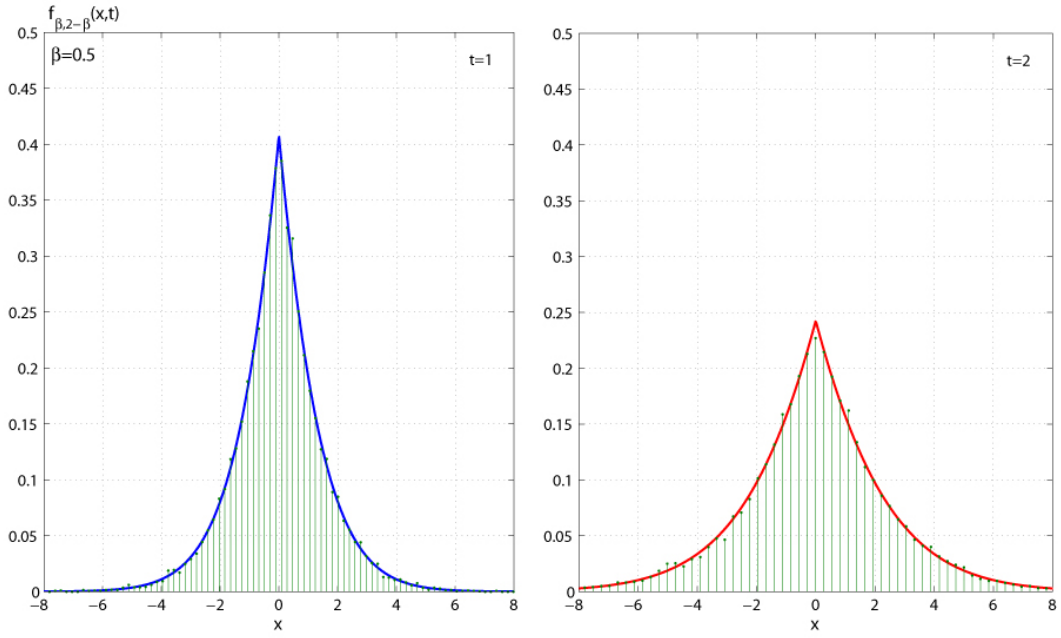

Figure 8.9: Histograms of two 15000 running simulations of $B_{2-\beta,\beta}(t)$ with $\beta = 0.5$ and exact distributions at different times $t = 2$ and $t = 1$.

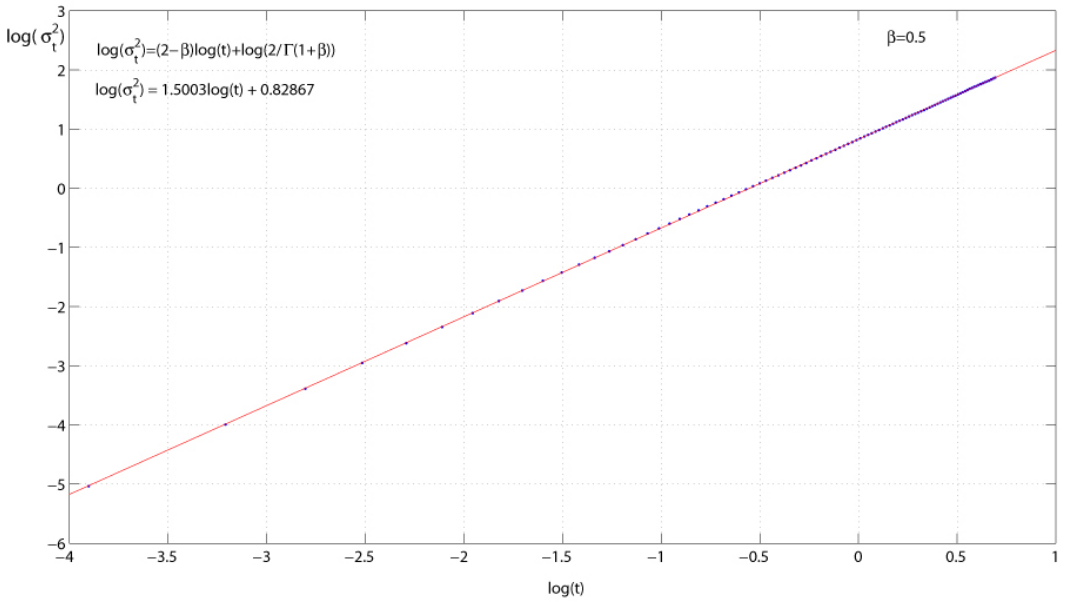

Figure 8.10: Sample variance in logarithmic scale and linear fitting ($N = 10^4$).

Bibliography

[1] J. Bardet, G. Lang, G. Oppenheim, A. Philippe, M.S. Taqqu, Generators of Long-Range Dependent Processes: A Survey, P. Doukan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and applications of long-range depenendece*, Birkäuser, Basel and Boston (2003), 579-623.

[2] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, 2-nd edn., Wiley, New York (1971). [1-st edn. (1966)].

[3] I.M. Gel'fand, G.E. Shilov Generalized Functions, Volume I *Academic Press* (1964)

[4] R. Gorenflo, F. Mainardi and A. Vivoli, *Continuous time random walk and parametric subordination in fractional diffusion*, Chaos, Solitons and Fractals, **34** (2007), 87-103. [E-print `http://arxiv.org/abs/cond-mat/0701126`]

[5] T. Hida, *Brownian Motion*, Springer Verlag , New York-Heidelberg-Berlin (1980).

[6] R. Gorenflo, F. Mainardi, D. Moretti and P. Paradisi, Time-fractional diffusion: a discrete random walk approach, *Nonlinear Dynamics* **29** (2002) 129-143.

[7] R. Gorenflo, F. Mainardi, D. Moretti, G. Pagnini and P. Paradisi, Discrete random walk models for space time fractional diffusion, *Chemical Physics***284** (2002) 521-541.

[8] T. Kolsrud, On a class of probabilistic integrodifferential equations, in: S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Editors), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, Vol I, Cambridge University Press, Cambridge (1990), pp. 168-172.

[9] H.H. Kuo, *White Noise Distribution Theory*, CRC Press, Boca Raton (1996).

[10] M. Meerschaert, D. Benson, H. Scheffler, B. Baeumer, Stochastic solution of space-time fractional diffusion equations, *Physical Review E* **65** (2002), 041103-1/4.

[11] K.S. Miller, S.G. Samko, Completely monotonic functions, *Integral Transforms and Special Functions*, **12** (2001), 389-402.

[12] A. Mura and F. Mainardi, A class of self similar stochastic processes with stationary increments to model anomalous diffusion in physics, *Submitted*, E-print, `http://arxiv.org/abs/0711.0665`.

[13] A. Mura, G. Pagnini, Characterization and simulation of a class of stochstic processes to model anomalous diffusion. E-print, `http://arxiv.org/abs/0801.4879`.

[14] V. Pipiras and M.S. Taqqu, Integration questions related to fractional Brownian motion, *Probability Theory and Related Fields*, **118** (2000), 251-291.

[15] V. Pipiras and M.S. Taqqu, Fractional calculus and its connections to fractional Brownian motion, in: P. Doukan, G. Oppenheim and M.S. Taqqu (Editors), *Long-range Dependence: Theory and Applications* Birkäuser, Basel and Boston (2002), pp. 165-201.

[16] V. Pipiras, M.S. Taqqu, Deconvolution of fractional Brownian motion, *Journal of Time Series Analysis* **23** (2002), 487-501.

[17] W.R. Schneider, Grey noise, in: S. Albeverio, G. Casati, U. Cattaneo, D. Merlini and R. Moresi (Editors), *Stochastic Processes, Physics and Geometry*, World Scientific, Singapore (1990), pp. 676-681.

[18] W.R. Schneider, Grey noise, in: S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Editors), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, Cambridge Univ. Press, Cambridge (1990), Vol. I, pp. 261-282.

[19] W.R. Schneider, Fractional diffusion, in: R. Lima, L. Streit and D. Vilela Mendes (Editors), *Dynamics and Stochastic Processes, Theory and Applications*, Springer Verlag, Heidelberg (1990), pp. 276-286. [Lecture Notes in Physics # 355]

[20] A. A. Stanislavsky, Subordinated Brownian motion and its fractional Fokker-Planck equation, *Physica Scripta* **67** (2003), 265-268.

9 Non-Markovian diffusion equations and processes

Brownian motion $B(t)$, $t \geq 0$, is a stochastic process with many properties (Def. 3.4). It is at the same time Gaussian, Markovian, has stationary increments and is self-similar .

We remember that, a process $X(t)$, $t \geq 0$, is said to be self-similar with self-similarity exponent H if, for all $a \geq 0$, the processes $X(at)$, $t \geq 0$, and $a^H X(t)$, $t \geq 0$, have the same finite-dimensional distributions (Def. 3.8).

Brownian motion is self-similar with exponent $H = 1/2$. In contrast, fractional Brownian motion $B_H(t)$, $t \geq 0$, is Gaussian, has stationary increments, is self-similar with self-similarity exponent $0 < H < 1$, but is not Markovian, unless $H = 1/2$, in which case the fractional Brownian motion becomes Brownian motion (Remark 3.22). When $1/2 < H < 1$, the increments of fractional Brownian motion have long-range dependence (Prop. 6.1).

Because Brownian motion is Markovian with stationary increments, its finite-dimensional distributions can be obtained from the marginal density function

$$f_B(x,t) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}, \quad x \in \mathbb{R} \tag{9.1}$$

at time $t \geq 0$ by using (eq. 3.22). This density function is the fundamental solution of the "standard" diffusion equation:

$$\partial_t u(x,t) = \partial_{xx} u(x,t), \tag{9.2}$$

which in integral form reads:

$$u(x,t) = u_0(x) + \int_0^t \partial_{xx} u(x,s) ds, \quad u_0(x) = u(x,0). \tag{9.3}$$

Thus, $f_B(x,t)$ is a solution of (eq. 9.3) with $u_0(x) = \delta(x)$, where $\delta(x)$ is the Dirac delta distribution.

Remark 9.1. Observe that, here we are following the physics convention of not including the factor $1/2$ in (eq. 9.2). Therefore, in the following, "standard" Brownian motion $B(t)$, $t \geq 0$, is such that, for each time $t \geq 0$, $B(t) \sim N(0, 2t)$.

In this chapter, we consider non-local, fractional and stretched modifications of the diffusion equation as we have anticipated in (eq. 7.13). These modified equations are indeed a generalization of time fractional equations, and will be called *Non-Markovian diffusion equations*, because, while they originate from a diffusion equation, the corresponding process, whose probability density function is a solution of these modified equations, will be typically non-Markovian.

To motivate the modifications, consider first the non-random process $l(t) = t$, $t \geq 0$, which depicts a non-random linear time evolution and let $f_l(\tau, t)$ denote its density function at time t. Therefore, one has $f_l(\tau, t) = \delta(\tau - t)$ where $\delta(x)$ is the Dirac distribution. It is natural to interpret $f_l(\tau, t)$ as the fundamental solution of the standard forward drift equation (eq. 9.4), namely,

$$\partial_t u(\tau, t) = -\partial_\tau u(\tau, t), \quad \tau, t \geq 0, \tag{9.4}$$

which have been written in integral form (eq. 9.5)

$$u(\tau,t) = u_0(\tau) - \int_0^t \partial_\tau u(\tau,s)ds, \quad u_0(\tau) = u(\tau,0). \tag{9.5}$$

We then consider the following generalization of the forward drift equation (eq. 9.5)

$$u(\tau,t) = u_0(\tau) - \int_0^t K(t-s)\partial_\tau u(\tau,s)ds, \ \ \tau,t \geq 0, \tag{9.6}$$

where $K(t)$, with $t \geq 0$, is a suitable kernel, chosen such that the fundamental solution of (eq. 9.6) is a probability density function at each $t \geq 0$. We refer to (eq. 9.6) as the *non-Markovian forward drift equation*.

The presence of the memory kernel K in (eq. 9.6) suggests a corresponding modification of the diffusion equation (eq. 9.3). Namely, we will consider the equation:

$$u(x,t) = u_0(x) + \int_0^t K(t-s)\partial_{xx} u(x,s)ds, \ \ x \in \mathbb{R}, \ \ t \geq 0. \tag{9.7}$$

We show that its fundamental solution turns out to be:

$$f(x,t) = \int_0^\infty G(x,\tau)h(\tau,t)d\tau, \tag{9.8}$$

where

$$G(x,t) = \frac{1}{\sqrt{4\pi t}}\exp(-x^2/4t), \tag{9.9}$$

and $h(\tau,t)$ is the fundamental solution of (eq. 9.6).

The solution (eq. 9.8) is a marginal (one-point) probability density function. We will consider different random processes whose marginal probability density function coincides with it. As illustration, consider the following examples.

Example 9.1. If we choose:

$$K(t) = \frac{t^{-1/2}}{\sqrt{\pi}}, \ \ t \geq 0, \tag{9.10}$$

we are in the fractional case of (eq. 9.24) with $\beta = 1/2$. Then, we have (see eq. 7.121 and eq. 7.126):

$$h(\tau,t) = \frac{1}{\sqrt{\pi t}}\exp\left(-\frac{\tau^2}{4t}\right), \ \ \tau \geq 0, \ \ t \geq 0, \tag{9.11}$$

as fundamental solution of (eq. 9.6). Now consider the process

$$D(t) = B(l(t)), \ \ t \geq 0, \tag{9.12}$$

where B is a "standard" Brownian motion and $l(t) \geq 0$ is a random time change (not necessarily increasing), independent of B, whose marginal density function is given by $h(\tau,t)$. One possible choice for the random time process is simply:

$$l(t) = |b(t)|, \ \ t \geq 0,$$

where $b(t)$, $t \geq 0$, is a "standard" Brownian motion. Such a random time process $l(t)$, $t \geq 0$, is self-similar of order $H = 1/2$.[1] Let now $B(t)$, $t \geq 0$, be another "standard" Brownian motion independent of $b(t)$. Thus, the process

$$D(t) = B(|b(t)|), \quad t \geq 0, \tag{9.13}$$

has marginal density defined by (eq. 9.8) with $h(\tau, t)$ given by (eq. 9.11).

But, $D(t)$ is not the only process with density function $f(x,t)$, given by (eq. 9.8). For example, the process

$$Y(t) = \sqrt{|b(1)|} B_{1/4}(t), \quad t \geq 0\,, \tag{9.14}$$

where $B_{1/4}$ is an independent fractional Brownian motion with self-similarity exponent $H = 1/4$, has the same one-dimensional probability density functions as the previous process $D(t)$, $t \geq 0$ (see eq. 9.40 with $\beta = 1/2$).

Example 9.2. The fractional Brownian motion in (eq. 9.14) has a self-similarity exponent $H < 1/2$. We know that the increments of such a process negatively correlated (Prop. 3.10). To allow for the presence of fractional Brownian motion $B_H(t)$ with $0 < H < 1$, we introduce a second (non-random) time change $t \to g(t)$, where $g(0) = 0$ and $g(t)$ is smooth and increasing, that is we consider the non-Markovian diffusion equation

$$u(x,t) = u_0(x) + \int_0^t g'(s) K\left(g(t) - g(s)\right) \partial_{xx} u(x,s) ds. \tag{9.15}$$

whose fundamental solution is now:

$$f(x,t) = \int_0^\infty G(x,\tau) h(\tau, g(t)) d\tau, \tag{9.16}$$

where h is the fundamental solution of (eq. 9.6).

Remark 9.2. Observe that (eq. 8.81) is of this type with $K(t) = t^{\beta-1}/\Gamma(\beta)$ and $g(t) = t^{\alpha/\beta}$.

If $K(t)$ is as in (eq. 9.10) and $g(t) = t^{2\alpha}$, with $0 < \alpha < 2$, then the processes:

$$D(t) = B(|b(t^{2\alpha})|), \quad t \geq 0,$$
$$Y(t) = \sqrt{|b(1)|} B_{\alpha/2}(t), \quad t \geq 0,$$

have a marginal density function defined by (eq. 9.16) with $h(\tau, t)$ as in (eq. 9.11), which is the fundamental solution of (eq. 9.15). In this case $Y(t)$ is defined through an independent fractional Brownian motion $B_{\alpha/2}$ with Hurst's parameter $H = \alpha/2$ and thus $0 < H < 1$. This is a special case of (eq. 9.65).

The preceding examples illustrate the themes pursued in this chapter. We will focus, however, not only on power-like kernels such as defined in (eq. 9.10), but also on exponential-like kernels such as:

$$K(t) = e^{-at}, \quad a \geq 0. \tag{9.17}$$

We also consider what happens when the Brownian motion $B(t)$, $t \geq 0$, is replaced by a more general linear (time-homogeneous) diffusion $Q(t)$, $t \geq 0$, governed by the Fokker-Planck equation,

$$\partial_t u(x,t) = \mathcal{P}_x u(x,t), \tag{9.18}$$

1) Another possible choice for a random time process with marginal density given by (eq. 9.11) is the *local-time* in zero of a "standard" Brownian motion [3]. In this case the time change process $l(t)$ is increasing.

where $\mathcal{P}_x$ is a linear operator independent of t acting on the variable $x \in \mathbb{R}$ (see eq. 7.11). In other words we consider the non-Markovian diffusion equation:

$$u(x,t) = u_0(x) + \int_0^t g'(s)K(g(t) - g(s))\mathcal{P}_x u(x,s)ds. \tag{9.19}$$

We show that its fundamental solution is:

$$f(x,t) = \int_0^\infty \mathcal{G}(x,\tau)h(\tau,g(t))d\tau, \tag{9.20}$$

where $\mathcal{G}(x,t)$ is the fundamental solution of (eq. 9.19), while $h(\tau,t)$ is the fundamental solution of (eq. 9.6). We also provide explicit solutions when $\mathcal{P}_x$ is the differential operator associated with Brownian motion with drift, when it is associated with Geometric Brownian motion and when the kernel $K(t)$ is the power kernel and the exponential kernel.

In order not to dwell on technicalities, we suppose implicitly, throughout the chapter, that we have sufficient regularity conditions, to justify the algebraic manipulations that are performed.

9.1 Historical notes

We begin with some historical remarks. Non-Markovian equations like (eq. 9.7), or more generally (eq. 9.19), are often encountered when studying physical phenomena related to relaxation and diffusion problems in complex systems (see Srokowsky [32] for examples).

Equations of the type (eq. 9.7) have been studied for example by Kolsrud [**?**]7. He obtained (eq. 9.8), but without providing specific examples. A similar study was done by Wyss [36] who, however, focused only on power-like kernels $K(t) = Ct^{\beta-1}$.

Sokolov [30] (see also Srokowsky [32]), studied the non-Markovian equation

$$\partial_t P(x,t) = \int_0^t k(t-s)L_x P(x,s)ds, \tag{9.21}$$

where L_x is a linear operator acting on the variable x. He provided a formal solution in the form of (eq. 9.20). Observe, however, that our equation (eq. 9.19) differs from (eq. 9.21), not only by the presence of the scaling function $g(t)$, but also by the choice of the memory kernel. Our kernel $K(t)$ and Sokolov's kernel $k(t)$ are related by the equation:

$$K(t) = \int_0^t k(s)ds \Rightarrow \widetilde{K}(s) = \widetilde{k}(s)/s, \;\; s > 0, \tag{9.22}$$

where the tilde indicates the Laplace transform. The suitability conditions for these memory kernels are thus not the same (these conditions are developed in Section 9.2). For example, consider the simple *exponential-decay* kernel e^{-at}, $a \geq 0$. This choice of the kernel is "safe" in the context of (eq. 9.19), i.e. for the choice $K(t) = e^{-at}$, but is "dangerous" if one considers (eq. 9.21) with the kernel $k(t) = e^{-at}$. In the case of (eq. 9.19), the *exponential-decay* kernel corresponds to a system for which non-local memory effects are initially negligible. In fact, $K(t) = e^{-at} \to 1$ as $t \to 0$ and thus the system appears Markovian at small times. On the other hand, the choice $k(t) = e^{-at}$ corresponds to the kernel $K(t) = a^{-1}(1 - e^{-at})$ which for small times behaves like t. In this case Sokolov [30] noticed that the corresponding equations are only reasonable in a restricted domain of the model parameters and for certain initial and boundary

conditions.

Our starting point is different from that of the previous authors. Instead of starting directly from the Fokker-Planck equations (eq. 9.18), we start from the forward drift equation (eq. 9.5) which is then generalized by introducing a memory kernel $K(t)$ (eq. 9.6). One is then naturally led to the non-Markovian diffusion equations (eq. 9.15 and eq. 9.19) after the introduction of the scaling function $g(t)$. In fact, in specific cases, it is sometimes simpler to solve first the non-Markovian forward drift equation (eq. 9.6) and then use the solution to solve the non-Markovian diffusion equation (eq. 9.15 or eq. 9.19) by using (eq. 9.16 or eq. 9.20). For instance, we have used (eq. 9.8) in order to solve the time-fractional diffusion equation (see Section 7.2.5).

The form of the solution (eq. 9.16 or eq. 9.20) has now a ready-made interpretation. For example, in (eq. 9.20) the function $\mathcal{G}(x,t)$ is the fundamental solution of the Markovian equation (eq. 9.18) and the function $h(\tau,t)$ is the fundamental solution of the non-Markovian equation (eq. 9.6) and it is these two solutions that contribute to (eq. 9.20) which is the fundamental solution of the non-Markovian diffusion equation (eq. 9.19).

The form (eq. 9.16 or eq. 9.20) of the solution has a natural interpretation in terms of *subordinated* processes (see eq. 9.12). According to Whitmore and Lee [16], the term "subordination" was introduced by Bochner [4,5]. It refers to processes of the form $Y(t) = X(l(t))$, $t \geq 0$, where $X(t)$, $t \geq 0$, is a Markov process and $l(t)$, $t \geq 0$, is a (non-negative) random time process independent of X. The marginal distribution of the subordinated process is clearly:

$$f_Y(x,t) = \int_0^\infty f_X(x,\tau) f_l(\tau,t) d\tau, \quad t \geq 0, \quad x \in \mathbb{R}, \tag{9.23}$$

where $f_X(x,t)$ and $f_l(\tau,t)$ represent the marginal density functions of the processes X and l. Therefore, (eq. 9.16 or eq. 9.20) can be interpreted in terms of subordinated processes, with (eq. 9.6) characterizing the random time process $l(t)$ and (eq. 9.18) characterizing the Markov parent process $X(t)$ (see Remark 7.37).

Remark 9.3. The stochastic interpretation through subordinated processes is very natural because $Y(t) = X(l(t))$ has a direct physical interpretation. For example, in equipment usage, $X(t)$ can be the state of a machine at time t and $l(t)$ the effective usage up to time t. In an econometric study, $X(t)$ may be a model for the price of a stock at time t. If $l(t)$ measures the total economic activity up to time t, the price of the stock at time t should not be described by $X(t)$ but by the subordinated process $Y(t) = X(l(t))$. The resulting subordinated process $Y(t)$ is in general non-Markovian. In this way, the non-local memory effects are attributable to the random time process $l(t)$ and to its dynamics which is in general non-local in time (eq. 9.6).

However, the solution of (eq. 9.19) represents only the marginal (one-point) density function of the process and therefore cannot characterize the full stochastic structure of the process. In fact, there are also processes that are not subordinated processes that serve as stochastic models for non-Markovian diffusion equations like (eq. 9.19) or (eq. 9.21). For instance, the grey Brownian motion of the previous chapter provides an example in the case of fractional type equations.

