Centrum voor Wiskunde en Informatica

REPORT *RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Information processes in filtered experiments

K.O. Dzhaparidze, P.J.C. Spreij, E. Valkeila

REPORT PNA-R0125 DECEMBER 31, 2001

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Information Processes in Filtered Experiments

Kacha Dzhaparidze
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Peter Spreij
Vrije Universiteit
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Esko Valkeila
University of Helsinki
P.O. Box 4, 00014 Helsinki, Finland

ABSTRACT
In this paper we give explicit representations for Kullback-Leibler information numbers between a priori and a posteriori distributions, when the observations come from a semimartingale. We assume that the distribution of the observed semimartingale is described in terms of the so-called triplet of predictable characteristics. We end by considering the corresponding notions in a model with a fractional noise.

2000 Mathematics Subject Classification: 60G07, 60H30, 62B15, 94A17.
Keywords and Phrases: Kullback-Leibler information, Hellinger process, posterior distribution, semimartingales, predictable characteristics, stochastic calculus, fractional Brownian motions.
Note: Work carried out under the project PNA3.3, 'Stochastic Processes and Applications' and under The Fifth Framework Programme of the European Commission through the Dynstoch research network.

1 Introduction

We study a statistical experiment with a filtration. About the parameter space of the experiment we make the assumption that a prior distribution can be defined on it. On this abstract parameter space the Kullback-Leibler information between a posterior and a prior distribution is defined. We begin with modelling observations by filtration and discuss some results of a general nature and afterwards we specify observations that either come to us in the form of a semimartingale or in the form of a fractional Brownian motion. Given these observations we define the posterior distribution on the parameter space and we study various information notions, specifically the information in the posterior given the prior and vice versa (in Bayesian terminology known as the *information from data*) between these two distributions on the parameter space.
Using the notions of arithmetic mean measure and geometric mean measure as they were developed in [5] (the latter generalizes a probability measure introduced by Grigelionis in [7]) we

are going to express explicitly the density process of the posterior distribution on the parameter space with respect to the prior distribution as a certain density process on the observation space. Consequently, relying on the general theory of processes (cf [18]), we are able to use the machinery of stochastic calculus to obtain representations of the information processes, like e.g. a Doob-Meyer decomposition.

The study of Hellinger integrals and Hellinger processes started for binary experiments in the series of papers [12], [13] and [17]. This theory took a complete form in the book [10] where the notions of Hellinger integrals and Hellinger processes were fully exploited. In the consequent papers [8] and [9] some of the results were generalized to a filtered experiment with a finite number of probability measures. In [7] some additional aspects of the latter experiment are discussed. These results were extended to an arbitrary parameter space in [5]. It turns out that properties of the Hellinger process are of fundamental importance to understand the Kullback-Leibler information processes between a posterior and a prior distribution on the parameter space. Therefore a considerable part of the present paper is devoted to Hellinger processes. To make the present paper self-contained we included some necessary results from [5].

The paper is organized as follows. In section 2 we summarize and further develop some notions and results from [5]. In section 3 we present explicit version of results by assuming that we observe a semimartingale. In particular we compute the Hellinger process for a given prior distribution and the triplet of predictable characteristics under both the arithmetic mean measure and the geometric mean measure. For the geometric mean measure the triplet is specified further when the collection of distributions of the observed process constitutes an exponential family. In section 4 we define the different information measures and show how we can use the results of section 3 to compute multiplicative and additive (Doob-Meyer) decompositions of the information processes. Finally in section 5 we investigate the precise form of the results of section 4 further for a number of examples involving discrete time independent processes, multivariate point processes, diffusions and processes driven by fractional Brownian motion. In the latter case we show how one can use the developed theory for diffusion processes by using a representation of fractional Brownian motion as a stochastic integral with respect to ordinary Brownian motion where the integral is a certain deterministic kernel.

In turns out that our formulas are closely related to results in [19] for the Shannon information that is contained in a received signal about the transmitted signal for both the case of diffusion observations and counting process observations.

2 Randomized experiments

2.1 Basic setup

We consider a filtered statistical experiment $(\Omega, \mathcal{F}, F, \{P_\theta\}_{\theta\in\Theta})$ under the following assumptions. There exists an equivalent probability measure Q for this experiment, so

$$\{P_\theta\}_{\theta\in\Theta} \sim Q, \tag{2.1}$$

the right continuous filtration $F = \{\mathcal{F}_t\}_{t\geq 0}$ starts from $\mathcal{F}_0 = \{\emptyset, \Omega\}$ Q-a.s., $\mathcal{F}_0$ contains all the Q-null sets of $\mathcal{F}$, and $\bigvee_t \mathcal{F}_t = \mathcal{F}_\infty = \mathcal{F}$.

For a F-stopping time T consider now the optional projections Q_T and $P_{\theta,T}$ of the probability measures Q and P_θ to the sub-σ-field $\mathcal{F}_T$. Since by (2.1) these projections are equivalent, we may define the *density process* $z(\theta, Q) = z(P_\theta, Q)$ by

$$z_T(\theta, Q) = \frac{dP_{\theta,T}}{dQ_T}.$$

We have $dP_\theta/dQ = z_\infty(\theta)$, since $\bigvee_t \mathcal{F}_t = \mathcal{F}$. The density process possesses the following properties (see [10], proposition III.3.5, for more details): for each $\theta \in \Theta$
(i) $\inf_t z_t(\theta, Q) > 0 \;\; Q - a.s.$
(ii) $\sup_t z_t(\theta, Q) < \infty \;\; Q - a.s.$
(iii) the density process $z(\theta, Q)$ is a (Q, F)-uniformly integrable martingale with $E_Q\{z_t(\theta, Q)\} = 1$, for all $t \in [0, \infty]$.
Due to these properties, for each $\theta \in \Theta$ the process

$$m(\theta, Q) = z_-(\theta, Q)^{-1} \cdot z(\theta, Q) \tag{2.2}$$

is a (Q, F)-local martingale, so that the density process is represented as the Doléans exponential $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ of this martingale.
We endow the parameter space Θ with a σ-algebra $\mathcal{A}$ and the measurable space $(\Theta, \mathcal{A})$ with a probability measure α. Define $\mathbf{Q}$ as the product measure $\mathbf{Q} = Q \times \alpha$ on $\mathbf{F} \doteq \mathcal{F} \otimes \mathcal{A}$, the product σ-algebra on $\mathbf{\Omega} = \Omega \times \Theta$, and the so-called mixture measure $\mathbf{P}$ on $\mathbf{F}$ by

$$\mathbf{P}(\mathbf{B}) = \int_{\mathbf{B}} z_\infty(\omega, \theta) Q(d\omega) \alpha(d\theta)$$

for any set $\mathbf{B} \in \mathbf{F}$. The Kullback-Leibler information in $\mathbf{P}$ with respect to $\mathbf{Q}$ is by definition $I(\mathbf{P}|\mathbf{Q}) = E_{\mathbf{Q}} \log\{d\mathbf{Q}/d\mathbf{P}\}$. In the sequel we assume that

$$0 < I(\mathbf{P}|\mathbf{Q}) < \infty. \tag{2.3}$$

For brevity, we denote by ϑ a random element of the parametric space $(\Theta, \mathcal{A})$ distributed according to the measure α. In these terms, we may also write $I(\mathbf{P}|\mathbf{Q}) = E_\alpha I(P_\vartheta|Q) = \int_\Theta I(P_\theta|Q)\alpha(d\theta)$, where $I(P_\theta|Q)$ is the Kullback-Leibler information in P_θ with respect to Q.
In the Bayesian setup this measure is called the prior (or *a priori*) probability. By means of the Bayes formula we may define at each stopping time T the posterior (or *a posteriori*) probability α^T, that for each $A \in \mathcal{A}$ is

$$\alpha^T(A) = \frac{\int_A z_T(\theta, Q)\alpha(d\theta)}{\int_\Theta z_T(\theta, Q)\alpha(d\theta)}. \tag{2.4}$$

We will return to this subject in section 4.

2.2 The arithmetic and geometric mean measures

The notions of *arithmetic mean measure* $\bar{P}_\alpha$ and *geometric mean measure* G_α are basic for the present theory. They are defined on the aforementioned filtered space $(\Omega, \mathcal{F}, F)$. For $B \in \mathcal{F}$ we set

$$\bar{P}_\alpha(B) \doteq \int_\Theta P_\theta(B)\alpha(d\theta).$$

The following simple lemma allows us to use $\bar{P}_\alpha$ as a measure equivalent to whole family $\{P_\theta\}_{\theta\in\Theta}$:

Lemma 2.1. *Assume (2.1). Then the measures $\bar{P}_\alpha$ and Q are equivalent and $\frac{d\bar{P}_\alpha}{dQ} = a(\alpha, Q)$.*

Proof. First note that the a-mean measure $\bar{P}$ is dominated by Q and the identity of our assertion holds. In particular, $Q(\frac{d\bar{P}}{dQ} = 0) = 0$. Therefore it suffices to show that $Q \ll \bar{P}$, i.e. that $\bar{P}(\frac{d\bar{P}}{dQ} = 0) = 0$. For then $\frac{dQ}{d\bar{P}} := 1/\frac{d\bar{P}}{dQ} < \infty$ $\bar{P} - a.s.$, so that for each $B \in \mathcal{F}$ we have $Q(B) = \int_B \frac{dQ}{d\bar{P}} d\bar{P}$. Suppose the contrary $\bar{P}(\frac{d\bar{P}}{dQ} = 0) > 0$. By the definition of $\bar{P}$ we have $P_\theta(\frac{d\bar{P}}{dQ} = 0) > 0$ at least for a certain θ. But since $P_\theta \sim Q$ we get $Q(\frac{d\bar{P}}{dQ} = 0) > 0$ which contradicts $\bar{P} \ll Q$. □

The corresponding density process $z(\bar{P}_\alpha, Q)$ is referred to as the *arithmetic mean process* and denoted by $a(\alpha, Q) = z(\bar{P}_\alpha, Q)$. This term is explained by the simple fact that $a(\alpha, Q) = \int_\Theta z(\theta, Q)\alpha(d\theta)$.
Notice that for the special choice of $Q = \bar{P}_\alpha$, we have $a(\alpha, \bar{P}_\alpha) = 1$. Consequently, equation (2.4) is equivalent to

$$\frac{d\alpha^T}{d\alpha}(\theta) = z_T(\theta, \bar{P}_\alpha). \tag{2.5}$$

Parallel to the statements (i) - (iii) of section 2.1 on the density processes, the following properties of the arithmetic mean process can be stated:

Proposition 2.2. *Assume (2.1). The arithmetic mean process $a = a(\alpha, Q)$ possesses the following properties:*
(i) $\inf_t a_t > 0$ *Q-a.s.*
(ii) $\sup_t a_t < \infty$ *Q-a.s.*
(iii) a is a (Q, F)-uniformly integrable martingale with $E_Q\, a_t = 1$ for all $t \geq 0$.

Proof. In view of lemma 2.1 it suffices to refer again to [10], section III.3, proposition 3.5. □

Due to these properties, the arithmetic mean process $a(\alpha, Q)$, viewed as a density process, may be represented as a Doléans exponential of a certain (Q, F)-local martingale. We postpone this till section 3.2 in which this martingale will be given the form (3.6) involving certain posterior characteristics of observations.

To define the geometric mean measure we introduce yet another process $g(\alpha, Q)$ called the *geometric mean process* and associated with the density process $z(\theta, Q)$ by

$$g(\alpha, Q) = e^{E_\alpha \log z(\vartheta, Q)}. \tag{2.6}$$

By Jensen's inequality the geometric mean process is dominated by the a-mean process identically, i.e.

$$g(\alpha, Q) \leq a(\alpha, Q) \tag{2.7}$$

so that the geometric mean process also possesses property (ii) of proposition 2.2. As for the lower bound, we have assumed (2.3) in order to guarantee that the geometric mean process has property (i) of proposition 2.2 as well. It will be shown in the next proposition that under the present conditions the geometric mean process is a (Q, F)-supermartingale of class (D).

Proposition 2.3. *Assume (2.1) and (2.3). The geometric mean process $g = g(\alpha, Q)$ possesses the following properties:*
(i) $\inf_t g_t > 0$ *Q-a.s.*
(ii) $\sup_t g_t < \infty$ *Q-a.s.*
(iii) g is a (Q, F)-supermartingale of class (D) with $g_0 = 1$.

