Nonparametric Bayesian inference for Gamma-type Lévy subordinators

Belomestny, D.; Gugushvili, S.; Schauer, M.; Spreij, P.

DOI
10.4310/CMS.2019.v17.n3.a8

Publication date
2019

Document Version
Submitted manuscript

Published in
Communications in Mathematical Sciences

NONPARAMETRIC BAYESIAN INFERENCE FOR GAMMA-TYPE LÉVY SUBORDINATORS*

DENIS BELOMESTNY†, SHOTA GUGUSHVILI‡, MORITZ SCHAUER§, AND PETER SPREIJ¶

Abstract. Given discrete time observations over a growing time interval, we consider a nonparametric Bayesian approach to estimation of the Lévy density of a Lévy process belonging to a flexible class of infinite activity subordinators. Posterior inference is performed via MCMC, and we circumvent the problem of the intractable likelihood via the data augmentation device, that in our case relies on bridge process sampling via Gamma process bridges. Our approach also requires the use of a new infinite-dimensional form of a reversible jump MCMC algorithm. We show that our method leads to good practical results in challenging simulation examples. On the theoretical side, we establish that our nonparametric Bayesian procedure is consistent: in the low frequency data setting, with equispaced in time observations and intervals between successive observations remaining fixed, the posterior asymptotically, as the sample size $n \to \infty$, concentrates around the Lévy density under which the data have been generated. Finally, we test our method on a classical insurance dataset.

Keywords. Bridge sampling; Data augmentation; Gamma process; Lévy process; Lévy density; MCMC; Metropolis-Hastings algorithm; Nonparametric Bayesian estimation; Posterior consistency; Reversible jump MCMC; Subordinator; θ-subordinator

AMS subject classifications. Primary: 62G20, Secondary: 62M30

1. Introduction In this paper, to the best of our knowledge for the first time in the literature, we study the problem of nonparametric Bayesian inference for infinite activity subordinators, i.e., Lévy processes with non-decreasing sample paths. In the last two decades, Lévy processes have received a lot of attention, mainly due to their numerous applications in mathematical finance and insurance, but also in natural sciences; see, e.g., Barndorff-Nielsen et al. (2001). As a matter of fact, thanks to their ability to reproduce stylised features of financial time series distributions, Lévy processes have become a fundamental building block for modelling asset prices with jumps, see Cont and Tankov (2004). By the Lévy-Khintchine formula, the law of a Lévy process is uniquely determined by the so-called Lévy triplet, which hence encodes all the probabilistic information on the process. Since the Lévy triplet involves an infinite-dimensional object, the Lévy measure of the process, this provides natural motivation for studying nonparametric inference procedures for Lévy processes, where the objects of inference are elements of some function spaces.

We term the class of increasing infinite activity Lévy processes that we study θ-subordinators. Our model generalises the well-known Gamma process, which is a popular risk model, see Dufresne et al. (1991), and also forms a building block for more general Lévy models, like the Variance-Gamma (VG) process, that finds many appli-

*Received date, and accepted date (The correct dates will be entered by the editor).

†Duisburg-Essen University, Thea-Leymann-Str. 9, D-45127 Essen, Germany, and National Research University, Higher School of Economics, Moscow, Russian Federation, denis.belomestny@uni-due.de

‡Biometris, Wageningen University & Research, Postbus 16, 6700 AA Wageningen, The Netherlands, gugushvili@gmail.com

§Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands, m.r.schauer@math.leidenuniv.nl

¶Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands, and Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, Nijmegen, The Netherlands, spreij@uva.nl

cations in finance, see, e.g., Madan and Seneta (1990). The family of θ-subordinators also overlaps with the class of self-decomposable Lévy processes, that likewise have important applications in finance, see, e.g., Carr et al. (2007).

We specifically concentrate on estimation of the Lévy triplet of a θ-subordinator. On the computational side, our Bayesian procedure circumvents the problem of the intractable likelihood for θ-subordinators via the data augmentation device, which relies on bridge process sampling via Gamma process bridges, and also employs an infinite-dimensional form of the reversible jump algorithm. On the theoretical side, we establish that our procedure is consistent: as the sample size grows to infinity, the posterior asymptotically concentrates around the parameters of the Lévy processes under which the data have been generated. We test our algorithm on simulated and real data examples. In particular we fit a θ-subordinator to a benchmark dataset in insurance theory, large fire losses in Denmark, and study the question whether a risk model based on a Gamma process is adequate for modelling this dataset.

1.1. Literature overview

To provide further motivation for a nonparametric Bayesian approach to inference in Lévy processes and to highlight some associated challenges, in this subsection we supply an overview of the literature on the subject.

The problem of nonparametric inference for Lévy processes has a long history, going back to Rubin and Tucker (1959) and Basawa and Brockwell (1982). Revival of interest in it dates around the year 2003, with contributions Buchmann and Grübel (2003), Buchmann and Grübel (2004) and van Es et al. (2007), as well as numerous later publications; see also Ilhe et al. (2015) for a further extension. Very recent works Coca (2017) and Duval and Mariucci (2017) provide an extensive list of references.

In general, there are two major strands of mathematical statistics literature dealing with inference for Lévy processes, or more generally semimartingales. The first considers the so-called high frequency setup where asymptotic properties of the corresponding estimators are studied under the assumption that observations are made at an increasing frequency in time. In the second strand of the literature, times between successive observations are assumed to be fixed (the so-called low frequency setup) and the asymptotic analysis is done under the premise that the observational horizon tends to infinity.

The last decade witnessed a tremendous advance in the area of statistics for high frequency financial data, due to the development of new mathematical methods to analyse these data, as well as increasing availability of such data. We refer to the recent book Aït-Sahalia and Jacod (2014) for a comprehensive treatment of modern statistical methods for high frequency data. At the same time, progress was achieved also in statistical inference for Lévy-driven models based on low frequency data, see, e.g., Belomestny et al. (2015) for an overview and references. The latter situation is more challenging, as e.g. it becomes quite difficult to distinguish between small jumps of a Lévy process and the Brownian increments. This often leads to rather slow, logarithmic convergence rates for resulting estimators, see, e.g., Belomestny and Reiß (2006), Gugushvili (2009) and Gugushvili (2012). Hence accurate nonparametric inference for Lévy processes typically requires very large amounts of data, which may not always be available in practice. Fortunately, in many cases there is additional (prior) information about the structure of the parameters, which can be used to improve the estimation quality under limited data. To account for this prior information, the Bayesian estimation framework is quite appealing. Furthermore, the Bayesian approach provides automatic uncertainty quantification in parameter estimates through the spread of the posterior distribution of the parameters. Also, in some fields, such as e.g. climate and weather science, Bayesian approaches are thought to be default (see, e.g., Berliner et al. (1999)), and studying them would go together with common practices in those fields. On the other hand, there are also some formidable challenges in applying the nonparametric Bayesian methodology to inference in Lévy processes. Firstly, the underlying process is usually observed at discrete time instances, while Lévy models are formulated in continuous time. This gives rise to complications that are typical in inference for discretely observed continuous time stochastic processes. Secondly, Bayesian estimation in its simplest, pristine form requires knowledge of the likelihood of observations, and hence of marginal densities of the underlying Lévy process; these, however, are rarely available in closed form. Thirdly, devising valid MCMC algorithms in infinite-dimensional settings is a highly non-trivial task. Cf. recent works on nonparametric Bayesian inference in diffusion models, such as Beskos et al. (2008) and van der Meulen et al. (2014).

The literature on nonparametric Bayesian inference for Lévy processes is very recent and also rather scarce, the only available works being Nickl and Söhl (2017a), Gugushvili et al. (2015) and Gugushvili et al. (2018). These deal with a particular case of compound Poisson processes, concentrate exclusively on theoretical aspects (with the exception of the latter paper), and do not appear to admit an obvious extension to other classes of Lévy processes. In fact, compound Poisson processes are rather special among Lévy processes, and are of limited applicability in many practically relevant cases. Hence there is space for improvement. On the positive side, the practical results we obtained in this work demonstrate great potential of Bayesian methods for inference in Lévy processes. Our approach is aimed at developing an applicable statistical methodology, which we substantiate by theoretical results, and also test via challenging examples. At the same time, we admit there remain several unresolved theoretical and practical issues, such as derivation of posterior contraction rates or practical fine-tuning of the prior we use. However, upon careful reading this should come as no surprise given the sheer complexity of our undertaking, where several topics would have merited to be subjects of independent research projects. We view our work as the first substantial step made in the direction of studying inference problems for Lévy processes via nonparametric Bayesian methods. It is our hope that our contribution will generate additional interest in this statistically and mathematically fascinating topic.

1.2. Structure of the paper The rest of the paper is organised as follows: in Section 2 we describe in detail the statistical problem we are dealing with and our nonparametric Bayesian approach to it. Posterior inference in our setting is performed through MCMC sampling, and Section 3 provides a detailed exposition of our sampling algorithm. In Section 4 we establish the fact that our approach is consistent in the frequentist sense: asymptotically, as the sample size $n \to \infty$, the posterior measure concentrates around the Lévy triplet under which the data used in the estimation procedure has been generated. In Section 5 we test the practical performance of our method via simulation on a challenging example. In Section 6 we further generalise our basic model from Section 2 and detail changes and extensions this requires in designing an MCMC sampler in comparison to the one from Section 3. This new sampler is tested in simulations in Section 7. In Section 8 we apply our methodology on an insurance dataset. Possible extensions of our inferential approach to more general Lévy models are discussed in Section 9. Finally, in Appendices A and B we state and prove some technical results used in the main body of the paper, while in Appendix C we provide some additional analyses to substantiate our modelling approach in Section 8.

2. Statistical problem and approach

In this section we introduce in detail the statistical problem we are dealing with and describe our approach to tackle it.

2.1. Statistical problem Consider a univariate Lévy process $X = (X_t\colon t \geq 0)$ with generating Lévy triplet $(\gamma, 0, \nu)$, where $\nu([1,\infty))$ is finite and

$$\gamma = \int_0^1 x\nu(\mathrm{d}x) < \infty. \tag{2.1}$$

Hence X has no Gaussian component and the law $\mathbb{P}_\nu$ of X is entirely determined by ν. By the Lévy-Khintchine formula, see Theorem 8.1 in Sato (1999), the characteristic function ϕ_{X_1} of X_1 admits the unique representation of the type

$$\phi_{X_1}(z) = \exp\left(\mathrm{i}\gamma z + \int_{\mathbb{R}} \left(e^{\mathrm{i}zx} - 1 - \mathrm{i}zx\mathbf{1}_{|x|\leq 1}\right)\nu(\mathrm{d}x)\right).$$

We also assume that the Lévy measure ν admits the representation

$$\nu(\mathrm{d}x) = \frac{\beta}{x} e^{-\alpha x - \theta(x)} \mathrm{d}x, \quad x > 0, \tag{2.2}$$

where α and $\theta\colon [0,\infty) \to \mathbb{R}$ are parameters to be estimated, while β is a known or unknown parameter. It follows that X is a pure jump process with non-decreasing sample paths, or put another way, a subordinator with zero drift, cf. Sections 2.6.1–2.6.2 in Kyprianou (2006). One may call this class of Lévy processes Gamma-type subordinators, because X is a Gamma process when $\theta \equiv 0$, but we prefer to simply refer to it as θ-subordinators.

Assume that the process X is observed at discrete time instances $0 = t_0 < t_1 < \cdots < t_n = T$, so our observations are $X^{(n)} = (X_{t_i}\colon i \in \{0,\ldots,n\})$. Our aim is nonparametric Bayesian estimation for the parameter triple (α, β, θ). This requires specification of the likelihood and the prior in our model, that are next combined via Bayes' formula to form the posterior distribution. This latter encodes all the necessary inferential information within the Bayesian setup. By Theorem 27.7 in Sato (1999), marginal distributions of X possess densities with respect to the Lebesgue measure. With $p_h(x;\beta,\alpha,\theta)$ denoting the density of an increment $X_{t+h} - X_t$, the likelihood

$$\prod_{i=1}^{n} p_{t_i - t_{i-1}}(X_{t_i} - X_{t_{i-1}}; \beta, \alpha, \theta)$$

is in general intractable, as the marginal densities of X are not known in closed form, except some special cases. This complicates a computational approach to Bayesian inference. We will circumvent this obstacle by employing the concept of data augmentation, see Tanner and Wong (1987). Specifically, we will propose a suitable nonparametric prior distribution $\pi(\beta, \alpha, \theta)$ on the parameter triple (β, α, θ), and derive a Metropolis-Hastings algorithm relying on data augmentation to sample from the posterior distribution. Details of our approach are given in the following subsections.

2.2. Likelihood We first consider the problem where β is known and fixed. All processes and their laws in this section are restricted to the time interval $[0,T]$ for a fixed $T > 0$. Note that for any two Lévy measures ν and ν_0 given by (2.2) with parameters

β,α,θ and β,α_0,θ_0, respectively, provided $\theta(0)=\theta_0(0)=0$ and both functions θ and θ_0 are Lipschitz in some neighbourhood of zero, we have

$$\nu \text{ and } \nu_0 \text{ are equivalent,}$$
$$d^2_{\mathcal{H}}(\nu,\nu_0)=\frac{1}{2}\int\limits_{(0,\infty)}(\sqrt{\mathrm{d}\nu}-\sqrt{\mathrm{d}\nu_0})^2<\infty, \tag{2.3}$$

where $d_{\mathcal{H}}(\cdot,\cdot)$ is the Hellinger distance between two (infinite) measures. By assumption (2.1) and property (2.3), together with Theorem 33.1 in Sato (1999), it follows that the laws $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu_0}$ of $X=(X_t\colon t\in[0,T])$ are equivalent. Furthermore, Theorem 33.2 in Sato (1999) implies that a.s.

$$U_T=\log\left(\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\nu_0}}(X)\right)=\sum_{(s,\Delta X_s)\in(0,T]\times\{\Delta X_s>0\}}\phi(\Delta X_s)-T\int\limits_{(0,\infty)}(e^{\phi(x)}-1)\nu_0(\mathrm{d}x),$$

where $\Delta X_s=X_s-X_{s-}$, and

$$\phi(x)=\log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(x)\right)=-(\alpha x+\theta(x)-\alpha_0 x-\theta_0(x)),\quad x>0.$$

We can also write the log-likelihood ratio U_T as

$$U_T=\int_{(0,T]}\int_{(0,\infty)}\phi(x)\mu(\mathrm{d}s,\mathrm{d}x)-T\int\limits_{(0,\infty)}(\nu-\nu_0)(\mathrm{d}x),$$

where the jump measure μ is defined by

$$\mu((0,t]\times B)=\#\{s\colon (s,\Delta X_s)\in(0,t]\times B\}$$

for any Borel subset B of $(0,\infty)$. We can view $\mathbb{P}_{\nu_0}$ as the dominating measure for $\mathbb{P}_\nu$. From the inferential point of view the specific choice of the dominating measure is immaterial. A convenient choice of ν_0 for the theoretical development in Section 4 is to actually take ν_0 to be the 'true' Lévy measure ν_0 with parameters α_0 and θ_0 (recall that β is fixed and assumed to be known).

2.3. Gamma processes

We temporarily specialise to the case of a Gamma process. A Gamma process is an example of a pure jump Lévy process with non-decreasing sample paths. Its Lévy triplet is given by $(\gamma,0,\nu)$, where

$$\gamma=\int_0^1 x\nu(\mathrm{d}x),\quad \nu(\mathrm{d}x)=\frac{\beta}{x}e^{-\alpha x}\mathrm{d}x,\quad x>0,$$

see Example 8.10 in Sato (1999). Making the dependence on parameters explicit, we also refer to X as a Gamma(β,α) process. The distribution of X_t, $t\in[0,T]$, is gamma with rate parameter α and shape parameter βt, so that

$$X_t\sim p_t(x;\beta,\alpha)=\frac{\alpha^{\beta t}x^{\beta t-1}e^{-\alpha x}}{\Gamma(\beta t)},\quad x>0, \tag{2.4}$$

where Γ denotes the gamma function.

2.4. Data augmentation and bridge sampling

By using the data augmentation technique, we can utilise existence of a closed-form likelihood for a continuously observed Lévy path, see Subsection 2.2, to define a Metropolis-Hastings algorithm to sample from the posterior given the discrete observations $X^{(n)}$. This treats the unobserved path segments between two consecutive observation times as missing data and augments the state space of the algorithm to sample from the joint posterior of missing data and unknown parameters. Specifically, this requires the ability to sample from the conditional distribution of the missing data given the parameters and the observations.

Consider again the Lévy process $X=(X_t\colon t\in[0,T])$ with fixed parameters β, α, θ, and denote the corresponding law by $\mathbb{P}$. Conditional on the observations $X_{t_{i-1}}$ and X_{t_i} and the parameters, by the independent increments property of a Lévy process, the process can be sampled on each time interval $[t_{i-1},t_i]$ independently. Samples from the conditional distribution on these intervals connect the observations in the form of so-called bridges. It suffices to describe the construction for a single bridge from 0 to T. A Gamma process $\widetilde{X}=(\widetilde{X}_t\colon t\in[0,T])$ shares with the Wiener process a remarkable property that samples from the conditional distribution can be obtained through a simple transformation of the unconditional path, see Yor (2007). For the Wiener process W conditional on $W_T=w_T$ for a number w_T, this transformation takes the form

$$t\mapsto W_t+\frac{t}{T}(w_T-W_T),\quad t\in[0,T].$$

For the Gamma process, the corresponding transformation takes a multiplicative form: define for a path $X=(X_t\colon t\in[0,T])$ a map g_{x_T} by

$$g_{x_T}(X)=(x_TX_t/X_T\colon t\in[0,T]). \tag{2.5}$$

Then $\widetilde{\mathbb{P}}^\star=g_{x_T}\circ\widetilde{\mathbb{P}}$, where $\widetilde{\mathbb{P}}$ denotes the law of $\widetilde{X}$, defines a factorisation of the conditional distribution $\widetilde{\mathbb{P}}^\star$ of $\widetilde{X}$ under the law $\widetilde{\mathbb{P}}$ given $\widetilde{X}_T=x_T$. This result in combination with a Metropolis-Hastings step can be used to sample from the conditional distribution of a θ-subordinator given the observations and parameters.