9.2 The non-Markovian forward drift equation

We start with the following generalization of (eq. 9.5), namely:

$$u(\tau,t) = u_0(\tau) - \int_0^t K(t-s)\partial_\tau u(\tau,s) ds, \quad \tau, t \geq 0, \tag{9.24}$$

where $K(t)$, with $t \geq 0$, is a suitable chosen kernel. We then choose a random time process $l(t)$ such that, for each $t \geq 0$, its marginal density $f_l(\tau, t)$ is the fundamental solution of (eq. 9.24). Observe that (eq. 9.24) is "non-local" because $u(\tau, t)$ involves $u(\tau, s)$ at all $0 \leq s \leq t$. We have called equation (9.24 *non-Markovian forward drift equation*.
It is convenient to work with Laplace transforms. We indicate by $\mathcal{L}\{\varphi(x,t); t, s\}$ the Laplace transform of the function φ with respect to t evaluated in $s \geq 0$, namely:

$$\mathcal{L}\{\varphi(x,t); t, s\} = \int_0^\infty e^{-ts}\varphi(x,t)dt, \quad s \geq 0. \tag{9.25}$$

If the function φ depends only on the variable t we write simply $\widetilde{\varphi}(s)$, because in this case there is no ambiguity concerning the integration variable. In particular we let $\widetilde{K}(s)$ denote the Laplace transform of the kernel K.

Prop. 9.1. *Let $f_l(\tau, t)$ denote the fundamental solution of (eq. 9.24). Then,*

$$\mathcal{L}\{f_l(\tau,t); t, s\} = \frac{1}{s\widetilde{K}(s)}\exp\left(-\frac{\tau}{\widetilde{K}(s)}\right), \quad \tau, s \geq 0, \tag{9.26}$$

and zero for $\tau < 0$.

Proof: we take the Laplace transform with respect to the variable t in (eq. 9.24):

$$\partial_\tau \widetilde{u}(\tau, s) = \frac{u_0(\tau)}{s\widetilde{K}(s)} - \frac{\widetilde{u}(\tau, s)}{\widetilde{K}(s)}, \tag{9.27}$$

thus (eq. 9.26) is a solution, in the distributional sense, when $u_0(\tau) = \delta(\tau)$. Indeed the general solution of (eq. 9.27) with $u_0(\tau) = \delta(\tau)$ is:

$$\varphi(\tau, s) = \frac{\theta(\tau)}{s\widetilde{K}(s)}\exp\left(-\frac{\tau}{\widetilde{K}(s)}\right) + C\exp\left(-\frac{\tau}{\widetilde{K}(s)}\right), \quad \tau \in \mathbb{R},$$

where C is a real constant. Since we require $\varphi(\tau, t) = 0$ for $\tau < 0$, we get $C = 0$, i.e. (eq. 9.26). □

9.2.1 Suitability conditions on the kernel K

We must choose the kernel K such that the fundamental solution of (eq. 9.24) is a probability density in $\tau \geq 0$. We observe that if $f_l(\tau, t)$ satisfies (eq. 9.24) and (eq. 9.26), then it is automatically normalized for each $t \geq 0$. In fact, for a function $\varphi(x, t)$ for which it is always possible to change the order of integration, one has:

$$\int_{\mathbb{R}} \varphi(x,t)dx = 1 \iff \int_{\mathbb{R}} \widetilde{\varphi}(x,s)dx = s^{-1}. \tag{9.28}$$

Since (eq. 9.26) satisfies the right-hand side of (eq. 9.28), we get $\int_{\mathbb{R}_+} f_l(\tau, t)d\tau = 1$. One still needs, however, to choose the kernel K such that $f_l(\tau, t) \geq 0$ for all $\tau, t \geq 0$.

In order to get a suitable condition on the kernel K, we make use of the notion of completely monotone function. Recall that a function $\varphi(t)$ is *completely monotone* if it is non-negative and possesses derivatives of any order and:

$$(-1)^k \frac{d^k}{dt^k}\varphi(t) \geq 0, \quad t > 0, \quad k \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}. \tag{9.29}$$

We observe that as $t \to 0$, the limit of $d^k\varphi(t)/dt^k$ may be finite or infinite. Typical non-trivial examples are $\varphi(t) = \exp(-at)$, with $a > 0$, $\psi(t) = 1/t$ and $\phi(t) = 1/(1+t)$. A less trivial example is provided by the Mittag-Leffler function $E_\beta(-t)$, $t \geq 0$, for $0 < \beta \leq 1$ (see Prop. 8.8) .

It is easy to show that if φ and ψ are completely monotone then their product $\varphi\psi$ is as well. Moreover, if φ is completely monotone and ψ is positive with first derivative completely monotone then the function $\varphi(\psi)$ is completely monotone.

We have the following characterization of completely monotone functions [9]:

Lemma 9.1. *A function $\varphi(s)$, defined on the positive real line, is completely monotone if and only if is on the form:*

$$\varphi(s) = \int_0^\infty e^{-ts} F(dt), \quad s \geq 0,$$

where F is a finite or infinite non-negative measure on the positive real semiaxis.

Hence, to ensure that $f_l(\tau, t) \geq 0$ for all $\tau, t \geq 0$, it is enough to require that the function defined in (eq. 9.26) must be completely monotone, as a function of s, for any $\tau \geq 0$, and thus that the kernel K satisfies the following:

Suitability conditions

- $s\widetilde{K}(s)$ is positive with first derivative completely monotone,
- $1/\widetilde{K}(s)$ is positive with first derivative completely monotone.

Indeed, we can view (eq. 9.26) as the product of the two completely monotone functions $1/u$ and $\exp(-\tau u)$, the first evaluated at $u = s\widetilde{K}(s)$ and the second evaluated at $u = 1/\widetilde{K}(s)$.

9.2.2 Examples

Example 9.3 (*β-power kernel*)**.** If we choose:

$$K(t) = \frac{t^{\beta-1}}{\Gamma(\beta)},$$

we get $\widetilde{K}(s) = s^{-\beta}$. In this case $s\widetilde{K}(s) = s^{1-\beta}$ is positive and has first derivative $(1-\beta)s^{-\beta}$ completely monotone if and only if $0 < \beta \leq 1$. Moreover, $1/\widetilde{K}(s) = s^\beta$ is positive with first derivative $\beta s^{\beta-1}$ completely monotone if and only if $0 < \beta \leq 1$. Therefore, a good choice for the kernel K is:

$$K(t) = \frac{t^{\beta-1}}{\Gamma(\beta)}, \quad 0 < \beta \leq 1. \tag{9.30}$$

Remark 9.4. In this way, we have incidentally proved that $\mathcal{M}_\beta(\tau, t)$, $0 < \beta \leq 1$, which is the fundamental solution of (eq. 9.24) with the power kernel choice, namely the time-fractional forward drift equation (eq. 7.110), defines indeed a probability density function in $\tau \geq 0$ for any $t \geq 0$ (see property 1. below eq. 7.122).

Example 9.4 (*Exponential decay kernel*)**.** Choosing:

$$K(t) = \exp(-at), \quad a \geq 0, \tag{9.31}$$

we get $s\widetilde{K}(s) = s/(s+a)$ which is positive with first derivative $a(s+a)^{-2}$ completely monotone for any $a \geq 0$. Moreover, $1/\widetilde{K}(s) = (s+a)$ is positive if $a \geq 0$ with first derivative completely monotone.

Example 9.5 (*β-power with exponential decay kernel*)**.** Choosing:

$$K(t) = \frac{t^{\beta-1}}{\Gamma(\beta)}\exp(-at), \;\; 0 < \beta \leq 1, \;\; a \geq 0, \tag{9.32}$$

we have $\widetilde{K}(s) = (s+a)^{-\beta}$. Therefore, $s\widetilde{K}(s) = s(s+a)^{-\beta}$ which is positive if $a \geq 0$ with first derivative $(s+a)^{-\beta}(1-\beta s(s+a)^{-1})$ completely monotone if $0 < \beta \leq 1$. Moreover, $1/\widetilde{K}(s) = (s+a)^{\beta}$ is positive if $a \geq 0$ with first derivative $\beta(s+a)^{\beta-1}$ completely monotone if $0 < \beta \leq 1$.

Example 9.6 (*Distributed order power-kernel*)**.** The choice

$$K(t) = \int_0^1 p(\beta)\frac{t^{\beta-1}}{\Gamma(\beta)}d\beta, \tag{9.33}$$

leads to Rieaman-Liouville distributed order fractional equations (see eq. 7.143). In this case, one has $\widetilde{K}(s) = \int_0^1 p(\beta)s^{-\beta}d\beta$. Then, one finds $s\widetilde{K}(s) = \int_0^1 p(\beta)s^{1-\beta}d\beta$, which is positive with first derivative $\int_0^1 p(\beta)(1-\beta)s^{-\beta}d\beta$. The latter can be seen as a linear superposition of completely monotone functions and it is then completely monotone as well. Moreover, it is also possible to show that $1/\widetilde{K}(s)$ is always positive with first derivative completely monotone.

The following theorem states that a self-similar random time process $l(t)$, $t \geq 0$, is associated with the power kernel $K(t)$ (Example 9.3):

Theorem 9.1. *If the time change process $l(t)$, $t \geq 0$, is self-similar (for instance of order $H = \beta$), with marginal probability density $f_l(\tau, t)$ satisfying (eq. 9.26), then we must have:*

$$K(t) = C\frac{t^{\beta-1}}{\Gamma(\beta)}, \;\; 0 < \beta \leq 1, \tag{9.34}$$

for some positive constant C.

Proof: The self-similarity condition entails that for any $\tau, t \geq 0$ and for any $a > 0$:

$$a^{-\beta}f_l(a^{-\beta}\tau, t) = f_l(\tau, at).$$

If we take the Laplace transform and set $\widetilde{f}(\tau, s) = \mathcal{L}\{f_l(\tau, t); t, s\}$, we have:

$$a^{-\beta}\widetilde{f}_l(a^{-\beta}\tau, s) = \frac{1}{a}\widetilde{f}_l\left(\tau, \frac{s}{a}\right).$$

Using (eq. 9.26) we get that for any $\tau, s \geq 0$ and $a > 0$:

$$\frac{a^{-\beta}}{\widetilde{K}(s)}\exp\left(-\frac{a^{-\beta}\tau}{\widetilde{K}(s)}\right) = \frac{1}{\widetilde{K}(\frac{s}{a})}\exp\left(-\frac{\tau}{\widetilde{K}(\frac{s}{a})}\right).$$

Since this relation is valid for any choice of $\tau \geq 0$ and $s \geq 0$, putting $\tau = 0$ and $s = a$, we get:

$$\frac{a^{-\beta}}{\widetilde{K}(a)} = \frac{1}{\widetilde{K}(1)}.$$

Thus, for any $a > 0$:

$$\widetilde{K}(a) = \widetilde{K}(1)a^{-\beta},$$

which is the Laplace transform of (eq. 9.34). If we add moreover the condition of complete monotonicity we find: $0 < \beta \leq 1$ as indicated in Example 9.3. □

9.3 Non-Markovian diffusion equation

We focus here on the non-Markovian diffusion equation introduced in (eq. 9.15). There are two ingredients:

1. The fundamental solution of (eq. 9.24), denoted here by $h(\tau,t)$ and defined by (eq. 9.26).
2. The fundamental solution $G(x,t)$, defined by (eq. 9.9), of the standard diffusion equation which is the one-dimensional density of the "standard" Brownian motion.

The following theorem combines these two ingredients and provides the fundamental solution of a corresponding non-Markovian diffusion equation.

Theorem 9.2. *Let $h(\tau,t)$ denote the fundamental solution of (eq. 9.24), so that by Proposition 9.1, one has:*

$$\mathcal{L}\{h(\tau,t);t,s\} = \frac{1}{s\widetilde{K}(s)}\exp\left(-\frac{\tau}{\widetilde{K}(s)}\right), \quad \tau,s \geq 0, \tag{9.35}$$

for a suitable choice of K. Let g be a strictly increasing function with $g(0) = 0$ and let $G(x,t)$ be defined by (eq. 9.9). Then,

$$f(x,t) = \int_0^\infty G(x,\tau)h(\tau,g(t))d\tau, \tag{9.36}$$

is the fundamental solution of the non-Markovian diffusion equation:

$$u(x,t) = u_0(t) + \int_0^t g'(s)K\left(g(t)-g(s)\right)\partial_{xx}u(x,s)ds. \tag{9.37}$$

Proof: see Section 9.4. □

We have immediately the following:

Cor. 9.1. *If $H(x,t)$ is a solution of the standard diffusion equation with initial condition $H(x,0) = u_0(x)$, then the function:*

$$u(x,t) = \int_0^\infty H(x,\tau)h(\tau,g(t))d\tau \tag{9.38}$$

is a solution of (eq. 9.37).

Proof: If, for any $t \geq 0$, the function $f(x,t)$ defined in (eq. 9.36) is the fundamental solution of (eq. 9.37) then a general solution is given by:

$$u(x,t) = \int_{\mathbb{R}} f(x-y,t)u_0(y)dy = \int_{\mathbb{R}}\int_0^\infty G(x-y,\tau)u_0(y)h(\tau,g(t))d\tau dy$$

$$= \int_0^\infty \left(\int_{\mathbb{R}} G(x-y,\tau)u_0(y)dy\right)h(\tau,g(t))d\tau = \int_0^\infty H(x,\tau)h(\tau,g(t))d\tau. \quad \square$$

We observe that:

A. The equation (9.35) states that $h(\tau,t)$ is the fundamental solution of (eq. 9.24).

B. While $G(x,t)$ is the fundamental solution of the standard diffusion equation obtained when $u_0(x) = \delta(x)$, the general solution, denoted $H(x,t)$ in the above theorem, results from a general initial condition $u_0(x)$.

Many physical phenomena, especially related to relaxation processes in complex systems, are described by non-Markovian "master equations" like (eq. 9.37). $K(t)$ is a memory kernel and $g(t)$ is just a "time scaling" function. Such equations are often argued by phenomenological considerations and some times can be more or less rigorously derived starting from a microscopic description [6, 38, 14, 32].

9.3.1 The stochastic representation is not unique

The solution of the non-Markovian diffusion equation can be viewed as the marginal density function of the subordinated process (see eq. 9.12)

$$D(t) = B(l(g(t))), \quad t \geq 0,$$

since its marginal density is:

$$f_D(x,t) = \int_0^\infty G(x,\tau) f_l(\tau, g(t)) d\tau.$$

Here, for each $t \geq 0$, $D(t) \sim f_D(x,t)$, $B(t) \sim G(x,t)$ and $l(t) \sim f_l(\tau,t)$. In the notation of Theorem 9.2, we have $f_D(x,t) = f(x,t)$ and $f_l(\tau,t) = h(\tau,t)$. The Laplace transform of $f_l(\tau,t)$ with respect to t is given by (eq. 9.35).

This stochastic representation is not unique (see Example 9.1, Example 9.2 and examples below). Indeed, the non-Markovian diffusion equation characterizes only the marginal, that is one-point, probability density function. However, processes with a different dependence structure can have the same marginal density $f(x,t)$. Additional requirements could be imposed so as to specify the stochastic model more precisely.

Example 9.7. If we require the random time process $l_\beta(t)$, $t \geq 0$, to be self-similar of order β, then in view of Theorem 9.1, the kernel must be chosen as in (eq. 9.34) with $0 < \beta \leq 1$. Then, we obtain the stretched time-fractional diffusion equation (eq. 8.81). We will study this case more in details in Section 9.5. Here we just observe that if we consider a "standard" fractional Brownian motion $B_{\beta/2}$ of order $\beta/2$, then $f(x,t)$ is also the marginal distribution of the grey Brownian motion (eq. 8.87 with $\alpha = \beta$),

$$Y(t) = \sqrt{l_\beta(1)} B_{\beta/2}(t), \tag{9.39}$$

where $B_{\beta/2}(t)$ is assumed to be independent of $l_\beta(1)$.

In fact, because $l_\beta(t)$, $t \geq 0$, is self-similar of order $H = \beta$, one has:

$$D(t) = B(l_\beta(t)) \stackrel{d}{=} \sqrt{l_\beta(t)} B(1) \stackrel{d}{=} \sqrt{l_\beta(1)} t^{\beta/2} B(1) \stackrel{d}{=} \sqrt{l_\beta(1)} t^{\beta/2} B_{\beta/2}(1) \stackrel{d}{=} \sqrt{l_\beta(1)} B_{\beta/2}(t) = Y(t), \tag{9.40}$$

where $\stackrel{d}{=}$ denotes here the equality of the marginal distributions.

Both $D(t)$, $t \geq 0$, and $Y(t)$, $t \geq 0$, are self-similar processes with Hurst's exponent $H = \beta/2$. However, while $Y(t)$, $t \geq 0$, has always stationary increments, this is not in general true in the case of the process $D(t)$, $t \geq 0$.

9.4 Non-Markovian Fokker-Planck equation

We considered up to now processes of the type $B(l(g(t)))$, where B is a "standard" Brownian motion. What happens if we replace B by a more general diffusion? Namely, what happens if instead of starting with the standard diffusion equation (eq. 9.2) we start with a more general Markovian Fokker-Planck equation:

$$\partial_t u(x,t) = \mathcal{P}_x u(x,t), \quad x \in \mathbb{R}, \quad t \geq 0, \tag{9.41}$$

where $\mathcal{P}_x$ is a linear operator, independent of t, acting on the variable x? We have the following generalization of Theorem 9.2:

Theorem 9.3. *Suppose that $h(\tau,t)$ is a probability density function satisfying (eq. 9.26)*

$$\mathcal{L}\{h(\tau,t);t,s\} = \frac{1}{s\widetilde{K}(s)}\exp\left(-\frac{\tau}{\widetilde{K}(s)}\right), \quad \tau,s \geq 0, \tag{9.42}$$

for a suitable choice of K. Let g be a strictly increasing function with $g(0) = 0$ and $\mathcal{G}(x,t)$ be the fundamental solution of (eq. 9.41). Then the fundamental solution of the integral equation:

$$u(x,t) = u_0(t) + \int_0^t g'(s)K\left(g(t) - g(s)\right)\mathcal{P}_x u(x,s)ds \tag{9.43}$$

is

$$f(x,t) = \int_0^\infty \mathcal{G}(x,\tau)h(\tau,g(t))d\tau. \tag{9.44}$$

We provide two versions of the proof. The first starts with the solution $f(x,t)$ in (eq. 9.44) and verifies that it satisfies (eq. 9.43) (existence). The second starts from the partial integro-differential equation (eq. 9.43) and derives the solution $f(x,t)$ under certain assumptions stated below (eq. 9.49) (uniqueness).

Proof 1: For first we observe that

$$\mathcal{L}\{f(x,t);g(t),s\} = \frac{1}{s\widetilde{K}(s)}\mathcal{L}\{\mathcal{G}(x,t);t,\widetilde{K}(s)^{-1}\}. \tag{9.45}$$

With the change of variables $g(s) = z$, we write:

$$u(x,g^{-1}(w)) = u_0(x) + \int_0^w K\left(w-z\right)\mathcal{P}_x u(x,g^{-1}(z))dz, \quad w = g(t). \tag{9.46}$$

We want to show that (eq. 9.44) with the choice (eq. 9.42) solves (eq. 9.43). If we take the Laplace transform of (eq. 9.43) using (eq. 9.46), we get:

$$\mathcal{L}\{u(x,t);g(t),s\} = \frac{u_0(x)}{s} + \widetilde{K}(s)\mathcal{P}_x\mathcal{L}\{u(x,t);g(t),s\}$$

that is:

$$s\mathcal{L}\{u(x,t);g(t),s\} - u_0(x) = s\widetilde{K}(s)\mathcal{P}_x\mathcal{L}\{u(x,t);g(t),s\}. \tag{9.47}$$

Now if we substitute on (eq. 9.47) a solution on the form of (eq. 9.44)

$$u(x,t) = \int_0^\infty \mathcal{H}(x,\tau)h(\tau,g(t))d\tau, \tag{9.48}$$

we have:

$$\widetilde{K}(s)^{-1}\mathcal{L}\{\mathcal{H}(x,t);t,\widetilde{K}(s)^{-1}\} = u_0(x) + \mathcal{P}_x\mathcal{L}\{\mathcal{H}(x,t);t,\widetilde{K}(s)^{-1}\}$$

i.e. we have, with obvious notations:

$$\tau\widetilde{\mathcal{H}}(x,\tau) = u_0(x) + \mathcal{P}_x\widetilde{\mathcal{H}}(x,\tau),$$

in which one readily recognizes the Laplace transform of the Markovian Fokker-Planck equation with the same initial condition $u_0(x)$. Therefore:

$$\partial_t\mathcal{H}(x,t) = \mathcal{P}_x\mathcal{H}(x,t), \quad \mathcal{H}(x,0) = u_0(x).$$

This argument shows not only that (eq. 9.44) is the fundamental solution of (eq. 9.43), but also that a general solution is given by (eq. 9.48) (see Corollary 9.1). This result is summarized in Corollary 9.2 (see below).

Proof 2: We now start from (eq. 9.43) and we use integral transforms in order to get the fundamental solution. Let $\mathcal{F}$ denote the Fourier transform operator and let:

$$(\mathcal{F}\varphi)(k,t) = \widehat{\varphi}(k,t) = \int_{\mathbb{R}} e^{ikx}\varphi(x,t)dx.$$

Since $\widehat{u}_0(k) = 1$, and since $(\mathcal{F}\mathcal{P}_x u)(k,t) = (\mathcal{F}\mathcal{P}_x\mathcal{F}^{-1}\mathcal{F}u)(k,t) = \widehat{\mathcal{P}}_k\widehat{u}(k,t)$, where $\widehat{\mathcal{P}}_k = (\mathcal{F}\mathcal{P}_x\mathcal{F}^{-1})_k$ denotes the Fourier transform of the operator $\mathcal{P}_x$, we have:

$$\widehat{u}(k, g^{-1}(w)) = 1 + \int_0^w K(w-z)\widehat{\mathcal{P}}_k\widehat{u}(k, g^{-1}(z))dz.$$

Taking the Laplace transform we have:

$$\mathcal{L}\{\widehat{u}(k, g^{-1}(w)); w, s\} = s^{-1} + \widehat{\mathcal{P}}_k\widetilde{K}(s)\mathcal{L}\{\widehat{u}(k, g^{-1}(w)); w, s\},$$

which is the same as:

$$\mathcal{L}\{\widehat{u}(k,t); g(t), s\} = s^{-1} + \widehat{\mathcal{P}}_k\widetilde{K}(s)\mathcal{L}\{\widehat{u}(k,t); g(t), s\}.$$

Therefore:

$$\left(\widetilde{K}(s)^{-1} - \widehat{\mathcal{P}}_k\right)\mathcal{L}\{\widehat{u}(k,t); g(t), s\} = s^{-1}\widetilde{K}(s)^{-1}.$$

Denoting $1(k) = 1$, we have:

$$\mathcal{L}\{\widehat{u}(k,t); g(t), s\} = \frac{1}{s\widetilde{K}(s)}\left(\widetilde{K}(s)^{-1} - \widehat{\mathcal{P}}_k\right)^{-1} 1(k), \tag{9.49}$$

where we suppose that the operator $\left(\widetilde{K}(s)^{-1} - \widehat{\mathcal{P}}_k\right)^{-1}$ is well defined and acts on the constant function $1(k) = 1$.