Proof. Property (i) is an immediate consequence of (2.3) and Jensen's inequality and (ii) follows from equation (2.7).
As for property (iii) we have that the g-mean process is indeed of class (D), since it is dominated by a process of class (D), a (Q, F)-uniformly integrable martingale a (see (2.7)). It remains to show that $E_Q\{g_t|\mathcal{F}_s\} \leq g_s$ for $s \leq t$. To this end apply first the Jensen inequality and then interchange the integration order: on the set $\{g_s > 0\}$ of full Q-measure

$$\begin{aligned} E_Q\left\{\frac{g_t}{g_s}|\mathcal{F}_s\right\} =& E_Q\{e^{E_\alpha \log \frac{z_t(\vartheta,Q)}{z_s(\vartheta,Q)}}|\mathcal{F}_s\} \leq E_Q\{E_\alpha \frac{z_t(\vartheta, Q)}{z_s(\vartheta, Q)}|\mathcal{F}_s\} \\ =& E_\alpha\{E_Q \frac{z_t(\vartheta, Q)}{z_s(\vartheta, Q)}|\mathcal{F}_s\} = 1. \end{aligned}$$

□

These properties of $g(\alpha, Q)$ allow us to characterize in the next theorem its compensator. In this theorem we define the Hellinger process of order α, denoted traditionally by $h(\alpha)$.

Theorem 2.4. *Assume (2.1) and (2.3). There exists a (unique up to Q-indistinguishability) predictable finite-valued increasing process $h(\alpha)$ starting from the origin $h_0(\alpha) = 0$, so that*

$$M(\alpha, Q) = g(\alpha, Q) + g_-(\alpha, Q) \cdot h(\alpha) \tag{2.8}$$

is a (Q, F)-uniformly integrable martingale. Moreover, two Hellinger processes $h(\alpha)$ determined under two different dominating measures Q and Q' are Q- and Q'-indistinguishable.

Proof. By the Doob-Meyer decomposition there exists a (unique up to Q-indistinguishability) increasing finite-valued predictable process A such that $g - A$ is a (Q, F)-uniformly integrable martingale. By proposition 2.3, property (ii), on the set $\{\sup_t g_t < \infty\}$ we can put $h(\alpha) = (1/g_-) \cdot A$ which satisfies the requirements of the theorem.
We show the uniqueness of the Hellinger process as follows. Assume $Q \ll Q'$. From $g(\alpha, Q') = Z\, g(\alpha, Q)$ and (2.8) we get

$$g(\alpha, Q') = Z\, g(\alpha, Q) = Z\, \{M(\alpha, Q) - g_-(\alpha, Q) \cdot h(\alpha)\}$$

so that by the Itô formula

$$g(\alpha, Q') = Z\, M(\alpha, Q) - \{g_-(\alpha, Q) \cdot h(\alpha)\} \cdot Z - Z_- g_-(\alpha, Q) \cdot h(\alpha).$$

The latter equation implies the desired result as the first two terms are Q'-martingales and the last term equals by $g(\alpha, Q') = Z\, g(\alpha, Q)$ to $g_-(\alpha, Q') \cdot h(\alpha)$. Thus similarly to (2.8)

$$g(\alpha, Q') + g_-(\alpha, Q') \cdot h(\alpha)$$

is a Q'-martingale. The proof is complete by the same reasoning as before. $\square$

The notions of the Hellinger process and the Hellinger integral of order α are closely related (see corollary 3.13 below). At a F-stopping time T, the Hellinger integral of the family of probability measures $\{P_{\theta,T}\}_{\theta\in\Theta}$, is defined according to [10], section IV.1, as the Q-expectation of the g-mean process evaluated at T:

$$H_T(\alpha) = E_Q\{g_T(\alpha, Q)\}. \tag{2.9}$$

This is called the *Hellinger integral of order* α. Its definition is independent of the dominating measure Q.

We are now in the position to define the *geometric mean measure* G_α with the help of the ratio

$$\zeta(\alpha, Q) = \frac{g(\alpha, Q)}{\mathcal{E}(-h(\alpha))} \tag{2.10}$$

as a density process, where $\mathcal{E}(-h(\alpha))$ is the Doléans-Dade exponential of $-h(\alpha)$.

Theorem 2.5. *Assume (2.1) and (2.3). Then the ratio (2.10) is a local martingale under Q and, with $M(\alpha, Q)$ as in (2.8), the following relations are valid:*

$$\zeta(\alpha, Q) = 1 + \frac{1}{\mathcal{E}(-h(\alpha))} \cdot M(\alpha, Q) \tag{2.11}$$

and

$$\zeta(\alpha, Q) = \mathcal{E}(\frac{1}{(1-\Delta h)g_-} \cdot M(\alpha, Q)). \tag{2.12}$$

Proof. Apply theorem 2.5.1 of [**?**] to the positive supermartingale $g(\alpha, Q)$ with the Doob-Meyer decomposition as in (2.8). This also yields formula (2.12). The expression (2.11) is a direct consequence of the Itô formula applied to $g(\alpha, Q)/\mathcal{E}(-h(\alpha))$ and the definition of $h(\alpha)$. It is now clear that $\zeta(\alpha, Q)$ is a Q-local martingale. $\square$

It is our purpose to use $\zeta(\alpha, Q)$ as a density process, for which it is necessary that $\zeta(\alpha, Q)$ is a martingale under Q. Since it is a nonnegative process, it is also a supermartingale, hence a sufficient condition for $\zeta(\alpha, Q)$ to become a martingale is $E_Q\zeta(\alpha, Q) \equiv 1$. In [7] this equality is assumed to hold.
As is well known, in general a positive local martingale is not necessarily a martingale. However, in a discrete time setting more can be said. Then it is shown in [11] that a nonnegative local martingale is in fact a martingale. So working in discrete time one obtains $E_Q\zeta(\alpha, Q) \equiv 1$. Other cases will be treated in the examples of section 5.

If we assume that $\zeta(\alpha,Q)$ is uniformly integrable, there is a nonnegative random variable $\zeta_\infty(\alpha,Q)$ with expectation 1 such that $E_Q\{\zeta_\infty(\alpha,Q)|\mathcal{F}_t\} = \zeta_t(\alpha,Q)$. We will often need this property, and therefore we will state this, in the same spirit as in [7], as an assumption. Since the nonnegative supermartingale $\zeta(\alpha,Q)$ has a limit a.s. for $t\to\infty$, call it $\zeta_\infty(\alpha,Q)$, we use it as a Radon-Nikodym derivative to define a new measure G_α on $(\Omega,\mathcal{F})$, so for all $B\in\mathcal{F}$ we have $G_\alpha(B) = E_Q 1_B \zeta_\infty(\alpha,Q)$. Alternatively, in terms of a density we have

$$\frac{dG_\alpha}{dQ} = \frac{g_\infty(\alpha,Q)}{\mathcal{E}(-h(\alpha))_\infty}, \tag{2.13}$$

with $g_\infty(\alpha,Q)$ the Q-a.s. limit of $g_\infty(\alpha,Q)$ for $t\to\infty$ and likewise $\mathcal{E}(-h(\alpha))_\infty$ the Q-a.s. limit of $\mathcal{E}(-h(\alpha))_t$ for $t\to\infty$. Clearly, both limits exist.
Notice that G_α is independent of the choice of the underlying measure Q and that in general G_α is a subprobability measure. When G_α is a probability measure, we call it the geometric mean measure.

Lemma 2.6. *Assume (2.3). Then the measure G_α is equivalent to Q.*

Proof. We have $G_\alpha \ll Q$ by construction. That $Q \ll G_\alpha$ follows from the first assertion of proposition 2.3. □

A sufficient condition for existence of G_α as a probability measure is given in the next proposition. It is in terms of the Hellinger process and we will return to it in section 5 when we treat examples. Furthermore, in section 3.5 we will see that G_α becomes a probability measure if the collection $\{P_\theta\}_{\theta\in\Theta}$ is an exponential family. Notice that the sufficient condition is satisfied if $h_\infty(\alpha)$ is $\bar{P}_\alpha$ (or Q)-a.s. bounded and in particular if it is deterministic and finite.

Proposition 2.7. *Assume that $E_{\bar{P}_\alpha}\{1/\mathcal{E}(-h(\alpha))_\infty\} < \infty$. Then the process $\zeta(\alpha,Q)$ is a uniformly integrable martingale under Q and hence G_α is a probability measure.*

Proof. If we use $\bar{P}_\alpha$ as the dominating measure, then the geometric mean is bounded above by the arithmetic mean, which for equals one, i.e. $a(\alpha,\bar{P}_\alpha) = 1$. Hence $\zeta(\alpha,\bar{P}_\alpha)$ is dominated by the $\bar{P}_\alpha$-integrable random variable $1/\mathcal{E}(-h(\alpha))_\infty$ and is therefore $\bar{P}_\alpha$-uniformly integrable. The conclusion now follows. □

Let us now agree upon the following notations. If $\{X(\theta)\}_{\theta\in\Theta}$ is a certain parametric family of processes, then $a(X) = E_\alpha X(\vartheta)$ and (for a nonnegative family) $g(X) = e^{E_\alpha \log X(\vartheta)}$ denote its arithmetic and geometric mean processes, respectively. Denote by $\phi(X) = a(X) - g(X)$ the difference of the arithmetic and geometric process and note that this difference process is homogeneous in the sense that if C is a process independent of θ, then

$$\phi(CX) = C\phi(X). \tag{2.14}$$

Note also that if the continuous part $X(\vartheta)^c$ possesses the *variance process*

$$v(X^c) \doteq var_\alpha X(\vartheta)^c = E_\alpha|X(\vartheta)^c|^2 - |E_\alpha X(\vartheta)^c|^2 \tag{2.15}$$

that is a (Q, F)-submartingale of class (D), then the compensator of $v(X^c)$ is given by

$$\tilde{v}(X^c) \doteq a(\langle X^c \rangle) - \langle a(X^c) \rangle. \tag{2.16}$$

In these terms the following general statement can be made.

Proposition 2.8. *Let $\{X(\theta)\}_{\theta\in\Theta}$ be a parametric family of (Q,F)-semimartingales with $\Delta X(\theta) > -1$ for all θ. Let its arithmetic mean process $a(X) = E_\alpha X(\vartheta)$ be a (Q,F)-semimartingale and $a_-(X) = E_\alpha X_-(\vartheta)$. Suppose that the increasing processes $a(\langle X^c \rangle)$ and $a(\sum_{s\le \cdot}(\Delta X_s - \log(1+\Delta X_s)))$ are finite-valued.*
Then the g-mean process $g(\mathcal{E}) = \exp E_\alpha\{\log \mathcal{E}(X(\vartheta))\}$ of the family of the Doléans exponentials $\{\mathcal{E}(X(\theta))\}_{\theta\in\Theta}$ is well-defined and

$$g(\mathcal{E}) = \mathcal{E}\{a(X) - \tfrac{1}{2}\tilde{v}(X^c) - \sum_{s\le\cdot} \phi_s(1+\Delta X)\} \tag{2.17}$$

where $\tilde{v}(\cdot) = a(\langle\cdot\rangle) - \langle a(\cdot)\rangle$ and $\phi(\cdot) = a(\cdot) - g(\cdot)$.

Proof. See [5, proposition 4.5]. □

Throughout we well use common notions and facts of the general theory of stochastic processes as developed e.g. in [10] or [18]. To describe, for instance, the discontinuous parts of processes in question, we associate with the jumps of a càdlàg process X an integer-valued random measure μ^X defined on $\mathbb{R}_+ \times E$ precisely following this theory, where $\mathbb{R}_+$ is the domain of the time component and E that of the space component (the range of the jumps of X), usually taken to be $\mathbb{R}\setminus\{0\}$. The same is applied to the notion of the compensator of the random measure μ^X with respect to a underlying measure. When this measure is the dominating measure Q, it is denoted as usual by ν. The latter occurs already in the beginning of the next section, together with $\nu(\theta)$ and $\bar{\nu}$ the compensators with respect to the measure $P_\theta, \theta\in\Theta$, and the arithmetic mean measure $\bar{P}_\alpha$, respectively.