Analogously, we denote by $\mathbb{P}^\star$ the conditional distribution of X under the law $\mathbb{P}$ given $X_T=x_T$. Here and later we use a superscript star to denote the conditional distributions, suppress the dependence on x_T in the notation and write for example $\mathbb{P}^\star(\mathrm{d}X)$ for integration with respect to the conditional distribution. By conditioning,

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\widetilde{\mathbb{P}}^\star}(g_{x_T}(X))=\frac{\widetilde{p}(x_T)}{p(x_T)}\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\widetilde{\mathbb{P}}}(g_{x_T}(X)), \tag{2.6}$$

where p and $\widetilde{p}$ are the densities of X_T under $\mathbb{P}$ and $\widetilde{\mathbb{P}}$, respectively. Note that $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\widetilde{\mathbb{P}}}(g_{x_T}(X))$ is the continuous-time likelihood, which is known in closed form. Hence $\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\widetilde{\mathbb{P}}^\star}$ is also known in closed form up to an unknown proportionality constant $\frac{\widetilde{p}(x_T)}{p(x_T)}$, and the ratio of Radon-Nikodym derivatives $\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\widetilde{\mathbb{P}}^\star}(X^\circ)/\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\widetilde{\mathbb{P}}^\star}(X)$, with X° denoting a proposal in the MCMC algorithm, is given by formula (2.11) below. This allows us to use samples distributed according to $\widetilde{\mathbb{P}}^\star$, i.e. Gamma$(\beta,\alpha)$ bridges, as proposals for the augmented segment that follows the intractable conditional distribution $\mathbb{P}^\star$.

2.5. Prior To define the prior, we consider a subclass of processes defined in (2.2), where the parameter θ in the Lévy measure ν has the following form. Fix a sequence

$$0<b_1<\cdots<b_N<\infty,$$

set for convenience $b_0=0$ and $b_{N+1}=\infty$, and define bins B_k by

$$B_k=[b_k,b_{k+1}),\quad k=0,\dots,N.$$

Given bins B_k, assume the function θ is piecewise linear, i.e.,

$$\theta(x)=\sum_{k=1}^{N}(\rho_k+\theta_k x)\mathbf{1}_{B_k},\tag{2.7}$$

where $\rho_k\in\mathbb{R}$, $k=1,\dots,N$, $\theta_k\in\mathbb{R}$, $k=1,\dots,N$, and $\theta_N>-\alpha$. Together with α, the parameter θ_k determines the slope of the function $\theta(x)+\alpha x$ on the bin B_k, while ρ_k gives the intercept. The process X with the law $\mathbb{P}_\nu$ can be viewed as a Gamma process with rate parameter α and shape parameter β, subjected to local deviations in the behaviour of jumps of sizes falling in bins B_k compared to what of a Gamma process. The parameters θ_k,ρ_k quantify the extent of these local deviations on the bin B_k.

We equip α,θ_k,ρ_k with independent priors. Note that these priors on α,θ_k,ρ_k implicitly define a prior on the Lévy measure ν as well. The specific form of the prior is not crucial for many arguments that follow, but is convenient computationally. In fact, theoretical results in Section 4 can be derived for other series priors as well. However, the local linear structure in (2.7) (which also means that the prior could be rewritten as series prior with basis functions with compact support) is important to derive some simple update formulae below.

For a realisation ν from the implicit prior on ν as above in the present section, let us work out the integral

$$\nu(B_k)=\int_{b_k}^{b_{k+1}}\frac{\beta}{x}e^{-(\alpha+\theta_k)x-\rho_k}\mathrm{d}x,$$

which enters the expression for the likelihood in Subsection 2.2. To that end remember the definition of the exponential integral, $E_1(z)=\int_z^\infty t^{-1}e^{-t}\mathrm{d}t$, see, e.g., §15.09 in Jeffreys and Swirles (1999) for its basic properties. Then a change of the integration variable gives

$$\nu(B_k)=\beta e^{-\rho_k}\{E_1((\theta_k+\alpha)b_k)-E_1((\theta_k+\alpha)b_{k+1})\},\quad k=1,\dots,N.\tag{2.8}$$

Observe that $\nu(B_N)=\beta e^{-\rho_N}E_1((\theta_k+\alpha)b_N)$. Similar to the case of ν,

$$\nu_0(B_k)=\beta\{E_1(\alpha_0 b_k)-E_1(\alpha_0 b_{k+1})\},\quad k=1,\dots,N.$$

Also here remark that $\nu_0(B_N)=\beta E_1(\alpha_0 b_N)$. For future reference in Subsection 2.6, note that for any α,α',

$$\lim_{x\to 0}\{E_1(\alpha x)-E_1(\alpha' x)\}=\log\left(\frac{\alpha'}{\alpha}\right),\tag{2.9}$$

which follows from the formula for Frullani's integral, see §12.16 in Jeffreys and Swirles (1999).

2.6. Likelihood expressions for parameter updates

The following expressions will be used in Section 3 to construct the Metropolis-Hastings algorithm to sample from the posterior of α,θ_k,ρ_k. Define random variables

$$\mu_T(B_k)=\mu((0,T]\times B_k)=\#\{s\colon (s,\Delta X_s)\in(0,T]\times B_k\},$$

that for each $k=1,\ldots,N$, give the number of jumps of X, whose sizes fall into the bin B_k. Consider two laws $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu^\circ}$, where the Lévy measure ν is given by (2.2) and (2.7), while ν° is given by (2.2) with coefficients α°, $\theta_1^\circ,\ldots,\theta_N^\circ$, $\rho_1^\circ,\ldots,\rho_N^\circ$ instead of the coefficients α, $\theta_1,\ldots,\theta_N$, $\rho_1,\ldots,\rho_N$. The two laws $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu^\circ}$ are equivalent, since each is equivalent to $\mathbb{P}_{\nu_0}$. We have the following expression for the log-likelihood,

$$\begin{aligned}\log\frac{\mathrm{d}\mathbb{P}_{\nu^\circ}}{\mathrm{d}\mathbb{P}_\nu}(X)=&-(\alpha^\circ-\alpha)\sum_{\substack{\Delta X_s\in B_0,\\0<s\le T}}\Delta X_s-\sum_{k=1}^N(\theta_k^\circ+\alpha^\circ-\theta_k-\alpha)\sum_{\substack{\Delta X_s\in B_k,\\0<s\le T}}\Delta X_s\\&-\sum_{k=1}^N(\rho_k^\circ-\rho_k)\mu_T(B_k)-T\sum_{k=0}^N(\nu^\circ-\nu)(B_k),\end{aligned}\tag{2.10}$$

where $\nu(B_k),k=1,\ldots,N$, can be evaluated using (2.8), and an analogous formula holds for $\nu^\circ(B_k)$, whereas by (2.9)

$$(\nu^\circ-\nu)(B_0)=\beta\log\left(\frac{\alpha}{\alpha^\circ}\right)-\beta\{E_1(\alpha^\circ b_1)-E_1(\alpha b_1)\}.$$

Finally, for the ratio of Radon-Nikodym derivatives with respect to the law of a Gamma process $\mathbb{P}_{\tilde\nu}$ with the same parameter β we have

$$\begin{aligned}\log\left(\frac{\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\tilde\nu}}(X^\circ)}{\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\tilde\nu}}(X)}\right)=&-\sum_{k=1}^N\theta_k\left(\sum_{\substack{\Delta X_s^\circ\in B_k,\\0<s\le T}}\Delta X_s^\circ-\sum_{\substack{\Delta X_s\in B_k,\\0<s\le T}}\Delta X_s\right)\\&-\sum_{k=1}^N\rho_k(\mu_T^\circ(B_k)-\mu_T(B_k))\end{aligned}\tag{2.11}$$

for $X^\circ=(X_t^\circ\colon t\in[0,T])$ and $X=(X_t\colon t\in[0,T])$ with $X_T=X_T^\circ$, where $\mu_T^\circ(B_k)$ is defined analogously to $\mu_T(B_k)$ using X° instead X. Note that in this situation the righthand side is independent of the choice of the α parameter of the Gamma process measure used as the dominating measure.

3. Sampling the posterior

Using the usual convention in Bayesian statistics, denote the prior density of the parameters $\vartheta=(\alpha,\theta_1,\rho_1,\ldots,\theta_N,\rho_N)$ by $\pi(\vartheta)$, and use a similar generic notation $q(\vartheta;\vartheta^\circ)$ for the density of the corresponding (joint) proposal kernel evaluated in ϑ°, e.g. for a random move from ϑ to ϑ°. We first describe the Metropolis–Hastings algorithm to sample from the posterior in continuous time and next make a remark about the discretisation below.

- Initialise the parameters α, θ_k, ρ_k, $k=1$, $\ldots$, N, with their starting values. Initialise the segments $(X_t\colon t_{i-1}\le t\le t_i)$ with Gamma(β,α) bridges connecting observations $X_{t_{i-1}}$ and X_{t_i}, $i=1,\ldots,n$, using (2.5).

- Repeat the following steps:
  (i) Independently, for each $i=1,\dots,n$:

    (a) Sample Gamma(β,α) bridge proposals $(X_t^\circ\colon t_{i-1}\le t\le t_i)$ connecting observations $X_{t_{i-1}}$ and X_{t_i} using (2.5).
    (b) Sample $U_i\sim U[0,1]$. If

$$\frac{\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\nu_0}}(X^\circ)}{\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}_{\nu_0}}(X)}\ge U_i, \tag{3.1}$$

    set X_t to X_t° on $t_{i-1}\le t\le t_i$, otherwise keep X_t on $t_{i-1}\le t\le t_i$.

  (ii) Independently of step (i), propose $\vartheta^\circ\sim q(\vartheta;\cdot)$ and let ν° denote the corresponding Lévy measure. Sample $U\sim U[0,1]$. If

$$\frac{\mathrm{d}\mathbb{P}_{\nu^\circ}}{\mathrm{d}\mathbb{P}_\nu}(X)\frac{\pi(\vartheta^\circ)}{\pi(\vartheta)}\frac{q(\vartheta^\circ;\vartheta)}{q(\vartheta;\vartheta^\circ)}\ge U$$

  replace ϑ by ϑ°, otherwise retain ϑ.

Note that Step (i)(b) is the accept-reject step based on (2.11). Note that while we formulate the

3.1. Discretisation

The Metropolis-Hastings algorithm described above assumes one can sample the various processes and their bridges in continuous time. In practice it is possible to simulate the relevant processes only on a discrete grid of time points, which, however, can be made arbitrarily fine. In general it is preferable to work with a finite-dimensional approximation of a valid MCMC algorithm with infinite-dimensional state space instead of just an MCMC algorithm targeting a finite-dimensional approximation of the (joint) posterior, because the latter approach might have a singularity (resulting e.g. in vanishing acceptance probabilities) with growing dimension; see Beskos et al. (2008) for an extended perspective. We now outline how our original algorithm can be discretised. Consider a discrete time grid $t_{i,j}=t_{i-1}+\frac{j}{m}(t_i-t_{i-1})$ (and t_n) for $i=1,\dots,n$, $j=0,\dots,m-1$. Formula (2.5) remains valid also for discretised Gamma processes, and those are readily obtained by sampling from the distribution of their increments. On the other hand, in the likelihood expressions of Subsection 2.6 and in (3.1) we approximate the sum of jumps of the process X with sizes in B_k, $k\ge 0$, by the sum of the increments of X falling in B_k,

$$\sum_{\substack{\Delta X_s\in B_k,\\ 0<s\le T}}\Delta X_s\approx\sum_i\sum_j(X_{t_{i,j}}-X_{t_{i,j-1}})\mathbf{1}_{[X_{t_{i,j}}-X_{t_{i,j-1}}\in B_k]}. \tag{3.2}$$

4. Posterior consistency

In this section we study asymptotic frequentist properties of our nonparametric Bayesian procedure. The only comparable works for Lévy processes available in the literature are Gugushvili et al. (2015), Gugushvili et al. (2018) and Nickl and Söhl (2017a), but they deal with the class of compound Poisson processes, which is quite different from the class of θ-subordinators considered in this work. Arguments in favour of studying frequentist asymptotics for Bayesian procedures have been already given in the literature many times, and will not be repeated here; see, e.g., Wasserman (1998).

Our main result in this section is that under suitable regularity conditions, with growing sample size, our nonparametric Bayesian approach consistently recovers the parameters of interest. Thereby it stands on a solid theoretical ground.

4.1. Main results

Recall the setup of Section 2, which is complemented as follows. In this section we assume that the process X is observed at equidistant times t_i, $i=1,\ldots,n$. Without loss of generality we assume that our observations are $X_1,\ldots,X_n$. This assumption, which we did not require in earlier sections, implies that the increments of the process are independent and identically distributed. This way we can develop our arguments without the additional technical burden caused by non-i.i.d. increments. We denote the increments by $\mathcal{Z}_n=\{Z_1,\ldots,Z_n\}$, where $Z_i=X_i-X_{i-1}$, $i=1,\ldots,n$, and assume that under the true Lévy density v_0, $Z_1\sim\mathbb{Q}_{v_0}$. In general, $\mathbb{Q}_v$ will stand for the law of the increment Z_1 under the Lévy density v. Furthermore, we introduce the law $\mathbb{P}_{v_0}$ of $(X_t\colon t\in[0,1])$ under the true Lévy density v_0. The law of this path under the Lévy density v will be denoted by $\mathbb{P}_v$. For our asymptotic results, we will let the number of bins N depend on the sample size n, and write N_n instead. The prior Π_n below will be defined on a special class of Lévy densities, V_n. These are the densities that on the bins $B_k=(b_{k-1},b_k]$, $k=1,\ldots,N$, $b_0=0$, $b_1=\underline{b}$, $b_N=\overline{b}$, have the form $v(x)=\frac{\beta_0}{x}\exp(-\alpha x-\theta_k(x))$, with $\theta_k(x)=\rho_k+\theta_k x$, with the special choice $\rho_0=\theta_0=0$ and $\beta_0=1$. So, with the above notation,

$$V_n=\left\{v\colon v_{|B_k}(x)=\frac{\beta_0}{x}\exp(-\alpha x-\theta_k(x)),k=1,\ldots,N\right\}.$$

Below we present our first condition, and we comment on it and give additional explanations after it, as well as a few further comments after Condition 2.

Condition 1. *Let the function θ_0 have a compact support on the interval $[\underline{b},\overline{b}]$ where the boundary points $0<\underline{b}<\overline{b}<\infty$ are known, $\|\theta_0\|_\infty<\overline{\theta}$, and suppose θ_0 is λ-Hölder continuous, $|\theta_0(x)-\theta_0(y)|\leq L|x-y|^\lambda$ $(\lambda\in(0,1],\ L>0)$. Suppose also that $\alpha_0\in[\underline{\alpha},\overline{\alpha}]$ with known boundary points $0<\underline{\alpha}<\overline{\alpha}<\infty$. Finally, assume that the parameter β_0 is known and, without loss of generality, equal to* 1.

The assumption of known β requires some further comments. As we already remarked elsewhere, the parameter β plays a role similar to the dispersion coefficient σ of a stochastic differential equation driven by a Wiener process. Derivation of nonparametric Bayesian asymptotics for the latter class of processes (all of which is a recent work) historically proceeded from the assumption of a known σ to the one where σ is unknown and has to be estimated; see van der Meulen and van Zanten (2013), Gugushvili and Spreij (2014) and Nickl and Söhl (2017b). In that sense the fact that at this stage we assume β is known does not appear unexpected or unnatural. This assumption assists in derivation of useful bounds on the Kullback-Leibler and Hellinger distances between marginals of θ-subordinators under different Lévy triplets, which in general is the key to establishing consistency properties of nonparametric Bayesian procedures. We achieve this by reducing some of the intractable computations for these marginals to calculations involving laws of continuously observed θ-subordinators, for which we need precisely to assume that the parameter β is known; otherwise the corresponding laws are singular, which would yield only trivial and useless bounds.

Condition 2. *The coefficients θ_i, $i=1,\ldots N-1$, are equipped with independent uniform priors on the known interval $[-\overline{\theta},\overline{\theta}]$, $\overline{\theta}>0$. Likewise, the coefficients ρ_i, $i=1,\ldots,N-1$, are independent uniform on the interval $[-\overline{\theta},\overline{\theta}]$, whereas α is uniform on $[\underline{\alpha},\overline{\alpha}]$, $\overline{\alpha}>0$. We* *assume that all priors are independent. Implicitly, this defines a prior Π_n on the class of Lévy densities V_n, which are realisations from the prior.*

The assumption in Condition 2 that various priors are uniform can be relaxed to the assumption that they are supported on compacts and have densities bounded away from zero there. In fact, other assumptions in Conditions 1 and 2 can be relaxed at the cost of extra technical arguments in the proofs, but we do not strive for full generality in this work: a clean, readable presentation of our results and conciseness in the proofs is our primary goal.

Theorem 4.1 is our first main result in this section. Said shortly, it implies that our Bayesian procedure is consistent in probability; this in turn implies the existence of consistent Bayesian point estimates, see, e.g., Ghosal et al. (2000), pp. 506–507. We use the notation $\Pi_n(\mathrm{d}v \,|\, \mathcal{Z}_n)$ for the posterior measure. Also, $\mathbb{Q}_{v_0}^n$ denotes the law of the sample $\mathcal{Z}_n$ under the true Lévy density v_0 and $\mathbb{Q}_{v_0}^\infty$ denotes the law of the infinite sample $Z_1, Z_2, \ldots$ under the true Lévy density v_0.