Observe that the Fokker-Planck equation (eq. 9.41) is obtained from (eq. 9.43) by setting $K(t) = 1$, for each $t \geq 0$, that is $\widetilde{K}(s) = s^{-1}$, and $g(t) = t$, for each $t \geq 0$. In this case (eq. 9.49) becomes:

$$\mathcal{L}\{\widehat{\mathcal{G}}(k,t); t, s\} = (s - \widehat{\mathcal{P}}_k)^{-1}1(k). \tag{9.50}$$

where $\mathcal{G}(x,t)$ is the fundamental solution. Taking the inverse Fourier transform, we get:

$$\mathcal{L}\{\mathcal{G}(x,t); t, s\} = \mathcal{F}^{-1}\left\{(s - \widehat{\mathcal{P}}_k)^{-1}1(k)\ ;\ k, x\right\}, \tag{9.51}$$

where:

$$\mathcal{F}^{-1}\left\{\varphi(k,s)\ ;\ k, x\right\} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ikx}\varphi(k,s)dk. \tag{9.52}$$

Replacing s by $\widetilde{K}(s)^{-1}$ in (eq. 9.51), one has:

$$\mathcal{L}\{\mathcal{G}(x,t); t, \widetilde{K}(s)^{-1}\} = \mathcal{F}^{-1}\left\{(\widetilde{K}(s)^{-1} - \widehat{\mathcal{P}}_k)^{-1}1(k)\ ;\ k, x\right\}. \tag{9.53}$$

Going back to (eq. 9.49) and inverting the Fourier transform we obtain in view of (eq. 9.53):

$$\mathcal{L}\{u(x,t); g(t), s\} = \frac{1}{s\widetilde{K}(s)}\mathcal{F}^{-1}\left\{\left(\widetilde{K}(s)^{-1} - \widehat{\mathcal{P}}_k\right)^{-1}1(k)\ ;\ k, x\right\} = \frac{1}{s\widetilde{K}(s)}\mathcal{L}\{\mathcal{G}(x,t); t, \widetilde{K}(s)^{-1}\}.$$

that is (eq. 9.45). □

Remark 9.5. If the Markovian process is a Brownian motion one has $\mathcal{P}_x = \dfrac{\partial^2}{\partial x^2}$. The Fourier transform of $\mathcal{P}_x$ is $\widehat{\mathcal{P}}_k = -k^2$ and (eq. 9.49) becomes:

$$\mathcal{L}\{\widehat{u}(k,t); g(t), s\} = \frac{1}{s\widetilde{K}(s)} \left(\widetilde{K}(s)^{-1} + k^2\right)^{-1} 1(k),$$

where

$$\left(\widetilde{K}(s)^{-1} + k^2\right)^{-1} 1(k) = \frac{1}{\left(\widetilde{K}(s)^{-1} + k^2\right)},$$

which is well defined because $\widetilde{K}(s)^{-1}$ is positive.

Cor. 9.2. *If $\mathcal{H}(x,t)$ is a general solution of the Markovian Fokker-Planck equation (eq. 9.41) with initial condition $\mathcal{H}(x,0) = u_0(x)$, then the function:*

$$u(x,t) = \int_0^\infty \mathcal{H}(x,\tau) h(\tau, g(t)) d\tau \tag{9.54}$$

is a general solution of (eq. 9.43).

From a stochastic point of view, $f(x,t)$ could be seen as the marginal distribution at time t of the subordinated process:

$$\mathcal{D}(t) = Q(l(g(t))) \tag{9.55}$$

where Q is the diffusion governed by the Fokker-Planck equation (9.41) and $l(t)$ is the random time process, independent of $Q(t)$, with marginal distributions defined by $h(\tau,t)$.

9.5 Examples involving standard Brownian motion

In the following examples, we consider stochastic models where the operator $\mathcal{P}_x$ in (eq. 9.41) is ∂_{xx}, namely the operator corresponding to standard Brownian motion. We will study more general operators in the next section. We shall choose various kernels $K(t)$ and various stretching functions $g(t)$. We let $h(\tau,t)$ denote the fundamental solution of the non-Markovian forward drift equation (eq. 9.24). Since the corresponding stochastic models are not unique, we will mainly focus on the subordinated process $B(l(t))$, $t \geq 0$. However, we also give examples of other appropriate stochastic models.

9.5.1 Time fractional diffusion equation

Let $g(t) = t$ and consider the *β-power* kernel:

$$K(t) = \frac{t^{\beta-1}}{\Gamma(\beta)}, \quad 0 < \beta \leq 1. \tag{9.56}$$

Let $h(\tau,t)$ denote the fundamental solution of the non-Markovian forward drift equation (eq. 9.24) with kernel (9.56). Namely, one recovers the time-fractional forward drift equation (eq. 7.110). One has (eq. 7.121):

$$h(\tau,t) = t^{-\beta} M_\beta(\tau t^{-\beta}) := \mathcal{M}_\beta(\tau,t), \quad \tau, t \geq 0. \tag{9.57}$$

Remark 9.6. In view of Theorem 9.1, such a kernel arises if one requires $h(\tau,t)$ to be the marginal density function of a self-similar random time process $l(t)$ of order β.

Then, inserting (eq. 9.56) in (eq. 9.37), we obtain the time-fractional diffusion equation of order β,

$$u(x,t) = u_0(t) + \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} \partial_{xx} u(x,s) ds. \tag{9.58}$$

which has been studied in Section (7.2.5). Using Theorem 9.2 and (eq. 7.127) together with (eq. 7.126) and (eq 9.57) we recover the fundamental solution of the time fractional diffusion equation (eq. 7.140):

$$f(x,t) = \int_0^\infty G(x,\tau) h(\tau,t) d\tau = \frac{1}{2} \int_0^\infty \mathcal{M}_{1/2}(|x|,\tau) \mathcal{M}_\beta(\tau,t) d\tau$$

$$= \frac{1}{2} \mathcal{M}_{\beta/2}(|x|,t) = \frac{1}{2} t^{-\beta/2} M_{\beta/2}(|x| t^{-\beta/2}). \tag{9.59}$$

Several plots of the M-function are presented in Chapter 7. In Figure (7.7) the function $h(\tau,t) = \mathcal{M}_\beta(\tau,t)$ is drawn at a fixed time $t = 1$ and for different values of the parameter β. Figure (7.8) presents the plot of $f(x,t) = \frac{1}{2}\mathcal{M}_{\beta/2}(|x|,t)$ at a fixed time $t = 1$ and for different values of β. Moreover, in Figure (7.9) is shown the time evolution of $f(x,t)$ for fixed $\beta = 1/2$. Finally, Figure (7.10) highlights and compares the queue behaviors for different choices of the time-derivative order β.

9.5.1.1 Stochastic interpretations of the solution

From a stochastic point of view, the function $h(\tau,t)$ in (eq. 9.57) can be regarded as the marginal distribution of

$$l_\beta(t), \quad t \geq 0,$$

where $l_\beta(t)$, $t \geq 0$, is a H-ss random time with $H = \beta$. We have that for each integer $m \geq 0$:

$$E(l_\beta(t)^m) = \frac{m!}{\Gamma(\beta m + 1)} t^{\beta m}. \tag{9.60}$$

In fact, from (eq. 7.114), for each integer $m \geq 0$, we have :

$$\int_0^\infty \tau^m s^{\beta-1} e^{-\tau s^\beta} d\tau = m! s^{-m\beta-1},$$

which, inverting the Laplace transform, gives (eq. 9.60).

For instance, with the suitable conventions [7], $l_\beta(t)$, $t \geq 0$, can be viewed as the local time in zero at time t of a $d = 2(1-\beta)$ dimensional Bessel process [23]. Alternatively, in the context of Continuous time random walks, it can be taken as the inverse of the totally skewed β-stable Lèvy process (see Chapter 1, Rem. 1.13). The function $f(x,t)$ in (eq. 9.59) is then the marginal density function of

$$D(t) = B\left(l_\beta(t)\right),$$

which is self-similar with $H = \beta/2$. In this case, because $l_\beta(t)$ is self-similar of order β, we immediately have an example of a different process with the same marginal distribution of $D(t)$ (see Example 9.7). In fact, if we consider a "standard" fractional Brownian motion $B_{\beta/2}$ of order $\beta/2$, then $f(x,t)$ can also be seen as the marginal distribution of the grey Brownian motion

$$Y(t) = \sqrt{l_\beta(1)} B_{\beta/2}(t), \tag{9.61}$$

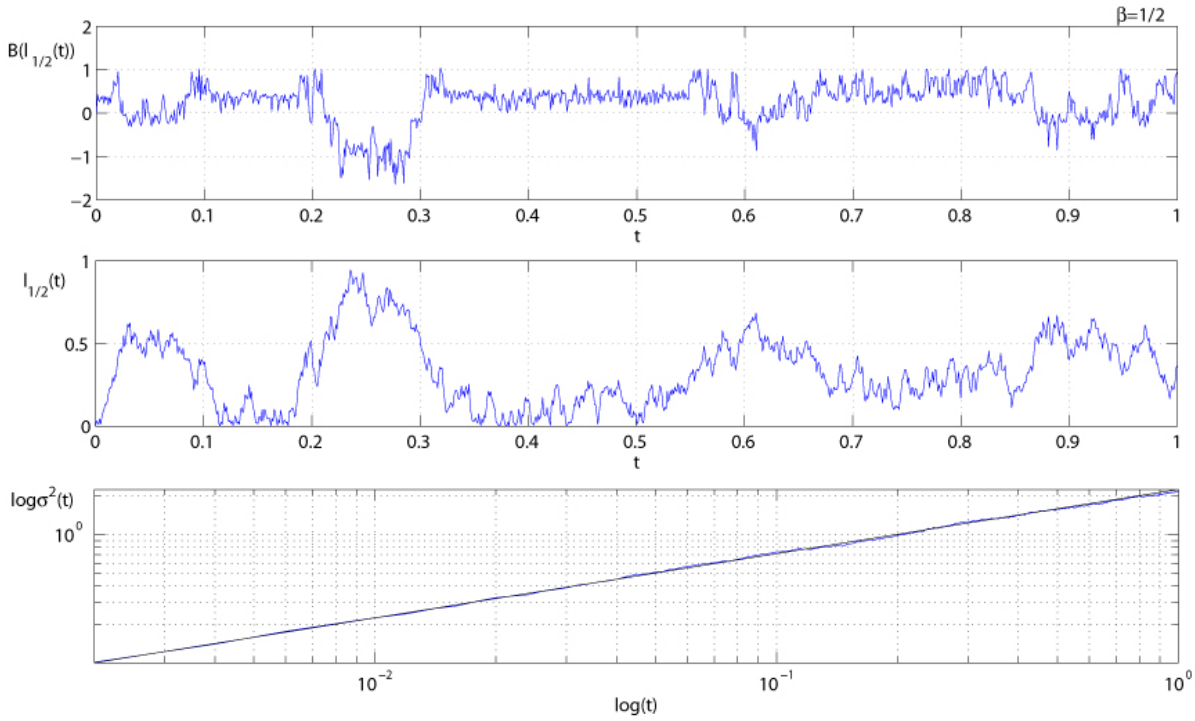

Figure 9.1: Trajectory of the process $B(l_\beta(t))$ (top panel), with $0 < t < 1$ and $\beta = 1/2$. The random time process is chosen to be $l_{1/2}(t) = |b(t)|$ where $b(t)$ is a "standard" Brownian motion (see Example 9.1). The corresponding trajectory of the random time process is presented in the middle panel. The estimated variance, computed on a sample of dimension $N = 5000$, is presented in logarithmic scale in the bottom panel and fits perfectly the theoretical curve $2t^{1/2}/\Gamma(3/2)$.

where $B_{\beta/2}(t)$ is assumed to be independent of $l_\beta(1)$ (see Example 9.7 and eq. 8.87). From (eq. 9.60) one can derive immediately all the moments for the processes $D(t)$ and $Y(t)$ (see also eq. 7.142). For any integer $m \geq 0$

$$\begin{cases} E(D(t)^{2m+1}) = E(Y(t)^{2m+1}) = 0; \\ E(D(t)^{2m}) = E(Y(t)^{2m}) = \dfrac{2m!}{\Gamma(\beta m + 1)} t^{\beta m}. \end{cases} \tag{9.62}$$

Because $0 < \beta < 1$, the variance grows slower than linearly with respect to time. That is, we have slow-anomalous diffusion for any β (eq. 7.142). Moreover, the increments of the fractional Brownian motion $B_{\beta/2}(t)$ do not have long-range dependence. In contrast, the next example allows for the presence of long-range dependence (which means fast diffusion) through the introduction of a scaling function $g(t) = t^{\alpha/\beta}$ (see also Example 9.2).

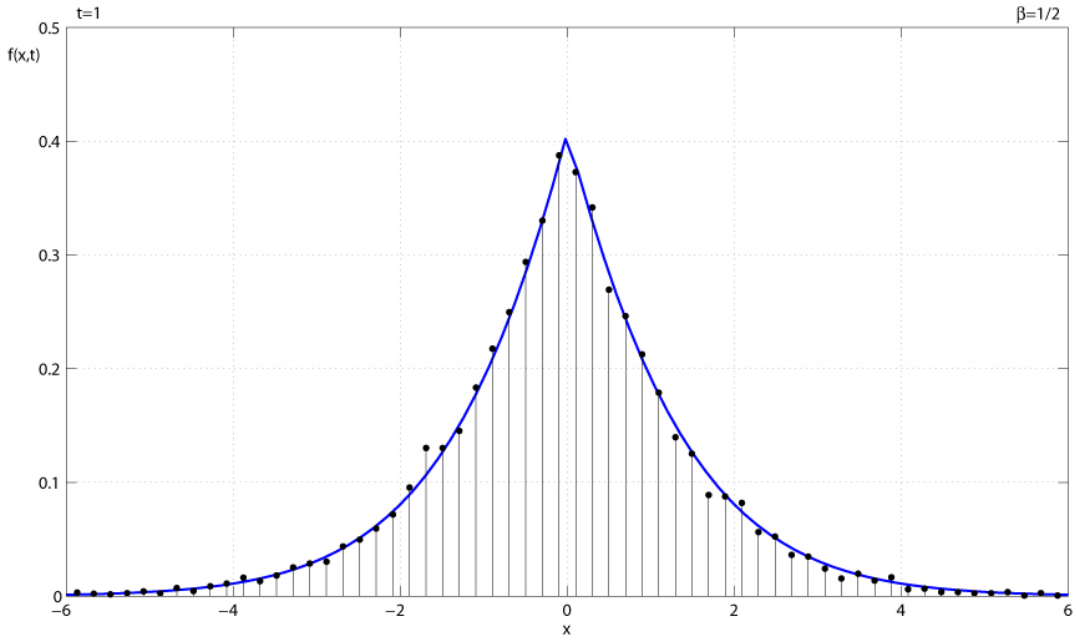

Figure 9.2: Marginal density function $f(x,t) = \frac{1}{2}\mathcal{M}_{1/4}(|x|,t)$ of the process $B(l_{1/2}(t))$ at time $t = 1$ and $x \in [-5,5]$. The histogram is evaluated over $N = 10^4$ simulated trajectories of the process $B(|b(t)|)$ (Figure 9.1).

9.5.2 "Stretched" time fractional diffusion equation

If in the setup of Section 9.5.1, where the kernel $K(t)$ is given by (eq. 9.56), we introduce a scaling time

$$g(t) = t^{\alpha/\beta}$$

with $\alpha > 0$, then the integral equation (eq. 9.58) is replaced by (eq. 9.37), namely

$$u(x,t) = u_0(t) + \frac{1}{\Gamma(\beta)}\frac{\alpha}{\beta}\int_0^t s^{\frac{\alpha}{\beta}-1}\left(t^{\frac{\alpha}{\beta}} - s^{\frac{\alpha}{\beta}}\right)^{\beta-1}\partial_{xx}u(x,s)ds. \tag{9.63}$$

Remark 9.7. We remember that the above "stretched" time-fractional diffusion equation is the master equation of the generalized grey Brownian motion marginal density (eq. 8.81) provided that $0 < \alpha < 2$.

Using (eq. 9.57) we find:

$$h(\tau, g(t)) = g(t)^{-\beta}M_\beta(\tau g(t)^{-\beta}) = t^{-\alpha}M_\beta(\tau t^{-\alpha}).$$

Then, by using (eq. 9.59), we recover (eq. 8.82). We indicate with $f_{\alpha.\beta}(x,t)$ the fundamental solution of (eq. 9.63), namely:

$$f_{\alpha,\beta}(x,t) = f(x,g(t)) = \frac{1}{2}t^{-\alpha/2}M_{\beta/2}(|x|t^{-\alpha/2}), \quad t \geq 0. \tag{9.64}$$

In this case, we are able to study a "full-ranged" anomalous diffusion. In fact, the function $f_{\alpha,\beta}(x,t)$, $t \geq 0$, is the marginal distribution of processes like

$$D(t) = B\left(l_\beta(t^{\alpha/\beta}))\right), \quad t \geq 0.$$

The time change process $l_\beta(t^{\alpha/\beta})$ is self-similar of order $H = \alpha$ and the process $D(t)$ is then self-similar with $H = \alpha/2$. In the case $0 < \alpha < 2$, the function $f_{\alpha,\beta}(x,t)$ is also the marginal density of the ggBm

$$\mathcal{Y}(t) = \sqrt{l_\beta(1)} B_{\alpha/2}(t), \quad t \geq 0, \quad 0 < \alpha < 2, \tag{9.65}$$

where $B_{\alpha/2}(t)$ is a "standard" fBm of order $H = \alpha/2$ independent of $l_\beta(1)$.

In any case, suppose $0 < \alpha < 2$, then, for any integer $m \geq 0$,

$$\begin{cases} E(D(t)^{2m+1}) = E(\mathcal{Y}(t)^{2m+1}) = 0; \\ E(D(t)^{2m}) = E(\mathcal{Y}(t)^{2m}) = \dfrac{2m!}{\Gamma(\beta m + 1)} t^{\alpha m}. \end{cases} \tag{9.66}$$

Therefore, one has slow-dffusion when $0 < \alpha < 1$ (the variance grows slower than linearly in time) and fast-diffusion when $1 < \alpha < 2$ (the variance grows faster than linearly in time) and the increments of the process $\mathcal{Y}(t)$ exhibit long-range dependence.

Graphical representations

It is interesting to study $f_{\alpha,\beta}(x,t)$ from a graphical point of view:

- In Figure 9.3 we have shown the plots of $f_{\beta,\beta}(x,t)$ (slow diffusion) and $f_{2-\beta,\beta}(x,t)$ (fast diffusion) for fixed time $t = 2$ and for different values of the β parameter. Here, we recognize the common Gaussian limit ($\beta = 1$).
- In Figure 9.4 are presented the plots of $f_{\beta,\beta}(x,t)$ and $f_{2-\beta,\beta}(x,t)$ for $\beta = 1/2$ and for different times $t \in [1,2]$. It is possible to observe the slow (left plot) and the fast (right plot) diffusion.
- In Figure 9.5 is shown how the model approaches the normal diffusion.
- In Figure 9.6 we present a model for an "ultra" slow-diffusion $\alpha = 0.1$, and a model for an "ultra" fast-diffusion $\alpha = 0.9$, with fixed $\beta = 0.8$.
- In Figure 9.7 we have shown the diffusion in the chaotic domain, that is for $\alpha = 1$. We have chosen $\beta = 0.4$ and $\beta = 0.8$. Here, the diffusion is "normal", in the sense that the variance grows linearly with the time, but it is not Gaussian.
- In Figure 9.8 we have displayed two examples of Gaussian but not "normal" diffusions. That is, $\beta = 1$, $0 < \alpha < 2$. Slow diffusion is on the left and fast diffusion is on the right.
- In Figure 9.9 and Figure 9.10 is presented the variance plot against time, for different βs.

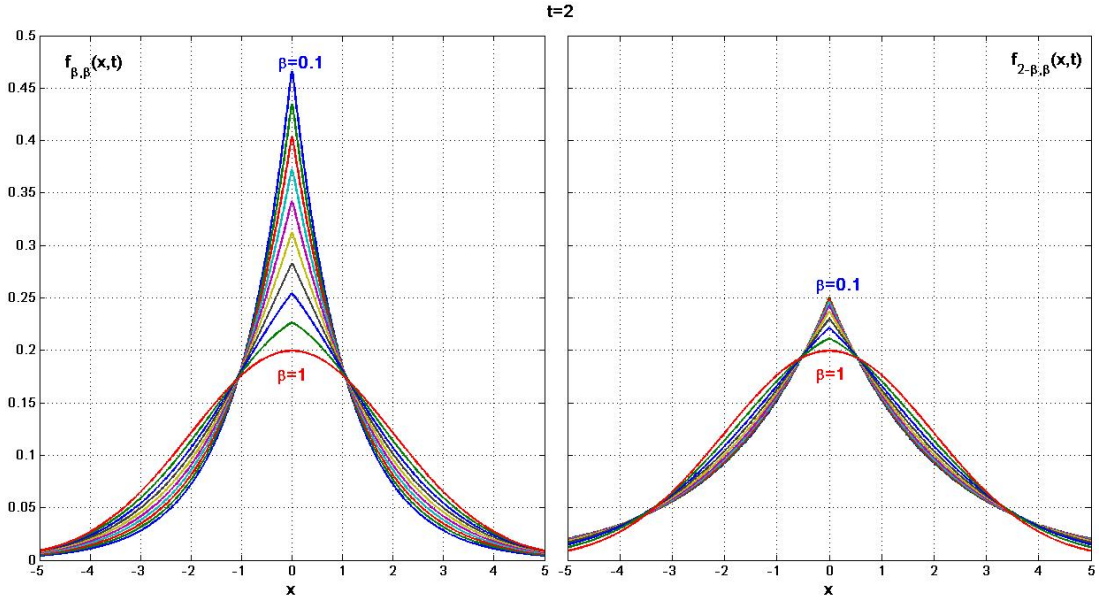

Figure 9.3: *PDF*s at fixed time $t = 2$ for slow (left plot) and fast (right plot) diffusions for different values of the β parameter.

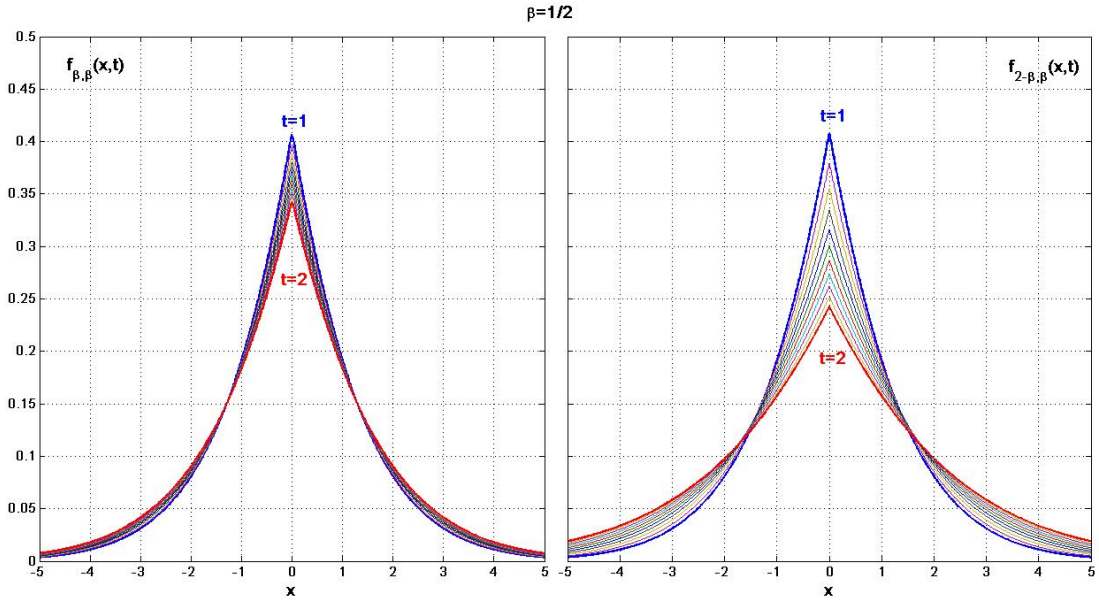

Figure 9.4: *PDF*s for slow (left plot) and fast (right plot) $\beta = 1/2$-diffusion for $t \in [1, 2]$.