3 Semimartingale observations

3.1 Characteristics w.r.t. the arithmetic mean measure

Suppose that we observe a semimartingale X defined on $(\Omega, \mathcal{F}, F, Q)$, i.e. a (Q,F)-semimartingale, with the triplet of predictable characteristics $T = (B, C, \nu)$. This and all the triplets considered in the present paper are related to a fixed truncation function $\hbar : \mathbb{R}\to\mathbb{R}$, a bounded function with a compact support so that $\hbar(x) = x$ in a vicinity of the origin. By the Girsanov theorem for semimartingales (see [10, Theorem III.3.24] or [18, Theorem IV.5.3]) X is also a (P_θ, F)-semimartingale for each $\theta\in\Theta$. Denote by $T(\theta) = (B(\theta), C(\theta), \nu(\theta))$ the corresponding triplet of predictable characteristics. It is related to the triplet T as follows:

$$\begin{cases} B(\theta) &= B + \beta(\theta)\cdot C + (Y(\theta)-1)\,\hbar\cdot\nu \\ C(\theta) &= C \\ \nu(\theta) &= Y(\theta)\cdot\nu \end{cases} \tag{3.1}$$

with certain processes $\beta(\theta) = \beta(\theta, Q)$ and $Y(\theta) = Y(\theta, Q)$ so that $|\beta(\theta)|^2 \cdot C_t < \infty$ and $(Y(\theta) - 1)\, \hbar \cdot \nu_t < \infty$ Q-a.s. for all $t \geq 0$. In [18, Lemma IV.5.6, p. 231] one can find the relationship of these processes to the density process $z(\theta, Q)$.
Under the present circumstances the observation of X constitute a semimartingale with respect to the arithmetic mean measure $\bar{P}_\alpha$, as well. The following theorem, taken over from [5, Section 3.3] (a generalization of a result by Kolomiets [14] that also can be found in [10, Theorem III.3.40] or [18, Theorem IV.5.4]), relates the triplet under $\bar{P}_\alpha$ to the triplets $T(\theta), \theta \in \Theta$:

Theorem 3.1. *Assume (2.1). Let X be a (P_θ, F)-semimartingale for each $\theta \in \Theta$ with the triplet $T(\theta)$ of predictable characteristics. Then it is a $(\bar{P}_\alpha, F)$-semimartingale as well, with the triplet $\bar{T} = (\bar{B}, \bar{C}, \bar{\nu})$ where*

$$\begin{cases} \bar{B} & = E_\alpha\{z_-(\vartheta, \bar{P}_\alpha) \cdot B(\vartheta)\} \\ \bar{C} & = C \\ \bar{\nu} & = E_\alpha\{z_-(\vartheta, \bar{P}_\alpha) \cdot \nu(\vartheta)\}. \end{cases} \tag{3.2}$$

Proof. See [5, Theorem 3.3]. □

This theorem yields an important corollary.

Corollary 3.2. *Under the conditions of theorem 3.1 the process $\bar{B}$ can be represented as $\bar{B} = B + \bar{\beta} \cdot C + (\bar{Y} - 1)\, \hbar \cdot \nu$, where the local characteristics $\bar{\beta}$ and $\bar{Y}$ with respect to the arithmetic mean measure $\bar{P}_\alpha$ are given by the posterior expectations of $\beta(\vartheta)$ and $Y(\vartheta)$: for each $t > 0$*

$$\bar{\beta}_t = E_{\alpha^{t-}} \beta_t(\vartheta) \quad \textit{and} \quad \bar{Y}_t = E_{\alpha^{t-}} Y_t(\vartheta). \tag{3.3}$$

Proof. In view of the identity (2.5) the definitions (3.3) are equivalent to

$$\bar{\beta} = E_\alpha\{z_-(\vartheta, \bar{P}_\alpha)\beta(\vartheta)\} \quad \text{and} \quad \bar{Y} = E_\alpha\{z_-(\vartheta, \bar{P}_\alpha)Y(\vartheta)\}. \tag{3.4}$$

By (3.1) and (3.2) $\bar{B} = B + \bar{\beta} \cdot C + (\bar{Y} - 1)\, \hbar \cdot \nu$ with $\bar{\beta}$ and $\bar{Y}$ as in (3.4). This confirms the desired assertion. □

Observe that the conditional expectations in equations (3.3) and (3.4) are precisely those that one encounters in the innovations representation in problems of nonlinear filtering. This is linked to the subject of this section by taking ϑ as the state (process or random variable) and X as the observations process. See [18, section 4.10] for a treatment of the case with semimartingale observation and state processes.

3.2 Arithmetic mean process as an exponential

Assume (2.1) and (2.3). For each $\theta \in \Theta$ let the density process be represented as the Doléans exponential $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ of the (Q, F)-local martingale $m(\theta, Q)$ given by (2.2). Upon further specification of the randomized experiment in question, one can assign to this martingale explicit form in terms of the triplet of predictable characteristics $T = (B, C, \nu)$ of the observed (Q, F)-semimartingale X. Assume therefore the setup of section 3.1. In addition to (2.1), assume that all (Q, F)-local martingales have the representation property relative to X. Then for each fixed $\theta \in \Theta$ the (Q, F)-local martingale (2.2) gets the form

$$m(\theta, Q) = \beta(\theta) \cdot X^c + \{Y(\theta) - 1 + \frac{\hat{Y}(\theta) - \hat{1}}{1 - \hat{1}}\} * (\mu^X - \nu) \tag{3.5}$$

where $\beta(\theta) = \beta(\theta, Q)$ and $Y(\theta) = Y(\theta, Q)$ are the same as in section 3.1. According to the usual 'hat' notation the processes $\hat{1} = \hat{1}(Q)$ and $\hat{Y}(\theta) = \hat{Y}(\theta, Q)$ are associated with the third characteristics ν and $\nu(\theta)$ (cf (3.1)) so that

$$\hat{1}_t(\omega) = \nu(\omega; \{t\} \times E) \quad \text{and} \quad \hat{Y}_t(\omega, \theta) = \int_E Y_t(\omega, \theta, x)\nu(\omega, \{t\}, dx) = \nu(\omega, \theta; \{t\} \times E),$$

with usually $E = \mathbb{R} \setminus \{0\}$, as was noted in subsection 2.2.
As we know, the arithmetic mean process is a certain density process, namely $a(\alpha, Q) = z(\bar{P}_\alpha, Q)$ with nice properties summarized in proposition 2.2. Departing from the representation property (3.5), we are now going to present this density process as a Doléans exponential of a certain (Q, F)-local martingale and to link it to that in (3.5).

Theorem 3.3. *Assume (2.1), (2.3) and the representation property (3.5). Then the arithmetic mean process is the Doléans exponential $a(\alpha, Q) = \mathcal{E}(\bar{m})$ of a (Q, F)-local martingale*

$$\bar{m} = \bar{\beta} \cdot X^c + \{\bar{Y} - 1 + \frac{\hat{\bar{Y}} - \hat{1}}{1 - \hat{1}}\} * (\mu^X - \nu) \tag{3.6}$$

where $\bar{\beta}$ and $\bar{Y}$ are given by (3.3).

Proof. Since the density process $a(\alpha, Q) = z(\bar{P}_\alpha, Q)$ possesses the properties given in proposition 2.2, it is indeed representable as an exponential, say $a(\alpha, Q) = \mathcal{E}(\bar{m})$. A (Q, F)-local martingale $\bar{m}$ involved has the presumed form by the assumption of the representation property, like the one displayed in (3.6). The only question remains, how to identify $\bar{\beta}$ and $\bar{Y}$ in the integrands. But from section 3.1 we already know the answer: they must be of the form (3.3), due to Girsanov's transformation and the formula (3.2) for the triplet of predictable characteristics $\bar{T}$ under the arithmetic mean measure $\bar{P}_\alpha$. The proof is complete. □

Theorem 3.3 has important consequence: it allows us to express the density (2.5) of the posterior with respect to the prior as a Doléans exponential.

Corollary 3.4. *Assume (2.1), (2.3) and assume the representation property (3.5). Then at each stopping time T the density of the posterior α^T with respect to the prior α is a Doléans exponential at each $\theta \in \Theta$*

$$\frac{d\alpha^T}{d\alpha}(\theta) = \mathcal{E}(m(\theta, \bar{P}_\alpha))_T$$

with $m(\theta, \bar{P}_\alpha)$ a $(\bar{P}_\alpha, F)$-local martingale defined by

$$m(\theta, \bar{P}_\alpha) = \big(\beta(\theta) - \bar{\beta}\big) \cdot X^{c, \bar{P}_\alpha} + \{\frac{Y\,(\theta)}{\bar{Y}} - 1 + \frac{\hat{Y}\,(\theta) - \hat{\bar{Y}}}{1 - \hat{\bar{Y}}}\} * \big(\mu^X - \bar{\nu}\big) \tag{3.7}$$

where $X^{c, \bar{P}_\alpha} = X^c - \bar{\beta} \cdot C$ is the continuous local martingale part of X under $\bar{P}_\alpha$.

Proof. By (2.5) it is required to show $z(\theta, \bar{P}_\alpha) = \mathcal{E}(m(\theta, \bar{P}_\alpha))$, that is to show $\mathcal{E}(m(\theta, Q)) = \mathcal{E}(m(\theta, \bar{P}_\alpha))\, \mathcal{E}(\bar{m})$. Using the well-known multiplication rule for Doléans exponentials, it suffices to verify $m(\theta, Q) = m(\theta, \bar{P}_\alpha) + \bar{m} + [m(\theta, \bar{P}_\alpha), \bar{m}]$. For the continuous parts this is easily verified. It is then enough to identify the jumps on the both sides and to verify the relation

$$1 + \Delta m(\theta, \bar{P}_\alpha) = \frac{1 + \Delta m(\theta, Q)}{1 + \Delta \bar{m}}. \tag{3.8}$$

To this end, observe first that

$$\begin{aligned} 1 + \Delta m(\theta, Q) =& 1 + \{Y(\theta; \cdot, \Delta X) - 1\} I_{\{\Delta X \neq 0\}} - \frac{\hat{Y}(\theta) - \hat{1}}{1 - \hat{1}} I_{\{\Delta X = 0\}} \\ =& Y(\theta; \cdot, \Delta X) I_{\{\Delta X \neq 0\}} + \frac{1 - \hat{Y}(\theta)}{1 - \hat{1}} I_{\{\Delta X = 0\}} \end{aligned} \tag{3.9}$$

(basically, we only need to recall the definition of the stochastic integral $W * (\mu^X - \nu)$: it is any purely discontinuous local martingale having the jumps $W(\cdot, \cdot, \Delta X) 1_{\{\Delta X \neq 0\}} - \hat{W}$, cf [10, definition II.1.27] or [18, theorem 3.5.1]). Substitute then θ by ϑ and take on both sides the expectation with respect to the posterior α^- to get

$$1 + \Delta \bar{m} = \bar{Y}(\cdot; \cdot, \Delta X) I_{\{\Delta X \neq 0\}} + \frac{1 - \hat{\bar{Y}}}{1 - \hat{1}} I_{\{\Delta X = 0\}} \tag{3.10}$$

(one may derive this directly from (3.6), of course). Finally, apply this device to (3.7). We get

$$1 + \Delta m(\theta, \bar{P}_\alpha) = \frac{Y(\theta; \cdot, \Delta X)}{\bar{Y}(\cdot; \cdot, \Delta X)} I_{\{\Delta X \neq 0\}} + \frac{1 - \hat{Y}(\theta)}{1 - \hat{\bar{Y}}} I_{\{\Delta X = 0\}}.$$

The last three relations imply (3.8). The proof is complete. □

3.3 Representation of Hellinger processes

Assume again (2.1), (2.3) and the representation (3.5) of a (Q, F)-local martingale $m(\theta, Q)$ for each $\theta \in \Theta$. By applying to the latter martingale the notations upon which we have agreed at the end of section 2.1, we may introduce the process

$$V = \frac{1}{2} v(m^c) + \sum_{s \leq \cdot} \phi_s(1 + \Delta m) \tag{3.11}$$

assumed to be a (Q, F)-submartingale. We have written m as a shorthand notation for $m(\vartheta, Q)$. Then *its compensator $\tilde{V}$ and the Hellinger process $h(\alpha)$ are Q-indistinguishable.* As is shown in [5, section 4.5], this statement is an easy consequence of the general proposition 2.8 applied to $m(\vartheta, Q)$. Therefore we do not dwell upon this here. Instead, we are going now to present in the next theorem the compensator $\tilde{V}$ in terms of the triplet of predictable characteristics of the observations (cf [5, theorem 5.3]; the proof is reproduced below, since the basic arguments are needed anew in the subsequent sections).

Theorem 3.5. *Along with the conditions (2.1) and (2.3) assume the representation property (3.5). Then*

$$h(\alpha) = \frac{1}{2}v(\beta)\cdot C + \phi(Y)\cdot\nu + \sum_{s\leq\cdot}\phi_s(1-\hat{Y}). \tag{3.12}$$

Proof. The first term in (3.11) is compensated as follows. The compensator $\tilde{v}(m^c)$ of the variance process $v(m^c)$ is $\tilde{v}(m^c) = v(\beta)\cdot C$. This is easily seen by applying (2.15) and (2.16) to $m(\theta, Q)^c = \beta(\theta)\cdot X^c$. Next, we have to show that the second term in (3.11) is compensated by the sum of the last two terms in (3.12), i.e. that

$$\sum_{s\leq\cdot}\phi_s(1+\Delta m) - \{\phi(Y)\cdot\nu + \sum_{s\leq\cdot}\phi_s(1-\hat{Y})\} \tag{3.13}$$

is a (Q, F)-local martingale. But this claim holds true, in view of lemma 3.6 below upon noting that ϕ is homogeneous (see equation (2.14)). □

Now we formulate a lemma with the computational tool that we needed in the course of proving theorem 3.5 and that we will also use in the proof of theorem 4.4.