THEOREM 4.1. *Assume that Conditions 1 and 2 hold and that $N_n \to \infty$ and $N_n/n \to 0$ as $n \to \infty$. Let $d_{\mathcal{H}}$ be the Hellinger metric. Then, for any fixed $\epsilon, \varepsilon > 0$,*

$$\mathbb{Q}_{v_0}^n\left(\Pi_n(v\colon d_{\mathcal{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) > \epsilon \,|\, \mathcal{Z}_n) > \varepsilon\right) \to 0$$

as $n \to \infty$.

Before proceeding further, we recall the definition of the Kullback-Leibler divergence $\mathcal{KL}$ and the discrepancy $\mathcal{V}$ for two probability measures $\mathbb{P} \ll \mathbb{Q}$:

$$\mathcal{KL}(\mathbb{P}, \mathbb{Q}) = \int \log\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right) \mathrm{d}\mathbb{P}, \quad \mathcal{V}(\mathbb{P}, \mathbb{Q}) = \int \log^2\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right) \mathrm{d}\mathbb{P}.$$

Here $\log^2$ stands for the square of the natural logarithm.

Proof of Theorem 4.1. The technical results needed in the proof are collected in Appendix A. Write $B(\epsilon) = \{v\colon d_{\mathcal{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \le \epsilon\}$ and note that

$$\Pi_n(B(\epsilon)^c \,|\, \mathcal{Z}_n) = \frac{\int_{B(\epsilon)^c} \prod_{i=1}^n \frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i) \Pi_n(\mathrm{d}v)}{\int \prod_{i=1}^n \frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i) \Pi_n(\mathrm{d}v)} = \frac{\mathrm{Num}_n}{\mathrm{Den}_n}.$$

We will treat the numerator and denominator separately. We start with the denominator. Define the set

$$K(\delta) = \{v\colon \mathcal{KL}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \le \delta, \mathcal{V}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \le \delta\},$$

where $\delta > 0$ is a fixed number. Let $\widetilde{\Pi}_n$ be a restriction of the prior Π_n to the set $K(\delta)$ normalised to have the total mass 1. We can write

$$\mathrm{Den}_n \ge \Pi_n(K(\delta)) \int_{K(\delta)} \prod_{i=1}^n \frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i) \widetilde{\Pi}_n(\mathrm{d}v).$$

By a standard argument as in Ghosal et al. (2000), p. 525, using Lemmas A.5 and A.7, on the sequence of events

$$A_n = \left\{ \int_{K(\delta)} \prod_{i=1}^n \frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i) \widetilde{\Pi}_n(\mathrm{d}v) \ge e^{-Cn\delta} \right\}$$

of $\mathbb{Q}_{v_0}^n$-probability tending to 1 as $n\to\infty$,

$$\frac{1}{\mathrm{Den}_n}\lesssim (c\delta)^{-2N_n}e^{Cn\delta}\lesssim e^{\overline{\delta}n}, \tag{4.1}$$

for $\overline{\delta}=2C\delta$, where for two sequences $\{a_n\}$ and $\{b_n\}$ of positive real numbers the notation $a_n\lesssim b_n$ indicates that there exists a constant $C>0$ that is independent of n such that $a_n\leq Cb_n$. We also used the fact that $N_n/n\to 0$. For future use remember that $\overline{\delta}$ can be made arbitrarily small by choosing δ small. This finishes bounding the term Den_n. Now we turn to Num_n. By Lemma A.10, on the sequence of events

$$B_n=\left\{\sup_{v\in B(\epsilon)^c}\prod_{i=1}^n\frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i)<\exp(-c_1n\epsilon^2)\right\}$$

of $\mathbb{Q}_{v_0}^n$-probability tending to 1 as $n\to\infty$, we have

$$\mathrm{Num}_n\leq\exp(-c_1n\epsilon^2). \tag{4.2}$$

The statement of the theorem now follows by choosing δ small enough, so that $\overline{\delta}<c_1\epsilon^2$. Indeed, for all big n one has on $A_n\cap B_n$ by combining the bounds (4.1) and (4.2) that $\Pi_n(v\colon d_{\mathcal{H}}(\mathbb{Q}_{v_0},\mathbb{Q}_v)>\epsilon\,|\,\mathcal{Z}_n)\leq\varepsilon$. Hence,

$$\mathbb{Q}_{v_0}^n\left(\Pi_n(v\colon d_{\mathcal{H}}(\mathbb{Q}_{v_0},\mathbb{Q}_v)>\epsilon\,|\,\mathcal{Z}_n)>\varepsilon\right)\leq\mathbb{Q}_{v_0}^n(A_n^c\cup B_n^c)\to 0,$$

which proves the theorem. □

The theorem has the following corollary that we will use in the proof of Theorem 4.2: a fixed ϵ can be replaced with a sufficiently slowly decaying ϵ_n.

Corollary 4.1. *For every fixed $\varepsilon>0$, there exists a sequence $\epsilon_n\to 0$, possibly depending on ε, such that*

$$\mathbb{Q}_{v_0}^n\left(\Pi_n(v\colon d_{\mathcal{H}}(\mathbb{Q}_{v_0},\mathbb{Q}_v)>\epsilon_n\,|\,\mathcal{Z}_n)>\varepsilon\right)\to 0$$

as $n\to\infty$.

Proof. The result follows from Lemma ⟨22⟩ on p. 181 in Pollard (2002). □

The metric for v, in which posterior convergence occurs in Theorem 4.1, is defined indirectly, in terms of the distance between the corresponding laws $\mathbb{Q}_v,\mathbb{Q}_{v_0}$. However, we will show that the theorem implies posterior consistency also in another and perhaps more natural metric for v. Let $\rightsquigarrow$ denote weak convergence of finite Borel measures and δ_0 be the Dirac measure at zero. The following proposition holds, as a consequence of Theorem 2 in Gnedenko (1939), see Appendix A for its proof. Note that in our setting the first component of the Lévy triplet is completely determined by the Lévy density, cf. (2.1).

Proposition 4.1. *Define for Lévy triplets $(\gamma_n,0,\nu_n)$, $(\gamma,0,\nu)$ finite Borel measures*

$$\widetilde{\nu}_n(\mathrm{d}x)=\gamma_n\delta_0(\mathrm{d}x)+(x^2\wedge 1)\nu_n(\mathrm{d}x),\quad \widetilde{\nu}(\mathrm{d}x)=\gamma\delta_0(\mathrm{d}x)+(x^2\wedge 1)\nu(\mathrm{d}x),$$

where we assume ν_n and ν are on $(0,\infty)$, and $\gamma_n=\int_0^1 x\nu_n(\mathrm{d}x)$ and $\gamma=\int_0^1 x\nu(\mathrm{d}x)$ are finite. Then $\mathbb{Q}_{v_n}\rightsquigarrow\mathbb{Q}_v$ if and only if $\widetilde{\nu}_n\rightsquigarrow\widetilde{\nu}$.

The following is our second main theoretical result, in which the metric for posterior contraction is defined directly for the Lévy density v (equivalently, Lévy measure ν). As the Lévy density uniquely determines the corresponding Lévy measure, in the theorem below as well as in its proof we will somewhat abuse the notation by considering posterior probabilities of certain sets of Lévy measures.

THEOREM 4.2. *Let $d_{\mathcal{W}}$ be any distance that metrises weak convergence of finite (signed) Borel measures. Then, for any fixed $\epsilon, \varepsilon > 0$,*

$$\mathbb{Q}^n_{v_0}(\Pi_n(\nu\colon d_{\mathcal{W}}(\widetilde{\nu}_0,\widetilde{\nu}) > \epsilon \,|\, \mathcal{Z}_n) > \varepsilon) \to 0$$

as $n \to \infty$.

Since the Lévy measures we consider are infinite in any neighbourhood of zero, using some weight function to convert them into finite measures does not appear to be an unnatural idea, cf. Comte and Genon-Catalot (2011) for a similar approach.

Proof of Theorem 4.2. Note that Hellinger consistency in Theorem 4.1 also holds when we replace $d_{\mathcal{H}}$ with $d_{\mathcal{W}}$ there, since Hellinger consistency implies consistency in any metric metrising weak convergence. The proof of the theorem is by contradiction. Assume that the statement of the theorem fails, so that there exist $\epsilon, \varepsilon, \delta > 0$, such that

$$\mathbb{Q}^n_{v_0}(\Pi_n(\nu\colon d_{\mathcal{W}}(\widetilde{\nu}_0,\widetilde{\nu}) > \epsilon \,|\, \mathcal{Z}_n) > \varepsilon) \geq \delta \tag{4.3}$$

along a subsequence of n, again denoted by n for economy of notation. On the other hand, by Theorem 4.1 and Corollary 4.1 we know that for any $\varepsilon', \delta' > 0$ there exists a sequence $\epsilon_n \to 0$, such that for all n large enough,

$$\mathbb{Q}^n_{v_0}(\Pi_n(v\colon d_{\mathcal{W}}(\mathbb{Q}_{v_0},\mathbb{Q}_v) \leq \epsilon_n \,|\, \mathcal{Z}_n) > 1-\varepsilon') \geq 1-\delta'. \tag{4.4}$$

Take $\delta' = \delta/2$. Then the elementary relation

$$P(A\cap B) = P(A)+P(B)-P(A\cup B) \geq P(A)+P(B)-1$$

together with (4.3)–(4.4) imply that the intersection of the events

$$A_n = \{\Pi_n(\nu\colon d_{\mathcal{W}}(\widetilde{\nu}_0,\widetilde{\nu}) > \epsilon \,|\, \mathcal{Z}_n) > \varepsilon\},$$
$$B_n = \{\Pi_n(v\colon d_{\mathcal{W}}(\mathbb{Q}_{v_0},\mathbb{Q}_v) \leq \epsilon_n \,|\, \mathcal{Z}_n) > 1-\varepsilon'\}$$

for all n large enough has $\mathbb{Q}^n_{v_0}$-probability at least $\delta/2$. In formula,

$$\mathbb{Q}^n_{v_0}(A_n\cap B_n) \geq \delta/2. \tag{4.5}$$

Let now $\varepsilon' = \varepsilon/2$, and suppose $\omega \in A_n \cap B_n$. Then by the same argument as above, for the realisation $\mathcal{Z}_n(\omega)$, the intersection of two sets

$$A' = \{\nu\colon d_{\mathcal{W}}(\widetilde{\nu}_0,\widetilde{\nu}) > \epsilon\}, \quad B'_n = \{v\colon d_{\mathcal{W}}(\mathbb{Q}_{v_0},\mathbb{Q}_v) \leq \epsilon_n\}$$

must have posterior mass at least $\varepsilon/2$, for all n large enough. Note that by this fact it also holds that

$$A_n\cap B_n = \left\{\Pi_n(A'\cap B'_n \,|\, \mathcal{Z}_n)\mathbf{1}_{[A_n\cap B_n]} \geq \varepsilon/2\right\}.$$

for all n large enough. However, by Proposition 4.1 the intersection $A'\cap B'_n$ is an empty set for $n\to\infty$, so that

$$\Pi_n(A'\cap B'_n \,|\, \mathcal{Z}_n)\mathbf{1}_{[A_n\cap B_n]} \to 0, \quad \mathbb{Q}^\infty_{v_0}\text{-a.s.}$$

But then, as $n \to \infty$,

$$\mathbb{Q}^n_{v_0}(A_n\cap B_n) = \mathbb{Q}^n_{v_0}\left(\Pi_n(A'\cap B'_n \,|\, \mathcal{Z}_n)\mathbf{1}_{[A_n\cap B_n]} \geq \varepsilon/2\right) \to 0.$$

This contradicts (4.5). The proof is completed. □

5. Example: Sum of two Gamma processes

Insurance theory, operational loss models, or more generally risk processes furnish a natural field of application for subordinators. In particular, a risk model based on Gamma process was extensively studied from a probabilistic point of view in the widely cited work Dufresne et al. (1991). On the other hand, a given risk process may itself be a result of conflation of several heterogeneous factors, for instance due to population heterogeneity. We may assume that individual risk processes can be modelled through independent Gamma processes. This is conceptually similar to using convolutions of gamma distributions in, e.g., storage models; see Mathai (1982). The cumulative risk process is again a Lévy process, though not necessarily gamma, as sums of independent Gamma processes are not necessarily Gamma. However, such sums can be closely approximated through θ-subordinators, as we will now demonstrate. It is enough to consider the particular case of a sum of two independent Gamma processes, the general case being only notationally more complex. Thus, let $\widetilde{X}=(\widetilde{X}_t\colon t\geq 0)$ and $\hat{X}=(\hat{X}_t\colon t\geq 0)$ be two independent Gamma processes with parameters (β_1,α_1) and (β_2,α_2). Let the process $X=(X_t\colon t\geq 0)$ be their sum, $X_t=\widetilde{X}_t+\hat{X}_t$. Its Lévy density is given by

$$v(x)=\frac{\beta_1}{x}e^{-\alpha_1 x}+\frac{\beta_2}{x}e^{-\alpha_2 x}.$$

The process X can be viewed as a mixture of phenomena happening at different time scales (slow and fast). For $x\to\infty$, the behaviour of v is determined by $\beta_1+\beta_2$ and $\min(\alpha_1,\alpha_2)$. On the hand, consider the equation

$$\frac{\beta_1}{x}e^{-\alpha_1 x}+\frac{\beta_2}{x}e^{-\alpha_2 x}=\frac{\beta_1+\beta_2}{x}e^{-\theta(x)-\alpha x},$$

where $\alpha>0$ will be chosen later on. Solving for θ, we get

$$\theta(x)=-\log\left(\frac{\beta_1 e^{-\alpha_1 x}+\beta_2 e^{-\alpha_2 x}}{\beta_1+\beta_2}\right)-\alpha x. \tag{5.1}$$

Now note that for $x\to 0$,

$$-\log\left(\frac{\beta_1 e^{-\alpha_1 x}+\beta_2 e^{-\alpha_2 x}}{\beta_1+\beta_2}\right)\approx\frac{\beta_1\alpha_1+\beta_2\alpha_2}{\beta_1+\beta_2}x.$$

We then take

$$\alpha=\frac{\beta_1\alpha_1+\beta_2\alpha_2}{\beta_1+\beta_2}.$$

This choice of α implies that the function θ is negligibly small in a neighbourhood of zero ($\theta(x)$ behaves as x^2 for x small). It then follows that the Lévy density of a sum of two independent Gamma processes can be closely approximated by the Lévy measure of the type (2.2), where θ is piecewise linear as in (2.7). Thus, θ-subordinators can be used to approximate, to an arbitrary degree of accuracy, sums of independent Gamma processes. For an illustration, see Figure 5.1, that plots the function $x\mapsto -\log(xv(x))$ together with the corresponding slope α at $x=0$, and the asymptote $\min(\alpha_1,\alpha_2)x-\mathrm{const}$ for Example 5.1 below.

We now consider a numerical example. All the computations in this work are performed using the software package **Bridge** (Schauer et al. (2017)) available for the Julia programming language, see Bezanson et al. (2017).

Example 5.1. For the simulation of the synthetic data we chose $\alpha_1=2.0$, $\beta_1=0.4$, $\alpha_2=0.2$, $\beta_2=0.04$. For these parameters the behaviour sample paths of both components is neither too similar nor too far apart (as judged by consulting Figure 5.1),

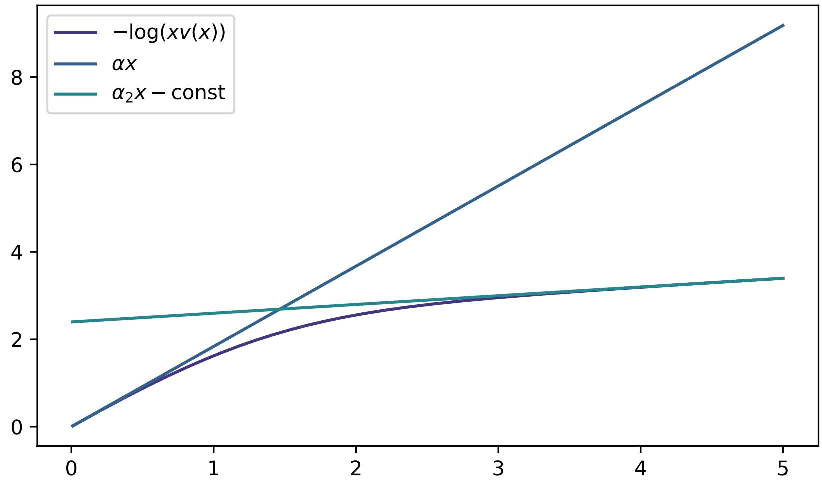

FIGURE 5.1. *The function $x \mapsto -\log(xv(x))$ in Example 5.1 together with the corresponding slope α at $x=0$, and the asymptote $\alpha_2 x - \mathrm{const} = \min(\alpha_1, \alpha_2)x - \mathrm{const}$.*

making this an interesting statistical problem. We simulated the process up to time $T=2000$ and took $n=10000$ observations at distance 0.2.

For the prior we chose $N=3$ with grid points $b=[1,2,4]$, $\alpha \sim \mathrm{Gamma}(2,1)$, $\theta_k \sim N(0,10)$ and $\rho_k \sim N(0,50)$, $k \geq 1$, conditional on the realisation fulfilling $\theta(x) \to \infty$ for $x \to \infty$.

In the data augmentation step we took intermediate points at distance 0.01.

In the Gibbs sampler in each step new Gamma bridges are proposed in the data augmentation step, followed by a parameter update Metropolis-Hastings step with normal random walk proposals. For the joint parameter update, using independent standard normal (Gaussian) innovations $Z_\alpha, Z_\theta, Z_\rho$ of appropriate dimensions, we set

$$\begin{aligned} \alpha &= \alpha + \sigma_\alpha Z_\alpha, \\ \theta^\circ &= \theta + \sigma_\theta Z_\theta - (\alpha^\circ - \alpha), \\ \rho^\circ &= \rho + \sigma_\rho Z_\rho, \end{aligned}$$

where $\sigma_\alpha = \sigma_\theta = 0.025$, $\sigma_\rho = 0.15$.