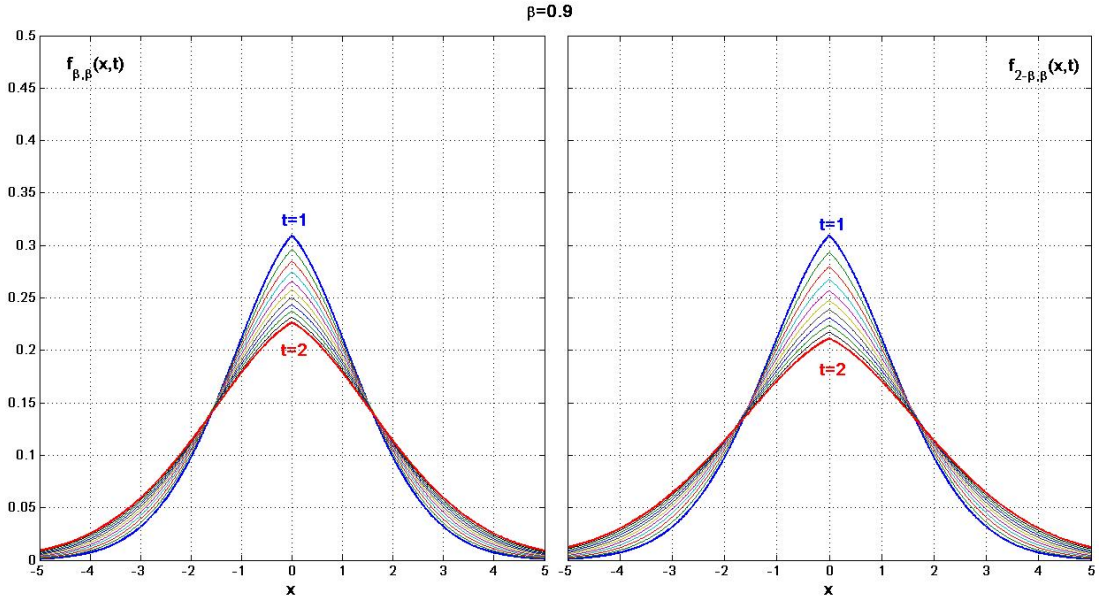

Figure 9.5: *PDF*s for slow (left plot) and fast (right plot) $\beta = 0.9$-diffusion for $t \in [1, 2]$.

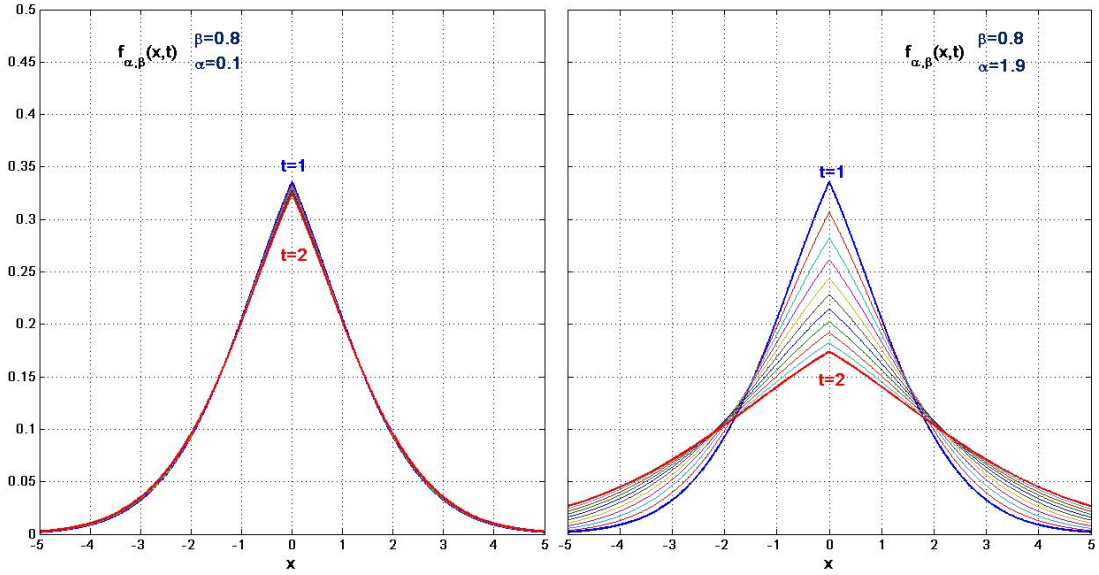

Figure 9.6: *PDF*s for "ultra" slow-diffusion ($\alpha = 0.1$, $\beta = 0.8$) on the left, and "ultra" fast diffusion ($\alpha = 0.9$, $\beta = 0.8$) on the right, for $t \in [1, 2]$.

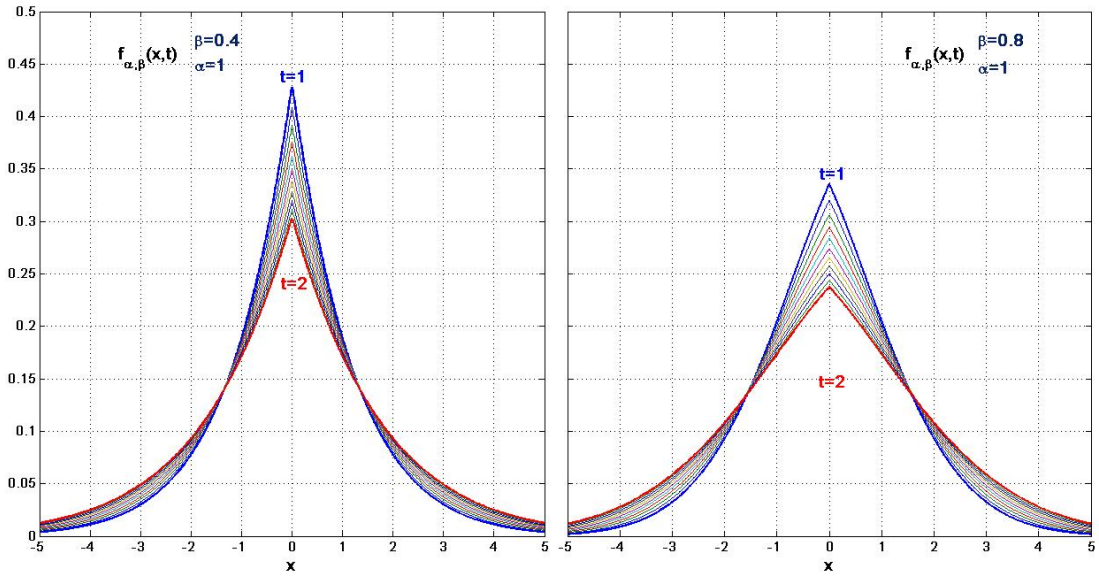

Figure 9.7: *PDF*s in the chaotic region: $\beta = 0.4$ on the left, and $\beta = 0.8$ on the right, for $t \in [1, 2]$.

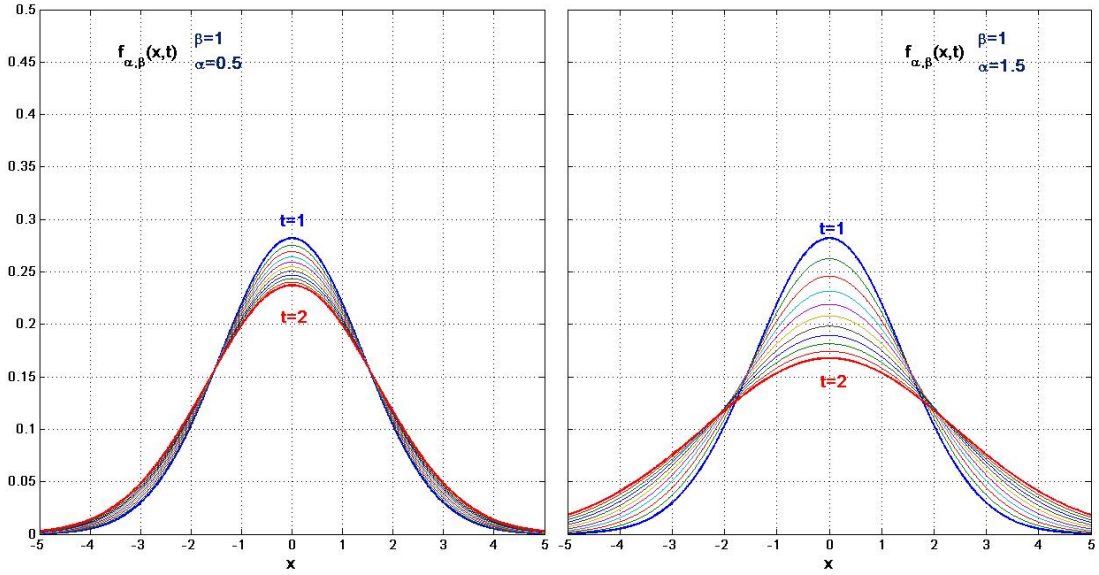

Figure 9.8: *PDF*s in the Gaussian region ($\beta = 1$) with $\alpha = 0.5$ on the left, and $\alpha = 1.5$ on the right, for $t \in [1, 2]$.

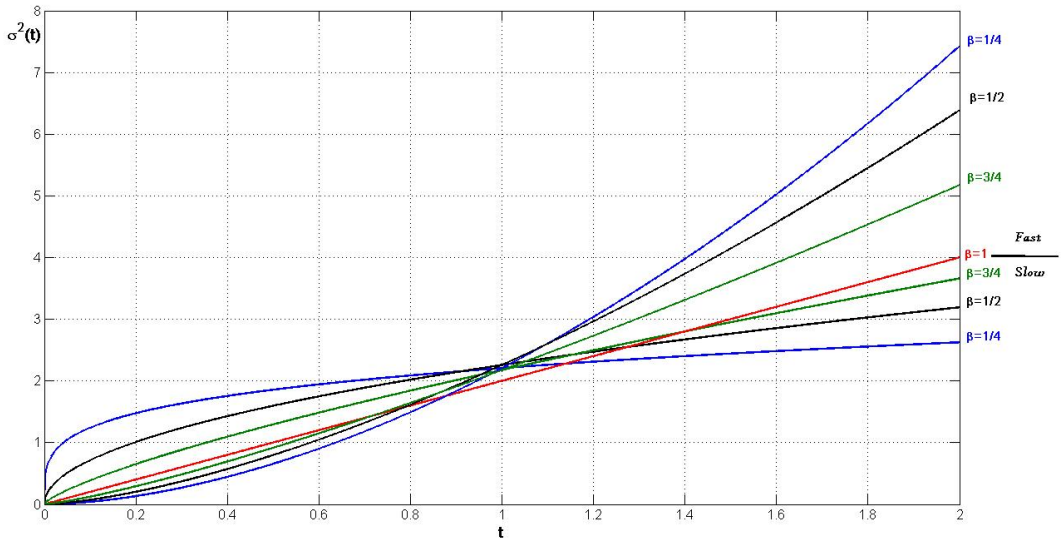

Figure 9.9: $\sigma^2(t) = \frac{2}{\Gamma(1+\beta)} t^\beta$ (slow diffusion) and $\sigma^2(t) = \frac{2}{\Gamma(1+\beta)} |t|^{2-\beta}$ (fast diffusion) for $\beta = [1/4, 1/2, 3/4, 1]$ and $t \in [0, 2]$. Linear scale.

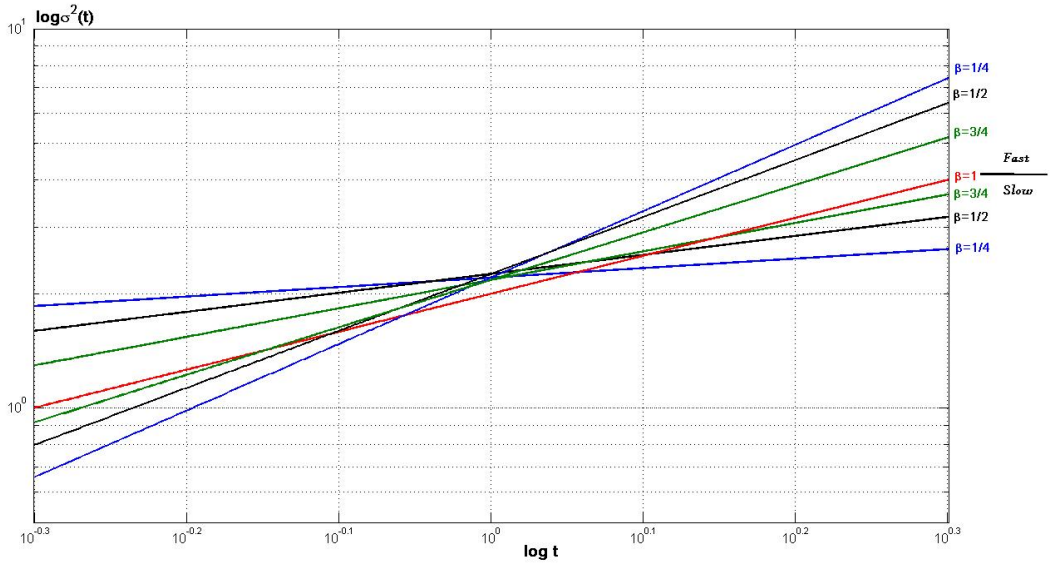

Figure 9.10: $\sigma^2(t) = \frac{2}{\Gamma(1+\beta)} t^\beta$ (slow diffusion) and $\sigma^2(t) = \frac{2}{\Gamma(1+\beta)} |t|^{2-\beta}$ (fast diffusion) for $\beta = [1/4, 1/2, 3/4, 1]$ and $t \in [0, 2]$. Logarithmic scale.

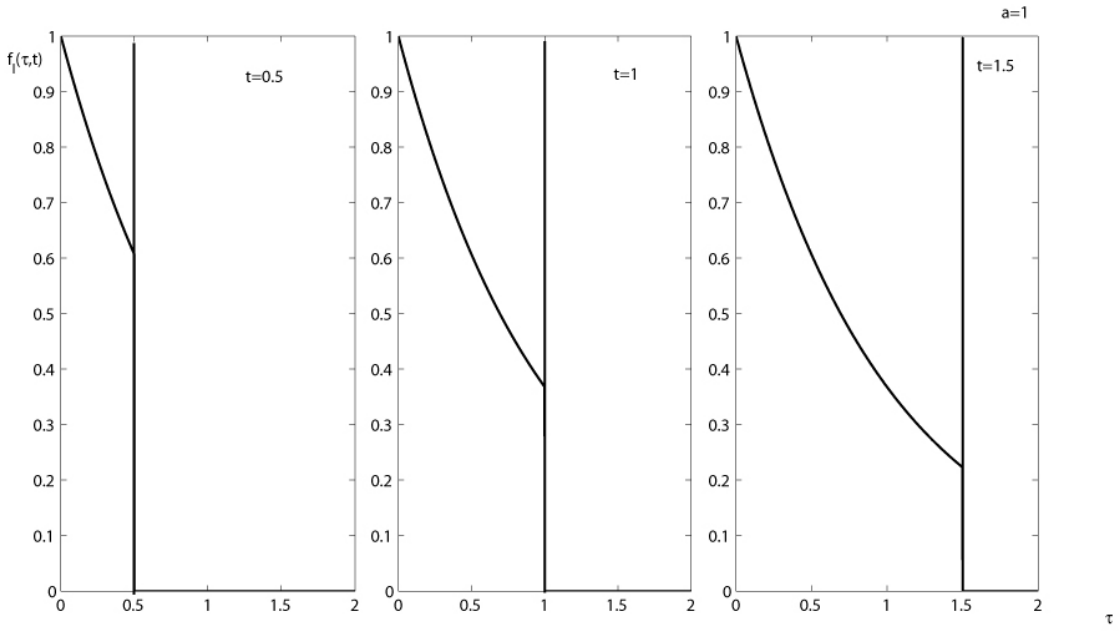

Figure 9.11: Plots of the marginal density of the random time $f_l(\tau, t)$ (eq. 9.69) as a function of τ at times $t = [0.5, 1, 1.5]$, and with $a = 1$. The vertical line corresponds to a point mass (delta function).

9.5.3 Exponential decay kernel

Let $g(t) = t$ and choose the exponential decay kernel (Example 9.4):

$$K(t) = \exp(-at), \;\; a \geq 0, \;\; t \geq 0. \tag{9.67}$$

Then, we obtain the following equation:

$$u(x,t) = u_0(x) + \int_0^t e^{-a(t-s)} \partial_{xx} u(x,s) ds. \tag{9.68}$$

In this case, $\widetilde{K}(s) = (s+a)^{-1}$ and the marginal distribution of the random time process $l(t)$, $t \geq 0$, is defined by (eq. 9.26):

$$\mathcal{L}\{f_l(\tau,t);t,s\} = \frac{s+a}{s} e^{-\tau(s+a)}, \;\; \tau \geq 0.$$

Therefore,

$$f_l(\tau,t) = e^{-\tau a}\left(\delta(\tau - t) + a\theta(t-\tau)\right) = e^{-ta}\delta(\tau - t) + ae^{-\tau a}\theta(t-\tau), \tag{9.69}$$

where $\theta(x)$ is the step function (eq. 7.116). A graphical representation of the time-evolution of $f_l(\tau, t)$ is presented in Figure 9.11.

Remark 9.8. The function $f_l(\tau, t)$ defined in (eq. 9.69) is the fundamental solution, in the sense of distributions, of the *"exponential" forward drift* equation:

$$u(\tau,t) = u_0(\tau) - \int_0^t e^{-a(t-s)} \partial_\tau u(\tau,s) ds.$$

This follows from Proposition 9.1. To check it directly, we note that $f_l(\tau,0)=\delta(\tau)$ and for any $t>0$:

$$-\int_0^t e^{-a(t-s)}\partial_\tau f_l(\tau,s)ds=-\int_0^t e^{-a(t-s)}\partial_\tau\left(e^{-as}\delta(\tau-s)+ae^{-a\tau}\theta(s-\tau)\right)ds$$

$$=-\int_0^t e^{-a(t-s)}\left(-e^{-as}\delta'(\tau-s)+ae^{-a\tau}\delta(s-\tau)-a^2e^{-a\tau}\theta(s-\tau)\right)ds$$

$$=e^{-at}\delta(\tau-t)+ae^{-at}\theta(t-\tau)+ae^{-a(t+\tau)}\theta(t-\tau)(e^{at}-e^{a\tau})$$

$$=e^{-at}\delta(\tau-t)+ae^{-a\tau}\theta(t-\tau)=f_l(\tau,t),$$

where we have used the fact that:

$$\int_0^t\delta'(\tau-s)ds=\delta(t-\tau).$$

We observe that when $a\to 0$ we recover the forward drift equation (eq. 9.5) and indeed $f_l(\tau,t)=\delta(\tau-t)$.

As noted in Example 9.4, (eq. 9.69) actually defines a probability density for any $t\geq 0$. The following proposition provides its moments.

Prop. 9.2. *For each integer $m\geq 0$ one has:*

$$E(l(t)^m)=\frac{m!}{a^m}\left(1-e^{-at}\right)+e^{-at}\left(t^m-\sum_{k=1}^m\frac{m!}{k!}t^ka^{k-m}\right). \tag{9.70}$$

Proof: for any $t\geq 0$, we must evaluate:

$$\int_0^\infty\tau^m f_l(\tau,t)d\tau=e^{-at}t^m+a\int_0^t\tau^me^{-a\tau}d\tau,$$

where we have used (eq. 9.69). In order to evaluate the exponential integral in the above equation we write:

$$a\int_0^t\tau^me^{-a\tau}d\tau=(-1)^ma\partial_a^m\left[(1-e^{-at})(a^{-1})\right]=(-1)^ma\sum_{k=0}^m\binom{m}{k}\partial_a^k(1-e^{-at})\partial_a^{m-k}(a^{-1})$$

$$=\sum_{k=0}^m(-1)^k\frac{m!}{k!}a^{k-m}\partial_a^k(1-e^{-at})=\frac{m!}{a^m}(1-e^{-at})-\sum_{k=1}^m\frac{m!}{k!}t^ka^{k-m}e^{-at}$$

thus one has (eq. 9.70). □

The function $f_l(\tau,t)$ can be written:

$$f_l(\tau,t)=e^{-at}\delta(\tau-t)+(1-e^{-at})\varphi(\tau,t),\quad \tau,t\geq 0,\quad a\geq 0, \tag{9.71}$$

where:

$$\varphi(\tau,t)=a\frac{e^{-a\tau}\theta(t-\tau)}{1-e^{-at}},\quad \tau,t\geq 0,\quad a\geq 0. \tag{9.72}$$

Because $f_l(\tau,t)$ is a probability density, then so is $\varphi(\tau,t)$. The corresponding random time process $l(t)$, $t\geq 0$, can then be chosen to be:

$$l(t)=b_tt+(1-b_t)j(t),\quad t\geq 0, \tag{9.73}$$

where b_t, $t\geq 0$, is a stochastic process such that, for any fixed $t\geq 0$, b_t is a Bernoulli random variable with $Pr(b_t=1)=e^{-at}$ and $Pr(b_t=0)=1-e^{-at}$, and $j(t)$, $t\geq 0$, is a stochastic process, independent of b_t, with marginal distribution given by $\varphi(\tau,t)$.

Remark 9.9. The random time $l(t)$ defined by (eq. 9.73) cannot be increasing everywhere. This is due to the fact that b_t and $j(t)$ are independent and $Pr(j(t) < t) = 1$ for any $t \geq 0$. Indeed, suppose that $l(t)$ is increasing. This implies that for any $t \geq 0$ and $\epsilon > 0$:

$$1 = Pr(l(t+\epsilon) \geq l(t) \big| \, b_t = 1) = Pr(l(t+\epsilon) \geq t)$$

$$= Pr(l(t+\epsilon) \geq t \big| \, b_{t+\epsilon} = 1) Pr(b_{t+\epsilon} = 1) + Pr(l(t+\epsilon) \geq t \big| \, b_{t+\epsilon} = 0) Pr(b_{t+\epsilon} = 0)$$

$$= e^{-a(t+\epsilon)} + \left(1 - e^{-a(t+\epsilon)}\right) Pr(j(t+\epsilon) \geq t)$$

$$= 1 - \left(1 - e^{-a(t+\epsilon)}\right) Pr(j(t+\epsilon) < t)$$

therefore taking $\epsilon \to 0$ we get $1 = e^{-at}$ with $a, t \geq 0$, which is a contradiction as soon as $a \neq 0$ and $t > 0$.

On the other hand, a trivial example of an increasing process with marginal distribution given by (eq. 9.69) is:

$$\bar{l}(t) = \min(X, t), \quad t \geq 0, \tag{9.74}$$

where X is an exponentially distributed random variable: $X \sim a e^{-a\tau}$, $\tau \geq 0$.

We now turn to (eq. 9.68). We have the following result:

Prop. 9.3. *The fundamental solution of (eq. 9.68) is:*

$$f(x,t) = e^{-at} G(x,t) + (1 - e^{-at}) \phi(x,t), \tag{9.75}$$

with:

$$\phi(x,t) = \frac{\sqrt{a}}{4(1-e^{-at})} \left\{ e^{x\sqrt{a}} \mathrm{Erf}\left(\frac{x}{2\sqrt{t}} + \sqrt{at}\right) - e^{-x\sqrt{a}} \mathrm{Erf}\left(\frac{x}{2\sqrt{t}} - \sqrt{at}\right) - 2\sinh(|x|\sqrt{a}) \right\}, \tag{9.76}$$

where $\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$ *and where* $\mathrm{Erf}(-x) = -\mathrm{Erf}(x)$.