Lemma 3.6. *Let $m = m(\theta, Q)$ be given by (3.5) and let for a certain function f the process $\sum_{s\leq\cdot} f(1+\Delta m_s)$ be a special semimartingale. Then this process has compensator*

$$f(Y)\cdot\nu + \sum_{s\leq\cdot} f\Big(\frac{1-\hat{Y}_s}{1-\hat{1}_s}\Big)(1-\hat{1}_s)$$

and the local martingale in its semimartingale decomposition takes the form

$$\{f(Y) - f\Big(\frac{1-\hat{Y}}{1-\hat{1}}\Big)\} * (\mu^X - \nu). \tag{3.14}$$

Proof. See [18, theorem 3.5.1]. □

Remark 3.7. The explicit expression for the (Q, F)-local martingale (3.13) is then (use again lemma 3.6 and the fact that ϕ is homogeneous)

$$\{\phi(Y) - \frac{\phi(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu). \tag{3.15}$$

3.4 Characteristics w.r.t. the geometric mean measure

In this section we compute the predictable characteristics of the observe process under the geometric mean measure and we give an explicit expression for the multiplicative decomposition of the geometric mean process $g(\alpha, Q)$.

Suppose once more that the observations constitute a semimartingale X that possesses the triplet of predictable characteristics $T = (B, C, \nu)$ with respect to the probability measure Q and the triplet $T(\theta) = (B(\theta), C(\theta), \nu(\theta))$ with respect to the probability measure $P_\theta, \theta \in \Theta$, cf (3.1). In the next theorem a characterization is given for the density process $z(G_\alpha, Q)$ which is defined at each $t \geq 0$ by $z_t(G_\alpha, Q) = E_Q\{\frac{dG_\alpha}{dQ}|\mathcal{F}_t\}$, provided that G_α is a probability measure.

Theorem 3.8. *Assume (2.1), (2.3) and the representation property (3.5). Let the geometric mean measure G_α be a probability measure. Then the density process $z(G_\alpha, Q)$ may be presented as a Doléans exponential*

$$z(G_\alpha, Q) = \mathcal{E}(\frac{1}{1-\Delta h(\alpha)} \cdot N(\alpha, Q)) \tag{3.16}$$

where

$$N(\alpha, Q) = a(\beta) \cdot X^c + \{g(Y) - \frac{g(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu) \tag{3.17}$$

is a (Q, F)-local martingale that is simply related to $M(\alpha, Q)$ defined by (2.8):

$$M(\alpha, Q) = g_-(\alpha, Q) \cdot N(\alpha, Q). \tag{3.18}$$

Proof. Relation (3.16) follows from (2.12) and equation (3.18).
Next we verify that $N(\alpha, Q)$ has the representation (3.17). Since the martingale $M(\alpha, Q)$ in theorem 2.4 can be expressed as

$$M = g_-(z) \cdot \{a(m) + \frac{1}{2}\{v(m^c) - \tilde{v}(m^c)\} - (V - \tilde{V})\},$$

it follows that N is the difference of $a(m) = a(m^c) + \{a(Y) - \frac{a(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu)$ and the local martingale in (3.13) that is according to remark 3.7 given by $\{\phi(Y) - \frac{\phi(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu)$, cf (3.15). Since $\phi = a - g$, the difference results in $a(m^c) + \{g(Y) - \frac{g(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu)$, which coincides with the right hand side of (3.17). □

Remark 3.9. Of course the decomposition (3.16) is also valid for the process $\zeta(\alpha, Q)$ of equation (2.10) without condition (3.5). We only needed it to specify the martingale in (3.17).

Theorem 3.10. *Assume (2.1), (2.3) and the representation property (3.5). If the geometric mean measure G_α is a probability measure, then the triplet of predictable characteristics $T^{G_\alpha} = (B^{G_\alpha}, C^{G_\alpha}, \nu^{G_\alpha})$ of X with respect to G_α is*

$$\begin{cases} B^{G_\alpha} & = a(B) + (Y^{G_\alpha} - a(Y))\hbar \cdot \nu \\ C^{G_\alpha} & = C \\ \nu^{G_\alpha} & = Y^{G_\alpha} \cdot \nu \ \ \textit{with} \ \ Y^{G_\alpha} = \frac{g(Y)}{1-\Delta h(\alpha)}. \end{cases} \tag{3.19}$$

Proof. We use theorem 3.8 and focus first on the third characteristic ν^{G_α}. By Girsanov's theorem for random measures (see [10, Theorem III.3.17]) we need to calculate the so-called "conditional M^P_μ-expectation" of $\Delta N(\alpha, Q)$, because it yields Y^{G_α}. The definition of this operation is given prior to the aforementioned theorem on p.157 of [10] and the rule for calculations in theorem 4.20 on p.170. According to this rule Y^{G_α} has to be related to the integrand $g(Y) - \frac{g(1-\hat{Y})}{1-\hat{1}}$ in the discontinuous part of $N(\alpha, Q)$ as follows: $g(Y) - \frac{g(1-\hat{Y})}{1-\hat{1}} = U + \frac{\hat{U}}{1-\hat{1}}$ with $Y^{G_\alpha} - 1 = \frac{U}{1-\Delta h(\alpha)}$. All we need then is to verify that the postulated $Y^{G_\alpha} = \frac{g(Y)}{1-\Delta h(\alpha)}$ indeed satisfies this relationship. This is accomplished by means of simple algebra upon noting that by equation (3.12) and by $\phi = a - g$ we have

$$1 - \Delta h(\alpha) = \hat{g}(Y) + g(1 - \hat{Y}). \tag{3.20}$$

Observe that the latter identity yields $1 - \Delta h(\alpha) + \Delta N(\alpha, Q) = g(Y(\cdot, \cdot, \Delta X))$, hence by (3.16) we have for $m(G_\alpha, Q) \doteq z_-(G_\alpha, Q)^{-1} \cdot z(G_\alpha, Q)$ that $1 + \Delta m(G_\alpha, Q) = \frac{g(Y(\cdot,\cdot,\Delta X))}{1-\Delta h(\alpha)}$. Since the second equality in (3.19) is trivial, we finally prove the first one. According to Girsanov's theorem III.3.24 in [10] we have

$$B^{G_\alpha} = B + a(\beta) \cdot C + (Y^{G_\alpha} - 1)\hbar \cdot \nu, \tag{3.21}$$

since $\langle N(\alpha, Q), X^c \rangle = a(\beta) \cdot C$. On the other hand, from (3.1) we obtain $a(B) = B + a(\beta) \cdot C + (a(Y) - 1)\hbar \cdot \nu$. We get the desired result by subtracting the two expressions. □

Remark 3.11. Application of the Girsanov theorem to the change of measure from Q to G_α yields (like in (3.1)) that B^{G_α} is given by $B^{G_\alpha} = B + \beta^{G_\alpha} \cdot C + (Y^{G_\alpha} - 1)\hbar \cdot \nu$. Comparing this to (3.21), we obtain that the local characteristic β^{G_α} is the arithmetic mean of the $\beta(\theta)$, i.e. $\beta^{G_\alpha} \cdot C = a(\beta) \cdot C$.

By (3.16), the multiplicative decomposition of the geometric mean process $g(\alpha, Q)$ resulting from (2.10), can be given a specific form.

Corollary 3.12. *Assume (2.1), (2.3) and the representation property (3.5). Then the geometric mean process possesses the multiplicative decomposition*

$$g(\alpha, Q) = \mathcal{E}(\frac{1}{1 - \Delta h(\alpha)} \cdot N(\alpha, Q))\, \mathcal{E}(-h(\alpha)), \tag{3.22}$$

with $N(\alpha, Q)$ as in (3.17).
If G_α is taken as the dominating measure, then the above identity can be replaced with

$$g(\alpha, G_\alpha) = \mathcal{E}(-h(\alpha)). \tag{3.23}$$

Proof. Combine equation (2.10) and theorem 3.8 to get (3.22), whereas (3.23) immediately follows from (2.10). □

Another important consequence is the following useful representation of the Hellinger integral that has been defined by (2.9).0

Corollary 3.13. *Assume (2.1), (2.3) and the representation property (3.5). Then at a stopping time T the Hellinger integral and the Hellinger process are related as follows:*

$$H_T(\alpha) = E_{G_\alpha} \mathcal{E}(-h(\alpha))_T.$$

Proof. Substitute Q in (2.9) by G_α and apply (3.23). □

3.5 Exponential families of experiments

In this section we apply some of the preceding results to a so called exponential family of measures and characterize these in terms of geometric mean measures. We confine ourselves to finite dimensional parameter sets, although extension to infinite dimensional parameter space is possible, at the cost of increasing technical complexity. So, let Θ be a subset of $\mathbb{R}^k$. We assume that Θ is convex. Then the expectation (if it exists) $\theta^\circ_\alpha \doteq E_\alpha \vartheta$ belongs to Θ for any measure α on $(\Theta, \mathcal{A})$ and hence $P_{\theta^\circ_\alpha}$ belongs to the parametric family $\{P_\theta\}_{\theta\in\Theta}$. It is already here that extension to infinite dimensional parameter spaces would become cumbersome, since one has to use integrals with values in an infinite dimensional space, like Pettis integrals, that presupposes that Θ is a subset of a (reflexive) Banach space, see e.g. [24].

We continue to work with an observed process X that is a semimartingale under all probability measures involved and we give necessary and sufficient conditions for the simple identity $G_\alpha = P_{\theta^\circ_\alpha}$, valid for all a priori distributions α such that $\theta^\circ_\alpha = E_\alpha\vartheta$ exists and is finite. In particular, validity of $G_\alpha = P_{\theta^\circ_\alpha}$ implies that G_α is a probability measure (in general this is not guaranteed, as was already pointed out). As we will show, this identity holds iff $\{P_\theta\}_{\theta\in\Theta}$ is an exponential family. In the book [15] (in particular chapter 11, that deals with semimartingales) exponential families are defined in a number of specific cases, see also [23, chapter 2]. Exponential families of a general nature in a static context have been studied in [21]. The definition below is much in spirit of [15, chapter 11]. We exclude, however, the so-called curved exponential families, for this would carry us too far afield.

The following notation and terminology will be in use. Vectors will always be understood as column vectors and $^\top$ denotes transposition.

A function f on Θ is called *affine* if both f and $-f$ are convex. For an affine function f there exist a unique $c \in \mathbb{R}$ and $\dot f \in \mathbb{R}^k$ such that $f(\theta) = c + \theta^\top \dot f$. Notice that $f(\theta^\circ_\alpha) = E_\alpha f(\vartheta)$ if f is affine and if $\theta^\circ_\alpha = E_\alpha\vartheta$ exists.

A positive function f is called *log-affine* if $l \doteq \log f$ is affine. In this case there exist unique $c > 0$ and $\dot l \in \mathbb{R}^k$ such that $f(\theta) = c\exp\{\theta^\top \dot l\}$. If f is log-affine, then $\exp\{E_\alpha \log f(\theta)\} = f(\theta^\circ_\alpha)$ if $\theta^\circ_\alpha = E_\alpha\vartheta$ exists.

Definition 3.14. Consider a family $\{P_\theta\}_{\theta\in\Theta}$ of probability measures with the associated triplet (3.1). It is said that this family is exponential with respect to a dominating measure Q if $\beta(\theta)$ and $Y(\theta)$ can be chosen to satisfy the following two requirements.
(i) At each instant $t \geq 0$ $\beta_t(\cdot)$ is an affine function of $\theta \in \Theta$ identically for all $\omega \in \Omega$.
(ii)Either the processes $Y(\theta)$ identically vanish, or at each instant $t \geq 0$ and $x \in E$

$$Y(\cdot; t, x) = \frac{l(\cdot; t, x)}{1_{\{\hat{1}<1\}} + \hat{l}_t(\cdot)} \tag{3.24}$$

where l is a certain log-affine function of $\theta \in \Theta$ identically for all $\omega \in \Omega$.

Remark 3.15. Note that if $\hat{1} < 1$, then (3.24) is equivalent to $\frac{Y(\theta;t,x)}{1-\hat{Y}_t(\theta)} = l(\theta; t, x)$, and that in the quasi-continuous case $Y(\cdot)$ itself is log-affine in $\theta \in \Theta$.