The MCMC algorithm was run for 200000 iterations. Figure 5.2 shows trace plots and running averages of the posterior samples of the parameters α and θ_1, θ_2, θ_3, ρ_1, ρ_2, ρ_3. Figure 5.3 shows marginal Bayesian credible bands for the function $\theta(x) + \alpha x$ contrasted with the true function given by (5.1). As evidenced by the size of the marginal posterior bands, for bins chosen as indicated the observations do contain information about the Lévy density on each bin.

6. Estimation of β

Thus far we assumed the parameter β in (2.2) is known. In practice such an assumption cannot always be justified, and the question arises how to adapt our Bayesian

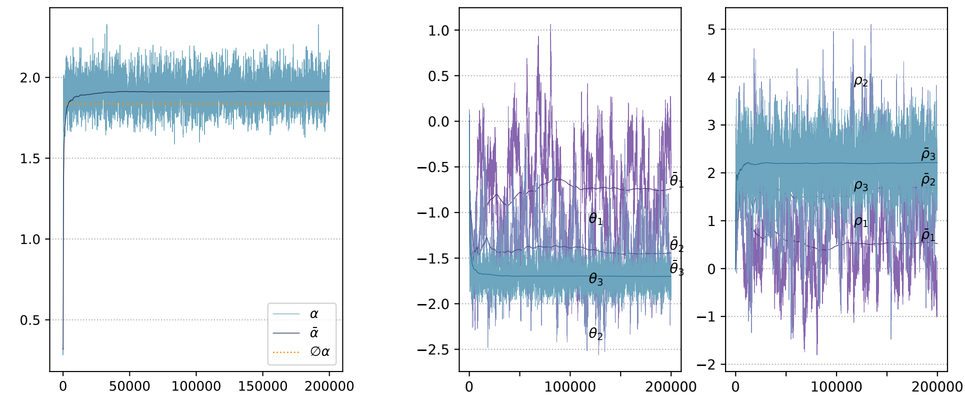

FIGURE 5.2. *Trace plots of the parameters for Example 5.1. First panel: trace and running average of samples of α, $\alpha = \frac{\beta_1\alpha_1+\beta_2\alpha_2}{\beta_1+\beta_2}$ is marked as yellow line. Second panel: trace and running average of samples of θ_1, θ_2, θ_3. Last panel: trace and running average of samples of ρ_1, ρ_2, ρ_3. Running averages of posterior samples of parameters are indicated by decorating the parameter with a bar.*

computational methodology to the case of an unknown β. It should be noted that when viewed from a Bayesian data augmentation point of view, the parameter β is rather different from the parameters α,θ: knowledge of β is required in order to write down the likelihood of a continuously observed process X. As we noted before, in a sense, the parameter β_0 plays a role similar to the diffusion coefficient of the stochastic differential equation driven by the Wiener process. Over the years, computational methods for handling the case of the unknown diffusion coefficient have been developed in the literature, see, e.g., van der Meulen and Schauer (2017) and references therein. The basic idea of one such approach is that the laws of the bridge proposals can be understood as push forwards of the laws of some underlying random processes. For Gamma process bridges (our bridge proposals) such a push forward map is given by (2.5) and $\widetilde{\mathbb{P}}_\beta$ is the law of a Gamma process with parameter β. In the case of diffusion processes, where the bridge proposals are defined as strong solutions of stochastic differential equations, the law $\widetilde{\mathbb{P}}$ of the driving Brownian motion serves this purpose as a single law common to all models with different diffusion coefficients σ^2. In our Lévy setting the laws are different – and mutually singular – but are chosen in such a way that Metropolis-Hastings steps from one law $\widetilde{\mathbb{P}}_\beta$ to another $\widetilde{\mathbb{P}}_{\beta^\circ}$ can be balanced.

We now move to providing details of our approach. Making use of the Markov property of a Lévy process, we can restrict our attention to the case of a single bridge segment from 0 at time $t=0$ to x_T at time $t=T$. A generalisation to several bridges is straightforward. Since in our MCMC sampler for the posterior in an update step for the parameter β, we will keep all other parameters fixed, in this section we can assume all the parameters except β are known and fixed. In what follows, $\mathbb{P}_\beta$ denotes the law of a Lévy process with Lévy measure

$$\nu(\mathrm{d}x) = \frac{\beta}{x}\mathrm{e}^{-\alpha x-\theta(x)},$$

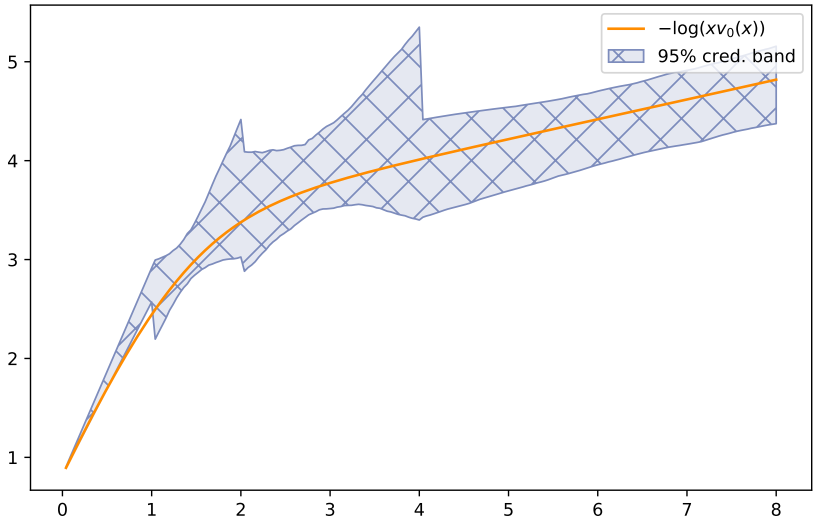

FIGURE 5.3. *Marginal Bayesian credible bands for Example 5.1 for the function $\theta(x)+\alpha x$, based on all samples. Orange: truth from equation (5.1).*

and $\widetilde{\mathbb{P}}_\beta$ denotes the law of a Gamma(β,α) process $\widetilde{X}$, both defined on $[0,T]$. Next, p_β and $\widetilde{p}_\beta$ denote marginal densities of X_T and $\widetilde{X}_T$; furthermore, conditional laws (under $\mathbb{P}_\beta$ and $\widetilde{\mathbb{P}}_\beta$) of the full Lévy path given $X_T = x_T$ are denoted $\mathbb{P}^\star_\beta$ and $\widetilde{\mathbb{P}}^\star_\beta$. The map g defined in (2.5) is written as $g_{x_T} = g_{0,x_T}$. Table 6.1 summarises the notation for easy reference.

Process	Law	Marginal density at $t=T$	Law conditional on $X_T = x_T$
X	$\mathbb{P}_\beta$	p_β	$\mathbb{P}^\star_\beta$
$\widetilde{X}$	$\widetilde{\mathbb{P}}_\beta$	$\widetilde{p}_\beta \sim \text{Gamma}(t\beta,\alpha)$	$\widetilde{\mathbb{P}}^\star_\beta$
$\widetilde{X}^\circ$	$\mathbb{Q}_{\beta,\beta^\circ}(\widetilde{X};\cdot)$	–	–

TABLE 6.1. *Notation chart for Section 6.*

Let β be equipped with a prior distribution Π assumed to be given by a density π. With $\Psi_\beta = \frac{\mathrm{d}\mathbb{P}_\beta}{\mathrm{d}\widetilde{\mathbb{P}}_\beta}$, the joint posterior of (β, X) given $X_T = x_T$ can be factorised as

$$\begin{aligned}\Pi((\mathrm{d}\beta,\mathrm{d}X)\mid x_T) &\propto \pi(\beta)p_\beta(x_T)\frac{\mathrm{d}\mathbb{P}^\star_\beta}{\mathrm{d}\widetilde{\mathbb{P}}^\star_\beta}(X)\widetilde{\mathbb{P}}^\star_\beta(\mathrm{d}X)\mathrm{d}\beta \\ &= \pi(\beta)\widetilde{p}_\beta(x_T)\Psi_\beta(X)\widetilde{\mathbb{P}}^\star_\beta(\mathrm{d}X)\mathrm{d}\beta,\end{aligned} \tag{6.1}$$

where the second equality follows from (2.6).

Define a measure

$$\Lambda(\mathrm{d}\beta,\mathrm{d}\widetilde{X})=\pi(\beta)\widetilde{p}_\beta(x_T)\Psi_\beta\big(g_{x_T}(\widetilde{X})\big)\widetilde{\mathbb{P}}_\beta(\mathrm{d}\widetilde{X})\mathrm{d}\beta. \tag{6.2}$$

Then $\Pi((\mathrm{d}\beta,\mathrm{d}X)\,|\,x_T)$ is proportional to the image measure of $\Lambda(\mathrm{d}\beta,\mathrm{d}\widetilde{X})$ under $(\beta,\widetilde{X})\mapsto(\beta,g_{x_T}(\widetilde{X}))$, because $g_{x_T}(\widetilde{X})\sim\widetilde{\mathbb{P}}^\star_\beta$ for $\widetilde{X}\sim\widetilde{\mathbb{P}}_\beta$. Note that Λ does not involve the intractable density p_β, and Ψ_β is analytically known, cf. (2.11).

We define a Metropolis-Hastings chain with Λ as its invariant measure, from which samples $(\beta,g_{x_T}(\widetilde{X}))$ of the joint posterior in (6.1) are obtained. As g_{x_T} is not invertible, this is a data augmentation procedure, only that $\widetilde{X}$, unlike the augmented path, can hardly be interpreted as an unobserved object.

Let $\widetilde{X}$ be a $\mathrm{Gamma}(\beta,\alpha)$ process and assume that a proposal density for β° is given by $q(\beta;\beta^\circ)$. For a given β°, if $\beta^\circ>\beta$, set $\widetilde{X}^\circ_t=\widetilde{X}_t+\widetilde{X}'_t$, where $\widetilde{X}'\sim\widetilde{\mathbb{P}}'$ is an independent $\mathrm{Gamma}(\beta^\circ-\beta,\alpha)$ process. If $\beta^\circ<\beta$, then set

$$\widetilde{X}^\circ_t=\sum_{\substack{\Delta\widetilde{X}_s>0\\ s\le t}}U_s\Delta\widetilde{X}_s,$$

where U_s is an independent collection of $\mathrm{Bernoulli}(\beta^\circ/\beta)$ random variables indexed by a countable set $\{s\colon\Delta\widetilde{X}_s>0\}$. By Lemma 6.1 (i) and (ii) ahead, $\widetilde{X}^\circ$ is a $\mathrm{Gamma}(\beta^\circ,\alpha)$ process with law $\widetilde{\mathbb{P}}_{\beta^\circ}$. Denote the probability kernel for a transition from $\widetilde{X}$ to $\widetilde{X}^\circ$ (conditional on β and β°), which is implied by the preceding construction, by $\mathbb{Q}_{\beta,\beta^\circ}(\widetilde{X};\cdot)$.

We will show that proposing a move from β to β° from q and subsequently from $\widetilde{X}$ to $\widetilde{X}^\circ$ and accepting it with acceptance probability $A((\beta,\widetilde{X}),(\beta^\circ,\widetilde{X}^\circ))$ to be derived below, is a reversible move for Λ. By Tierney (1998), this follows if detailed balance

$$\begin{aligned}\Lambda(\mathrm{d}\beta,\mathrm{d}\widetilde{X})q(\beta;\beta^\circ)&\mathbb{Q}_{\beta,\beta^\circ}((\beta,\widetilde{X});(\mathrm{d}\beta^\circ,\mathrm{d}\widetilde{X}^\circ))A((\beta,\widetilde{X}),(\beta^\circ,\widetilde{X}^\circ))\mathrm{d}\beta^\circ\\ &=\Lambda(\mathrm{d}\beta^\circ,\mathrm{d}\widetilde{X}^\circ)q(\beta^\circ;\beta)\mathbb{Q}_{\beta^\circ,\beta}((\beta^\circ,\widetilde{X}^\circ);(\mathrm{d}\beta,\mathrm{d}\widetilde{X}))A((\beta^\circ,\widetilde{X}^\circ),(\beta,\widetilde{X}))\mathrm{d}\beta\end{aligned}$$

holds. By (6.2) and Lemma 6.2 given below, the lefthand side is equal to

$$\pi(\beta)\widetilde{p}_\beta(x_T)\Psi_\beta\big(g_{x_T}(\widetilde{X})\big)q(\beta;\beta^\circ)\mu((\mathrm{d}\beta,\mathrm{d}\widetilde{X}),(\mathrm{d}\beta^\circ,\mathrm{d}\widetilde{X}^\circ))A((\beta,\widetilde{X}),(\beta^\circ,\widetilde{X}^\circ))$$

with μ defined in Lemma 6.2 ahead. Therefore, choosing

$$A((\beta,\widetilde{X}),(\beta^\circ,\widetilde{X}^\circ))=\max\left(\frac{\pi(\beta^\circ)}{\pi(\beta)}\frac{\widetilde{p}_{\beta^\circ}(x_T)}{\widetilde{p}_\beta(x_T)}\frac{\Psi_{\beta^\circ}(g_{x_T}(\widetilde{X}^\circ))}{\Psi_\beta(g_{x_T}(\widetilde{X}))}\frac{q(\beta^\circ;\beta)}{q(\beta;\beta^\circ)},1\right)$$

can be seen to make the expressions on both sides of the last display equal, thanks to (6.2) and Lemma 6.2 together with the symmetry of μ established in Lemma 6.2.

Lemma 6.1. *Let* $\widetilde{X}_t=\sum_{s\le t\colon\Delta\widetilde{X}_s>0}\Delta\widetilde{X}_s$ *be a* $\mathrm{Gamma}(\beta,\alpha)$ *process.*

(i) If $\beta^\circ>\beta$ *and* X' *is an independent* $\mathrm{Gamma}(\beta^\circ-\beta,\alpha)$ *process, then*

$$\widetilde{X}^\circ_t=\widetilde{X}_t+X'_t,$$

is a $\mathrm{Gamma}(\beta^\circ,\alpha)$ *process.*

(ii) If $\beta^\circ < \beta$ and U_s is a countable collection of Bernoulli(β°/β) *random variables indexed by $\{s\colon \Delta\widetilde{X}_s > 0\}$, then*

$$\widetilde{X}^\circ_t = \sum_{\substack{\Delta\widetilde{X}_s>0\\ s\le t}} U_s \Delta\widetilde{X}_s$$

is a Gamma(β°,α) *process.*

Proof. We sketch the proof. The first part is straightforward. The second part is more involved, but is a standard technique to sample Lévy processes by thinning marked Poisson point processes, see the rejection method in Rosiński (2001); it could also be derived from the proof of Lemma 6.2. □

LEMMA 6.2 (Transdimensional balance). *For $\beta,\beta^\circ > 0$,*

$$\widetilde{\mathbb{P}}_\beta(\mathrm{d}\widetilde{X})\mathbb{Q}_{\beta,\beta^\circ}(\widetilde{X};\mathrm{d}\widetilde{X}^\circ) = \widetilde{\mathbb{P}}_{\beta^\circ}(\mathrm{d}\widetilde{X}^\circ)\mathbb{Q}_{\beta^\circ,\beta}(\widetilde{X}^\circ;\mathrm{d}\widetilde{X}) \tag{6.3}$$

holds, and

$$\mu((\mathrm{d}\beta,\mathrm{d}\widetilde{X}),(\mathrm{d}\beta^\circ,\mathrm{d}\widetilde{X}^\circ)) = \mathrm{d}\beta\mathrm{d}\beta^\circ\widetilde{\mathbb{P}}_\beta(\mathrm{d}\widetilde{X})\mathbb{Q}_{\beta,\beta^\circ}(\widetilde{X};\mathrm{d}\widetilde{X}^\circ)(=\mu((\mathrm{d}\beta^\circ,\mathrm{d}\widetilde{X}^\circ),(\mathrm{d}\beta,\mathrm{d}\widetilde{X})))$$

defines a symmetric measure.

Proof. Without loss of generality, assume $\beta^\circ > \beta$. The process $\widetilde{X}$ is determined by the jump times $J^i = \{s\colon \Delta\widetilde{X}_s \in [u_i,v_i)\}$ and jump sizes $\Delta\widetilde{X}_s$, $s\in J^i$ on all disjoint strips $[0,T]\times[u_i,v_i)$, where $(0,\infty) = \bigcup_{i=1}^\infty [u_i,v_i)$ with $v_0=\infty$, $v_i = 1/i$, $u_i = 1/(i+1)$. Similar to J^i, denote by $J^{i,\circ}$ the jump times of $\widetilde{X}^\circ$ with their sizes in $[u_i,v_i)$. The number of jumps $|J^i|$ is Poisson(βc^i) distributed, with density written as $p^i_\beta(|J^i|)$, where

$$c^i = T\widetilde{\nu}([u_i,v_i))/\beta = T\widetilde{\nu}^\circ([u_i,v_i))/\beta^\circ.$$

Conditional on $|J^i|$, the elements of J^i are independent uniforms on $[0,T]$, and $\Delta\widetilde{X}_s$, $s\in J^i$, are independently

$$T\widetilde{\nu}(\cdot)|_{[u_i,v_i)}/(\beta c^i) = T\widetilde{\nu}^\circ(\cdot)|_{[u_i,v_i)}/(\beta^\circ c^i) \tag{6.4}$$

distributed; note that either side of (6.4) does not depend on β, which cancels from the formulae. Let $q^i_{\beta,\beta^\circ}(n;n^\circ)$ denote the counting density of moving from $|J^i| = n$ to $|J^{i,\circ,}| = n^\circ$ under $\mathbb{Q}_{\beta,\beta^\circ}(\widetilde{X};\cdot)$. This is well defined, as $|J^{i,\circ,}|$ under $\mathbb{Q}_{\beta,\beta^\circ}(\widetilde{X};\cdot)$ only depends on $\widetilde{X}$ through $|J^i|$.