Proof: by Theorem 9.2 and (eq. 9.72), the fundamental solution of (eq. 9.68) is:

$$f(x,t) = \int_0^\infty G(x,\tau) f_l(\tau,t) d\tau = e^{-at} G(x,t) + (1 - e^{-at}) \phi(x,t), \tag{9.77}$$

where

$$\phi(x,t) = \int_0^\infty G(x,t) \varphi(\tau,t) d\tau.$$

We have that:

$$\phi(x,t) = \frac{a}{1-e^{-at}} \int_0^\infty G(x,\tau) e^{-a\tau} \theta(t-\tau) d\tau.$$

One has to evaluate:

$$\chi(x,t) = \int_0^t \frac{e^{-x^2/4\tau} e^{-a\tau}}{\sqrt{4\pi\tau}} d\tau, \quad x \in \mathbb{R}, t \geq 0. \tag{9.78}$$

First we observe that:

$$\chi(0,t) = \int_0^t \frac{e^{-a\tau}}{\sqrt{4\pi\tau}} d\tau = \frac{1}{2\sqrt{a}} \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{at}} e^{-y^2} dy = \frac{1}{2\sqrt{a}} \mathrm{Erf}(\sqrt{at}).$$

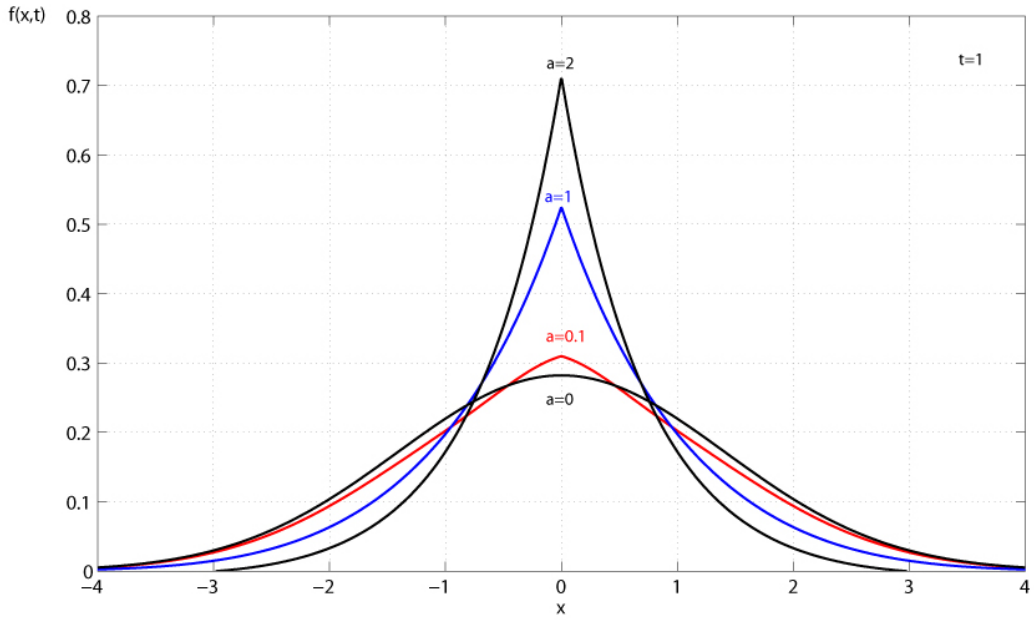

Figure 9.12: Plot of the fundamental solution $f(x,t)$ (eq. 9.75) at time $t = 1$, for different values of the parameter $a = [0, 0.1, 1, 2]$. When $a = 0$ we have the standard Gaussian density.

after the change of variables $y = \sqrt{a\tau}$. Because $\text{Erf}(-u) = -\text{Erf}(u)$ we can write:

$$\chi(0,t) = \frac{1}{4\sqrt{a}} \left\{ \text{Erf}(\sqrt{at}) - \text{Erf}(-\sqrt{at}) \right\}. \tag{9.79}$$

Now, for any $x \in \mathbb{R}$:

$$\chi(x,t) = \frac{1}{4\sqrt{a}} \left\{ e^{x\sqrt{a}} \,\text{Erf}\left(\frac{x}{2\sqrt{t}} + \sqrt{at}\right) - e^{-x\sqrt{a}} \,\text{Erf}\left(\frac{x}{2\sqrt{t}} - \sqrt{at}\right) \right\} - \frac{1}{2\sqrt{a}} \sinh(|x|\sqrt{a}), \tag{9.80}$$

because:

$$\begin{aligned}
&\frac{d}{d\tau}\left[\frac{1}{4\sqrt{a}}\left\{e^{x\sqrt{a}}\,\text{Erf}\left(\frac{x}{2\sqrt{\tau}} + \sqrt{a\tau}\right) - e^{-x\sqrt{a}}\,\text{Erf}\left(\frac{x}{2\sqrt{\tau}} - \sqrt{a\tau}\right)\right\}\right] \\
&= \frac{1}{4\sqrt{a}}\left\{\frac{2}{\sqrt{\pi}} e^{x\sqrt{a}} \exp\left(-\left[\frac{x}{2\sqrt{\tau}} + \sqrt{a\tau}\right]^2\right)\left(-\frac{x}{4}\tau^{-3/2} + \frac{\sqrt{a}}{2\sqrt{\tau}}\right)\right\} \\
&-\frac{1}{4\sqrt{a}}\left\{\frac{2}{\sqrt{\pi}} e^{-x\sqrt{a}} \exp\left(-\left[\frac{x}{2\sqrt{\tau}} - \sqrt{a\tau}\right]^2\right)\left(-\frac{x}{4}\tau^{-3/2} - \frac{\sqrt{a}}{2\sqrt{\tau}}\right)\right\} \\
&= \frac{1}{\sqrt{4\pi\tau}} \exp\left(-\frac{x^2}{4\tau} - a\tau\right).
\end{aligned}$$

Moreover, because $\text{Erf}(\pm\infty) = \pm 1$, we have (eq. 9.80), which actually reduces to (eq. 9.79) when $x = 0$. Therefore, the fundamental solution of (eq. 9.68) is:

$$f(x,t) = e^{-at}G(x,t) + \frac{\sqrt{a}}{4}\left\{e^{x\sqrt{a}}\,\text{Erf}\left(\frac{x}{2\sqrt{t}} + \sqrt{at}\right) - e^{-x\sqrt{a}}\,\text{Erf}\left(\frac{x}{2\sqrt{t}} - \sqrt{at}\right)\right\} - \frac{\sqrt{a}}{2}\sinh(|x|\sqrt{a}), \tag{9.81}$$

which can be rewritten as (eq. 9.75). □

Remark 9.10. With the choice (eq. 9.73) the process:

$$B(l(t)) = B(b_t t + (1-b_t)j(t)), \quad t \geq 0,$$

has marginal density (9.75). Observe that, for any $t \geq 0$:

$$B(b_t t + (1-b_t)j(t)) \stackrel{d}{=} b_t B(t) + (1-b_t)B(j(t)),$$

which naturally corresponds to (eq. 9.75).

Remark 9.11. We observe that (eq. 9.75) reduces to $G(x,t)$ when $a = 0$ (see Figure 9.12), which is as expected because the memory kernel disappears. For small times, the non-local memory effects are negligible and the process appears Markovian.

For large times we have:

$$\lim_{t\to\infty} f(x,t) = \lim_{t\to\infty} \phi(x,t) = \overline{\phi}(x),$$

where:

$$\overline{\phi}(x) = \frac{\sqrt{a}}{2}(\cosh(x\sqrt{a}) - \sinh(|x|\sqrt{a})) = \frac{\sqrt{a}}{2}e^{-|x|\sqrt{a}}, \quad x \in \mathbb{R}. \tag{9.82}$$

Remark 9.12. In view of (eq. 9.69), it is always possible to choose the random time process $l(t)$, $t \geq 0$, such that it becomes stationary at large times, in the sense of finite-dimensional densities. With this choice, the subordinated process $B(l(t))$ tends to a stationary process with asymptotic marginal distribution given by (eq. 9.82). For instance, if we look at (eq. 9.74), as $t \to \infty$ we have $\bar{l}(t) = X \sim ae^{-a\tau}$, $\tau \geq 0$. A less trivial example can be constructed by replacing the random variable X with a stationary process $X(t)$, $t \geq 0$, such that for each $t \geq 0$ the random variable $X(t)$ has an exponential distribution with mean $E(X(t)) = a^{-1}$. The resulting process $l(t) = \min(X(t), t)$ is not increasing, has marginal distribution defined by (eq. 9.69) and tends to $X(t)$ for large t.

Remark 9.13. To obtain an idea on how fast the stationary regime is reached, one can look at the variance of the subordinated process. Using (eq. 9.70) with $m = 1$, we find:

$$E(B(l(t))^2) = \frac{2}{a}(1 - e^{-at}), \tag{9.83}$$

which, for large times, tends exponentially to $2/a$ (i.e. the variance of eq. 9.82).

9.5.4 Exponential decay kernel with logarithmic scaling time

What happen if we choose an exponential kernel $K(t) = e^{-at}$ and a logarithmic scaling time? That is:

$$g(t) = \log(t+1), \quad t \geq 0. \tag{9.84}$$

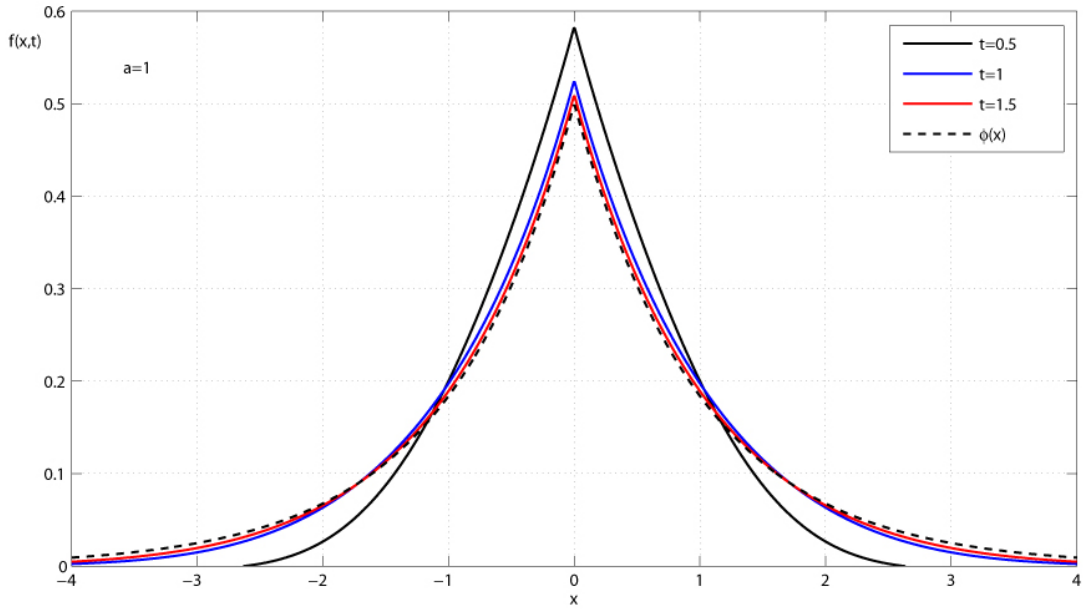

Figure 9.13: Plot of the fundamental solution $f(x,t)$ (eq. 9.75) at time $t = [0.5, 1, 1.5]$, and $a = 1$. The dashed line represents the asymptotic distribution $\overline{\varphi}(x)$ (eq. 9.82)

Since $g'(t)K(g(t) - g(s)) = (t+1)^{-a}(s+1)^{a-1}$, we get:

$$u(x,t) = u_0(x) + \frac{1}{(t+1)^a}\int_0^t (s+1)^{a-1}\partial_{xx}u(x,s)ds. \tag{9.85}$$

Its fundamental solution is:

$$f(x,t) = \frac{1}{(t+1)^a}G(x, \log(t+1)) - \frac{\sqrt{a}}{2}\sinh(|x|\sqrt{a})$$

$$+\frac{\sqrt{a}}{4}\left\{e^{x\sqrt{a}}\,\mathrm{Erf}\left(\frac{x}{2\sqrt{\log(t+1)}} + \sqrt{a\log(t+1)}\right) - e^{-x\sqrt{a}}\,\mathrm{Erf}\left(\frac{x}{2\sqrt{\log(t+1)}} - \sqrt{a\log(t+1)}\right)\right\} \tag{9.86}$$

Remark 9.14. As in Remark 9.12, consider a random time process $l(t)$, $t \geq 0$, with marginal distribution defined by (eq. 9.69), that becomes stationary for large times. The subordinated process $B(l(\log(t+1)))$, $t \geq 0$, has marginal density function defined by $f(x,t)$ of (eq. 9.86). Observe that in this case the random time process $l(\log(t+1))$ is no longer asymptotically stationary. This is because the translational time-invariance is broken by the logarithmic transformation. However, we can always consider a random time process $l^*(t)$, $t \geq 0$, with the same marginal distribution of $l(\log(t+1))$, which becomes stationary for large times. Thus, the process $B(l^*(t))$ has still marginal density function defined by $f(x,t)$ but becomes stationary as $t \to \infty$, in the sense of finite dimensional distribution, with asymptotic marginal distribution given by (eq. 9.82).

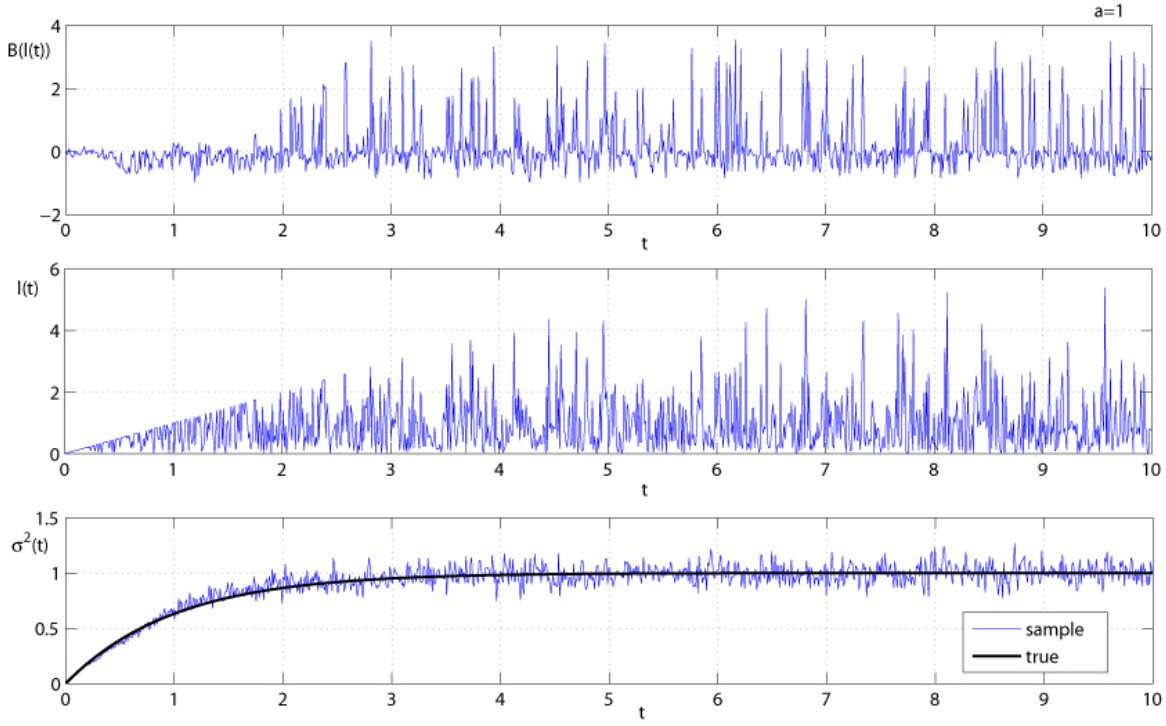

Figure 9.14: Trajectory of the process $B(l(t))$ (top panel), with $0 < t < 10$, $E(B(1)) = 1$, $l(t) = \min(t, X(t))$ where $X(t)$ is an exponential White Noise with mean one. The corresponding trajectory of the random time $l(t)$ process is presented in the middle panel. The estimated variance is computed on a sample of dimension $N = 500$. The smooth black line in the bottom panel corresponds to $\sigma^2(t)$ given by (eq. 9.83) and the stationary value is $\lim_{t\to\infty} \sigma^2(t) = 1$.

Remark 9.15. While $B(l(t)), t \geq 0$, satisfies (eq. 9.83) and thus has a variance which tends exponentially fast to the limit value $2/a$, here the stationary regime is reached more slowly. Indeed, the variance of the subordinated process is:

$$E(B(l^*(t))^2) = \frac{2}{a}\left(1 - \frac{1}{(t+1)^a}\right), \tag{9.87}$$

which, for large times, converges to the stationary value $2/a$ with a power-like behavior.

9.6 Examples involving other diffusions

We shall now consider examples of fractional and stretched Fokker-Planck equations involving other diffusion operators than $\mathcal{P}_x = \partial_{xx}$ which corresponds to standard Brownian motion. We will choose $K(t) = t^{\beta-1}/\Gamma(\beta)$ and $g(t) = t^{\alpha/\beta}$ as in Section 9.5.1 and consider the partial integro-differential equation:

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)}\frac{\alpha}{\beta}\int_0^t s^{\frac{\alpha}{\beta}-1}\left(t^{\frac{\alpha}{\beta}} - s^{\frac{\alpha}{\beta}}\right)^{\beta-1}\mathcal{P}_x u(x,s)ds, \quad 0 < \beta \leq 1, \quad \alpha > 0. \tag{9.88}$$

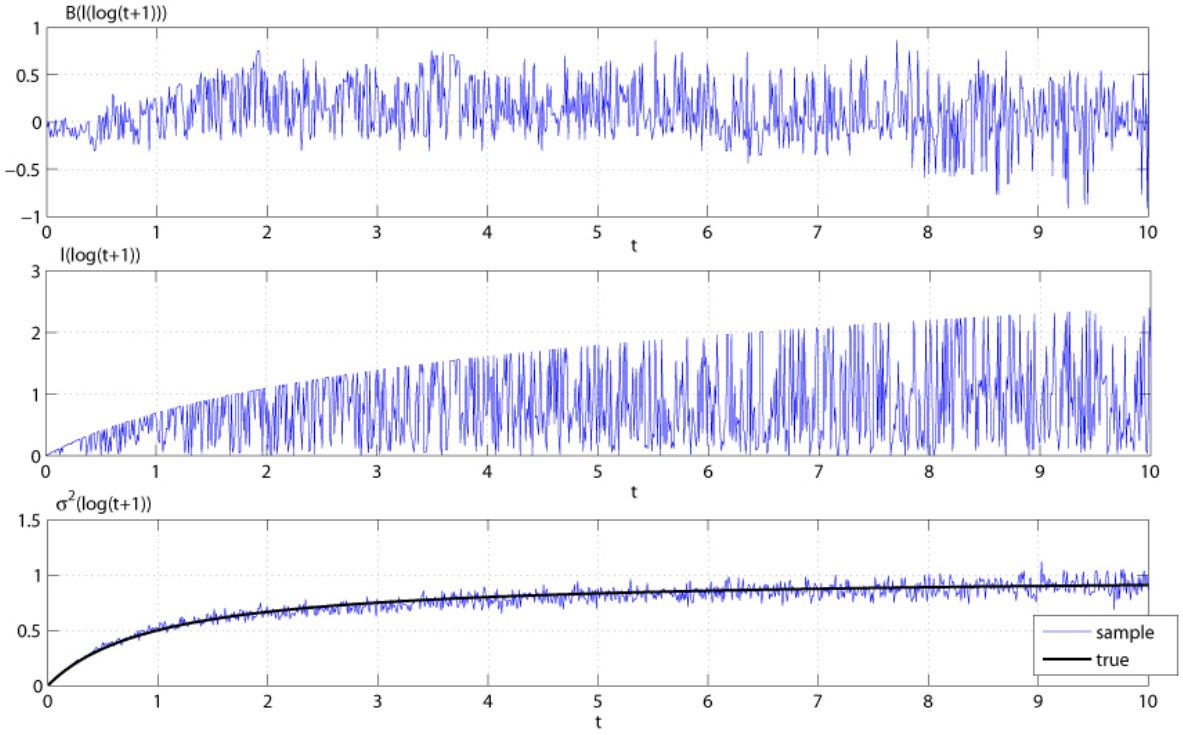

Figure 9.15: Trajectory of the process $B(l(\log(t+1)))$ (top panel), with $0 < t < 10$, $E(B(1)) = 1$, $l(t) = \min(t, X(t))$ where $X(t)$ is an exponential White Noise with mean one. The corresponding trajectory of the random time process $l(\log(t+1))$ is presented in the middle panel. The estimated variance is computed on a sample of dimension $N = 500$. The smooth black line in the bottom panel corresponds to (eq. 9.83). The stationary value is $\lim_{t\to\infty} \sigma^2(\log(t+1)) = 1$. The stationary regime is achieved more slowly than in the case of Figure 9.14.

Its fundamental solution is the marginal density of the process:

$$\mathcal{D}(t) = Q(l_\beta(t^{\alpha/\beta})), \tag{9.89}$$

where $Q(t)$, $t \geq 0$, is the stochastic diffusion associated to $\mathcal{P}_x$ and $l_\beta(t)$, $t \geq 0$, is a suitable self-similar random time process. One has the following particular cases:

- When $\alpha = \beta$ and $0 < \beta \leq 1$, (eq. 9.88) becomes the *"time-fractional" Fokker-Planck equation*:

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} \mathcal{P}_x u(x,s) ds, \quad 0 < \beta \leq 1, \tag{9.90}$$

whose fundamental solutions are the marginal distributions of the process:

$$\mathcal{D}(t) = Q(l_\beta(t)),$$

and are given by:

$$f_{\mathcal{D}}(x,t) = \int_0^\infty f_Q(x,\tau) f_{l_\beta}(\tau,t) d\tau, \tag{9.91}$$

where

$$f_{l_\beta}(\tau,t) = t^{-\beta} M_\beta(\tau t^{-\beta}), \quad \tau, t \geq 0, \quad 0 < \beta \leq 1 \tag{9.92}$$

and $f_Q(x,t)$ is the probability density of $Q(t)$.

- When $\beta = 1$ and $\alpha > 0$ we get a *"time-stretched" Fokker-Planck equation*:

$$u(x,t) = u_0(x) + \int_0^t \alpha s^{\alpha-1} \mathcal{P}_x u(x,s) ds, \quad 0 < \beta \leq 1. \tag{9.93}$$

In this case $f_l(\tau,t) = \delta(\tau - t^\alpha)$ and we get:

$$f_{\mathcal{D}}(x,t) = f_Q(x,t^\alpha),$$

which corresponds to the "stretched" diffusion:

$$\mathcal{D}(t) = Q(t^\alpha), \quad \alpha > 0.$$

- The case $\alpha = \beta = 1$ is trivial and corresponds merely to the Markovian case where the equation is:

$$u(x,t) = u_0(x) + \int_0^t \mathcal{P}_x u(x,s) ds$$

whose fundamental solution is the density function of $D(t) = Q(t)$, namely the Markovian process.

In the following subsections we study the above equations under particular choices of the Fokker-Planck operator $\mathcal{P}_x$. In all the cases considered, we also give the results involving the exponential decay kernel.

9.6.1 Brownian motion with drift

Let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ be given. Consider a linear diffusion $B^{(\mu,\sigma)} := B^{(\mu)}$ on $\mathbb{R}$ satisfying the stochastic differential equation (eq. 4.60 with a=0):

$$dB^{(\mu)}(t) = \mu dt + \sigma dB(t), \quad t \geq 0, \tag{9.94}$$

where $B(t)$, $t \geq 0$, is a "standard" Brownian motion. Then, $B^{(\mu)}(t)$, $t \geq 0$, is a Brownian motion with drift μ (eq. 4.63), namely

$$B^{(\mu)}(t) = \mu t + \sigma B(t), \quad t \geq 0. \tag{9.95}$$

The marginal density function of $B^{(\mu)}$(t), $t \geq 0$, is:

$$f_{B^{(\mu)}}(x,t) = \frac{1}{|\sigma|\sqrt{4\pi t}} \exp\left(-\frac{(x-\mu t)^2}{\sigma^2 4t}\right), \quad t \geq 0, \quad x \in \mathbb{R}, \tag{9.96}$$

which is the fundamental solution of the Fokker-Planck equation (eq. 4.61), with a=0 and with a "standard" Wiener process B_t, namely:

$$\partial_t u(x,t) = -\mu \partial_x u(x,t) + \sigma^2 \partial_{xx} u(x,t), \quad t \geq 0. \tag{9.97}$$

9.6.1.1 The β-power kernel.

We consider the *"fractional" Fokker-Planck equation* (see eq. 9.90):

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} \left(-\mu \partial_x u(x,s) + \sigma^2 \partial_{xx} u(x,s)\right) ds, \;\; 0 < \beta \leq 1. \tag{9.98}$$

Its fundamental solution can be regarded as the marginal density function of the process:

$$D(t) = B^{(\mu)}(l_\beta(t)), \;\; t \geq 0, \;\; 0 < \beta \leq 1, \tag{9.99}$$

where the process $l_\beta(t)$, $t \geq 0$, is a self-similar random time process with parameter $H = \beta/2$, independent of $B^{(\mu)}$, such that its marginal distribution is given by (eq. 9.92).