The main result of this section is

Theorem 3.16. *(i) If $\{P_\theta\}_{\theta\in\Theta}$ is an exponential family and $\theta^\circ_\alpha = E_\alpha\vartheta$ exists, then G_α is a probability measure and the identity $G_\alpha = P_{\theta^\circ_\alpha}$ holds.*
(ii) If the identity $G_\alpha = P_{\theta^\circ_\alpha}$ holds for all α with finite support, then $\{P_\theta\}_{\theta\in\Theta}$ is an exponential family.

Proof. (i) Assume that our family of measures is exponential, i.e. by definition 3.14, at $\theta^\circ_\alpha = E_\alpha\vartheta$ we have $\beta(\theta^\circ_\alpha) = a(\beta) = \beta^{G_\alpha}$ and (excluding the trivial case of vanishing Y) $Y(\theta^\circ_\alpha) = \frac{l(\theta^\circ_\alpha)}{1_{\{\hat{1}<1\}}+\hat{l}(\theta^\circ_\alpha)} = \frac{g(l)}{1_{\{\hat{1}<1\}}+\hat{g}(l)}$. Thus to show that $Y(\theta^\circ_\alpha) = Y^{G_\alpha}$ as well, it suffices to verify

$$\frac{g(Y)}{\hat{g}(Y) + g(1-\hat{Y})} = \frac{g(l)}{1_{\{\hat{1}<1\}} + \hat{g}(l)}, \tag{3.25}$$

cf definition of Y^{G_α} in (3.19) in conjunction with (3.20). But due to (3.24) this follows from $1-\hat{Y}(\theta) = 1_{\{\hat{1}<1\}}/(1_{\{\hat{1}<1\}} + \hat{l}(\theta))$ and $g(Y) = g(l)/g(1_{\{\hat{1}<1\}} + \hat{l})$, since the former implies that $g(1-\hat{Y}) = 1_{\{\hat{1}<1\}}/g(1_{\{\hat{1}<1\}} + \hat{l})$ and the latter that $\hat{g}(Y) = \hat{g}(l)/g(1_{\{\hat{1}<1\}} + \hat{l})$. Indeed, these substitutions conform the identity (3.25).
(ii) Since now $G_\alpha = P_{\theta^\circ_\alpha}$ for all α with finite support, we have $a(\beta) = \beta(\theta^\circ_\alpha)$ and $Y^{G_\alpha} = Y(\theta^\circ_\alpha)$. Since this holds for all such α, the first of these equalities implies that $\beta(\cdot)$ is affine in θ. So we now focus on the second one and distinguish two cases. Fix a time t. In the first case we assume that $\hat{1}_t = 1$. By (2.1) and [10, theorem III.3.17] then also all $\hat{Y}_t(\theta) = 1$. Hence we have $\frac{g(Y)}{\hat{g}(Y)} = Y(\theta^\circ_\alpha)$. Choosing x_0 such that $Y(\theta^\circ_\alpha; t, x_0) > 0$, we obtain that (we now and henceforth in this part of the proof suppress the fixed time instant t)

$$g(\frac{Y(\cdot;x)}{Y(\cdot;x_0)}) = \frac{g(Y(\cdot;x))}{g(Y(\cdot;x_0))} = \frac{Y(\theta^\circ_\alpha;x)}{Y(\theta^\circ_\alpha;x_0)}.$$

Since the extreme sides of this double equation are the same for all α with finite support, we must have that $l(\cdot;x) \doteq \frac{Y(\cdot;x)}{Y(\cdot;x_0)}$ is log-affine in θ. Hence $Y(\cdot;x) = l(\cdot;x)Y(\cdot;x_0)$ and by taking hats and using that $\hat{Y}_t(\theta) = 1$ we finally obtain $Y(\cdot;x) = \frac{l(\cdot;x)}{\hat{l}(\theta)}$, which is what we had to show for this case (cf (3.24)).
In the second case we assume that $\hat{1}_t < 1$. In view of remark 3.15 it is now sufficient to prove that $\frac{Y}{1-\hat{Y}}$ (again we suppress the fixed time instant t) is log-affine in θ. From $Y^{G_\alpha} = Y(\theta^\circ_\alpha)$ we deduce that

$$g(\frac{Y}{1-\hat{Y}}) = \frac{Y(\theta^\circ_\alpha)}{1-\hat{Y}(\theta^\circ_\alpha)}.$$

But since this equality now holds true for all α with finite support, we conclude that indeed $\frac{Y}{1-\hat{Y}}$ is log-affine in θ. □

Remark 3.17. The statements of theorem 3.16 can be summarized by saying that the geometric mean measures G_α are equal to $P_{\theta^\circ_\alpha}$ for all α such that $E_\alpha\vartheta$ exists iff they are equal to $P_{\theta^\circ_\alpha}$ for all α with finite support.

We recall the multiplicative decomposition of the geometric mean process given by (3.22) in conjunction with (3.16):

$$g(\alpha, Q) = z(G_\alpha, Q)\mathcal{E}(-h(\alpha)). \tag{3.26}$$

We may now take $Q = P_{\theta^\circ_\alpha}$ as a dominating probability measure to get the following characterization of an exponential family.

Proposition 3.18. *The family of measures $\{P_\theta\}_{\theta\in\Theta}$ is exponential iff for all α on $(\Theta, \mathcal{A})$ such that $\theta^\circ_\alpha = E_\alpha\vartheta$ exists we have $g(\alpha, P_{\theta^\circ_\alpha}) = \mathcal{E}(-h(\alpha))$.*

Proof. This follows directly from theorem 3.16 and equation (3.26) upon the substitution $Q = P_{\theta^\circ_\alpha}$. □

4 Information quantities

4.1 Information in the posterior given the prior

Let us turn back to the Bayes formula (2.4). Recall that, using the arithmetic mean measure $\bar{P}_\alpha$ as a dominating measure, we may present this formula as identity (2.5) of section 2.2. This representation proves to be useful, since the process $z(\theta, \bar{P}_\alpha)$ is a martingale with respect to $\bar{P}_\alpha$. Define at a stopping time $T > 0$ the Kullback-Leibler information in the posterior probability measure α^T with respect to the prior α by

$$I(\alpha^T|\alpha) = E_\alpha \log \frac{d\alpha}{d\alpha^T}(\vartheta) \tag{4.1}$$

that is a non-negative quantity by the Jensen inequality. It is simply related to the arithmetic and geometric mean processes as follows:

$$e^{-I(\alpha^T|\alpha)} = \frac{g_T(\alpha, Q)}{a_T(\alpha, Q)} = g_T(\alpha, \bar{P}_\alpha). \tag{4.2}$$

Observe that the information $I(\alpha^T|\alpha)$ depends only on the prior α but not on the choice of a dominating measure Q. By (2.5) we have

$$E_{\bar{P}_\alpha} I(\alpha^T|\alpha) = E_\alpha I(P_{\vartheta,T}|\bar{P}_{\alpha,T}). \tag{4.3}$$

In view of the propositions 2.2 and 2.3 we have

Proposition 4.1. *Assume (2.1) and (2.3). Let $I(\alpha^\cdot|\alpha)$ be the information process starting from zero, $I(\alpha^0|\alpha) = 0$, and at $t > 0$ defined by (4.1). Then it possesses the following properties:*
(i) $\inf_t I(\alpha^t|\alpha) > 0$ *Q-a.s.*
(ii) $\sup_t I(\alpha^t|\alpha) < \infty$ *Q-a.s.*
(iii) $e^{-I(\alpha^\cdot|\alpha)}$ is a $(\bar{P}_\alpha, F)$-supermartingale of class (D).

Proof. In view of the relationship (4.2), this is a direct consequence of the propositions 2.2 and 2.3. □

4.2 Information in the prior given the posterior

The previous considerations rely on the condition (2.3) concerning the Kullback-Leibler information $I(\mathbf{P}|\mathbf{Q})$ in $\mathbf{P}$ given $\mathbf{Q}$. Now we need to look at $I(\mathbf{Q}|\mathbf{P})$, called sometimes the *relative entropy* in $\mathbf{P}$ given $\mathbf{Q}$ (the term used in the theory of large deviations to characterize this quantity as *the average relative entropy in the experiment given a dominating measure* Q ; cf e.g. [4], section 1.4; for a different, statistical context, see e.g. [16]). Contrary to (2.3), we then will need the condition $0 < I(\mathbf{Q}|\mathbf{P}) < \infty$. Actually, we only apply this to the particular dominating measure $\bar{P}_\alpha$, so it suffices to require

$$0 < I(\bar{\mathbf{P}}_\alpha|\mathbf{P}) < \infty, \tag{4.4}$$

where $\bar{\mathbf{P}}_\alpha$ is the product measure $\bar{P}_\alpha \times \alpha$ on $\mathbf{\Omega}$. The latter condition is indeed implied by the former, since $I(\mathbf{Q}|\mathbf{P}) = I(Q|\bar{P}) + I(\bar{\mathbf{P}}|\mathbf{P})$.
Define at a stopping time $T > 0$ the relative entropy in the prior given the posterior with

$$I(\alpha|\alpha^T) \doteq E_{\alpha^T} \log \frac{d\alpha^T}{d\alpha}(\vartheta). \tag{4.5}$$

In Bayesian statistics this quantity is called *information from data* (see [3, Definition 2.26]). The expression $E_{\bar{P}_\alpha} I(\alpha|\alpha^T)$ is called *expected utility from data.* By taking into consideration (2.5) we get the following representation:

$$I(\alpha|\alpha^T) = E_\alpha \left\{ z_T\left(\vartheta, \bar{P}_\alpha\right) \log z_T\left(\vartheta, \bar{P}_\alpha\right)\right\}$$

so that the expected utility from data at the stopping time T equals to

$$E_{\bar{P}_\alpha} I(\alpha|\alpha^T) = E_\alpha I(\bar{P}_{\alpha,T}|P_{\vartheta,T}). \tag{4.6}$$

Notice that also the information from data process $I(\alpha|\alpha^\cdot)$ is a $(\bar{P}_\alpha, F)$-submartingale. Indeed this follows from the fact that $z(\theta, \bar{P}_\alpha)$ is a $(\bar{P}_\alpha, F)$-martingale and that $I(\alpha|\alpha^\cdot) = E_\alpha \ell(z(\vartheta, \bar{P}_\alpha))$ where $\ell(x) = x\log x$ is a convex function of $x \in \mathbb{R}_+$.
It is easily seen that at $T = \infty$ the expected utility from data is nothing else but the relative entropy in (4.4), this clarifies its necessity in the present context.

4.3 Representation of a posterior information

The information $I(\alpha^T|\alpha)$ in the posterior α^T with respect to the prior α satisfies identity (4.2), therefore we have

Theorem 4.2. *Assume is (2.1), (2.3), the representation property (3.5) and that G_α is a probability measure. Then the information $I(\alpha^T|\alpha)$ at a stopping time $T > 0$ may be presented as follows:*

$$e^{-I(\alpha^T|\alpha)} = z_T(G_\alpha, \bar{P}_\alpha)\, \mathcal{E}(-h(\alpha))_T \tag{4.7}$$

where the density process $z(G_\alpha, \bar{P}_\alpha)$ of the geometric mean measure G_α with respect to the arithmetic mean measure $\bar{P}_\alpha$ is the Doléans exponential

$$z(G_\alpha, \bar{P}_\alpha) = \mathcal{E}(\frac{1}{1-\Delta h(\alpha)} \cdot N(\alpha, \bar{P}_\alpha))$$

with

$$N(\alpha, \bar{P}_\alpha) = \big(a(\beta) - \bar{\beta}\big) \cdot X^{c,\bar{P}_\alpha} + \{g\left(\tfrac{Y}{\bar{Y}}\right) - g\left(\tfrac{1-\hat{Y}}{1-\hat{\bar{Y}}}\right)\} * (\mu^X - \bar{\nu}) \tag{4.8}$$

where $\bar{\beta}$, $\bar{Y}$ and $\bar{\nu}$ are predictable characteristics of the observed process X with respect to the arithmetic mean measure $\bar{P}_\alpha$ and $X^{c,\bar{P}_\alpha}$ the continuous local martingale in the semimartingale decomposition of X under $\bar{P}_\alpha$, cf (3.7).