On each strip it holds that

$$\begin{aligned}
p^i_\beta(|J^i|)q^i_{\beta;\beta^\circ}(|J^i|;|J^{i,\circ}|) &= \frac{(\beta c^i)^{|J^i|}e^{-\beta c^i}}{|J^i|!}\frac{((\beta^\circ-\beta)c^i)^{|J^{i,\circ}|-|J^i|}e^{-(\beta^\circ-\beta)c^i}}{(|J^{i,\circ}|-|J^i|)!}\\
&= \frac{(\beta^\circ c^i)^{|J^{i,\circ}|}e^{-\beta^\circ c^i}}{|J^{i,\circ}|!}\binom{|J^{i,\circ}|}{|J^i|}\left(\frac{\beta}{\beta^\circ}\right)^{|J^i|}\left(1-\frac{\beta}{\beta^\circ}\right)^{|J^i|-|J^{i,\circ}|}\\
&= p^i_{\beta^\circ}(|J^{i,\circ}|)q^i_{\beta^\circ;\beta}(|J^{i,\circ}|;|J^i|),
\end{aligned}$$

as the number of jumps of $\widetilde{X}'$ (the notation is as in Lemma 6.1 (i)) in $[u_i,v_i)$ has the Poisson$((\beta^\circ-\beta)c^i)$ distribution. Note that

$$\prod_{s\in J^{i,\circ}} p((t^\circ_s,\Delta\widetilde{X}^\circ_s)) = \prod_{s\in J^i} p((t_s,\Delta\widetilde{X}_s)) \prod_{s\in J^{i,\prime}} p((t'_s,\Delta\widetilde{X}'_s))$$

where we used that the joint density p is the same for all arguments by (6.4).

Therefore on each strip it holds that

$$\widetilde{\mathbb{P}}_\beta(\mathrm{d}\pi^i(\widetilde{X}))\mathbb{Q}_{\beta,\beta^\circ}(\pi^i(\widetilde{X});\mathrm{d}\pi^i(\widetilde{X}^\circ))=\widetilde{\mathbb{P}}_{\beta^\circ}(\mathrm{d}\pi^i(\widetilde{X}^\circ))\mathbb{Q}_{\beta^\circ,\beta}(\pi^i(\widetilde{X}^\circ);\mathrm{d}\pi^i(\widetilde{X})), \tag{6.5}$$

where $\pi^i\colon \widetilde{X}\mapsto(|J^i|,\{(t_s,\Delta\widetilde{X}_s),s\in J^i\})$. The statement of the lemma now follows from an application of Lemma B.1, by which (6.5) together with the independent increments property of the jump measure of a Lévy process gives (6.3) and thus also the symmetry of μ. □

The terminology 'transdimensional balance' for (6.3) is suggested by a connection to the transdimensional MCMC in Green (1995). In fact, note that for $\beta^\circ>\beta$, with $\widetilde{X}\sim\widetilde{\mathbb{P}}_\beta$ and $\widetilde{X}'$ as in Lemma 6.1, the proposal

$$\widetilde{X}^\circ_t=\begin{cases}\widetilde{X}_{t\beta^\circ/\beta} & t\le\frac{\beta}{\beta^\circ}T\\ \widetilde{X}_T+\widetilde{X}'_{\frac{\beta^\circ-\beta}{\beta^\circ}(t-\frac{\beta}{\beta^\circ}T)} & t>\frac{\beta}{\beta^\circ}T,\end{cases}$$

has also distribution $\widetilde{P}_\beta$. This closely resembles the 'standard template' given by Green (1995) for a transdimensional reversible jump move, although here all spaces are infinite-dimensional.

6.1. Discretisation In order to be able to employ the result of this section in practice, we now discuss how to perform steps (i) and (ii) of Lemma 6.1 for the approximations defined on the discrete time grid as introduced in Subsection 3.1. Step (i) is straightforward, noting that for $\beta^\circ>\beta$,

$$\widetilde{X}^\circ_{t+h}-\widetilde{X}^\circ_t\,|\,\widetilde{X}_{t+h}-\widetilde{X}_t\sim\widetilde{X}_{t+h}-\widetilde{X}_t+Z,$$

where $Z\sim\mathrm{Gamma}(h(\beta^\circ-\beta)\alpha)$.

For step (ii), when $\beta^\circ<\beta$, we use the following formula linking the law of the increments of the thinned process with the Beta distribution,

$$\widetilde{X}^\circ_{t+h}-\widetilde{X}^\circ_t\,|\,\widetilde{X}_{t+h}-\widetilde{X}_t\sim\left(\widetilde{X}_{t+h}-\widetilde{X}_t\right)Z,$$

where $Z\sim\mathrm{Beta}(h\beta^\circ,h(\beta-\beta^\circ))$.

7. Example: sum of two Gamma processes, unknown β

We revisit Example 5.1 from Section 5, but now additionally assuming the parameter β is unknown. We endow β with an independent uniform prior on the interval $[0.1,1000]$. To estimate β, we perform a transdimensional move, as explained in Section 6, at every 5th iteration in the otherwise unchanged algorithm from Section 5. Proposals for β° are obtained from a random walk with independent Gaussian increments, with standard deviation $\sigma_\beta=0.01$. No further tuning is necessary.

Figure 7.1 shows trace plots and running averages of the posterior samples of the parameters α and β. The data – for the parameter values considered – is informative for the parameter β and the Metropolis-Hastings chain sampling from the posterior of β mixes fast. While not covered by our posterior consistency result, the results of the numerical experiment indicate that the sampling procedure for β integrates seamlessly into the algorithm. Figure 7.2 shows trace plots and running averages of the posterior samples of θ_1, θ_2, θ_3 and of ρ_1, ρ_2, ρ_3. Figure 7.3 shows histograms of the posterior samples of α and β, whereas Figure 7.4 shows histograms of the posterior samples of θ_1, θ_2, θ_3 and of ρ_1, ρ_2, ρ_3. Figure 7.5 shows marginal Bayesian 95 % credible bands for the function $-\log(xv(x))$ contrasted with the true function $-\log(xv_0(x))$ given by (5.1). The conclusion is that we are able to recover the qualitative properties (as indicated by the asymptotes in Figure 5.1) of the process in both time scales from the discrete observations.

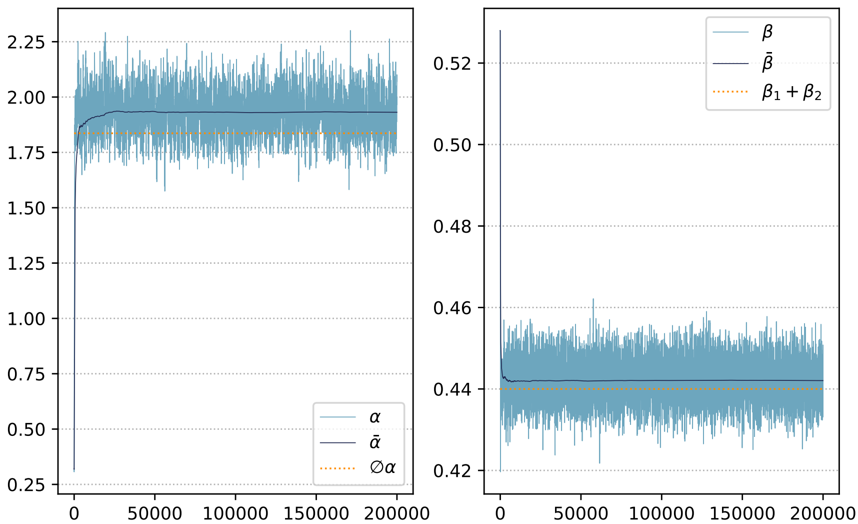

FIGURE 7.1. *Trace plots of the parameters α and β for Example 5.1. Left: trace and running average ($\bar{\alpha}$) of samples of α. The value $\o\alpha = \frac{\beta_1\alpha_1+\beta_2\alpha_2}{\beta_1+\beta_2}$ is marked as a dotted yellow line. Right: trace and running average of samples of β. The value $\beta_1+\beta_2$ is marked as a dotted yellow line.*

8. Danish data on fire losses

Over the last two decades there has been an increasing interest in applying Bayesian methods to insurance problems, see, e.g., Hong and Martin (2017a) and references therein. Hong and Martin (2017b) apply a Dirichlet process mixture prior to model the density of insurance claim sizes, and provide motivation for using a nonparametric Bayesian approach in the actuarial science. In this section we will apply our Bayesian approach to the Danish data on large fire losses. This dataset is a standard test example in extreme value theory, and from that point of view it has been a subject of several deep studies, such as McNeil (1997) and Resnick (1997). Our goals here are more modest, and aim at demonstrating the facts that firstly, θ-subordinators can be potentially used to capture some aggregate features of the Danish data on large fire losses, and secondly, statistical inference for real data modelled through such processes can be successfully performed using the Bayesian methodology developed in this paper. This can be viewed as a partial empirical investigation of the risk model based on Gamma processes from Dufresne et al. (1991). As observed in Hewitt and Lefkowitz (1979), a single standard distribution, such as the gamma, log-gamma or log-normal distribution, may not suffice to adequately model the distribution of individual insurance losses. For instance,

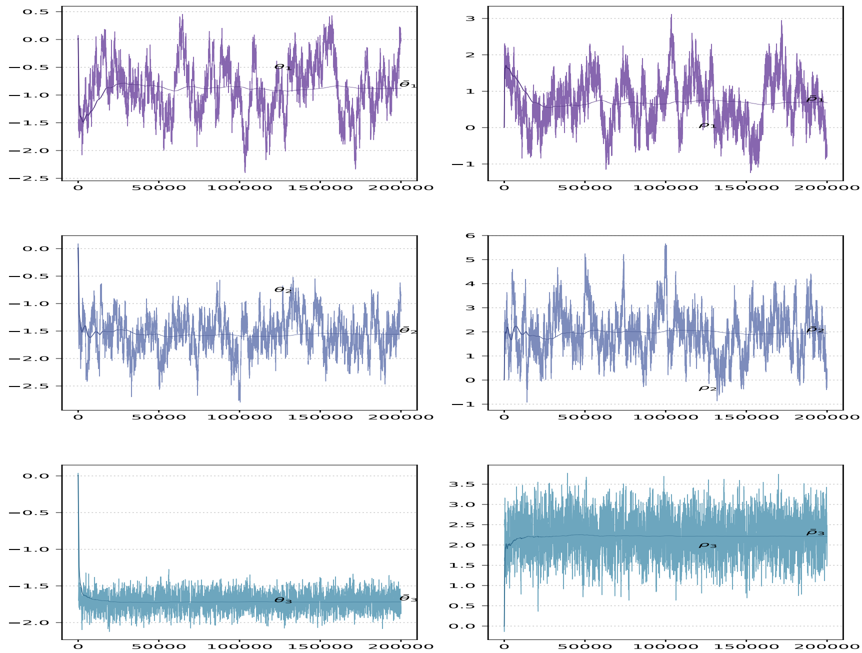

Figure 7.2. *Trace plots of the parameters for Example 5.1. Left column: trace and running average of samples θ_1, θ_2, θ_3. Right column: trace and running average of samples of ρ_1, ρ_2, ρ_3,*

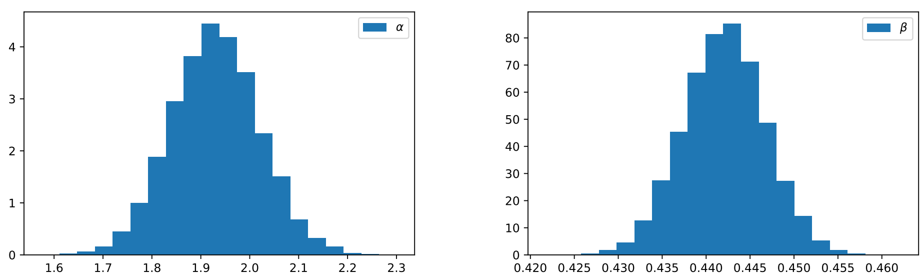

Figure 7.3. *Histograms of the posterior samples of α (left) and β (right) for Example 5.1.*

multimodality in claim size distribution may result from presence of hidden factors or due to existence of illegal practices, such as exaggeration of injuries and excessive treatment costs, that are well-documented in auto insurance; see, e.g., Rempala and Derrig (2005) and the references therein. Since allowing for greater flexibility, in particular multimodality, in claim size distribution modelling is likely to result in multimodality of marginal distributions of the cumulative risk process, using a θ-subordinator instead of a Gamma process to model evolution of the cumulative risk process over time a priori appears to be a sound approach.

8.1. Data description and visualisation

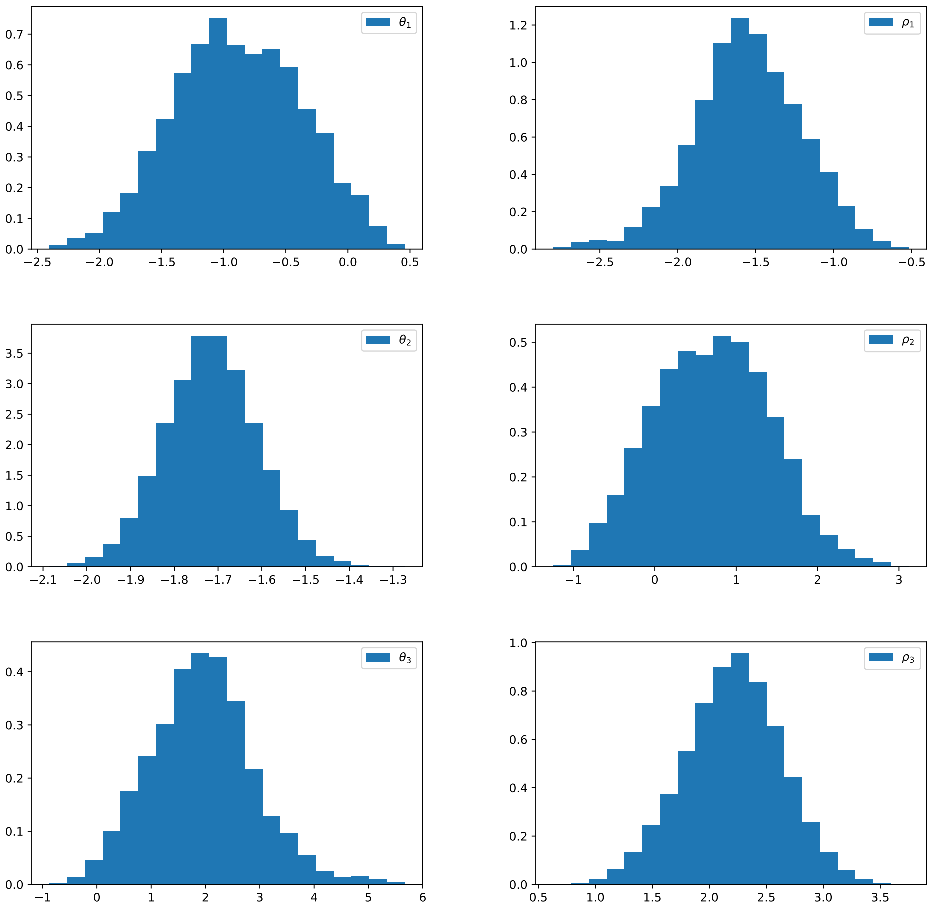

FIGURE 7.4. *Histograms of the posterior samples of the parameters for Example 5.1. Left column: parameters θ_1, θ_2, θ_3. Right column: parameters ρ_1, ρ_2, ρ_3.*

A succinct description of the Danish data on large fire losses can be found on p. 298 in Embrechts et al. (1997). The dataset (scaled for privacy reasons) comprises 2167 fire losses (adjusted suitably for inflation to reflect the 1985 values) in Denmark over the 10 year period starting on 6 January 1980 and ending on 30 December 1990, that exceed in size one million DKK, and that were registered by Copenhagen Reinsurance. The rationale for thresholding losses at one million DKK is given in McNeil (1997), pp. 119–120, and consists in the fact that in practice it is virtually impossible to collect exhaustive data on small losses: insurance is typically provided against significant losses, while small losses are dealt with by insured parties directly.

The data can be accessed through the **QRM** package in **R** under the name `danish`. The time plot of the data is given in the left panel of Figure 8.1. Presence of several exceedingly large losses is apparent from the plot, and therefore we use a logarithmic transformation to stabilise extreme variations in the data. Furthermore, this transforms observations on $[1,\infty)$ to observations on $[0,\infty)$, the support of the marginal distribu-

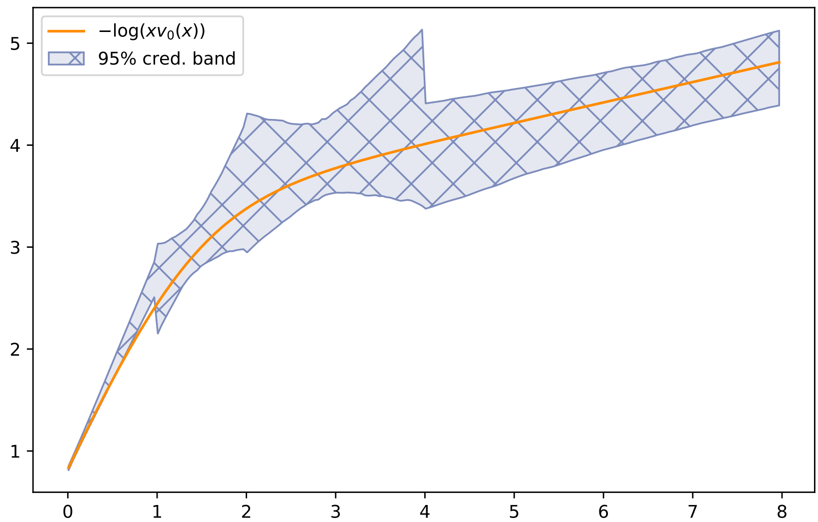

FIGURE 7.5. *Marginal Bayesian credible bands for Example 5.1 for the function* $-\log(xv(x))$ *based on all samples. Red: truth* $-\log(xv_0(x))$ *from equation* (5.1).

tions of a θ-subordinator. One feature of the data is that on numerous days no losses have been registered. This is not compatible with the behaviour of an infinite activity subordinator; in fact, such a subordinator X with probability one must have an infinite number of jumps in every finite time interval, and hence its increments must be strictly positive with probability one. A simple fix to this is to aggregate log losses over longer time periods than daily ones; aggregation over weekly periods (from Monday to Sunday) turned out to be sufficient (except few cases, where we had to aggregate data over periods of two weeks). The aggregated data on a logarithmic scale is displayed in the right panel of Figure 8.1. The idea of aggregation is a natural one, and embodies the fact that a probabilistic model unsuitable on a certain time scale may very well be appropriate on another time scale. In fact, already Albert Einstein in his classical paper on the Brownian motion observed that his model for displacement of a Brownian particle becomes inapplicable as the time interval between successive measurements of positions of a Brownian particle becomes increasingly small; see pp. 380–381 in Einstein (1906).