Prop. 9.4. *The fundamental solution of (eq. 9.98) is:*

$$f_D(x,t) = \int_0^\infty f_{B^{(\mu)}}(x,\tau) f_{l_\beta}(\tau,t) d\tau, \;\; t \geq 0, \quad x \in \mathbb{R},$$

i.e.

$$f_D(x,t) = \int_0^\infty \frac{1}{|\sigma|\sqrt{4\pi\tau}} \exp\left(-\frac{(x-\mu\tau)^2}{4\sigma^2\tau}\right) \mathcal{M}_\beta(\tau,t) d\tau, \;\; t \geq 0, \;\; x \in \mathbb{R}, \tag{9.100}$$

which is equal to:

$$f_D(x,t) = e^{\mu x/2\sigma^2} \frac{1}{2|\sigma|} \sum_{k=0}^\infty \frac{(-\mu^2 t^\beta/4\sigma^2)^k}{k!} t^{-\beta/2} H_{2,2}^{2,0}\left(|x\sigma^{-1}|t^{-\beta/2} \,\middle|\, \begin{matrix}(1/2,1/2),(1-\beta/2+\beta k,\beta/2)\\(0,1),(k+1/2,1/2)\end{matrix}\right), \tag{9.101}$$

where the Fox H-function is defined by (eq. 7.118).

Proof: In order to evaluate $f_D(x,t)$ we write:

$$f_D(x,t) = |\sigma|^{-1} e^{\mu x'/2\sigma} \int_0^\infty e^{-\mu^2\tau/4\sigma^2} G(x',\tau) \mathcal{M}_\beta(\tau,t) d\tau,$$

where $G(x,t)$ is the standard Gaussian density (eq. 9.9) and $x' = x/\sigma$. Then, we have to evaluate an integral of the form (see eq. 7.126):

$$\Phi(x,t) = \frac{1}{2} \int_0^\infty e^{-a\tau} \mathcal{M}_{1/2}(|x|,\tau) \mathcal{M}_\beta(\tau,t) d\tau, \quad x \in \mathbb{R}, \;\; t \geq 0, \;\; a \geq 0. \tag{9.102}$$

One has:

$$\begin{aligned}\Phi(x,t) &= \frac{1}{2} \int_0^\infty e^{-a\tau} \tau^{-1/2} M_{1/2}(|x|\tau^{-1/2}) t^{-\beta} M_\beta(\tau t^{-\beta}) d\tau \\ &= \frac{1}{2} \int_0^\infty \frac{1}{y} M_{1/2}\left(\frac{|x|}{y}\right) 2y e^{-ay^2} t^{-\beta} M_\beta(y^2 t^{-\beta}) dy.\end{aligned}$$

after th the change of variables $y = \sqrt{\tau}$. Because of the symmetry, it is enough to consider only the case $x \geq 0$. We get:

$$\Phi(x,t) = \frac{1}{2}(M_{1/2} \star Y_t)(x), \quad x \geq 0,$$

where

$$(\varphi \star \phi)(x) = \int_0^\infty \frac{1}{y} \varphi\left(\frac{x}{y}\right) \phi(y) dy$$

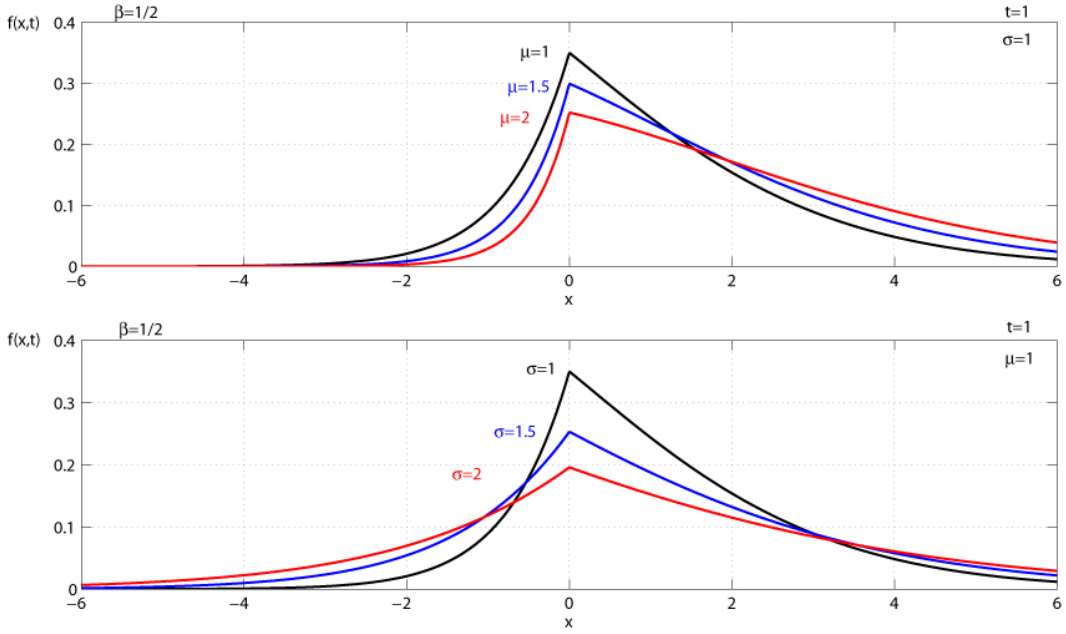

Figure 9.16: Plot of the fundamental solution $f(x,t)$ (eq. 9.101) with $\beta = 1/2$, at time $t = 1$, for different values of the parameters $\mu = [1, 1.5, 2]$ and $\sigma = [1, 1.5, 2]$.

indicates the Mellin convolution and where:

$$Y_t(x) = 2xe^{-ax^2}t^{-\beta}M_\beta(x^2t^{-\beta}), \quad x \geq 0, \quad t \geq 0. \tag{9.103}$$

Using the Mellin convolution theorem we get:

$$\mathcal{M}\{2\Phi(x,t);x,u\} = \mathcal{M}\{M_{1/2}(x);x,u\}\mathcal{M}\{Y_t(x);x,u\}. \tag{9.104}$$

Because of $M_{1/2}(x) = H^{1,0}_{1,1}\left(x\middle| \begin{matrix}(1,1/2)\\(1,1)\end{matrix}\right)$ (see eq. 7.121), and by using (eq. 7.118), this can be written as:

$$\mathcal{M}\{2\Phi(x,t);x,u\} = \frac{\Gamma(u)}{\Gamma(1/2+u/2)}\mathcal{M}\{Y_t(x);x,u\}. \tag{9.105}$$

We now evaluate:

$$\mathcal{M}\{Y_t(x);x,u\} = \int_0^\infty e^{-ax^2}2xt^{-\beta}M_\beta(x^2t^{-\beta})x^{u-1}dx.$$

After the change of variables $x^2t^{-\beta} = z$, we get

$$\mathcal{M}\{Y_t(x);x,u\} = \int_0^\infty (zt^\beta)^{\frac{1}{2}(u-1)}e^{-azt^\beta}M_\beta(z)dz$$

$$= t^{\frac{\beta}{2}(u-1)}\sum_{k=0}^{\infty}\frac{(-at^\beta)^k}{k!}\int_0^\infty z^{k-\frac{1}{2}+\frac{u}{2}}M_\beta(z)dz$$

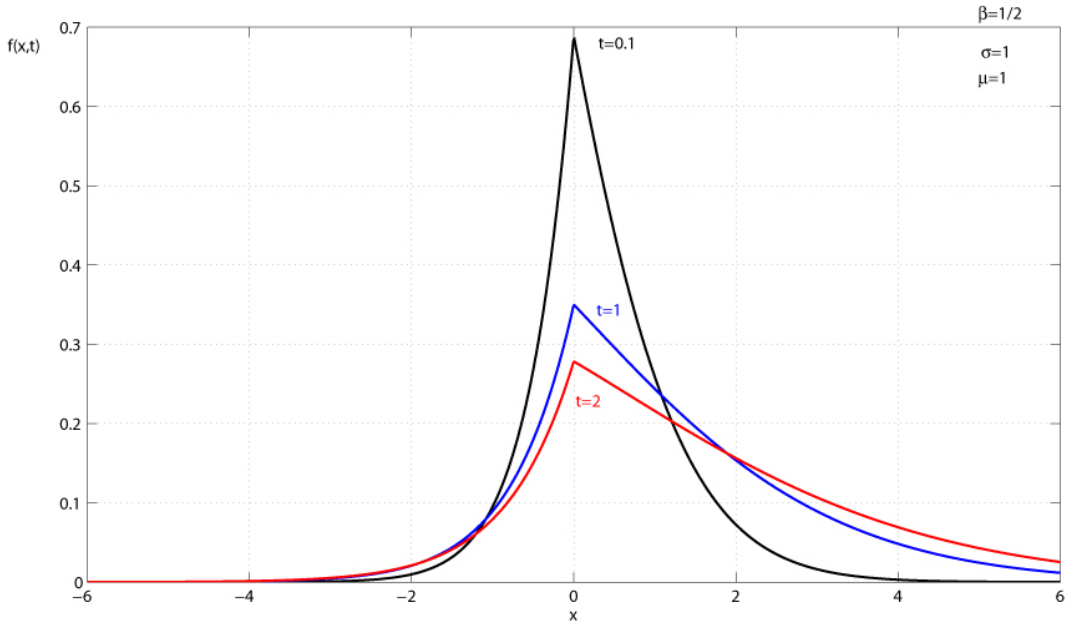

Figure 9.17: Plot of the fundamental solution $f(x,t)$ (eq. 9.101) with $\beta = 1/2$, $\mu = 1$, $\sigma = 1$, at times $t = [0.1, 1, 2]$.

$$= t^{\frac{\beta}{2}(u-1)} \sum_{k=0}^{\infty} \frac{(-at^{\beta})^k}{k!} \mathcal{M}\{M_{\beta}(x); x, k + 1/2 + u/2\}$$

$$= t^{\frac{\beta}{2}(u-1)} \sum_{k=0}^{\infty} \frac{(-at^{\beta})^k}{k!} \frac{\Gamma(1/2 + k + u/2)}{\Gamma(1 + \beta k - \beta/2 + \beta u/2)},$$

where we have used again (eq. 7.121) and (eq. 7.118). Thus:

$$\mathcal{M}\{2\Phi(x,t); x, u\} = \sum_{k=0}^{\infty} \frac{(-at^{\beta})^k}{k!} t^{\frac{\beta}{2}(u-1)} \frac{\Gamma(u)\Gamma(1/2 + k + u/2)}{\Gamma(1/2 + u/2)\Gamma(1 + \beta k - \beta/2 + \beta u/2)}.$$

Inverting the Mellin transform, (eq. 7.118) gives:

$$\Phi(x,t) = \frac{1}{2} \sum_{k=0}^{\infty} \frac{(-at^{\beta})^k}{k!} t^{-\beta/2} H_{2,2}^{2,0}\left(|x| t^{-\beta/2} \middle| \begin{matrix} (1/2, 1/2), (1 - \beta/2 + \beta k, \beta/2) \\ (0,1), (k + 1/2, 1/2) \end{matrix} \right), \tag{9.106}$$

with $x \in \mathbb{R}$ and $t \geq 0$. Therefore, the fundamental solution of (eq. 9.90) can be expressed as:

$$f_D(x,t) = e^{\mu x/2\sigma^2} \frac{1}{2|\sigma|} \sum_{k=0}^{\infty} \frac{(-\mu^2 t^{\beta}/4\sigma^2)^k}{k!} t^{-\beta/2} H_{2,2}^{2,0}\left(|x\sigma^{-1}| t^{-\beta/2} \middle| \begin{matrix} (1/2, 1/2), (1 - \beta/2 + \beta k, \beta/2) \\ (0,1), (k + 1/2, 1/2) \end{matrix} \right),$$

that is (eq. 9.101). □

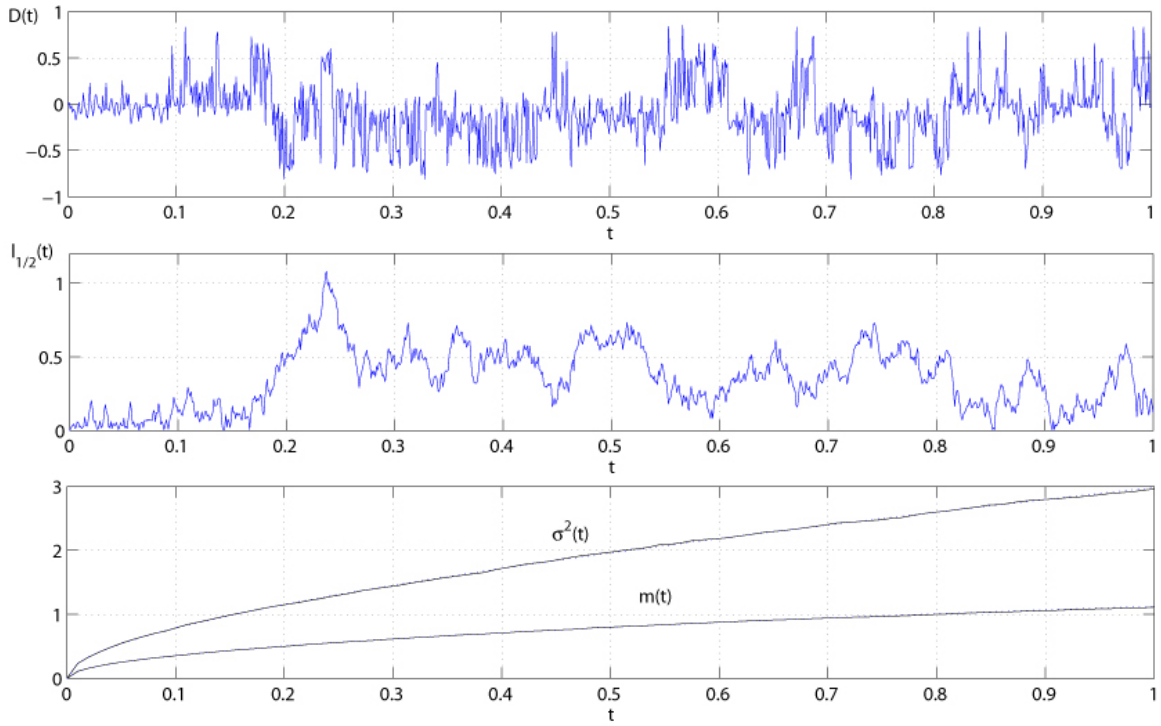

Figure 9.18: Trajectory of the process $D(t) = B^{(\mu)}(l_{1/2}(t))$ defined in (eq. 9.99) with $\beta = 1/2$ (top panel). The random time process is $l_{1/2}(t) = |b(t)|$, where $b(t)$ is a "standard" Brownian motion (middle panel). The variance and the mean are evaluated over a sample of size $N = 5 \cdot 10^4$ and fit the theoretical values (bottom panel).

When $\mu = 0$ and $\sigma = 1$, (eq. 9.101) reduces to:

$$f_D(x,t) = \frac{1}{2} t^{-\beta/2} H^{2,0}_{2,2}\left(|x| t^{-\beta/2} \,\middle|\, \begin{matrix} (1/2,1/2),(1-\beta/2,\beta/2) \\ (0,1),(1/2,1/2) \end{matrix} \right),$$

that is, using the reduction formula for the Fox H-function [18],

$$f_D(x,t) = \frac{1}{2} t^{-\beta/2} H^{1,0}_{1,1}\left(|x| t^{-\beta/2} \,\middle|\, \begin{matrix} (1-\beta/2,\beta/2) \\ (0,1) \end{matrix} \right) = \frac{1}{2} \mathcal{M}_{\beta/2}(|x|,t).$$

As expected, we recover in this case the fundamental solution of the time fractional diffusion equation (eq. 9.59).

Moreover, if we set $\beta = 1$ in (eq. 9.101) we have:

$$f_D(x,t) = e^{\mu x/2\sigma^2} \frac{1}{2|\sigma|} \sum_{k=0}^{\infty} \frac{(-\mu^2 t/4\sigma^2)^k}{k!} t^{-1/2} H^{2,0}_{2,2}\left(|x\sigma^{-1}| t^{-1/2} \,\middle|\, \begin{matrix} (1/2,1/2),(1/2+k,1/2) \\ (0,1),(1/2+k,1/2) \end{matrix} \right),$$

$$= e^{\mu x/2\sigma^2} \sum_{k=0}^{\infty} \frac{(-\mu^2 t/4\sigma^2)^k}{k!} \frac{1}{2|\sigma|} t^{-1/2} H^{1,0}_{1,1}\left(|x\sigma^{-1}| t^{-1/2} \,\middle|\, \begin{matrix} (1/2,1/2) \\ (0,1) \end{matrix} \right) = \frac{1}{|\sigma|\sqrt{4\pi t}} \exp\left(-\frac{(x-\mu t)^2}{4\sigma^2 t} \right),$$

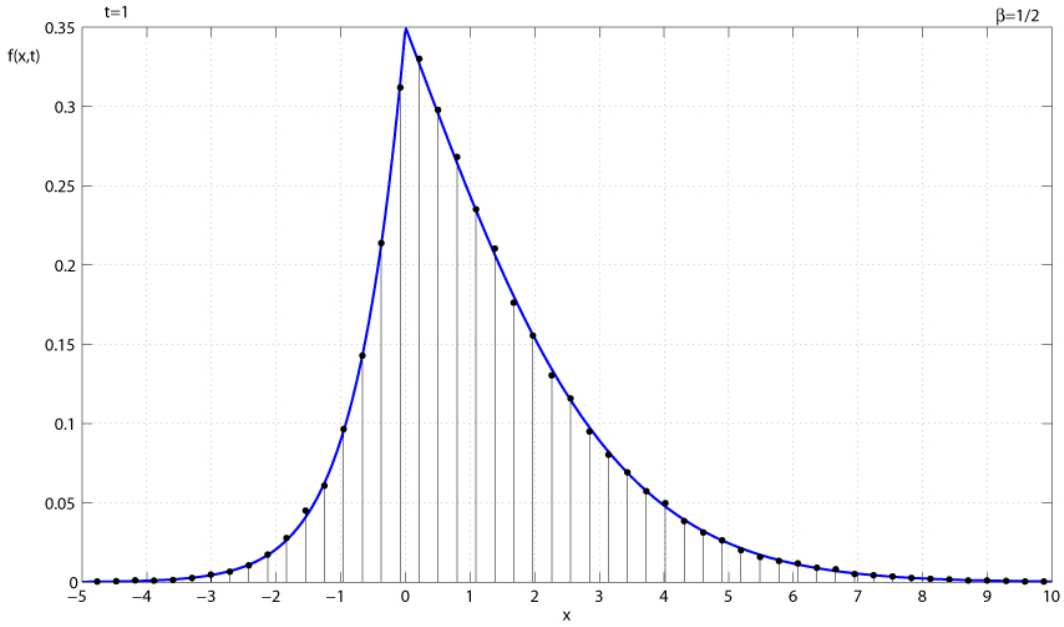

Figure 9.19: Marginal density function $f(x,t)$ of the process $B^{(\mu)}(l_{1/2}(t))$ at time $t = 1$ (eq. 9.101). The histogram is evaluated over $N = 10^5$ simulated trajectories (see Figure 9.18).

and we recover $f_{B^{(\mu)}}(x,t)$.

In Figure 9.16 and Figure 9.17 we have used (eq. 9.100) to plot the fundamental solution (eq. 9.101) with $\beta = 1/2$ for different values of the parameters μ and σ at fixed time and, for fixed parameters, at different times t. As expected, the fundamental solution is not symmetric in space with a time-growing skewness. Moreover, due to the presence of the positive taken drift term ($\mu = 1$), the probability to find the particle in the positive semi-axis increases with time (fig. 9.17).

In Figure 9.18 is presented a trajectory of the process $D(t) = B^{(\mu)}(l_\beta(t))$ with $\beta = 1/2$. Using (eq. 9.60) it is easy to write all the moments of the process:

$$E(D(t)^m) = \sum_{k=0}^{[m/2]} \binom{m}{2k} \frac{2k!(m-k)!}{k!} \sigma^{2k} \mu^{m-2k} \frac{t^{\beta(m-k)}}{\Gamma(\beta(m-k)+1)}, \quad 0 < \beta \leq 1, \tag{9.107}$$

where m is an integer greater than zero and $[a]$ indicates the integer part of a. Therefore, we have:

$$m(t) = E(D(t)) = \mu \frac{t^\beta}{\Gamma(\beta+1)}, \tag{9.108}$$

and

$$\sigma^2(t) = E(D(t)^2) - m(t)^2 = 2\mu^2 \frac{t^{2\beta}}{\Gamma(2\beta+1)} - \mu^2 \frac{t^{2\beta}}{\Gamma(\beta+1)^2} + 2\sigma^2 \frac{t^\beta}{\Gamma(\beta+1)}. \tag{9.109}$$

In the bottom panel of Figure 9.18 the mean and the variance have been estimated from a sample of trajectories of the process $B^{(\mu)}(l_{1/2})(t)$. Then, they have been compared with the theoretical values

given above. In Figure 9.19 we compare the theoretical density function $f(x,t)$ given by (eq. 9.101) at time $t = 1$ with an histogram evaluated over a sample of $N = 10^5$ trajectories.