Proof. Equation (4.7) follows from (3.26) and (4.2). Then, it suffices to substitute Q in (3.17) by $\bar{P}_\alpha$ and to verify that $N(\alpha, \bar{P}_\alpha)$ indeed has the asserted form, which we will do by following the same arguments as in the course of proving corollary 3.4. Firstly, the multiplication rule for Doléans exponentials is applied according to which the following identity has to hold: $N(\alpha, Q) = N(\alpha, \bar{P}_\alpha) + (1 - \Delta h(\alpha)) \cdot \bar{m} + [N(\alpha, \bar{P}_\alpha), \bar{m}]$. The comparison of the continuous parts is simple. As for the discontinuous parts, it suffices to equate the jumps and to verify that

$$1 + \frac{\Delta N(\alpha, \bar{P}_\alpha)}{1 - \Delta h(\alpha)} = \frac{1 + \Delta N(\alpha, Q)/(1 - \Delta h(\alpha))}{1 + \Delta \bar{m}}$$

like in (3.8). To this end use (3.10) and (3.20) and determine $\Delta N(\alpha, Q)/(1 - \Delta h(\alpha))$ and $\Delta N(\alpha, \bar{P}_\alpha)/(1 - \Delta h(\alpha))$ from (3.17) and (4.8), respectively, by following the same device as in the course of proving corollary 3.4. ☐

Remark 4.3. Under the conditions of theorem 4.2 we have

$$E_{\bar{P}_\alpha} I(\alpha^T|\alpha) = E_\alpha I(P_{\vartheta,T}|\bar{P}_{\alpha,T}) = I(G_{\alpha,T}|\bar{P}_{\alpha,T}) - E_{\bar{P}_\alpha} \log \mathcal{E}(-h(\alpha))_T. \tag{4.9}$$

The first identity is (4.3). The second one follows from (4.7).

4.4 Representation of the information from data

Suppose that the observed process X is a (Q, F)-semimartingale with the triplet of predictable characteristics $T = (B, C, \nu)$. As in section 3.3, assume the representation property for the density processes $z(\theta, Q)$.

Denote by $L(x, y)$ the function $L(x, y) = x \log \frac{x}{y}$. The function L may be used to compute Kullback-Leibler information with respect to a dominating measure, e.g. if for two equivalent measures P and Q the information $I(P|Q)$ is needed to be calculated in terms of a certain measure Q' that dominates both P and Q, then the following relation is applied: $I(P|Q) = E_{Q'} L(z(Q, Q'), z(P, Q'))$.

In the next theorem we will use the following notation, in the spirit of section 3.1: for a quantity f free of θ and g possibly depending on θ we write $\bar{L}(g, f) = E_{\alpha^-} L(g(\vartheta), f)$ (assuming of course the appropriate measurability and integrability conditions). Besides, we will use the *posterior variance* of $\beta(\vartheta)$ that is defined like in (2.15) as follows: $\bar{v}(\beta) = E_{\alpha^-}(\beta(\vartheta) - \bar{\beta})^2$. In the present circumstances we get the Doob-Meyer decomposition of the of the information from data process.

Theorem 4.4. *Assume (2.1), (4.4) and that (Q, F)-local martingales have the representation property relative to X. Then the nondecreasing finite-valued predictable process*

$$\tfrac{1}{2}\bar{v}(\beta) \cdot \langle X^c \rangle + \bar{L}(Y, \bar{Y}) \cdot \nu + \sum_{s \leq \cdot} \bar{L}(1 - \hat{Y}_s, 1 - \hat{\bar{Y}}_s) \tag{4.10}$$

compensates the information from data process $I(\alpha|\alpha^{\cdot})$ (cf (4.5)) to a $(\bar{P}_\alpha, F)$-martingale.

Proof. It will be seen below that the $(\bar{P}_\alpha, F)$-local martingale just mentioned is in fact the sum of two terms

$$E_\alpha \left\{ z_-(\vartheta, \bar{P}_\alpha) \log z_-(\vartheta, \bar{P}_\alpha) \cdot m\left(\vartheta, \bar{P}_\alpha\right) \right\} \tag{4.11}$$

and

$$\left(\frac{\bar{L}(Y, \bar{Y})}{\bar{Y}} - \frac{\bar{L}(1 - \hat{Y}, 1 - \hat{\bar{Y}})}{1 - \hat{\bar{Y}}}\right) * (\mu^X - \bar{\nu}). \tag{4.12}$$

It is not hard[1] to see that $I(\alpha|\alpha^{\cdot})$ is the sum of the three terms: the local martingale (4.11) plus the first term in (4.10) and the expression $\sum_{s \leq \cdot} E_{\alpha^{s-}} \ell(1 + \Delta m_s\left(\vartheta, \bar{P}_\alpha\right))$ with $\ell(x) = x \log x$. It is therefore sufficient to decompose the process in this third summand and to show that its martingale part is just (4.12), while the compensator may be identified with the last two terms in (4.10). To this end apply lemma 3.6 – substitute f in its assertion by ℓ to see that this compensator is given by $E_{\alpha^-} \ell(Y(\vartheta)/\bar{Y}) \cdot \bar{\nu} + \sum_{s \leq \cdot} E_{\alpha^{s-}} \ell\left((1 - \hat{Y}_s(\vartheta))/(1 - \hat{\bar{Y}}_s)\right)(1 - \hat{\bar{Y}}_s)$, equal indeed to the sum of the last two terms in (4.10). As for the martingale part, by the same lemma we get $E_{\alpha^-}\{\ell(Y(\vartheta)/\bar{Y}) - \ell\left((1 - \hat{Y}(\vartheta))/(1 - \hat{\bar{Y}})\right)\} * (\mu^X - \bar{\nu})$ that yields (4.12). The proof is complete. □

Remark 4.5. We obtain from theorem 4.4 that the expected utility from data at the stopping time T equals to

$$\begin{aligned} E_{\bar{P}_\alpha} I(\alpha|\alpha^T) =& E_\alpha I(\bar{P}_{\alpha,T}|P_{\vartheta,T}) \\ =& E_{\bar{P}_\alpha}\{\tfrac{1}{2}\bar{v}(\beta) \cdot \langle X^c \rangle_T + \bar{L}(Y, \bar{Y}) \cdot \nu_T + \sum_{s \leq T} \bar{L}(1 - \hat{Y}, 1 - \hat{\bar{Y}})\} \end{aligned}$$

The first identity is already known, see (4.6). The second one follows from (4.10).

5 Examples

5.1 Discrete observations

As confined to the special case of a discrete-time filtered space $(\Omega, \mathcal{F}, F = \{\mathcal{F}_n\}_{n \in \mathbb{N}})$, the present theory is quite straightforward. Let us therefore shortly review the results. Suppose that

[1] I guess

the present space is endowed with the family of probability measures $\{P_\theta\}_{\theta\in\Theta}$ that are all equivalent to a certain probability measure Q. Denote their restrictions to $\mathcal{F}_n$ by $\{P_{\theta,n}\}_{\theta\in\Theta}$ and Q_n. Often the n^{th} experiment is described by its outcomes, say vectors $(X_1,\dots,X_n)$ that generate the σ-algebra $\mathcal{F}_n$, and the above restrictions are viewed as their distributions. For each n and $\theta\in\Theta$ denote by $z_n(\theta,Q)$ the density of $P_{\theta,n}$ with respect to Q_n. The sequence of densities $\{z_n(\theta,Q)\}_{n\in\mathbb{N}}$ is related to the martingale sequence $\{m_n(\theta,Q)\}_{n\in\mathbb{N}}$ according to (2.2), i.e. $\Delta m_n(\theta,Q)=\Delta z_n(\theta,Q)/z_{n-1}(\theta,Q)$ with the convention $z_0(\theta,Q)\equiv 1$. Within this setup the condition (2.1) is equivalent to

$$\sum_{n=1}^{\infty} E_Q\left((\sqrt{1+\Delta m_n(\theta,Q)}-1)^2|\mathcal{F}_{n-1}\right)<\infty \quad P_\theta+Q \text{ a.s.} \tag{5.1}$$

(see [10, Theorem IV.2.36]). The arithmetic mean sequence $a(\alpha,Q)=\{a_n(\alpha,Q)\}_{n\in\mathbb{N}}$ is defined by $a_n(\alpha,Q)=E_\alpha z_n(\vartheta,Q)$. This is in fact the density (with respect to Q_n) of the restriction $\bar{P}_{\alpha,n}$ to $\mathcal{F}_n$ of the arithmetic mean measure, i.e. $a_n(\alpha,Q)=z_n(\bar{P}_\alpha,Q)$. The geometric mean sequence $g(\alpha,Q)=\{g_n(\alpha,Q)\}_{n\in\mathbb{N}}$ is defined by $g_n(\alpha,Q)=\prod_{i=1}^n\gamma_i(\alpha,Q)$, the product of the geometric means

$$\gamma_i(\alpha,Q)=e^{E_\alpha\log(1+\Delta m_i(\vartheta,Q))}=e^{E_\alpha\log\frac{z_i(\vartheta,Q)}{z_{i-1}(\vartheta,Q)}}. \tag{5.2}$$

The condition (2.3) is equivalent to

$$\sum_{n=1}^{\infty}E_\alpha E_Q\log(1+\Delta m_n(\vartheta,Q))>-\infty \tag{5.3}$$

since the sum on the left is identical to $-I(\mathbf{P}|\mathbf{Q})$. Compare this with condition (4.4) that now reads

$$\sum_{n=1}^{\infty}E_\alpha E_{P_\vartheta}\log(1+\Delta m_n(\vartheta,\bar{P}_\alpha))<\infty. \tag{5.4}$$

Obviously, the geometric mean sequence $g(\alpha,Q)$ has the multiplicative decomposition (3.22) in discrete time, with the Hellinger sequence of order α defined by

$$h_n(\alpha)=\sum_{i=1}^n E_Q\left\{1-\gamma_i(\alpha,Q)|\mathcal{F}_{i-1}\right\}.$$

Note that $E_Q h_\infty(\alpha)\le -E_Q\sum_{n=1}^\infty\log\gamma_n(\alpha,Q)<\infty$ by (5.3).
The density (with respect to Q_n) of the restriction $G_{\alpha,n}$ to $\mathcal{F}_n$ of the geometric mean measure G_α is

$$z_n(G_\alpha,Q)=\prod_{i=1}^n\frac{\gamma_i(\alpha,Q)}{E_Q\left\{\gamma_i(\alpha,Q)|\mathcal{F}_{i-1}\right\}}=\mathcal{E}_n(\frac{1}{1-\Delta h(\alpha)}\cdot N(\alpha,Q)) \tag{5.5}$$

where $N_n(\alpha,Q)=\sum_{i=1}^n\left(\gamma_i(\alpha,Q)-E_Q\left\{\gamma_i(\alpha,Q)|\mathcal{F}_{i-1}\right\}\right)$. Under the conditions (5.1) and (5.3) the geometric mean measure exists as a probability measure on each finite time interval. Indeed, we can apply a result in [11] that implies that every nonnegative discrete time local martingale

is in fact a martingale.
With the n^{th} experiment the posterior measure α^n is associated whose density with respect to the prior α is defined for each $\theta \in \Theta$ as follows

$$\frac{d\alpha^n}{d\alpha}(\theta) = \frac{z_n(\theta, Q)}{a_n(\theta, Q)} = z_n(\theta, \bar{P}_\alpha) = \prod_{i=1}^{n}(1 + \Delta m_i(\vartheta, \bar{P}_\alpha)),$$

cf (2.5). Then the Kullback-Leibler information in the posterior α^n with respect to the prior α is

$$I(\alpha^n|\alpha) = E_\alpha \log \frac{d\alpha}{d\alpha^n}(\vartheta) = -\sum_{i=1}^{n} \log \gamma_i(\alpha, \bar{P}_\alpha), \tag{5.6}$$

hence $E_{\bar{P}_\alpha} I(\alpha^n|\alpha) = -\sum_{i=1}^{n} E_{\bar{P}_\alpha} \log \gamma_i(\alpha, \bar{P}_\alpha)$. Note finally that in the present case the expected utility from data of size n (cf (4.9)) is

$$E_{\bar{P}_\alpha} I(\alpha|\alpha^n) = \sum_{i=1}^{n} E_{\bar{P}_\alpha} E_{\alpha^{i-1}} \ell(1 + \Delta m_i(\vartheta, \bar{P}_\alpha)), \tag{5.7}$$

that is well-defined, for condition (4.4) ensures the convergence of this series as $n \to \infty$.