According to the exploratory analysis of the transformed data that we supply in Appendix C, the data can be modelled as an i.i.d. sequence that follows a Gamma-like distribution, but perhaps is not genuinely Gamma. This suggests a possibility of using a θ-subordinator to model the data.

8.2. Modelling fire losses with a θ-subordinator

Because the sample size is much smaller compared to our simulation examples, we chose $N=1$ corresponding to a single grid point $b_1=2$ and four parameters α, β, θ_1, ρ_1. In light of Example 5.1 and in order to improve mixing of the chain, we use a reparameterisation $\alpha_1=\alpha+\theta_1$, $\beta_1=\beta\exp(-\rho)$, and work with four parameters α, β,

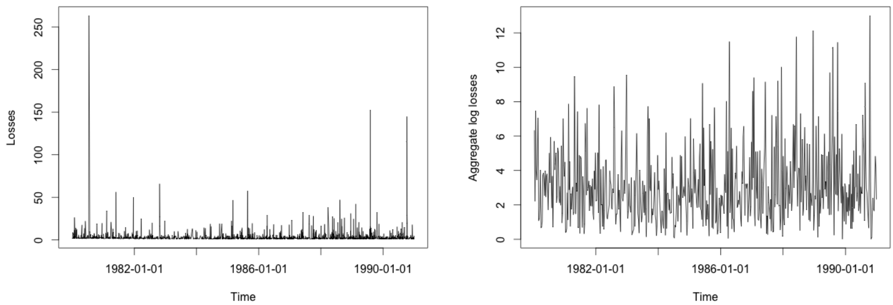

FIGURE 8.1. *Danish data on large fire losses.* Left*: original daily data (the unit is one million DKK).* Right*: logarithmically transformed and aggregated data.*

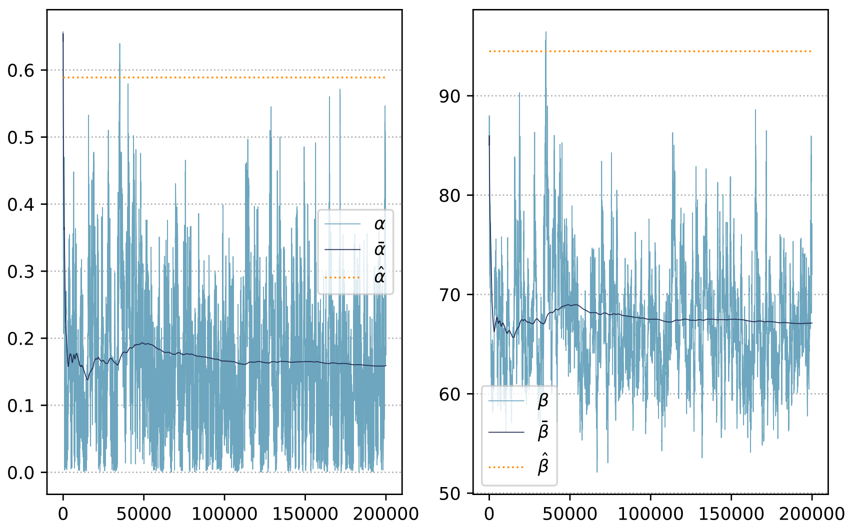

FIGURE 8.2. *Trace plots of the parameters α and β for the fire loss data. Left: trace and running average of samples of α. (The latter indicated by $\bar{\alpha}$.) The maximum likelihood estimate $\hat{\alpha}$ of α using a Gamma process model is marked as the dotted yellow line. Right: trace and running average of samples of β. (The latter indicated by $\bar{\beta}$.) The maximum likelihood estimate $\hat{\beta}$ of β using a Gamma process model is marked as the yellow dotted line.*

α_1, β_1, so that

$$v(x) = \begin{cases} \frac{\beta}{x}\exp(-\alpha x) & x \le b_1, \\ \frac{\beta_1}{x}\exp(-\alpha_1 x) & x > b_1. \end{cases}$$

A priori we equip these four parameters with independent Gamma distributions, with mean 0.75 and variance 0.36 for the parameters α, α_1, and mean 90 and variance 2500

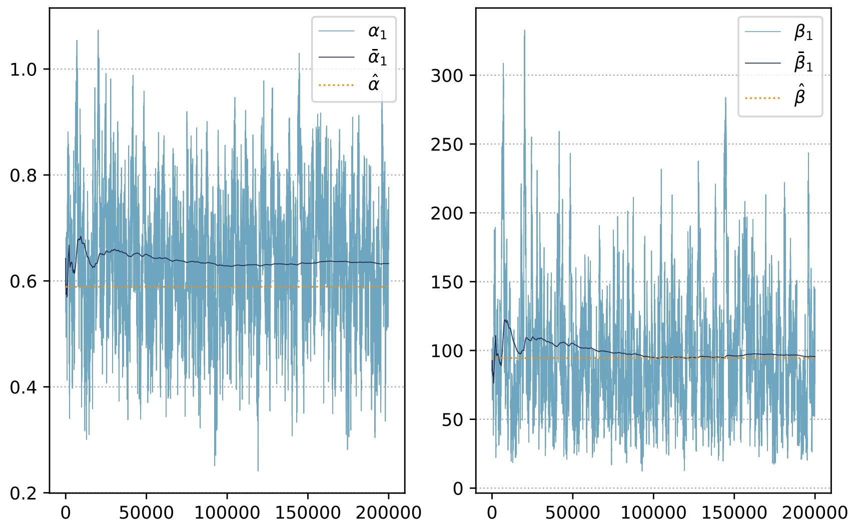

FIGURE 8.3. *Trace plots of the parameters used for the bin (b_1, ∞) for the fire loss data. Left: trace and running average of the samples of α_1. The maximum likelihood estimate of α using a Gamma process model is marked as yellow line. Right: trace and running average of the samples of β_1. The maximum likelihood estimate of β using a Gamma process model is marked as yellow line.*

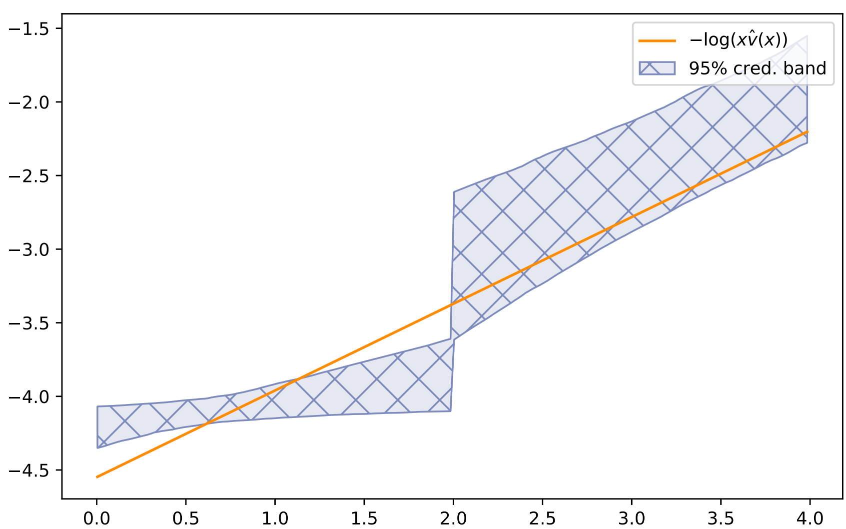

FIGURE 8.4. *Marginal Bayesian credible bands for the fire loss data for the function $-\log(xv(x))$ based on all samples. Yellow: maximum likelihood estimate $-\log(x\hat{v}(x))$ assuming a Gamma process.*

for the parameters β, β_1. In the data augmentation step we take intermediate points at distance 0.0192, corresponding to $m=1000$.

For the parameter updates we took independent Gaussian innovations with standard deviations $\sigma_\alpha=\sigma_{\alpha_1}=0.03$, $\sigma_\beta=1$ and $\sigma_{\beta_1}=6$, respectively. In the Gibbs sampler in each step new Gamma bridges are proposed in the data augmentation step, followed by a parameter update Metropolis-Hastings step cycling through updates of β in the first and second and the other parameters jointly in each of the remaining three of in total 5 stages. With these choices, the chains mix sufficiently well. The MCMC algorithm was run for 200000 iterations. Figure 8.2 shows trace plots and running averages of the posterior samples of the parameters α and β, whereas Figure 8.3 shows similar plots for the parameters α_1 and β_1.

Figure 8.4 shows the 95% marginal Bayesian credible band for the function $\theta(x)+\alpha x$ contrasted to the maximum likelihood estimate that assumes the observations come from a Gamma process. This plot suggests that modelling the losses with a Gamma process leads to overestimation of the number of small jumps and possibly of large jumps too; however, more data is necessary to make a definitive statement (unfortunately, as observed in Chavez-Demoulin et al. (2016), it is difficult for academia to gain access to the insurance data). In connection to this, we note that a difference in the estimates of the rate of decay of the Lévy density (value of α_1 in the model) has serious implications of practical relevance for the assessment of the risk of very large fire losses.

9. Outlook As a possible extension of the model studied in this paper, one can consider a class of increasing, infinite activity Lévy processes, which one can call (a,b,θ)-subordinators. Fix some $a\in[0,1)$, $b\geq 0$ and a non-decreasing, non-negative function θ on $\mathbb{R}_+$; then a Lévy process $(X_t)_{t\geq 0}$ is called an (a,b,θ)-subordinator, if the characteristic function of X_1 has the form

$$\varphi(z)=\mathrm{E}[e^{izX_1}]=\exp\left(\int_{\mathbb{R}}(e^{izx}-1)\,\nu(\mathrm{d}x)\right),\quad z\in\mathbb{R},$$

where the Lévy measure ν is given by

$$\nu(\mathrm{d}x)=\frac{b}{x^{1+a}}e^{-\theta(x)}\mathbf{1}_{(0,\infty)}(x)\,\mathrm{d}x. \tag{9.1}$$

On one hand, this model generalises the Gamma process with $a=0$ and $\theta(x)\equiv\lambda x$, $\lambda>0$. On the other hand, (a,b,θ)-subordinators cover the class of one-sided tempered stable processes, that have recently gained attention in physics and mathematical finance, see Rachev et al. (2011). Furthermore, the family of (a,b,θ)-subordinators overlaps with the class of self-decomposable Lévy processes, that likewise have important applications in finance, see, e.g., Carr et al. (2007).

In order to extend the inferential approach presented in the current work to this new model, we need to be able to sample from the distribution of X conditional on $X_T=x_T$. The problem of sampling from tempered stable bridges has been recently studied in Kim et al. (2016). Let us also mention the fact that the problem of estimating the stability index α is difficult from a Bayesian point of view due to singularity of the measures induced by two Lévy processes with different stability indices. However, several frequentist approaches to estimate α are available in the literature, see Belomestny and Reiß (2006). Also, our estimation approach can be conceivably extended to Gamma driven stochastic differential equation models.

Appendix A. Technical results for Section 4.

Proof of Proposition 4.1.

For ease of notation we put $\mu_n(\mathrm{d}x)=(x^2\wedge 1)\nu_n(\mathrm{d}x)$ and $\mu(\mathrm{d}x)=(x^2\wedge 1)\nu(\mathrm{d}x)$. Gnedenko's theorem, see, e.g., Theorem 2 in Gnedenko (1939), states that $\mathbb{Q}_{v_n}\rightsquigarrow\mathbb{Q}_v$ if and only if $\gamma_n\to\gamma$ and $\mu_n\rightsquigarrow\mu$, referred in this proof as Gnedenko's conditions. We show that these conditions are equivalent to $\widetilde{\nu}_n\rightsquigarrow\widetilde{\nu}$. Assume the latter and take the bounded and continuous function $f=1$. It then follows that $\gamma_n+\mu_n(\mathbb{R})\to\gamma+\mu(\mathbb{R})$. Next we show that $\gamma_n\to\gamma$. Let $f_\varepsilon(x)=(1-\frac{x}{\varepsilon})^+$ for $x\geq 0$ and $0<\varepsilon\leq 1$. Then

$$0\leq\int f_\varepsilon(x)\mu_n(\mathrm{d}x)=\int_0^\varepsilon f_\varepsilon(x)x^2\nu_n(\mathrm{d}x)\leq\int_0^\varepsilon x^2\nu_n(\mathrm{d}x)\leq\varepsilon\int_0^\varepsilon x\nu_n(\mathrm{d}x)\leq\varepsilon\gamma_n.$$

It follows that $\gamma_n\leq\int f_\varepsilon\mathrm{d}\widetilde{\nu}_n\leq(1+\varepsilon)\gamma_n$, and hence $\limsup\gamma_n\leq\int f_\varepsilon\mathrm{d}\widetilde{\nu}\leq(1+\varepsilon)\liminf\gamma_n$. Similar considerations yield $\gamma\leq\int f_\varepsilon\mathrm{d}\widetilde{\nu}\leq(1+\varepsilon)\gamma$, and a combination of these results yields $\max\{\limsup\gamma_n,\gamma\}\leq(1+\varepsilon)\min\{\liminf\gamma_n,\gamma\}$. Since ε is arbitrary, it follows that $\gamma_n\to\gamma$ and, in view of the earlier limit, also $\mu_n(\mathbb{R})\to\mu(\mathbb{R})$. Let f_0 be bounded and continuous such that $f_0(0)=0$. Then $\int f_0\mathrm{d}\mu_n=\int f_0\mathrm{d}\widetilde{\nu}_n\to\int f_0\mathrm{d}\widetilde{\nu}=\int f_0\mathrm{d}\mu$. Take now an arbitrary bounded and continuous function f, and let $f_0=f-f(0)$. Then, in view of the above, one has $\int f\mathrm{d}\mu_n=\int f_0\mathrm{d}\mu_n+f(0)\mu_n(\mathbb{R})\to\int f_0\mathrm{d}\mu+f(0)\mu(\mathbb{R})=\int f\mathrm{d}\mu$. Both of Gnedenko's conditions are thus satisfied. This shows one implication. Conversely, by assuming Gnedenko's conditions, one obtains by a simple addition that $\widetilde{\nu}_n\rightsquigarrow\widetilde{\nu}$. □

The next two lemmas bound the Kullback-Leibler divergence between two measures $\mathbb{Q}_{v_0},\mathbb{Q}_v$.

Lemma A.1. *We have $\mathcal{KL}(\mathbb{Q}_{v_0},\mathbb{Q}_v)\leq\mathcal{KL}(\mathbb{P}_{v_0},\mathbb{P}_v)$.*

Proof. This is the inequality stated on p. 12 in Gugushvili et al. (2015). The fact that there it is obtained in the context of the compound Poisson processes plays no role in our case: the result follows from the well-known inequality due to Csiszár (1963); cf. Lemma 2 and arguments preceding it in Gugushvili et al. (2015). □

Lemma A.2. *We have $\mathcal{KL}(\mathbb{Q}_{v_0},\mathbb{Q}_v)\lesssim|\alpha-\alpha_0|+\|\theta-\theta_0\|_\infty$. The constant in the inequality depends on α_0,θ_0 and known constants only.*

Proof. We will bound from above $\mathcal{KL}(\mathbb{P}_{v_0},\mathbb{P}_v)$, which by Lemma A.1 automatically yields an upper bound on $\mathcal{KL}(\mathbb{Q}_{v_0},\mathbb{Q}_v)$. By formula (A.1) in Cont and Tankov (2006),

$$\mathcal{KL}(\mathbb{P}_{v_0},\mathbb{P}_v)=\int_{x>0}v_0(x)\log\left(\frac{v_0(x)}{v(x)}\right)\mathrm{d}x+\int_{x>0}(v(x)-v_0(x))\mathrm{d}x=\mathrm{I}+\mathrm{II}.$$

We will separately bound the two terms. We start with the first one:

$$\mathrm{I}=(\alpha-\alpha_0)\int_{x>0}e^{-\alpha_0x-\theta_0(x)}\mathrm{d}x+\int_{\underline{b}\leq x\leq\overline{b}}\frac{1}{x}e^{-\alpha_0x-\theta_0(x)}(\theta(x)-\theta_0(x))\mathrm{d}x.$$

It follows that $|\mathrm{I}|\lesssim|\alpha-\alpha_0|+\|\theta-\theta_0\|_\infty$. The constant in the inequality depends on α_0,θ_0, and known constants.