9.6.1.2 Exponential-decay kernel

The *exponential decay kernel* case is straightforward. The non-Markovian Fokker-Planck equation is:

$$u(x,t) = u_0(x) + \int_0^t e^{-a(t-s)}(-\mu\partial_x u(x,s) + \sigma^2\partial_{xx}u(x,s))ds, \quad a \geq 0. \tag{9.110}$$

If we indicate by $\mathcal{G}(x,t)$ the fundamental solution of the Markovian equation; i.e. (eq. 9.96)

$$\mathcal{G}(x,t) = \frac{1}{|\sigma|\sqrt{4\pi t}}\exp\left(-\frac{(x-\mu t)^2}{4\sigma^2 t}\right), \quad t \geq 0, \quad x \in \mathbb{R}, \tag{9.111}$$

then, using (eq. 9.69), the fundamental solution of (eq. 9.110) is:

$$f(x,t) = e^{-at}\mathcal{G}(x,t) + (1-e^{-at})\Phi(x,t), \tag{9.112}$$

where:

$$\Phi(x,t) = \frac{a}{1-e^{-at}}e^{\mu x/2\sigma^2}\int_0^t \frac{e^{-x^2/4\sigma^2\tau}e^{-(a+\mu^2/4\sigma^2)\tau}}{|\sigma|\sqrt{4\pi\tau}}d\tau.$$

Using (eq. 9.78) and (eq. 9.80) we have:

Prop. 9.5. *The fundamental solution of (eq. 9.110) is:*

$$f(x,t) = e^{-at}\mathcal{G}(x,t) - \frac{ae^{\frac{\mu}{2\sigma^2}x}}{2|\sigma|\sqrt{a+\frac{\mu^2}{4\sigma^2}}}\sinh\left(|x\sigma^{-1}|\sqrt{a+\frac{\mu^2}{4\sigma^2}}\right) +$$

$$+\frac{ae^{\frac{\mu}{2\sigma^2}x}}{4|\sigma|\sqrt{a+\frac{\mu^2}{4\sigma^2}}}\left\{\exp\left(\frac{x}{|\sigma|}\sqrt{a+\frac{\mu^2}{4\sigma^2}}\right)\mathrm{Erf}\left(\frac{x}{2|\sigma|\sqrt{t}}+\sqrt{\left(a+\frac{\mu^2}{4\sigma^2}\right)t}\right)\right.$$

$$\left.-\exp\left(-\frac{x}{|\sigma|}\sqrt{a+\frac{\mu^2}{4\sigma^2}}\right)\mathrm{Erf}\left(\frac{x}{2|\sigma|\sqrt{t}}-\sqrt{\left(a+\frac{\mu^2}{4\sigma^2}\right)t}\right)\right\}. \tag{9.113}$$

When $t \to \infty$ we obtain the stationary distribution:

$$\overline{\phi}(x) = \frac{ae^{\frac{\mu}{2\sigma^2}x}}{2|\sigma|\sqrt{a+\frac{\mu^2}{4\sigma^2}}}\left(\cosh\left(x|\sigma^{-1}|\sqrt{a+\frac{\mu^2}{4\sigma^2}}\right)-\sinh\left(|x\sigma^{-1}|\sqrt{a+\frac{\mu^2}{4\sigma^2}}\right)\right)$$

that is:

$$\overline{\phi}(x) = \frac{a}{2|\sigma|\sqrt{a+\frac{\mu^2}{4\sigma^2}}}\exp\left(\mu x/2\sigma^2 - |x\sigma^{-1}|\sqrt{a+\frac{\mu^2}{4\sigma^2}}\right). \tag{9.114}$$

9.6.2 Geometric Brownian motion

Let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ be given. Consider a linear diffusion S on $\mathbb{R}$ defined by the stochastic differential equation (eq. 4.65):

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t), \quad t \geq 0, \tag{9.115}$$

where $B(t)$, $t \geq 0$, is a "standard" Brownian motion. The process $S(t)$, $t \geq 0$, is a Geometric Brownian motion. If S starts in x_0 at time $t = 0$ (i.e. $P(S(0) = x_0) = 1$), then a solution of (eq. 9.115) is (eq. 4.66), namely

$$S(t) = x_0 \exp\left[(\mu - \sigma^2/2)t + \sigma B(t)\right], \quad t \geq 0, \quad x_0 > 0. \tag{9.116}$$

The marginal density function of $S(t)$ is the log-normal distribution:

$$f_S(x,t) = \frac{1}{x|\sigma|\sqrt{4\pi t}} \exp\left(-\frac{\left(\log(x/x_0) - (\mu - \sigma^2/2)t\right)^2}{\sigma^2 4t}\right), \quad t \geq 0, \quad x \geq 0. \tag{9.117}$$

The function $f_S(x,t)$ is a solution of the Fokker-Planck equation (eq. 4.67):

$$\partial_t u(x,t) = \left[(2\sigma^2 - \mu) + (4\sigma^2 - \mu)x\partial_x + \sigma^2 x^2 \partial_{xx}\right] u(x,t), \quad x \geq 0, \tag{9.118}$$

with deterministic initial condition

$$u_0(x) = \delta(x - x_0), \quad x \geq 0, \quad x_0 > 0. \tag{9.119}$$

9.6.2.1 β-power kernel

If we introduce the β-power kernel $K(t) = \Gamma(\beta)^{-1} t^{\beta-1}$, $0 < \beta \leq 1$, in this setting we obtain the following *"fractional" Fokker-Planck* equation:

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)} \int_0^t (t-s)^{\beta-1} \left[(2\sigma^2 - \mu) + (4\sigma^2 - \mu)x\partial_x + \sigma^2 x^2 \partial_{xx}\right] u(x,s)ds, \quad x \geq 0. \tag{9.120}$$

A solution of the above equation with initial condition given by (eq. 9.119) is given by (see Corollary 9.2):

$$f_D(x,t) = \int_0^\infty f_S(x,\tau)\mathcal{M}_\beta(\tau,t)d\tau \tag{9.121}$$

which is the marginal distribution of the process

$$D(t) = S(l_\beta(t)), \quad t \geq 0, \quad 0 < \beta \leq 1, \tag{9.122}$$

starting almost surely in $x_0 > 0$, where $l_\beta(t)$, $t \geq 0$, is a self-similar random time process with $H = \beta/2$, independent of the geometric Brownian motion $S(t)$ and with marginal density function given by (eq. 9.92). It is easy to see that:

$$f_D(x,t) = \frac{1}{x|\sigma|} \exp\left(\frac{\log(x/x_0)(\mu - \sigma^2/2)}{4\sigma^2}\right) \frac{1}{2} \int_0^\infty e^{-a\tau} \mathcal{M}_{1/2}(|x'|,\tau)\mathcal{M}_\beta(\tau,t)d\tau,$$

where:

$$a = (\mu - \sigma^2/2)^2/4\sigma^2, \quad x' = \log(x/x_0)/\sigma.$$

We have the same integral as in (eq. 9.102). Therefore:

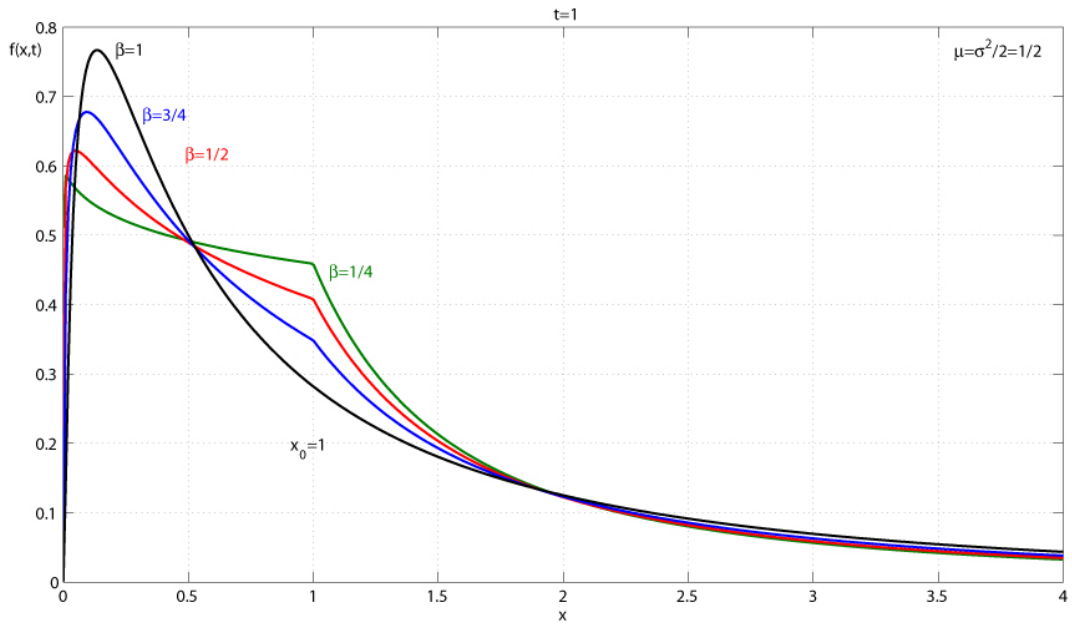

Figure 9.20: Plot of the fundamental solution $f(x,t)$ (eq. 9.123) at time $t = 1$, when $\mu = \sigma^2/2 = 1$ (eq. 9.124), $x_0 = 1$, for different values of the parameter $\beta = [1/4, 1/2, 3/4, 1]$. For $\beta = 1$ $f(x,t)$ reduces to the log-normal density (eq. 9.117). The angular point corresponds to the initial value $x_0 = 1$ and is due to the presence of $|\log(x/x_0)|$ in the solution.

Prop. 9.6. *for each $t \geq 0$:*

$$f_D(x,t) = \frac{1}{x|\sigma|}\exp\left(\frac{\log(x/x_0)(\mu-\sigma^2/2)}{4\sigma^2}\right)\times$$

$$\times\sum_{k=0}^{\infty}\frac{1}{k!}\left(-\frac{(\mu-\sigma^2/2)^2t^\beta}{4\sigma^2}\right)^k t^{-\frac{\beta}{2}}H^{2,0}_{2,2}\left(|x'|t^{-\frac{\beta}{2}}\,\middle|\,\begin{matrix}(1/2,1/2),(1-\beta/2+\beta k,\beta/2)\\(0,1),(k+1/2,1/2)\end{matrix}\right). \tag{9.123}$$

This result can be obtained directly from (eq. 9.101) because:

$$D(t) = x_0\exp(B^{(\mu')}(l_\beta(t))),$$

where $B^{(\mu')}$ is a Brownian motion with drift $\mu' = (\mu - \sigma^2/2)$. When $\beta = 1$ we recover (eq. 9.117). Moreover, if $\mu = \sigma^2/2$ (i.e. $\mu' = 0$) we have (see Figure 9.20):

$$f_D(x,t) = \frac{1}{x|\sigma|}t^{-\beta/2}M_{\beta/2}\left(\left|\frac{\log(x/x_0)}{\sigma}\right|t^{-\beta/2}\right),\quad x\geq 0,\quad t\geq 0, \tag{9.124}$$

which is the marginal probability density of:

$$D(t) = x_0e^{\sigma B(l_\beta(t))},\quad t\geq 0.$$

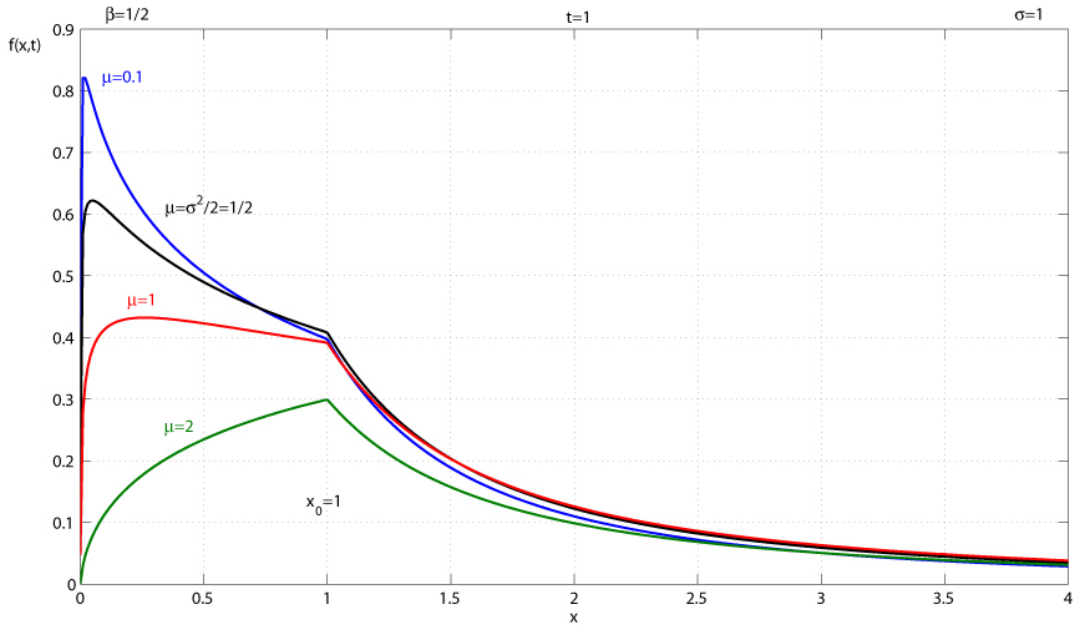

Figure 9.21: Plot of the fundamental solution $f(x,t)$ (eq. 9.123) with $\beta = 1/2, \sigma = 1, x_0 = 1$, at time $t = 1$, for different values of the parameter $\mu = [0.1, 1/2, 1, 2]$. For $\mu = 1/2$ we have (eq. 9.124), see also Figure 9.20.

- In Figure 9.20 we show the plot of the fundamental solution $f(x,t)$ in the particular case given by (eq. 9.124). Here we can see the behavior of the solution varying the parameter β. For $\beta = 1$ we recover the log-normal density (eq. 9.117) with $\mu = \sigma^2/2$.
- The behavior of the solution varying the drift parameter μ for fixed $\beta = 1/2$, $t = 1$, $\sigma = 1$ and $x_0 = 1$ is shown in Figure 9.21.
- In Figure 9.22 can be observed the time evolution of the fundamental solution for fixed parameters $\beta = 1/2$ and $\beta = 1/4$.
- Finally, in Figure 9.23 we can see a trajectory of the process $D(t) = S(l_\beta(t))$ with $\beta = 1/2$ (see eq. 9.122).

We shall now compute the mean and the variance of the process $D(t)$. We have that:

$$E(S(t)) = E\left(x_0 \exp\left[(\mu - \sigma^2/2)t + \sigma B(t))\right]\right) = x_0 \exp\left[(\mu - \sigma^2/2)t\right] E\left(e^{\sigma B(t)}\right).$$

Therefore, because:

$$E(e^{\sigma B(t)}) = \int_{\mathbb{R}} e^{\sigma x} G(x,t)dx = e^{\sigma^2 t} \frac{1}{\sqrt{4\pi t}} \int_{\mathbb{R}} e^{-\frac{(x-2\sigma t)^2}{4t}} dx,$$

we have:

$$E(S(t)) = x_0 \exp\left[(\mu + \sigma^2/2)t\right]. \tag{9.125}$$

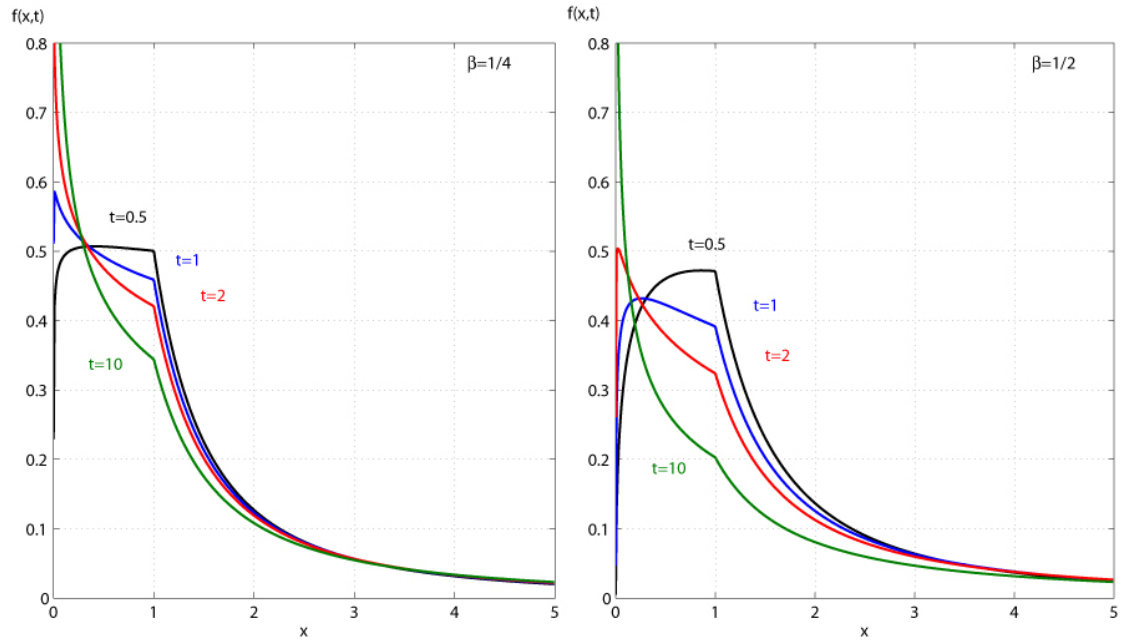

Figure 9.22: Plot of the fundamental solution $f(x,t)$ (eq. 9.123) with $\beta = 1/4$ (left panel) and $\beta = 1/2$ (right panel), at different times $t = [0.5, 1, 2, 10]$.

In the same way one has:

$$E(S(t)^2) = x_0^2 \exp\left[(2\mu + 3\sigma^2)t\right]. \tag{9.126}$$

Using the above equations we have:

$$E(S(l_\beta(t))) = x_0 E\left(\exp\left[(\mu + \sigma^2/2)l_\beta(t)\right]\right) = x_0 \sum_{k=0}^{\infty} \frac{(\mu + \sigma^2/2)^k}{k!} E(l_\beta(t)^k),$$

which, using (eq. 9.60), becomes:

$$E(S(l_\beta(t))) = x_0 \sum_{k=0}^{\infty} \frac{\left((\mu + \sigma^2/2)t^\beta\right)^k}{\Gamma(\beta k + 1)} = x_0 E_\beta((\mu + \sigma^2/2)t^\beta),$$

where $E_\beta(z) = \sum_{k=0}^{\infty} z^k/\Gamma(\beta k + 1)$ is the Mittag-Leffler function of order β (see Section 7.2.2). Similarly:

$$E(S(l_\beta(t))^2) = x_0^2 E\left(\exp\left[(2\mu + 3\sigma^2)l_\beta(t)\right]\right) = x_0^2 E_\beta((2\mu + 3\sigma^2)t^\beta).$$

Finally one has:

$$\begin{cases} m(t) = E(D(t)) = x_0 E_\beta((\mu + \sigma^2/2)t^\beta) \\ \sigma^2(t) = E(D(t)^2) - m(t)^2 = x_0^2 \left[E_\beta((2\mu + 3\sigma^2)t^\beta) - E_\beta((\mu + \sigma^2/2)t^\beta)^2\right]. \end{cases} \tag{9.127}$$

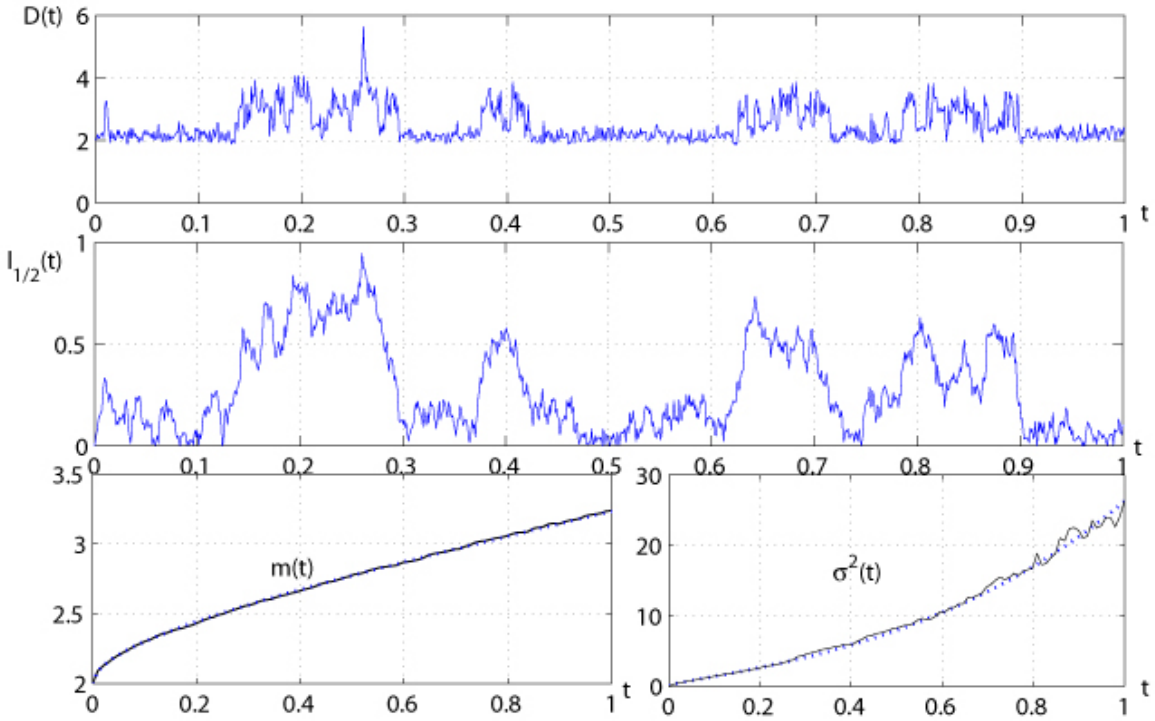

Figure 9.23: Trajectory of the process $D(t) = S(l_{1/2}(t))$ defined in (eq. 9.122) with $\beta = 1/2$ (top panel). The random time process is $l_{1/2}(t) = |b(t)|$ (middle panel). The variance and the mean are evaluated over a sample of size $N = 5 \cdot 10^4$ and are presented together with the theoretical functions (eq. 9.127) in the bottom panels.

9.6.2.2 Exponential-decay kernel

We now consider the *exponential decay kernel* $K(t) = e^{-at}$, $a \geq 0$. The non-Markovian Fokker-Planck equation is:

$$u(x,t) = u_0(x) + \int_0^t e^{-a(t-s)} \left[(2\sigma^2 - \mu) + (4\sigma^2 - \mu) x \partial_x + \sigma^2 x^2 \partial_{xx} \right] u(x,s) ds, \quad a \geq 0. \tag{9.128}$$

We denote by $\mathcal{G}(x,t)$ the fundamental solution of the Markovian equation; namely (eq. 9.117)

$$\mathcal{G}(x,t) = \frac{1}{x|\sigma|\sqrt{4\pi t}} \exp\left(- \frac{\left(\log(x/x_0) - (\mu - \sigma^2/2)t \right)^2}{\sigma^2 4t} \right), \quad x, t \geq 0. \tag{9.129}$$

Then, using (eq. 9.69), the fundamental solution of (eq. 9.110) is:

$$f(x,t) = e^{-at}\mathcal{G}(x,t) + (1 - e^{-at})\Phi(x,t), \tag{9.130}$$

where:

$$\Phi(x,t) = \frac{1}{1 - e^{-at}} \left[\frac{a}{|\sigma|x} e^{\log(\frac{x}{x_0})(\frac{2\mu - \sigma^2}{4\sigma^2})} \int_0^t G(x',\tau) e^{-a'\tau} d\tau \right], \quad x \geq 0$$

and:

$$\begin{cases} a' = \frac{(\mu-\sigma^2/2)^2}{4\sigma^2} + 4a, \quad a, \mu \geq 0, \ \ \sigma > 0, \\ x' = \log(x/x_0)/\sigma, \quad x \geq 0. \end{cases} \tag{9.131}$$

Thus, as in (eq. 9.81), we have:

Prop. 9.7.

$$f(x,t) = e^{-at}\mathcal{G}(x,t)$$
$$+\frac{a}{|\sigma|x}e^{\log(\frac{x}{x_0})(\frac{2\mu-\sigma^2}{4\sigma^2})}\left\{\frac{1}{4\sqrt{a'}}\exp(x'\sqrt{a'})\mathrm{Erf}\left(\frac{x'}{2\sqrt{t}}+\sqrt{a't}\right) - \exp(-x'\sqrt{a'})\mathrm{Erf}\left(\frac{x'}{2\sqrt{t}}-\sqrt{a't}\right)\right\}$$
$$-\frac{a}{4|\sigma|x\sqrt{a'}}e^{\log(\frac{x}{x_0})(\frac{2\mu-\sigma^2}{4\sigma^2})}\sinh(|x'|\sqrt{a'}). \tag{9.132}$$

The stationary distribution, obtained as $t \to \infty$, is:

$$\overline{\Phi}(x) = \frac{a}{4|\sigma|x\sqrt{a'}}\exp\left(\log\left(\frac{x}{x_0}\right)\left(\frac{2\mu-\sigma^2}{4\sigma^2}\right)\right)(\cosh(x'\sqrt{a'}) - \sinh(|x'|\sqrt{a'}). \tag{9.133}$$

Bibliography

[1] E. Barkai, CTRW pathways to the fractional diffusion equation, *Chem. Phys.* **284** (2002) 13–27.

[2] E. Barkai, R. Metzler and J. Klafter, From continuous time random walk to fractional Fokker-Planck equation, *Phys. Rev. E* **61** (2000) 132-138.

[3] L. Beghin and E. Orsingher, The distribution of the local time for "pseudoprocesses" and its connection with fractional diffusion equations, *Stochastic Process. Appl.* **115** (2005) 1017–1040.

[4] S. Bochner, *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley, 1955.

[5] S. Bochner, Subordination of non-Gaussian stochastic processes, *Proc. Nat. Acad. Sciences, USA* **48** (1962) 19–22.