Special case: Independent observations. Let $X_1, X_2, \ldots$ be a sequence of independent real-valued random variables. Suppose that under the measure P_θ for $\theta \in \Theta$ the random variable X_n possesses a probability density $f_n(\cdot, \theta)$ and under the measure Q a density $f_n(\cdot)$, all with respect to some σ-finite measure ρ. Then the condition (2.1) is equivalent to

$$\sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \left(\sqrt{f_n(x)} - \sqrt{f_n(x, \theta)}\right)^2 \rho(dx) < \infty \quad \forall \theta \in \Theta,$$

cf (5.1). Moreover, suppose $0 < \Gamma_n(\alpha) \doteq \int_{-\infty}^{\infty} \gamma_{\alpha,n}(x)\rho(dx) < 1$ for all $n \in \mathbb{N}$ where $\gamma_{\alpha,n} = \exp\{E_\alpha \log f_n(\cdot, \vartheta)\}$ (this is always less or equal 1 by Jensen's inequality but the equality is excluded by the assumption that ϑ is nondegenerate under α). The condition (2.3) is equivalent to

$$\sum_{n=1}^{\infty} \int_{-\infty}^{\infty} E_\alpha \log \left\{ \frac{f_n(x, \vartheta)}{f_n(x)} \right\} f_n(x)\rho(dx) > -\infty,$$

cf (5.3). In the present case the Hellinger integral and the Hellinger sequence are given by $H(\alpha, n) = \prod_{i=1}^{n} \Gamma_i(\alpha)$ and $h_n(\alpha) = \sum_{i=1}^{n}(1 - \Gamma_i(\alpha))$ with the relationship $H(\alpha, \cdot) = \mathcal{E}(-h(\alpha))$, since $h(\alpha)$ is deterministic, cf (2.9). For a sample of size n the posterior measure α^n is so that its density with respect to the prior α is

$$\frac{d\alpha^n}{d\alpha}(\theta) = f_1(X_1, \theta) \cdots f_n(X_n, \theta)/a_{\alpha,n}$$

where the denominator $a_{\alpha,n} = \int_\Theta f_1(X_1, \theta) \cdots f_n(X_n, \theta)\alpha(d\theta)$ is the density with respect to $\rho^{\otimes n}$ of the arithmetic mean measure $\bar{P}_\alpha$ restricted to $\mathcal{F}_n$. Note that the observations are

not independent relative to this measure and this causes considerable computational complications. For instance, the information in α^n given α amounts to $I(\alpha^n|\alpha) = \log\{a_{\alpha,n}/g_{\alpha,n}\}$ where $g_{\alpha,n} = \gamma_{\alpha,n}(X_1)\cdots\gamma_{\alpha,n}(X_n)$. For further calculations preceding formulas may be applied (for instance (5.6) and (5.7)) by taking into consideration that in the present example $1+\Delta m_n(\theta,\bar{P}_\alpha) = f_n(X_n,\theta)/a_{\alpha,n}$ and $\gamma_n(\alpha,\bar{P}_\alpha) = \gamma_{\alpha,n}(X_n)/a_{\alpha,n}$.
Calculations under the geometric mean measure G_α on the other hand are less cumbersome, since under this measure the X_n keep on being independent with densities with respect to ρ given by $\gamma_{\alpha,n}(\cdot)/\Gamma_n(\alpha)$. This statement is evident from (5.5).

5.2 Point processes

Suppose that observed is a d-dimensional counting process $(N^1,\dots,N^d)$. Under the probability measure P_θ for $\theta\in\Theta$ the cumulative intensity of the i^{th} component N^i is $\Lambda^i(\theta)$ and under the measure Q it is A^i, both positive increasing processes so that the densities $d\Lambda^i(\theta)/dA^i = Y^i(\theta)$ exist for all $i=1,\dots,d$ and $\theta\in\Theta$. The condition (2.1) is equivalent to

$$\sum_{i=1}^d \int_0^\infty \left(\sqrt{Y_s^i(\theta)}-1\right)^2 dA_s^i + \sum_{s\in\mathbb{R}} \left(\sqrt{1-\Delta\Lambda_s(\theta)}-\sqrt{1-\Delta A_s}\right)^2 < \infty$$

$P_\theta + Q$-a.s. for all $\theta\in\Theta$ (see [10, Theorem IV.2.1]). In the second term $\Lambda = \Lambda^1+\cdots+\Lambda^d$ and $A = A^1+\cdots+A^d$. The expression for the corresponding density process is well-known:

$$z_T(\theta,Q) = e^{-\Lambda_T(\theta)^c + A_T^c} \prod_{s\le T} \left(\tfrac{1-\Delta\Lambda_s(\theta)}{1-\Delta A_s}\right)^{1-\Delta N_s} \prod_{i=1}^d Y_s^i(\theta)^{\Delta N_s^i} \tag{5.8}$$

with $N = N^1+\cdots+N^d$. Moreover, assume that each density $Y^i(\theta)$ for $i=1,\dots,d$ satisfies $E_\alpha \log\{Y_s^i(\vartheta)/(1-\Delta\Lambda_s(\vartheta))\} > -\infty$ for all $s>0$. The Hellinger process of order α is given by

$$h(\alpha) = \sum_{i=1}^d \int_0^\cdot \phi_s(Y^i)dA_s^i + \sum_{s\le\cdot} \phi_s(1-\Delta\Lambda).$$

The condition (2.3) holds if the expression

$$\sum_{i=1}^d \int_0^T a_s(Y^i - 1 - \log Y^i)dA_s^i + \sum_{s\le T} a_s\left(\tfrac{1-\Delta\Lambda}{1-\Delta A} - 1 - \log\tfrac{1-\Delta\Lambda}{1-\Delta A}\right)(1-\Delta A_s) \tag{5.9}$$

evaluated at $T=\infty$, has a finite expectation with respect to Q.
The arithmetic mean measure $\bar{P}_\alpha$ assigns to the component N^i the intensity $\bar{\Lambda}^i$ that has the density $\bar{Y}^i$ with respect to A^i. This density is the predictable posterior expectation of $Y^i(\vartheta)$ as in (3.3).
The geometric mean measure G_α is a probability measure if $E_{\bar{P}_\alpha}\{1/\mathcal{E}_\infty(-h(\alpha))\} < \infty$, see proposition 2.7. According to (3.19) this measure assigns to the component N^i the intensity density (with respect to the same A^i) of the form $g(Y^i)/(1-\Delta h(\alpha))$.

Making use of the formula (2.5) in conjunction with (5.8) we get the density of the posterior α^T with respect to the prior α:

$$\frac{d\alpha^T}{d\alpha}(\theta) = e^{-\Lambda_T(\theta)^c + \bar{\Lambda}_T^c} \prod_{s \le T} \left(\frac{1 - \Delta\Lambda_s(\theta)}{1 - \Delta\bar{\Lambda}_s} \right)^{1-\Delta N_s} \prod_{i=1}^{d} \left(\frac{Y_s^i(\theta)}{\bar{Y}_s^i} \right)^{\Delta N_s^i},$$

that in turn yields the information $I(\alpha^T|\alpha)$. To get $E_{\bar{P}_\alpha} I(\alpha^T|\alpha)$, for instance, we have to take the expectation with respect to $\bar{P}_\alpha$ of the expression (5.9) with Y^i substituted by $Y^i/\bar{Y}^i$ and A^i by $\bar{\Lambda}^i$.

According to remark 4.5, the expected utility from the present data equals to

$$E_{\bar{P}_\alpha} I(\alpha|\alpha^T) = E_{\bar{P}_\alpha}\{\sum_{i=1}^{d} \int_0^T \bar{L}(Y_s^i, \bar{Y}_s^i) dA_s^i + \sum_{s \le T} \bar{L}(1 - \Delta\Lambda_s, 1 - \Delta\bar{\Lambda}_s)\}.$$

Finiteness of this expression for $T = \infty$ is just condition (4.4). The special case $d = 1$ with a continuous cumulative intensity process has been considered in [19]. In this special case the above expression for the expected utility from data reduces to equation (19.132) in [19]. The latter expression was derived in [19] for the Shannon information about a transmitted message ϑ that is contained in the received signal N. In that book, ϑ had been taken as a certain random process, a situation that is also covered in the present paper upon appropriate adjustments.

For the use of the arithmetic mean measure in model testing and for a discussion on computational problems see [2].

Special case: Cox' regression model. Assume that the compensators $\Lambda^i(\theta)$ and A^i are all absolutely continuous, so we can write $\Lambda^i(\theta) = \int_0^\cdot \lambda_s^i(\theta) ds$ and $A^i = \int_0^\cdot \lambda_s^i ds$. Then working under an exponential family of measures basically means that we deal with Cox' regression model, see [1, page 477]. So the intensity of the component N^i under P^θ is assumed to be modeled so that at each instant $t \ge 0$

$$\lambda_t^i(\theta) = \lambda_t^\circ e^{\theta^\top Z_t^i} \gamma_t^i$$

with θ a k-dimensional parameter vector, λ° a deterministic function called the *baseline hazard*, Z^i a certain k-dimensional predictable process of covariates and γ^i another (nonnegative) predictable processes served in survival analysis to model a censoring mechanism. Then we have $Y_t^i(\theta) = \lambda_t^\circ e^{\theta^\top Z_t^i} \gamma_t^i / \lambda_t^i$, which is clearly log-affine, so that indeed we deal with an exponential family.

The Hellinger process $h(\alpha)$ now takes the form

$$h(\alpha) = \int_0^\cdot \lambda_t^\circ \sum_{i=1}^{d} (E_\alpha e^{\vartheta^\top Z_t^i} - e^{E_\alpha \vartheta^\top Z_t^i}) \gamma_t^i dt.$$

Finally, observe that under the geometric mean measure G_α the compensators $\Lambda^{G,i}$ are given by $\Lambda^{G,i} = \int_0^\cdot \lambda_t^\circ e^{E_\alpha \vartheta^\top Z_t^i} \gamma_t^i dt$. The information processes can now be given a more specific form by making the appropriate substitutions.

5.3 Diffusion processes

Let the observed process X be defined so that under each measure $P_\theta, \theta \in \Theta$, the process $X - \int_0^\cdot \beta_s(\theta)ds$ is a Wiener process $W(\theta)$ with the intensity σ^2 that is free of the parameter θ. Then the condition (2.1) is equivalent to $\int_0^\infty \beta_s^2(\theta)ds < \infty \quad P_\theta + Q$-a.s. for all $\theta \in \Theta$ (see [10, Theorem IV.2.1]).

Suppose that at each instant $t > 0$ the drift $\beta_t(\vartheta)$ has non-vanishing variance with respect to α, denoted as above by $v(\beta_t)$. Then the Hellinger process is $h(\alpha) = \frac{\sigma^2}{2}\int_0^\cdot v(\beta_s)ds$ and the condition (2.3) is equivalent to $E_Q h_\infty(\alpha) < \infty$.

In the same vain it is easily seen that condition (4.4) in this context is satisfied if $E_\alpha E_{P_\vartheta}\int_0^\infty (\beta_s(\vartheta) - \bar{\beta}_s)^2 ds < \infty$. By applying theorem 4.4 we can rewrite this last condition as $E_{\bar{P}_\alpha}\int_0^\infty \bar{v}(\beta_s)ds$.

As we know from corollary 3.2, the arithmetic mean measure $\bar{P}_\alpha$ assigns to our observations the posterior characteristic $\bar{\beta}$, see (3.3), i.e. $\bar{W} \doteq X - \int_0^\cdot \bar{\beta}_s ds$ is a Wiener process. Assume now $E_{\bar{P}_\alpha}\exp\{\frac{\sigma^2}{2}\int_0^\infty v(\beta_s)ds\} < \infty$. Then the geometric mean measure G_α is a probability measure (see proposition 2.7) and under this measure $X - \int_0^\cdot a_s(\beta)ds$ is a Wiener process. Alternatively, under Novikov's condition $E_Q \exp\{\frac{\sigma^2}{2}\int_0^\infty a_s(\beta)^2 ds\} < \infty$, the measure G_α is a probability measure. A sufficient condition for this to hold is $E_\alpha E_Q \exp\{\frac{\sigma^2}{2}\int_0^\infty \beta_s(\vartheta)^2 ds\} < \infty$, which follows from Jensen's inequality. The latter condition is appealing as it says that the arithmetic mean of the $E_Q \exp\{\frac{\sigma^2}{2}\int_0^\infty \beta_s(\theta)^2 ds\}$ is finite. Finiteness of the latter expectation is just Novikov's condition for absolute continuity of P_θ with respect to Q, our basic condition (2.1).