Now we turn to II. We have

$$\begin{aligned}\mathrm{II}=&\int_{0<x<\underline{b}}\frac{1}{x}\left(e^{-\alpha x}-e^{-\alpha_0x}\right)\mathrm{d}x\\&+\int_{\underline{b}\leq x\leq\overline{b}}\frac{1}{x}\left(e^{-\alpha x-\theta(x)}-e^{-\alpha_0x-\theta_0(x)}\right)\mathrm{d}x\\&+\int_{\overline{b}<x<\infty}\frac{1}{x}\left(e^{-\alpha x}-e^{-\alpha_0x}\right)\mathrm{d}x.\end{aligned}$$

By the mean-value theorem, using also the facts that $\alpha_0,\alpha \geq \underline{\alpha}$, $x>0$, the first term on the right in the above display is up to a constant bounded in absolute value by $|\alpha-\alpha_0|$. A similar bound is true for the third term too. As far as the second term is concerned, notice that for any x,y,

$$|e^x-e^y|\leq \max(e^x,e^y)|x-y|,$$

so that for $x\in[\underline{b},\overline{b}]$ we have

$$\left|e^{-\alpha x-\theta(x)}-e^{-\alpha_0 x-\theta_0(x)}\right|\lesssim |\alpha-\alpha_0|x+\|\theta-\theta_0\|_\infty.$$

This in turn entails that

$$\left|\int_{\underline{b}\leq x\leq \overline{b}}\frac{1}{x}\left(e^{-\alpha x-\theta(x)}-e^{-\alpha_0 x-\theta_0(x)}\right)\mathrm{d}x\right|\lesssim |\alpha-\alpha_0|+\|\theta-\theta_0\|_\infty.$$

Combination of the above intermediate inequalities completes the proof. □

The next three lemmas bound the discrepancy $\mathcal{V}$ between two measures $\mathbb{Q}_{v_0},\mathbb{Q}_v$.

Lemma A.3. *We have*

$$\mathcal{V}(\mathbb{Q}_{v_0},\mathbb{Q}_v)\leq \mathcal{V}(\mathbb{P}_{v_0},\mathbb{P}_v)+4\mathcal{KL}(\mathbb{P}_{v_0},\mathbb{P}_v).$$

Proof. This is equation (21) in Gugushvili et al. (2015). The fact that in the original context it dealt with the compound Poisson process, plays no role in our case, the arguments go through without modification. □

Lemma A.4. *We have*

$$\mathcal{V}(\mathbb{P}_{v_0},\mathbb{P}_v)=\int_0^\infty v_0(y)\log^2\left(\frac{v(y)}{v_0(y)}\right)\mathrm{d}y \\ +\left(\int_0^\infty\left(1-\frac{v(y)}{v_0(y)}+\log\left(\frac{v(y)}{v_0(y)}\right)\right)v_0(y)\,\mathrm{d}y\right)^2.$$

Proof. It follows from Theorem 4 in Brockett et al. (1978) that

$$\begin{aligned}\phi(u)&:=\mathrm{E}_{\mathbb{P}_{v_0}}\left[\exp\left(\mathrm{i}u\log\left(\frac{\mathrm{d}\mathbb{P}_v}{\mathrm{d}\mathbb{P}_{v_0}}\right)\right)\right]\\&=\exp\left[\mathrm{i}u\int_0^\infty\left(1-\frac{v(x)}{v_0(x)}\right)v_0(x)\,\mathrm{d}x+\int_0^\infty\left(e^{\mathrm{i}ux}-1\right)v_0\circ g^{-1}(\mathrm{d}x)\right]\end{aligned}$$

with $g(x)=\log\left(\frac{v(x)}{v_0(x)}\right)$. We have

$$\phi'(u)=\left(\mathrm{i}\int_0^\infty\left(1-\frac{v(x)}{v_0(x)}\right)v_0(x)\,\mathrm{d}x+\mathrm{i}\int_0^\infty xe^{\mathrm{i}ux}(v_0\circ g^{-1})(\mathrm{d}x)\right)\phi(u)$$

and

$$\phi''(u)=-\left(\int_0^\infty x^2e^{\mathrm{i}ux}(v_0\circ g^{-1})(\mathrm{d}x)\right)\phi(u)$$

$$-\left(\int_0^\infty \left(1-\frac{v(x)}{v_0(x)}\right) v_0(x)\,\mathrm{d}x + \int_0^\infty x e^{\mathrm{i}ux}(v_0\circ g^{-1})(\mathrm{d}x)\right)^2 \phi(u).$$

As a result, we get that

$$\mathrm{E}_{\mathbb{P}_{v_0}}\left[\left(\log\left(\frac{\mathrm{d}\mathbb{P}_v}{\mathrm{d}\mathbb{P}_{v_0}}\right)\right)^2\right] = -\phi''(0) = \int_0^\infty x^2 (v_0\circ g^{-1})(\mathrm{d}x)$$
$$+\left(\int_0^\infty \left(1-\frac{v(x)}{v_0(x)}\right) v_0(x)\,\mathrm{d}x + \int_0^\infty x(v_0\circ g^{-1})(\mathrm{d}x)\right)^2.$$

Now note that by the change of the variable formula,

$$\int_0^\infty x(v_0\circ g^{-1})(\mathrm{d}x) = \int_0^\infty v_0(y)\log\left(\frac{v(y)}{v_0(y)}\right)\mathrm{d}y,$$
$$\int_0^\infty x^2(v_0\circ g^{-1})(\mathrm{d}x) = \int_0^\infty v_0(y)\log^2\left(\frac{v(y)}{v_0(y)}\right)\mathrm{d}y.$$

This completes the proof. □

The next result is used to bound from below the denominator in the posterior and is a simple restatement of Lemma 8.1 in Ghosal et al. (2000).

Lemma A.5. *Let $\widetilde{\Pi}$ be an arbitrary probability measure on the set*

$$K(\delta) = \{v\colon \mathcal{KL}(\mathbb{Q}_{v_0},\mathbb{Q}_v)\le\delta, \mathcal{V}(\mathbb{Q}_{v_0},\mathbb{Q}_v)\le\delta\},$$

where $\delta>0$ is any fixed number. Then for every constant $C>1$,

$$\mathbb{Q}_{v_0}^n\left(\int_{K(\delta)}\prod_{i=1}^n \frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i)\widetilde{\Pi}(\mathrm{d}v)\le e^{-Cn\delta}\right)\le\frac{1}{(C-1)^2 n\delta}.$$

The next lemma, together with Lemma A.2, is instrumental in verifying the prior mass condition, that is one of the key ingredients for derivation of posterior consistency.

Lemma A.6. *We have*

$$\mathcal{V}(\mathbb{Q}_{v_0},\mathbb{Q}_v)\lesssim|\alpha-\alpha_0|+\|\theta-\theta_0\|_\infty+|\alpha-\alpha_0|^2+\|\theta-\theta_0\|_\infty^2.$$

The constant in the inequality depends on α_0,θ_0 and known constants only.

Proof. The result follows from Lemmas A.2, A.3 and A.4 after some tedious calculations as in the proof of Lemma A.2. □

The next results deals with the prior mass condition.

Lemma A.7. *For every $\delta>0$ small enough and all n large,*

$$\Pi_n(K(\delta))\gtrsim(c\delta)^{2N_n}$$

for a constant c independent of n.

Proof. By Lemmas A.2 and A.6, there exists a constant $c>0$, such that

$$K(\delta)\subseteq\{|\alpha-\alpha_0|\vee|\alpha-\alpha_0|^2\le c\delta\}\cap\{\|\theta-\theta_0\|_\infty\vee\|\theta-\theta_0\|_\infty^2\le c\delta\}.$$

Since priors on α and θ are independent, we get that

$$\Pi_n(K(\delta))\ge\left[\Pi_n(|\alpha-\alpha_0|\le c\delta)\wedge\Pi_n(|\alpha-\alpha_0|^2\le c\delta)\right]$$

$$\times \left[\Pi_n(\|\theta-\theta_0\|_\infty \leq c\delta) \wedge \Pi_n(\|\theta-\theta_0\|_\infty^2 \leq c\delta)\right].$$

We will bound each of the terms on the right separately. For δ small enough,

$$\Pi_n(|\alpha-\alpha_0| \leq c\delta) \leq \Pi_n(|\alpha-\alpha_0|^2 \leq c\delta), \quad \Pi_n(\|\theta-\theta_0\|_\infty \leq c\delta) \leq \Pi_n(\|\theta-\theta_0\|_\infty^2 \leq c\delta),$$

so that it is sufficient to bound from below the terms on the left hand side of these two inequalities.

Note that since α is equipped with the uniform prior, $\Pi_n(|\alpha-\alpha_0| \leq c\delta) \asymp \delta$. On the other hand,

$$\begin{aligned}\Pi_n(\|\theta-\theta_0\|_\infty \leq c\delta) &= \Pi_n\left(\max_{1\leq k\leq N}\sup_{x\in B_k} |\theta(x)-\theta_0(x)| \leq c\delta\right) \\ &= \prod_{k=1}^{N_n} \Pi_n\left(\sup_{x\in B_k} |\theta(x)-\theta_0(x)| \leq c\delta\right).\end{aligned}$$

Consider a term

$$\Pi_n\left(\sup_{x\in B_k} |\theta(x)-\theta_0(x)| \leq c\delta\right) = \Pi_n\left(\sup_{x\in B_k} |\rho_k+\theta_k x-\theta_0(x)| \leq c\delta\right).$$

By the Hölder assumption on θ_0, we have by the triangle inequality

$$\begin{aligned}|\rho_k+\theta_k x-\theta_0(x)| &\leq |\rho_k+\theta_k b_k-\theta_0(b_k)| + L(x-b_k)^\lambda \\ &\leq |\rho_k+\theta_k b_k-\theta_0(b_k)| + L\Delta_n^\lambda,\end{aligned}$$

where Δ_n denotes the length of the bins, $\Delta_n = \bar{b}/N_n$. As $\Delta_n \to 0$ for $n\to\infty$, we can make it small enough to have (for any $c,\delta>0$) $L\Delta_n^\lambda \leq \delta/2$. It follows that for sufficiently small δ one has

$$\left\{\sup_{x\in B_k} |\theta(x)-\theta_0(x)| \leq c\delta\right\} \supset \left\{|\rho_k+\theta_k b_k-\theta_0(b_k)| \leq \frac{c\delta}{2}\right\}.$$

Furthermore, we have

$$\left\{|\rho_k+\theta_k b_k-\theta_0(b_k)| \leq \frac{c\delta}{2}\right\} \supset \left\{|\rho_k-\theta_0(b_k)| \leq \frac{c\delta}{4}\right\} \cap \left\{|\theta_k b_k| \leq \frac{c\delta}{4}\right\}.$$

Then by independence of θ_k and ρ_k,

$$\Pi_n\left(|\rho_k+\theta_k b_k-\theta_0(b_k)| \leq \frac{c\delta}{2}\right) \geq \Pi_n\left(|\rho_k-\theta_0(b_k)| \leq \frac{c\delta}{4}\right)\Pi_n\left(|\theta_k b_k| \leq \frac{c\delta}{4}\right).$$

As the interval $(\theta_0(b_k)-\frac{c\delta}{4}, \theta_0(b_k)+\frac{c\delta}{4})$ is contained in $[-\bar{\theta},\bar{\theta}]$ for all sufficiently small δ, the first factor on the right is bounded from below by a constant (independent of n and k) times δ. So is the second factor, because

$$\Pi_n\left(|\theta_k b_k| \leq \frac{c\delta}{4}\right) \geq \Pi_n\left(|\theta_k| \leq \frac{c\delta}{4\bar{b}}\right).$$

It follows that

$$\Pi_n\left(\sup_{x\in B_k} |\theta(x)-\theta_0(x)| \leq c\delta\right) \gtrsim \delta^2.$$

Thus, after an evident renaming of constants, $\Pi_n(K(\delta)) \gtrsim (c\delta)^{2N_n}$ for a constant c independent of n. □

The result of the next lemma is a variation on Lemma A.2. Its main use lies in establishing a certain metric entropy bound in Lemma A.9.

LEMMA A.8. *It holds that* $d_{\mathcal{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \lesssim |\alpha - \alpha_0| + ||\theta - \theta_0||_\infty$.

Proof. We first note that

$$d^2_{\mathcal{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \leq d^2_{\mathcal{H}}(\mathbb{P}_{v_0}\mathbb{P}_v),$$

see Gugushvili et al. (2015), p. 14.

Further, one has $d^2_{\mathcal{H}}(\mathbb{P}_{v_0}\mathbb{P}_v) = 1 - \exp(-h) \leq h$, see Theorème 1 in Mémin and Shiryaev (1985), where $h = \frac{1}{2}\int_0^\infty (\sqrt{v_0(x)} - \sqrt{v(x)})^2 \, \mathrm{d}x$. By a splitting procedure as in the proof of Lemma A.2, we get $d^2_{\mathcal{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \lesssim |\alpha - \alpha_0|^2 + ||\theta - \theta_0||^2_\infty$. Finally, use the inequality $\sqrt{x^2 + y^2} \leq x + y$ for $x, y \geq 0$. □

In the proof of Lemma A.10 below we need an auxiliary result. For any class of functions $\mathcal{F}$, recall the bracketing entropy $H_{[]}(u, \mathcal{F}) = \log N_{[]}(u, \mathcal{F})$, with $N_{[]}(u, \mathcal{F})$ the bracketing number under the Hellinger metric. Useful will be the inequality $H_{[]}(u, \mathcal{F}) \leq H_\infty(u/2, \mathcal{F})$, see Lemma 2.1 in van de Geer (2000), where $H_\infty(u, \mathcal{F}) = \log N_\infty(u, \mathcal{F})$, with $N_\infty(u, \mathcal{F})$ the covering number of $\mathcal{F}$ with balls of radius u under the supremum norm. For the latter we have the following result.

LEMMA A.9. *Let $\mathcal{F}_n$ be the set of probability measures $\mathbb{Q}_v$, where the Lévy densities v are elements of V_n. It holds that $H_\infty(u, \mathcal{F}_n) \asymp N_n \log(1 + \frac{1}{u})$, and hence there is $C > 0$ such that for all sufficiently small $\delta > 0$ and sufficiently large N_n (the number of bins), one has*

$$\int_0^\delta H^{1/2}_\infty(u, \mathcal{F}_n) \, \mathrm{d}u \leq C\sqrt{N_n}\,\delta \log^{1/2}\left(\frac{1}{\delta} + 1\right).$$

Proof. Starting point is the result of Lemma A.8. First we need a δ-cover of the interval $[\underline{\alpha}, \overline{\alpha}]$, for which the covering number needed is of order $\delta^{-1} + 1$. To cover a set of functions θ, it is sufficient to cover the bounded intervals to which the corresponding ρ_k and θ_k belong. Hence δ-covers for both are again of order $\delta^{-1} + 1$, and we have to do this on N_n bins separately. Altogether, this implies that a cover of size $O(\delta^{-1} + 1)^{2N_n + 1}$ is sufficient to cover the set $\mathcal{F}_n$. Hence $\int_0^\delta H^{1/2}_\infty(u, \mathcal{F}_n) \, \mathrm{d}u \asymp \sqrt{N_n} \int_0^\delta \log^{1/2}(u^{-1} + 1) \, \mathrm{d}u$. We now show that the latter integral is of order $\delta \log^{1/2}(1 + \frac{1}{\delta})$ for small δ. For this we assume that $\delta < \frac{1}{e-1}$, which entails $\log(1 + \frac{1}{\delta}) > 1 > \frac{1}{1+\delta}$, $\log(y+1) > 1$ and and $\frac{1}{y} < \frac{2}{y+1}$ for $y > \delta^{-1}$. These inequalities are used to show via lengthy but standard computations that

$$\int_0^\delta \log^{1/2}(u^{-1} + 1) \, \mathrm{d}u \leq 2\delta \log^{1/2}(\delta^{-1} + 1).$$

The result of the lemma follows. □

The next result is used to handle the numerator in Bayes' formula in our main result, Theorem 4.1.

LEMMA A.10. *Fix $\epsilon > 0$ and define $B(\epsilon) = \{v \in V_n : d_{\mathcal{H}}(\mathbb{Q}_{v_0}, \mathbb{Q}_v) \leq \epsilon\}$. Then there exist positive constants c_1, c_2, c_3, independent of n, such that*

$$\mathbb{Q}^n_{v_0}\left(\sup_{v \in B(\epsilon)^c} \prod_{i=1}^n \frac{\mathrm{d}\mathbb{Q}_v}{\mathrm{d}\mathbb{Q}_{v_0}}(Z_i) \geq \exp(-c_1 n\epsilon^2)\right) \leq c_3 \exp(-c_2 n\epsilon^2).$$

Proof. We will use Theorem 1 in Wong and Shen (1995). The main fact to establish is a bound on the entropy integral $\int_0^\epsilon H_{[]}^{1/2}(u,\mathcal{F}_n)\mathrm{d}u$ (the set $\mathcal{F}_n$ as in Lemma A.9) of the form $C\sqrt{n}\epsilon^2$. It follows from Lemma A.9 and the remarks preceding it, that $\int_0^\epsilon H_{[]}^{1/2}(u,\mathcal{F}_n)\mathrm{d}u \le C\sqrt{N_n}\epsilon\log^{1/2}(\frac{1}{\epsilon}+1)$. We want to choose N_n, so that

$$\sqrt{N_n}\epsilon\log^{1/2}\left(\frac{1}{\epsilon}+1\right)\lesssim\sqrt{n}\epsilon^2$$

for all n and all small enough ϵ. To that end it is enough to have

$$\frac{N_n}{n}\lesssim\frac{\epsilon^2}{\log(\frac{1}{\epsilon}+1)},$$

which in fact holds for all n large enough, since $N_n/n\to 0$ by assumption. Then Condition (3.1) in Wong and Shen (1995) is satisfied, and hence we can apply Theorem 1 of that paper, which yields the assertion. □

Appendix B. Technical lemma for Section 6.