[6] M. Dentz et al. Time behavior of solute transport in heterogeneous media: transition from anomalous to normal transport, *Adv. Water Resources* **27** (2004), 155–173.

[7] C. Donati-Martin, B. Roynette, P. Vallois and M. Yor, On constants related to the choice of the local time at 0, and the corresponding Itô measure for Bessel processes with dimension $d = 2(1-\alpha)$; $0 < \alpha < 1$, *Studia Scientiarum Mathematicarum Hungarica*, to appear (2007). [E-print `http://hal.archives-ouvertes.fr/hal-00141513/en/`]

[8] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, 3-rd edn., Wiley, New York, 1968. [1-st edn. 1957]

[9] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2-nd edn., Wiley, New York, 1971. [1-st edn. 1966]

[10] R. Gorenflo, Yu. Luchko and F. Mainardi, Analytical properties and applications of the Wright function, *Fractional Calculus and Applied Analysis* **2** (1999) 383-414. [E-print `http://arxiv.org/abs/math-ph/0701069`]

[11] R. Gorenflo and F. Mainardi, Fractional diffusion processes: Probability distributions and continuous time random walk, in: G. Rangarajan and M. Ding (Editors), *Processes with Long Range Correlations*, Springer Verlag, Berlin, 2003, pp. 148-161. [Lecture Notes in Physics, No 621]

[12] P. Grigolini, A. Rocco and B.J. West, Fractional calculus as a macroscopic manifestation of randomness, *Phys. Rev. E* **59** (1999) 2603-2613.

[13] R. Hilfer and L. Anton, Fractional master equations and fractal time random walks. *Phys. Rev. E* **51** (1995) R848–R851.

[14] V. M. Kenkre, R. S. Knox, Generalized-master-equation theory of excitation transfer, *Physical Review B* **9** (1974) 5279-5290.

[15] T. Kolsrud, On a class of probabilistic integrodifferential equations, in: S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Editors), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, Vol I, Cambridge University Press, Cambridge, 1990, pp. 168-172.

[16] M.L. T. Lee; G. A. Whitmore, Stochastic Processes Directed by Randomized Time, *Journal of Applied Probability* **30** (1993) 302-314

[17] F. Mainardi, Yu. Luchko and G. Pagnini, The fundamental solution of the space - time fractional diffusion equation, *Fractional Calculus and Applied Analysis* **4** (2001) 153-192. [E-print http://arxiv.org/abs/cond-mat/0702419]

[18] F. Mainardi, G. Pagnini and R. K. Saxena, Fox H functions in fractional diffusion, *J. Comput. Appl. Math.***178** (2005) 321-331.

[19] A.M. Mathai and R.K. Saxena, *The H-function with Applications in Statistics and Other Disciplines*, Wiley Eastern Ltd, New Delhi, 1978.

[20] M. Meerschaert, D. Benson, H. Scheffler, B. Baeumer, Stochastic solution of space-time fractional diffusion equations, *Physical Review E* **65** (2002) 041103/1-4.

[21] R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, *Phys. Reports* **339** (2000) 1-77.

[22] R. Metzler and J. Klafter, The restaurant at the end of the random walk: Recent developments in the description of anomalous transport by fractional dynamics, *J. Phys. A. Math. Gen.* **37** (2004) R161-R208.

[23] S.A. Molchanov and E. Ostrovskii, Symmetric stable processes as traces of degenerate diffusion processes, *Theory of Probability and its Applications* **14** (1969) 128-131.

[24] A. Mura, M.S. Taqqu and F. Mainardi, Non-Markovian diffusion equations and processes: analysis and simulations. *submitted: in revision*. [E-print http://arxiv.org/abs/0712.0240]

[25] A. Mura and F. Mainardi, A class of self-similar stochastic processes with stationary increments to model anomalous diffusion in physics, *submitted*. [E-print http://arxiv.org/abs/0711.0665]

[26] A. Mura and G. Pagnini, Characterization and simulation of a class of stochastic processes related to anomalous diffusion, [E-print http://arxiv.org/abs/0801.4879].

[27] A. Piryatinska, A.I. Saichev and W.A. Woyczynski, Models of anomalous diffusion: the subdiffusive case, *Physica A* **349** (2005) 375-420.

[28] W.R. Schneider, Grey noise, in: S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrøm (Editors), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, Vol 1, Cambridge Univ. Press, Cambridge (1990), pp. 261-282.

[29] W.R. Schneider, Fractional diffusion, in: R. Lima, L. Streit and D. Vilela Mendes (Editors), *Dynamics and Stochastic Processes, Theory and Applications*, Springer Verlag, Heidelberg, 1990, pp. 276-286. [Lecture Notes in Physics No 355]

[30] I.M. Sokolov, Solutions of a class of non-Markovian Fokker-Planck equations, *Phys. Rev. E* **66** (2002) 041101/1-5.

[31] H.M. Srivastava, K.C. Gupta and S.P. Goyal, *The H-Functions of One and Two Variables with Applications*, South Asian Publishers, New Delhi, 1982.

[32] T. Srokowski, Non-Markovian Lévy diffusion in non homogeneous media, *Physical Review E* **75** (2007) 051105/1-8. [E-print http://arxiv.org/abs/cond-mat/0611056]

[33] A. A. Stanislavsky, Subordinated Brownian Motion and its Fractional Fokker-Planck Equation, *Physica Scripta* **67** (2003) 265-268.

[34] M.S. Taqqu, Fractional Brownian motion and long-range dependence, in: P. Doukhan, G. Oppenheim and M.S. Taqqu (Editors), *Theory and Applications of Long-range Dependence*, Birkäuser, Boston - Basel - Berlin, 2003, pp. 5-38.

[35] G.H. Weiss, *Aspects and Applications of Random Walks*, North-Holland, Amsterdam, 1994.

[36] M.M. Wyss and W. Wyss, Evolution, its fractional extension and generalization, *Fractional Calculus and Applied Analysis* **4** (2001) 273-284.

[37] G.M. Zaslavsky, Chaos, fractional kinetics and anomalous transport, *Phys. Reports* **371** (2002) 461-580.

[38] R. Zwanzig, Memory effects in irreversible thermodynamics, *Physical Review* **124** (1961) 983-992.

10 Concluding remarks

From **Chapter 2** to **Chapter 4**, we brought the reader through the fundamental notions of *probability* and *stochastic processes*, *stochastic integration* and *stochastic differential equations* as well.

In particular, within the study of H-**sssi** processes, we focused on fractional Brownian motion (fBm) and its discrete-time increment process, the fractional Gaussian noise (fGn), which provide examples of non-Markovian Gaussian processes.

Fractional Gaussian noise, together with stationary FARIMA processes, is widely used in the modeling and estimation of long-range dependence (LRD). In fact, long-range dependence is, roughly speaking, characterized by the asymptotic power-like behavior of the autocovariance function. However, typically, stationary Markovian processes have autocovariance function which decreases exponentially to zero as the time lag becomes infinity, for instance AR(1) processes or Ornstein-Ulhenbeck processes. In contrast, many non-Markovian processes, just like fractional Gaussian noise and FARIMA processes, show an asymptotic power-like behavior of their autocovariance function.

Time series manifesting long-range dependence are often observed in nature especially in physics, meteorology, climatology, but also in hydrology, geophysics, economy and many others. Then, after having used in **Chapter 5** the theory of stochastic integration to provide integral representations useful to obtain many fundamental results about fGn and FARIMA (such as the autocorrelation and spectral density functions structure), in **Chapter 6** we went into more depth to the study of LRD, giving many real data examples, providing statistical analysis and introducing parametric methods of estimation. We have tried to give, even if briefly, the more exhaustive as possible treatment of this very interesting and fascinating topic, which has become fundamental within non-Markovian time series analysis.

In **Chapter 7**, we introduced the theory of fractional integrals and derivatives, which turns out to be very appropriate for studying and modeling systems which exhibit long-memory properties. In fact, for instance, FARIMA models can be seen as a discrete fractional integral of a white noise and, in the same way, fractional Brownian motion can be represented as a fractional integral of a Gaussian white noise.

After we introduced the basics concepts, we provided many examples and applications. For instance, in **Section 7.2.1** and **Section 7.2.3**, we have investigated the relaxation equation with distributed order time-fractional derivatives, both in the Riemann-Liouville and in the Caputo-Dzherbashyan sense. Such equations can be seen as simple models of a more general distributed order fractional evolution in Banach spaces, where the relaxation parameter λ is replaced by an operator A. Models described by such equations are characterized by a strong memory component and can be used to describe relaxation in complex systems, which deviates from the classical exponential Debye pattern.

Our interest has been focused on structural properties of the solutions, in particular on the asymptotic behavior at small and large times.

The two approaches give, in a certain sense, opposite behavior. For instance, we considered the case of double-order fractional relaxation. In the RL case, we found that the smallest order of occurring fractional differentiation determines the behavior near infinity, whereas the largest order governs the

behavior near zero. In the CD case, we found just the opposite behavior.

The topic deserves further studies in several directions in terms of integral transforms and special functions like those of Mittag-Leffler type.

From **Section 7.2.4** to **Section 7.2.6** we focused in the study of generalizations of the standard diffusion equation, by passing through the preliminary study of the fractional forward drift equation. Such generalizations have been obtained by using fractional integrals and derivatives of distributed orders. Therefore, after having outlined the basic theory of the Cauchy problem for the spatially one-dimensional and symmetric time-fractional diffusion equation (with its main equivalent formulations), we gave special attention on integral transform methods, which have been used in order to find the fundamental solutions. We have stressed the importance of Fourier, Laplace and Mellin transformations, of Mittag-Leffler and Wright type functions and of Fox-H functions as well.

Indeed, a natural first step in the construction of the fundamental solution has consisted in applying, in some order, the Fourier transform in space and the Laplace transform in time to the corresponding equation. This has yielded to the explicit form of the solution in the Fourier-Laplace domain. Then, we have had to invert the two transforms in sequence, in order to obtain the solution. In this way, we have led to a power series representation in the spatial variable with time-dependent coefficients. Finally, we have worked out to express the fundamental solution in terms of Laplace integrals, which are more suitable for numerical evaluations.

We have studied and illustrated in details the time-fractional diffusion of single order (where self-similarity holds true) and moreover we carried out the study of two simple but noteworthy cases of distributed order. Namely, the case of a superposition of two different orders β_1 and β_2 and the case of uniform order distribution.

In the first case, one of the orders dominates the time-asymptotics near zero, the other near infinity, but β_1 and β_2 change their roles when switching from the RL to the CD formulation. The case of uniform distributed order was remarkably different. The extreme orders now being (roughly speaking) 0 and 1. We met super-slow and super-fast variance time behaviors near zero and near infinity, again interchanging the behaviors between the RL and CD models. Clearly, one could actually observe these properties in all the figures we provided. For instance, one can easily notice the extremely slow growth of the variance, as $t \to \infty$, in the CD case.

In order to find a connection between the anomalous diffusion described by fractional type equations and the long-range dependence related fo self-similar stationary increments stochastic processes, in **Chapter 7** we introduced and studied the generalized grey Brownian motion $B_{\alpha,\beta}(t)$ (ggBm). We showed that the parametric class $\{B_{\alpha,\beta}(t),\ \ 0 < \alpha < 2,\ \ 0 < \beta \leq 1\}$ indeed provides stochastic models for the anomalous diffusions described by the following *stretched time-fractional diffusion equation*

$$u(x,t) = u_0(x) + \frac{1}{\Gamma(\beta)} \int_0^t \frac{\alpha}{\beta} s^{\alpha/\beta - 1} \left(t^{\frac{\alpha}{\beta}} - s^{\frac{\alpha}{\beta}} \right)^{\beta - 1} \frac{\partial^2}{\partial x^2} u(x,s) ds, \quad t \geq 0, \tag{10.1}$$

which for $\alpha = \beta$ actually reduces to the usual time-factional diffusion equation of order β. We have shown that:

When $0 < \alpha < 1$, the diffusion is slow. The increments of the process $B_{\alpha,\beta}(t)$ turn out to be negatively correlated and this implies that the trajectories are very "zigzaging" (antipersistent). The increments

form a stationary process which does not exhibit long-range dependence.

When $\alpha = 1$, the diffusion is normal. The increments of the process are *uncorrelated*, but not independent unless $\beta = 1$.

When $1 < \alpha < 2$, the diffusion is fast. The increments of the process $B_{\alpha,\beta}(t)$ are positively correlated, so that the trajectories are more regular (persistent). In this case the increments exhibits long-range dependence.

The ggBm is of course Non-Markovian. We have observed that non-Markovian equations like (eq. 10.1) are usually associated to subordinated stochastic processes $D(t) = B(l_\beta(t))$, where the parent Markov process $B(t)$ is a standard Brownian motion and the random time process $l_\beta(t)$ is a self-similar of order $H = \beta$ non-negative non-decreasing non-Markovian process. For example, the random time $l_\beta(t)$ can be taken to be related to the local time of a $d = 2(1-\beta)$-dimesional fractional Bessel process. Or alternatively, it can be the inverse of the totally skewed strictly β-stable process. However, we have seen that $\{B_{\alpha,\beta}(t),\ t \geq 0\}$ cannot be a subordinated process, in fact, for example, if $\beta = 1$ it reduces to a fractional Brownian motion. Therefore, we actually provided an example of a class of stochastic models associated to time-fractional diffusion equations, which are not subordinated processes.

We have remarked many times that, starting from a master equation which describes the dynamic evolution of a probability density function $f(x,t)$, it is always possible to define an equivalence class of stochastic processes with the same marginal density function $f(x,t)$. All these processes provide suitable stochastic models for the starting equation. Here, we focused on a subclass $\{B_{\alpha,\beta}(t),\ t \geq 0\}$ associated to the non-Markovian equation (eq. 10.1). This subclass is made up of processes with *stationary increments*. In this case, the memory effects are enclosed in the typical dependence structure of a H-sssi process (eq. 8.79), while, for instance in the case of a subordinated process, these are due to the non-Markovian property of the *random time process*.

The ggBm has been defined canonically in the so called grey noise space $(\mathcal{S}'(\mathbb{R}), \mathcal{B}, \mu_{\alpha,\beta})$, where the grey noise measure satisfies (eq. 8.61). However, the ggBm is an H-**sssi** process of order $H = \alpha/2$ and we have been able to provide a characterization of $B_{\alpha,\beta}(t)$ notwithstanding the underline probability space.

We also pointed out that that the generalized grey Brownian motion is a direct generalization of a Gaussian process. Indeed, it is Gaussian when $\beta = 1$ (indeed a fractional Brownian motion of order $\alpha/2$) and, fixed β, it is defined only by its covariance structure, which is a property of Gaussian processes. Inspired by this consideration, we have obtained an enlighten representation of ggBm. In fact, we found that it turns out to be merely a fractional Brownian motion with stochastic variance, that is $B_{\alpha,\beta}(t) = \Lambda_\beta X_\alpha(t)$, $t \geq 0$, where Λ_β is a suitable random variable independent of the fractional Brownian motion $X_\alpha = B_{\alpha/2}(t)$ (see eq. 8.87). For instance, one could take $\Lambda_\beta = \sqrt{l_\beta(1)}$, where $l_\beta(t)$ is the random time process associated to a suitable subordinated model. This representation has been useful when we focused on path simulations.

Clearly, in order to simulate ggBm trajectories, we have needed of method to generate the random variable L_β. For this purposes, in order to build a suitable random walk scheme, we have made use of finite difference approximation of non-local partial integro differential equation of fractional type. The construction has followed two steps: using the *Grünwald-Letnikov discretization* of Caputo-Dzherbashyan derivative, and then interpreting the corresponding finite difference scheme as a random walk scheme.

Following this idea, we have been able to simulate generalized grey Brownian motion trajectories both in the short and long-range dependence domain.

Moreover, we have observed that the process $\Lambda_\beta X_\alpha(t)$, is not ergodic[1], as, heuristically, follows by the multiplication with the random variable Λ_β. This was evident also looking at the simulated trajectories. Indeed, it is impossible with a single realization of the system $B_{\alpha,\beta}(t,\omega)$, $\omega \in \Omega$, to distinguish a ggBm from a fBm with variance $2\Lambda^2_\beta(\omega)t^\alpha$, where $\Lambda_\beta(\omega)$ indicates a single realization of the random variable Λ_β.

In **Chapter 9** we studied a certain type of Volterra equations, which actually represent a further generalization of fractional type equations. We found that the fundamental solution $f(x,t)$ of a non-Markovian diffusion equation of the form (eq. 9.43)

$$u(x,t) = u_0(t) + \int_0^t g'(s)K\left(g(t) - g(s)\right)\mathcal{P}_x u(x,s)ds, \quad x \in \mathbb{R}, \quad t \geq 0, \tag{10.2}$$

is

$$f(x,t) = \int_0^\infty \mathcal{G}(x,\tau)h(\tau,g(t))d\tau, \tag{10.3}$$

where $\mathcal{G}(x,t)$ is the fundamental solution of the Markovian equation (eq. 9.41) and $h(\tau,t)$ is the fundamental solution of the non-Markovian forward drift equation

$$u(\tau,t) = u_0(\tau) - \int_0^t K(t-s)\partial_\tau u(\tau,s)ds, \quad \tau,t \geq 0. \tag{10.4}$$

If the memory kernel $K(t)$ were chosen in a suitable way (see Section 9.2), the solution $f(\cdot,t)$ would preserve non-negativity and normalization for all $t \geq 0$. Thus, it could be interpreted as the marginal density function of a non-Markovian stochastic process. In view of (eq. 10.3), this stochastic process has been naturally interpreted as a subordinated process (eq. 9.55).

We focused on two kind of memory kernels: the power kernel $K(t) = t^{\beta-1}/\Gamma(\beta)$, $0 < \beta \leq 1$, and the exponential decay kernel $K(t) = e^{-at}$, $a \geq 0$.

The first provides what we called time-fractional Fokker-Planck equations (eq. 9.88). Such equation directly generalizes the fractional diffusion equation, which in fact arises by setting $\mathcal{P}_x = \partial_{xx}$, and actually corresponds to the choice of a standard Brownian motion as the parent Markov model.

We have also considered more complicated cases, namely Brownian motion with drift μ (see Section 9.6.1) and Geometric Brownian motion (see Section 9.6.2). In these cases, the fundamental solutions can be written in terms of a superposition of Fox H-functions (eq. 9.101 and eq. 9.123).

The exponential-decay kernel corresponds heuristically to a system in which the non-local memory effects are negligible for small times. In fact, the fundamental solution can always be written in the form of (eq. 9.75),

$$f(x,t) = e^{-at}\mathcal{G}(x,t) + (1 - e^{-at})\phi(x,t), \quad t \geq 0,$$

where $\mathcal{G}(x,t)$ is the fundamental solution of the Markovian equation, and where the function $\phi(x,t)$ is a probability density which becomes stationary for large times. Therefore, it is always possible to

1) Here we mean that the increment process, which is indeed stationary, is not ergodic. Namely, time averages do not necessary coincide with population averages.

find stochastic models that become stationary for large times and whose marginal density is given by (eq. 10.3).

However we pointed out again (see Subsection 9.3.1) that the stochastic representation cannot be unique, as we have seen in the case where $\mathcal{P}_x = \partial_{xx}$ and $g(t) = t$. In this case, $f(x,t)$ is the marginal density of $B(l(t))$, $t \geq 0$, where $B(t)$ is a standard Brownian motion and where $l(t)$ is a random time process satisfying (eq. 10.4). If the random time $l(t)$ is required to be self-similar of order β, then we have shown (Theorem 9.1) that the memory kernel must be a power function $K(t) = t^{\beta-1}/\Gamma(\beta)$ with $0 < \beta \leq 1$. We know that, in this case, the corresponding random time process $l(t) = l_\beta(t)$, can be the local time of a $d = 2(1-\beta)$-dimensional fractional Bessel process or, alternatively, the inverse of the totally skewed strictly β-stable process. However, $f(x,t)$ is also the marginal density of the grey Brownian motion $B_{\beta,\beta}(t) = \sqrt{l_\beta(1)} B_{\beta/2}(t)$, where $B_{\beta/2}$ is a fractional Brownian motion independent of the random time $l_\beta(t)$.

The solution of the "non-Markovian" equation (eq. 10.2) has been stated explicitly in all the cases considered. We computed it analytically and graphed it in particular cases. This solution is a marginal (one-point) density function. We have then presented various random processes whose marginal density function coincides with that solution.

I do believe that my peace of work is indeed providing a forward step in the study and comprehension of the relationships between stochastic processes and a certain class of integral-partial differential equation, which can be used in order to model anomalous diffusion and transport in statistical physics.

List of Figures

List of Tables

Tesi di dottorato in fisica, XX ciclo

Presentata da:
Antonio Mura

Relatore:
Prof. Francesco Mainardi
Coordinatorne:
Prof. Fabio Ortolani

NON-MARKOVIAN STOCHASTIC PROCESSES AND THEIR APPLICATIONS: FROM ANOMALOUS DIFFUSION TO TIME SERIES ANALYSIS.

This work provides a forward step in the study and comprehension of the relationships between stochastic processes and a certain class of integral-partial differential equation, which can be used in order to model anomalous diffusion and transport in statistical physics. In the first part, we brought the reader through the fundamental notions of *probability* and *stochastic processes*, *stochastic integration* and *stochastic differential equations* as well. In particular, within the study of H-**sssi** processes, we focused on *fractional Brownian motion* **(**fBm**)** and its discrete-time increment process, the *fractional Gaussian noise* **(**fGn**)**, which provide examples of *non-Markovian* Gaussian processes. The fGn, together with stationary FARIMA processes, is widely used in the modeling and estimation of *long-memory*, or *long-range dependence* **(**LRD**)**. Time series manifesting long-range dependence, are often observed in nature especially in physics, meteorology, climatology, but also in hydrology, geophysics, economy and many others. We deepely studied LRD, giving many real data examples, providing statistical analysis and introducing parametric methods of estimation. Then, we introduced the theory of fractional integrals and derivatives, which indeed turns out to be very appropriate for studying and modeling systems with long-memory properties. After having introduced the basics concepts, we provided many examples and applications. For instance, we investigated the relaxation equation with distributed order time-fractional derivatives, which describes models characterized by a strong memory component and can be used to model relaxation in complex systems, which deviates from the classical exponential Debye pattern. Then, we focused in the study of generalizations of the standard diffusion equation, by passing through the preliminary study of the fractional forward drift equation. Such generalizations have been obtained by using fractional integrals and derivatives of distributed orders. In order to find a connection between the anomalous diffusion described by these equations and the long-range dependence, we introduced and studied the *generalized grey Brownian motion* **(**ggBm**)**, which is actually a parametric class of H-**sssi** processes, which have indeed marginal probability density function evolving in time according to a partial integro-differential equation of fractional type. The ggBm is of course Non-Markovian. All around the work, we have remarked many times that, starting from a master equation of a probability density function $f(x,t)$, it is always possible to define an equivalence class of stochastic processes with the same marginal density function $f(x,t)$. All these processes provide suitable stochastic models for the starting equation. Studying the ggBm, we just focused on a subclass made up of processes with *stationary increments*. The ggBm has been defined canonically in the so called grey noise space. However, we have been able to provide a characterization notwithstanding the underline probability space. We also pointed out that that the generalized grey Brownian motion is a direct generalization of a Gaussian process and in particular it generalizes Brownain motion and fractional Brownain motion as well. Finally, we introduced and analyzed a more general class of diffusion type equations related to certain non-Markovian stochastic processes. We started from the forward drift equation, which have been made non-local in time by the introduction of a suitable chosen memory kernel $K(t)$. The resulting non-Markovian equation has been be interpreted in a natural way as the evolution equation of the marginal density function of a random time process $l(t)$. We then consider the subordinated process $Y(t)=X(l(t))$ where $X(t)$ is a Markovian diffusion. The corresponding time-evolution of the marginal density function of $Y(t)$ is governed by a non-Markovian Fokker-Planck equation which involves the memory kernel $K(t)$. We developed several applications and derived the exact solutions. Moreover, we considered different stochastic models for the given equations, providing path simulations.