According to corollary 3.13 the Hellinger processes $h(\alpha)$ are related to the Hellinger integrals evaluated at a certain stopping time T as follows: $H_T(\alpha) = E_{G_\alpha}\mathcal{E}_T(-h(\alpha)) = E_{G_\alpha}\exp\{-\frac{\sigma^2}{2}\int_0^T v(\beta_s)ds\}$. Combining (2.5) and corollary 3.2, we find that the density of the posterior α^T with respect to the prior α is given by

$$\frac{d\alpha^T}{d\alpha}(\theta) = \exp\{\int_0^T (\beta_s(\theta) - \bar{\beta}_s)d\bar{W}_s - \frac{\sigma^2}{2}\int_0^T (\beta_s(\theta) - \bar{\beta}_s)^2 ds\}.$$

Hence the information in α^T with respect to the prior α is according to (4.2) given by

$$I(\alpha^T|\alpha) = -\int_0^T (a(\beta_s) - \bar{\beta}_s)d\bar{W}_s + \frac{\sigma^2}{2}\int_0^T a((\beta_s - \bar{\beta}_s)^2)ds$$

and $E_{\bar{P}_\alpha} I(\alpha^T|\alpha) = \frac{\sigma^2}{2}E_{\bar{P}_\alpha}\int_0^T a((\beta_s - \bar{\beta}_s)^2)ds$. Finally, by Theorem 4.4 the expected utility from the data equals to $E_{\bar{P}_\alpha} I(\alpha|\alpha^T) = \frac{\sigma^2}{2}E_{\bar{P}_\alpha}\int_0^T \bar{v}(\beta_s)ds$. For related results see also [22] and [19]. In fact, equation (16.65) of [19] is nothing else but $E_{\bar{P}_\alpha} I(\alpha|\alpha^T)$, although the context is different. In [19] this formula has been derived for the Shannon information in the received signal X about the transmitted signal ϑ. As in the counting process example of the previous section, also this interpretation is covered in our general set up.

Assume now that the family $\{P_\theta\}_{\theta\in\theta}$ is exponential. Then β is affine, $\beta(\theta) = a + \theta^\top \dot{\beta}$, say. It follows that $v_s(\beta) = \dot{\beta}_s^\top \Gamma_\alpha \dot{\beta}_s$, where Γ_α, the covariance matrix of ϑ under α, is assumed to be finite.

Hence the Hellinger process becomes $\frac{\sigma^2}{2}\int_0^\cdot \dot{\beta}_s^\top \Gamma_\alpha \dot{\beta}_s ds$. Under the measure G_α the process X

has the compensator $\int_0^{\cdot}(a_s + E_\alpha \vartheta^\top \dot\beta_s)ds$. By making the relevant substitutions the information processes can now be given a more specific form.

5.4 Fractional processes

It is said that X is a fractional Brownian motion with self-similarity index $H \in (0,1)$ if it is a continuous centered Gaussian process with $X_0 = 0$ and with the covariance

$$EX_tX_s = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right)$$

at $s,t \geq 0$. For $H \neq \frac{1}{2}$ fractional Brownian motion is not a semimartingale and for $H = \frac{1}{2}$ it is the standard Brownian motion. H is also called Hurst index.
Denote by c_H the constant $\sqrt{2H\Gamma(\frac{3}{2}-H)/\Gamma(H+\frac{1}{2})\Gamma(2-2H)}$, where Γ is the gamma function and let $\sigma_H^2 = c_H^2/4H^2(2-2H)$. The following facts are taken from [20]:

Theorem 5.1. *Under the conditions of the present section we have*
(i) The process M defined by $M_t = \int_0^t m(t,s)dX_s$ is a continuous Gaussian martingale with independent increments, where at each instant $t > 0$ the kernel $m(t,s)$ is non-zero only if $s \in (0,t)$, when it equals to $s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H}/2HB(\frac{3}{2}-H, H+\frac{1}{2})$ with $B(a,b)$ the beta coefficient. Furthermore the quadratic variation of M is given by $\langle M\rangle_t = \sigma_H^2 t^{2-2H}$.
(ii) The process X defined by $X_t = \int_0^t z(t,s)dM_s$ is a fractional Brownian motion with self-similarity index H, where at each instant $t > 0$ the kernel $z(t,s)$ is non-zero only if $s \in (0,t)$, when it equals to $2Ht^{H-\frac{1}{2}}(t-s)^{H-\frac{1}{2}} - H(2H-1)\int_s^t u^{H-\frac{1}{2}}(u-s)^{H-\frac{3}{2}}du$.

Proof. See [20, Theorem 3.1 and Theorem 5.2]. □

The integrals of the kernels $z(\cdot,\cdot)$ and $m(\cdot,\cdot)$ with respect to M and X, respectively, is defined by integration by parts. Since the kernels are non-random, we have the identity $F^X = F^M$ between the basic filtration F^X generated by the observed fractional Brownian motion X from one hand and the filtration F^M generated by the Gaussian martingale M of theorem 5.1, assertion (i), on the other hand (we refer to [20] for more details).
Consider the following parametric model. Take Q to be the probability measure that makes X a fractional Brownian motion with self-similarity index H. Suppose that under the probability measure P_θ with $\theta \in \Theta$ the process $X(\theta) = X - \int_0^{\cdot}\beta_s(\theta)ds$ for some progressive process $\beta(\theta)$ is a fractional Brownian motion with self-similarity index H. Then the process $M(\theta) = \int_0^{\cdot} m(\cdot,s)dX_s(\theta)$ is a (P_θ, F)-Gaussian martingale with the same quadratic variation process as M, so that $\langle M(\theta)\rangle_t = \sigma_H^2 t^{2-2H}$.
Since the measures P_θ and Q are completely determined by the characteristics of the corresponding Gaussian martingales $M(\theta)$ and M, the change of measure is accomplished by an ordinary Girsanov transformation as in the diffusion case. So, we have a density process $z(\theta, Q) = \mathcal{E}(\rho(\theta)\cdot M)$, where the process $\rho(\theta)$ is such that $M(\theta) = M - \int_0^{\cdot}\rho_s(\theta)\,d\langle M\rangle_s$. But in view of theorem 5.1 we must have $\int_0^{\cdot}\rho_s(\theta)d\langle M\rangle_s = \int_0^{\cdot} m(\cdot,s)\beta_s(\theta)\,ds$. Therefore $\rho(\theta)$ satisfies the integral equation

$$\int_0^{\cdot} s^{\frac{1}{2}-H}(\cdot - s)^{\frac{1}{2}-H}\beta_s(\theta)ds = (2-2H)B(\tfrac{3}{2}-H, \tfrac{3}{2}-H)\int_0^{\cdot}\rho_s(\theta)s^{1-2H}ds. \tag{5.10}$$

Suppose that the solution ρ to equation (5.10) is so that $\int_0^\infty \rho_s^2(\theta) d\langle M\rangle_s < \infty$ $P_\theta + Q$-a.s for all $\theta \in \Theta$. This condition is equivalent to (2.1), see [10, Theorem IV.2.1]).
Switching to $\bar{P}_\alpha$ as the dominating measure, we likewise obtain

$$z(\theta, \bar{P}_\alpha) = \mathcal{E}((\rho(\theta) - \bar{\rho}) \cdot \bar{M}), \tag{5.11}$$

where $\bar{M} = \int_0^\cdot m(\cdot, s) dX_s - \int_0^\cdot \bar{\rho}_s d\langle M\rangle_s$ is a $(\bar{P}_\alpha, F)$-Gaussian martingale with angle bracket $\langle \bar{M}\rangle = \langle M\rangle$. Moreover, the Hellinger process of order α is similarly obtained $h(\alpha) = \frac{1}{2}\int_0^\cdot v(\rho_s) d\langle M\rangle_s$, provided that the variance process $v(\rho)$ is non-vanishing. These formulas follow directly from the corresponding formulas of the diffusion model. Note by the way that the process $X - \int_0^\cdot \bar{\beta}_s ds$ is $(\bar{P}_\alpha, F)$-fractional Brownian motion with Hurst index H.
The condition (2.3) is equivalent to $E_Q h_\infty(\alpha) < \infty$. Similarly, condition (4.4) is equivalent to $E_\alpha \int_0^\infty \rho_s(\vartheta)^2 d\langle M\rangle_s < \infty$. If, moreover, $E_{\bar{P}_\alpha} \exp\{h_\infty(\alpha)\} < \infty$, then the geometric mean measure G_α is a probability measure in view of proposition 2.7.
According to corollary 3.13 the Hellinger integral of order α is evaluated at a stopping time T as follows: $H(\alpha, T) = E_{G_\alpha} \exp\{-\frac{1}{2}\int_0^T v(\rho_s) d\langle M\rangle_s\}$. In virtue of (2.5), the equation (5.11) gives the density of the posterior α^T with respect to the prior α. We get, in particular, that $E_{\bar{P}_\alpha} I(\alpha^T|\alpha) = \frac{1}{2} E_{\bar{P}_\alpha} \int_0^T a\left((\rho_s - \bar{\rho}_s)^2\right) d\langle M\rangle_s$. Finally, by Theorem 4.4 the expected utility from data equals to $E_{\bar{P}_\alpha} I(\alpha|\alpha^T) = \frac{1}{2} E_{\bar{P}_\alpha} \int_0^T \bar{v}(\rho_s) d\langle M\rangle_s$. As a last remark we note that the family $\{P_\theta\}_{\theta\in\Theta}$ becomes exponential if the processes $\beta(\theta)$ above are affine functions of θ. It follows from equation (5.10) that then also the processes $\rho(\theta)$ are affine functions of θ, so that the density processes $z(\theta, Q)$ are of the desired form.

References

[1] P.K. Andersen, O. Borgan, R.D. Gill and N. Keiding (1993), *Statistical Models based on Counting Processes*, Springer.

[2] E. Arjas and D. Gasbarra (1997), On prequential model assessment in life history analysis, *Biometrika* **84**, 505-522.

[3] J.M. Bernardo and A.F.M. Smith (1994), *Bayesian Theory*, Wiley.

[4] P. Dupuis and R.S. Ellis (1997), *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley.

[5] K. Dzhaparidze, P.J.C. Spreij and E. Valkeila (1997), On Hellinger Processes for Parametric Families of Experiments. *Statistics and Control of Stochastic Processes, The Liptser Festschrift* (Yu. Kabanov et al., eds.). World Scientific, Singapore, 41-62.

[6] K. Dzhaparidze, P.J.C. Spreij and E. Valkeila (1999), Hellinger and information processes in filtered experiments, Part I: general concepts.

[7] B. Grigelionis (1990), Hellinger integrals and Hellinger processes for solutions of martingale problems, *Prob. Theory and Math. Stat.* **1**, 446-454 (B. Grigelionis et al. Eds.), VSP/Mokslas.

[8] J. Jacod (1989), Filtered statistical models and Hellinger processes, *Stochastic Processes and their Appl.* **32**, 3-45.

[9] J. Jacod (1989), Convergence of filtered statistical models and Hellinger processes, *Stochastic Processes and their Appl.* **32**, 47-68.

[10] J. Jacod and A.N. Shiryaev (1987), *Limit Theorems for Stochastic Processes*, Springer.

[11] J. Jacod and A.N. Shiryaev (1998), Local martingales and the fundamental asset pricing theorems in the discrete-time case, *Finance and Stochastics* **2**, no. 3, 259-273. Yu. Kabanov, R.Sh. Liptser and A.N. Shiryayev (1986), On the variational distance for probability measures defined on a filtered space, *Prob. Theory Rel. Fields* **71**, 19-36.

[12] Yu. Kabanov, R.Sh. Liptser and A.N. Shiryayev (1979, 1980), Absolute continuity and singularity of locally absolutely continuous probability distributions, *Math. USSR Sbornik* **35**, 631-680 (Part I), 36, 31-58 (Part II).

[13] Yu. Kabanov, R.Sh. Liptser and A.N. Shiryayev (1986), On the variational distance for probability measures defined on a filtered space, *Prob. Theory Rel. Fields* **71**, 19-36.

[14] E.I. Kolomiets (1984), Relations between triplets of local characteristics of semimartingales, *Russian Math. Surveys* **39**, no. 4, 123-124.

[15] U. Küchler and M. Sørensen (1997), *Exponential families of stochastic processes*, Springer.

[16] S. Kullback (1959), *Information Theory and Statistics*, Wiley.

[17] R.Sh. Liptser and A.N. Shiryayev (1983), On the problem of "predictable" criteria of contiguity, *Proc. 5th Japan-USSR Symp.*, Lecture Notes in Mathematics 1021, 384-418, Springer.

[18] R.Sh. Liptser and A.N. Shiryayev (1986), *Theory of Martingales*, Kluwer.

[19] R.S. Liptser and A.N. Shiryaev (1978), *Statistics of Random Processes II*, Springer.

[20] I. Norros, E. Valkeila and J. Virtamo (1999), An elementary approach to a Girsanov formula and other analytic results on fractional Brownian motions, *Bernoulli* **5**, 571-587.

[21] G. Pistone and C. Sempi (1995), An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Annals of Statistics* **23**, no. 5, 1543-1561.

[22] N.G. Polson and G.O. Roberts (1993), A utility based approach to information for stochastic differential equations, *Stochastic Processes and their Applications* **48**, 341-356.

[23] B.L.S. Prakasa Rao (1999), *Semimartingales and their Statistical Inference*, Chapmann & Hall.

[24] N.N. Vakhania, V.I. Tarieladze and S.A. Chobanyan (1987), *Probability Distributions on Banach Spaces*, Reidel.