LEMMA B.1. *Let I be a countable index set and $(E_i,\mathfrak{A}_i,\mathbb{P}_i)$, $i\in I$, a collection of probability spaces or σ-finite measure spaces. Denote the corresponding product measurable space with the product measure by $(E,\mathfrak{A},\mathbb{P})$. Let $\pi^J : x\in E\mapsto (x_i)_{i\in J}$ be the coordinate projections for $J\subset I$. Assume that $\mathbb{Q}(x,\mathrm{d}x^\circ)$ is a σ-finite transition measure with a localisation property*

$$\mathbb{Q}(x;\mathrm{d}\pi^{I_1\cup I_2}(\cdot))=\mathbb{Q}(\pi^{I_1}(x);\mathrm{d}\pi^{I_1}(\cdot))\otimes\mathbb{Q}(\pi^{I_2}(x);\mathrm{d}\pi^{I_2}(\cdot))$$

for all $x\in E$, $I_1,I_2\subset I$, $I_1\cap I_2=\varnothing$. Then the local balance condition

$$\mathbb{P}^i(\mathrm{d}x_i)\mathbb{Q}^i(x_i;\mathrm{d}x_i^\circ)=\mathbb{P}^i(\mathrm{d}x_i^\circ)\mathbb{Q}^i(x_i^\circ;\mathrm{d}x_i),$$

where $\mathbb{Q}^i(x_i;A)=\mathbb{Q}(\pi^i(x);(\pi^i)^{-1}(A))$ for $A\in\mathfrak{A}_i$, implies

$$\mathbb{P}(\mathrm{d}x)\mathbb{Q}(x;\mathrm{d}x^\circ)=\mathbb{P}(\mathrm{d}x^\circ)\mathbb{Q}(x^\circ;\mathrm{d}x). \tag{B.1}$$

Proof. A measure on E^2 can be written as a measure on $\widetilde{E}^2=\bigtimes_{i\in\mathbb{N}}E_i^2$ using an obvious change of coordinates. Denote the measure $\mathbb{P}(\mathrm{d}x)\mathbb{Q}(x,\mathrm{d}x^\circ)$ seen as a measure on $\widetilde{E}^2$ by μ. Then

$$\mu\left((\bigtimes_{i\le n}(A_i\times A_i^\circ))\times(\bigtimes_{i>n}E_i^2)\right)=\prod_{i\le n}\int_{A_i}\mathbb{Q}^i(x_i,A_i^\circ)\mathbb{P}^i(\mathrm{d}x),\quad A_i,A_i^\circ\in\mathfrak{A}_i$$

for all $n\in\mathbb{N}$. Therefore μ is a product measure. It is also a symmetric measure in the following sense: $\mu(s(A))=\mu(A)$ for $s(A)=\{(x_i^\circ,x_i)_{i\in I}:(x_i,x_i^\circ)_{i\in I}\in A\}$. This can be formally shown by the "good set principle": Let $\mathfrak{S}$ be the collection of sets such that $\mu(s(S))=\mu(S)$ holds for $S\in\mathfrak{S}$. First, $\bigtimes_{i\le n}(A_i\times A_i^\circ)\times(\bigtimes_{i>n}E_i^2))\in\mathfrak{S}$, so $\mathfrak{S}$ contains a generator which has the intersection property (π-system). Now $E\in\mathfrak{S}$, and also complements of sets in $\mathfrak{S}$ are in $\mathfrak{S}$, and countable unions of disjoint sets in $\mathfrak{S}$ are in $\mathfrak{S}$ as well: if $A_i\in\mathfrak{S}$ are disjoint sets and $A=\bigcup A_i$, then

$$\mu(A)=\sum\mu(A_i)=\sum\mu(s(A_i))=\mu(s(A)).$$

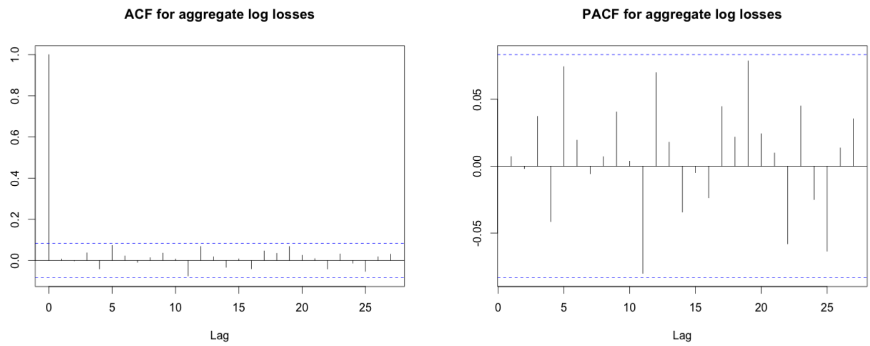

FIGURE C.1. *Logarithmically transformed and aggregated Danish data on large fire losses.* Left*: autocorrelation function.* Right*: partial autocorrelation function.*

Therefore $\mathfrak{S} = \mathfrak{A}$ by Dynkin's π-λ theorem. The balance equation (B.1) follows. □

Appendix C. Danish fire losses: exploratory data analysis.

In this appendix we perform an exploratory analysis of the Danish data on large fire losses. We primarily use graphical tools; these may look simple, but are commonly applied in similar analyses (see, e.g., McNeil (1997) and Resnick (1997)) and convey useful information that is not easily obtainable otherwise.

Figure C.1 gives the plots of the estimated autocorrelation and partial autocorrelation functions of logarithmically transformed and aggregated Danish fire losses. Both plots are compatible with the assumption that the data follow a white noise process. A more formal confirmation comes from the Box-Pierce and Ljung-Box tests, that we applied with 20 lags, and that yielded p-values 0.5847 and 0.5547, respectively (the tests are implemented in **R** via `Box.test`). This suggests that weekly data can indeed be modelled as an i.i.d. sequence.

We also produced the histogram of the weekly data, and fitted the Gamma distribution via the maximum likelihood method. The results are displayed in the left panel of Figure C.2, and provide a visual hint that a Gamma-type distribution yields a reasonable fit to the data. Since a histogram is a somewhat crude nonparametric estimator and is strongly dependent on the choice of the bin number (we used the default implementation in **R** via the command `hist`), we also visually compared the Gamma fit to a kernel density estimator, with bandwidth selected through cross-validation (we used the `density` in **R** with the Gaussian kernel), see the right panel of Figure C.2. Ignoring the edge effects near the boundary point of the support of the distribution, it appears that the two estimates are different e.g. in a neighbourhood of the mode of the Gamma density, with probability mass of the kernel density estimate shifted to the right. On the other hand, the tail behaviour of both estimates is similar.

Although evidence is not decisive, a further hint that the Gamma distribution is perhaps not entirely adequate for modelling the Danish fire losses data comes from the QQ-plot of empirical quantiles of the Danish fire losses data versus theoretical Gamma quantiles; see Figure C.3 (we used the command `qqPlot` from the `car` package in **R**).

Summarising the results of our exploratory data analysis, it appears that if aggregated over weekly (or in some exceptional cases over bi-weekly) periods, the logarithmi-

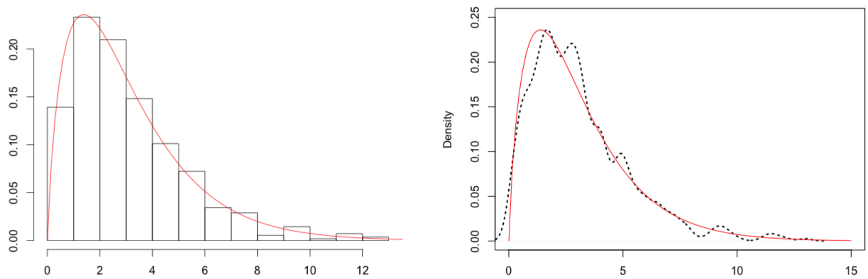

Figure C.2. *Logarithmically transformed and aggregated Danish data on large fire losses.* Left*: histogram with a superimposed gamma density evaluated at the maximum likelihood estimate.* Right*: kernel density estimate (dotted line) with the same superimposed gamma density (solid line) evaluated at the maximum likelihood estimate.*

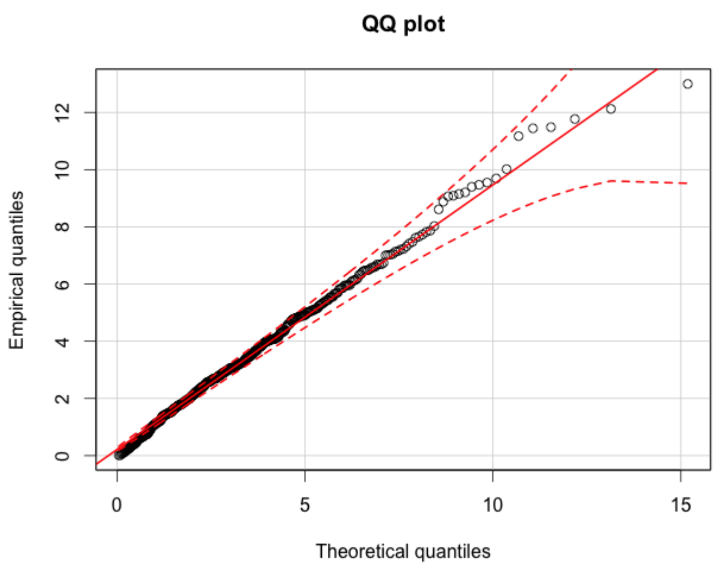

Figure C.3. *Logarithmically transformed and aggregated Danish data on large fire losses: QQ-plot of empirical quantiles versus theoretical gamma quantiles.*

cally transformed Danish fire losses data can be adequately modelled as a realisation of an i.i.d. sequence that follows a Gamma-like distribution, but perhaps is not genuinely Gamma.

Acknowledgement. The research leading to the results in this paper has received funding from the European Research Council under ERC Grant Agreement 320637. The research of the first author was supported by the Russian Academic Excellence Project “5-100” and the German Science Foundation research grant (DFG Sachbeihilfe) 406700014. The authors are grateful to anonymous referees for careful reading and

insightful comments that lead to improvements in the paper.

References.

Y. Aït-Sahalia and J. Jacod. *High-Frequency Financial Econometrics.* Princeton University Press, Princeton, 2014.

O.E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick (Eds.). *Lévy Processes: Theory and Applications.* Birkhäuser Boston, Inc., Boston, MA.

I.V. Basawa and P.J. Brockwell. Nonparametric estimation for nondecreasing Lévy processes. *J. Roy. Statist. Soc. Ser. B*, 44:262–269, 1982.

D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda and M. Reiß. *Lévy Matters. IV. Estimation for Discretely Observed Lévy Processes.* Lecture Notes in Mathematics, 2128. Lévy Matters. Springer, Cham, 2015.

D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models. *Finance Stoch.* 10:449–474, 2006.

L.M. Berliner, J.A. Royle, C.K. Wikle and R.F. Milliff. Bayesian methods in the atmospheric sciences. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), *Bayesian Statistics 6*, pages 83–100. Oxford University Press, 1999.

A. Beskos, G. Roberts, A. Stuart and J. Voss. MCMC methods for diffusion bridges. *Stoch. Dyn.*, 8:319–350, 2008.

J. Bezanson, A. Edelman, S. Karpinski and V.B. Shah. Julia: a fresh approach to numerical computing. *SIAM Rev.*, 59:65–98, 2017.

P. L. Brockett, W. N. Hudson and H. G. Tucker. The distribution of the likelihood ratio for additive processes. *J. Multivariate Anal.*, 8(2):233-243, 1978.

B. Buchmann and R. Grübel. Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.*, 31:1054–1074, 2003.

B. Buchmann and R. Grübel. Decompounding Poisson random sums: recursively truncated estimates in the discrete case. *Ann. Inst. Statist. Math.*, 56:743–756, 2004.

P. Carr, H. Geman, D.B. Madan and M. Yor. Self-decomposability and option pricing. *Mathematical Finance*, 17:31–57, 2007.

V. Chavez-Demoulin, P. Embrechts and M. Hofert. An extreme value approach for modeling operational risk losses depending on covariates. *J. Risk and Insur.*, 83:735–776, 2016.

A.J. Coca. Efficient nonparametric inference for discretely observed compound Poisson processes. *Probab. Theory Relat. Fields*, doi:10.1007/s00440-017-0761-5, 2017.

F. Comte and V. Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.*, 39:803–837, 2011.

R. Cont, Rama and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.

R. Cont and P. Tankov. Retrieving Lévy processes from option prices: regularization of an ill-posed inverse problem. *SIAM J. Control Optim.*, 45(1):1–25, 2006.

I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:85–108, 1963.

F. Dufresne, H. Gerber and E. Shiu. Risk theory with the gamma process. *ASTIN Bulletin*, 21:177–192, 1991.

C. Duval and E. Mariucci. Compound Poisson approximation to estimate the Lévy density. arXiv 1702.08787 **[math.PR]**, 2017.

A. Einstein. Zur Theorie der Brownschen Bewegung. *Ann. Phys.*, 324:371–381, 1906.

P. Embrechts, C. Klüppelberg and T. Mikosch. *Modelling Extremal Events. For Insurance and Finance.* Applications of Mathematics (New York), 33. Springer-Verlag, Berlin, 1997.

B. van Es, S. Gugushvili and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13:672–694, 2007.

S. Ghosal, J.K. Ghosh and A.W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28:500–531, 2000.

B. Gnedenko. To the theory of limiting theorems for sums of independent random variables (Russian). *Bull. Acad. Sci. URSS. Sér. Math. [Izvestia Akad. Nauk SSSR]*, 1939:181–232, 1939.

P.J. Green. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

S. Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.*, 21:321–343, 2009.

S. Gugushvili. Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48:282–307, 2012.

S. Gugushvili and P. Spreij. Non-parametric Bayesian drift estimation for stochastic differential equations. *Lith. Math. J.*, 54:127–141, 2014.

S. Gugushvili, F. van der Meulen and P. Spreij. Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Mod. Stoch. Theory Appl.*, 2:1–15, 2015.

S. Gugushvili, F. van der Meulen and P. Spreij. A non-parametric Bayesian approach to decompounding from high frequency data. *Stat. Inference Stoch. Process.*, 21:53–79, 2018.

Ch.C. Hewitt and B. Lefkowitz. Methods for fitting distributions to insurance loss data. In: *Proceedings of the Casualty Actuarial Society*, LXVI, pp. 139–160, 1979.

L. Hong and R. Martin. A review of Bayesian asymptotics in general insurance applications. *Eur. Actuar. J.*, 7:231–255, 2017a.

L. Hong and R. Martin. Dirichlet process mixture models for insurance loss data. *Scand. Actuar. J.*, 0:1–10, doi:10.1080/03461238.2017.1402086, 2017.

P. Ilhe, É. Moulines, F. Roueff, and A. Souloumiac. Nonparametric estimation of marks distribution of an exponential shot-noise process. *Electron. J. Statist.*, 9:3098–3123, 2015.

H. Jeffreys and B. Swirles. *Methods of Mathematical Physics.* Reprint of the third (1956) edition. Cambridge University Press, Cambridge, 1999.

Kim, Kyoung-Kuk, and Sojung Kim. *Simulation of tempered stable Lévy bridges and its applications.* Operations Research 64, no. 2 (2016): 495-509.

A.E. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Universitext. Springer-Verlag, Berlin, 2006.

D.B. Madan and E. Seneta. The Variance Gamma (V.G.) model for share market returns. *J. Bus.*, 63:511–524, 1990.

A.M. Mathai. Storage capacity of a dam with gamma type inputs. *Ann. Inst. Statist. Math.* 34, Part A:591–597, 1982.

A. McNeil. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27:117–137, 1997.

J. Mémin, A.N. Shiryaev. Distance de Hellinger-Kakutani des lois correspondants à deux processus à accroissements indépendants. *Z. Wahrsch. Verw. Gebiete*, 70:67–90, 1985.

S.A. van de Geer. *Empirical Processes in M-Estimation.* Cambridge University Press, Cambridge, UK, 2000.

F. van der Meulen and M. Schauer. Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.*, 11:2358–2396, 2017.

F. van der Meulen, M. Schauer and H. van Zanten. Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.*, 71:615–632, 2014.

F.H. van der Meulen and J.H. van Zanten. Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli*, 19:44–63, 2013.

M.H. Neumann and M. Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15:223–248, 2009.

R. Nickl and J. Söhl. Bernstein - von Mises theorems for statistical inverse problems II: Compound Poisson processes. arXiv:1709.07752 [**math.ST**], 2017a.

R. Nickl and J. Söhl. Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.*, 45:1664–1693, 2017b.

D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge Series in Statistical and Probabilistic Mathematics, 8. Cambridge University Press, Cambridge, 2002.

Rachev, Svetlozar T., Young Shin Kim, Michele L. Bianchi, and Frank J. Fabozzi. Financial models with Lévy processes and volatility clustering. Vol. 187. John Wiley & Sons, 2011.

G.A. Rempala and R.A. Derrig. Modeling hidden exposures in claim severity via the EM algorithm. *N. Am. Actuar. J.*, 9:108–128, 2005.

S. Resnick. Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin*, 27:139–151, 1997.

J. Rosiński. Series representations of Lévy processes from the perspective of point processes. In O.E. Barndorff-Nielsen, S.I. Resnick and T. Mikosch (eds.), *Lévy Processes: Theory and Applications*, pages 401–415. Birkhäuser Boston, Boston, MA, 2001.

H. Rubin and H.G. Tucker. Estimating the parameters of a differential process. *Ann. Math. Statist.*, 30:641–658, 1959.

K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, UK, 1999.

M. Schauer et al. Bridge 0.6.0. *Zenodo*, doi:10.5281/zenodo.891231, 2017.

A.V. Skorohod. *Случайные процессы с независимыми приращениями. (Russian) [Random Processes with Independent Increments]*. Izdat. "Nauka", Moscow, 1964.

M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. With discussion and with a reply by the authors. *J. Amer. Statist. Assoc.*, 82:528–550, 1987.

L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8:1–9, 1998.

L. Wasserman. Asymptotic properties of nonparametric Bayesian procedures. In: Dey D., Müller P., Sinha D. (eds), *Practical Nonparametric and Semiparametric Bayesian Statistics*, pages 293–304. Lecture Notes in Statistics, vol. 133. Springer, New York, NY.

W.H. Wong and X. Shen. Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.*, 23:339–362, 1995.

M. Yor. Some remarkable properties of Gamma processes. In: Fu M.C., Jarrow R.A., Yen J.-Y.J., Elliott R.J. (editors), *Advances in Mathematical Finance*, pages 37–47. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, 2007.