Information processes in filtered experiments
Part I: general concepts

Kacha Dzhaparidze[1], Peter Spreij[2] and Esko Valkeila[3]

January 31, 2001

Abstract

In this paper we define randomized filtered experiments with an abstract parameter space and analyze some properties of this concept. To that end we relate with the parametric family of density processes the arithmetic and geometric mean processes. The latter will be naturally linked to a generalized Hellinger process and Hellinger integrals. We also introduce and analyze the Kullback-Leibler information processes of a posterior distribution on the parameter space with respect to a prior (and vice versa) and give certain characterizations to their development in terms of the concepts given above.

keywords Kullback-Leibler information, relative entropy, information from data, expected utility from data, Hellinger processes and Hellinger integrals, posterior and prior distributions.

AMS-classification 60G07, 60H30, 62B15, 94A17

[1]Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam
[2]Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam
[3]Department of Mathematics, University of Helsinki, P.O. Box 4, 00014 Helsinki

1 Introduction

In this paper we study the Kullback-Leibler information between a posterior and a prior distribution on an abstract parameter space. Here we model the observations by a filtration and in our companion paper [8], referred throughout as Part II, we specialize this to observations in terms of semimartingales or of certain fractional processes.
For a proper analysis it is necessary to define arithmetic and geometric mean measures, (the latter generalizations of a probability measure introduced by Grigelionis in [9]) as well as the notions of Hellinger integral and Hellinger process with respect to an arbitrary, not necessarily finite, family of probability measures. To understand the dynamical behaviour of the Kullback-Leibler information we give a representation of this information in terms of the likelihood ratio between the arithmetic mean and the geometric mean measures and Hellinger process.
The study of Hellinger integrals and Hellinger processes started for binary experiments in the series of papers [15], [16] and [18]. This theory took a complete form in the book [13] where the notions of Hellinger integrals and Hellinger processes were fully exploited. In the consequent papers [11] and [12] some of the results were generalized to a filtered experiment with a finite number of probability measures. In [9] some additional aspects of the latter experiment are discussed (similar to results in section 5.5 below). These results were extended to an arbitrary parameter space in [7]. 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 [7].
The paper is organized as follows. After the introduction we treat in section 2 randomized statistical experiments in a static setting and introduce the notions of the information in the posterior given a prior, as well as the notion of relative entropy in the posterior given a prior. In section 3 we switch to a dynamic setting and in the sections 4 and 5 discuss a dynamic family of arithmetic and geometric mean measures. Finally, in section 6 we treat the dynamics of the information and entropy processes in terms of the previously defined concepts.
In this Part I we only present the results in a general form. For explicit representations in terms of observations that can be modeled by a semimartingale (involving the triplets of predictable characteristics) and examples, that also include fractional processes, we refer to Part II.

2 Randomized experiments

2.1 Statistical experiment

We consider a *statistical experiment* $(\Omega, \mathcal{F}, \{P_\theta\}_{\theta\in\Theta})$, where $\{P_\theta\}_{\theta\in\Theta}$ is a certain parametric family of probability measures defined on a measurable space $(\Omega, \mathcal{F})$ with a set of elementary events Ω and a σ-field $\mathcal{F}$. We suppose that each member of the family $\{P_\theta\}_{\theta\in\Theta}$ is equivalent to a certain probability measure Q, i.e.

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

and for each fixed $\theta \in \Theta$ we denote by p_θ the Radon-Nikodym derivative of P_θ with respect to Q:

$$p_\theta = \frac{dP_\theta}{dQ}. \tag{2.2}$$

So, for each $\theta \in \Theta$ and $B \in \mathcal{F}$

$$P_\theta(B) = \int_B p_\theta(\omega)Q(d\omega) = E_Q\{1_B p_\theta\}. \tag{2.3}$$

Here and elsewhere below we use the expectation sign E indexed by a probability measure. When we want to express dependence of p_θ on the dominating measure Q, we also write $p(\theta, Q)$.

2.2 Randomization

On the set of parameter values Θ define a σ-field $\mathcal{A}$ and consider a probability space $(\Theta, \mathcal{A}, \alpha)$ where α is a certain probability measure. In this way a statistical parameter ϑ is viewed as a random variable (on a possibly different probability space) with values in $(\Theta, \mathcal{A})$. The probability measure α determines the distribution of ϑ.
Consider now the direct product $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{Q})$ of two probability spaces $(\Omega, \mathcal{F}, Q)$ and $(\Theta, \mathcal{A}, \alpha)$, where $\boldsymbol{\Omega} = \Omega\times\Theta$, $\boldsymbol{\mathcal{F}} = \mathcal{F}\otimes\mathcal{A}$ and $\mathbf{Q} = Q\times\alpha$. Along with $\mathbf{Q}$ define on $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}})$ another probability measure $\mathbf{P}$ as follows: for each $\boldsymbol{B} \in \boldsymbol{\mathcal{F}}$

$$\mathbf{P}(\boldsymbol{B}) = \int_{\mathbf{B}} p(\omega,\theta)Q(d\omega)\alpha(d\theta) \doteq E_{\mathbf{Q}}\{1_{\mathbf{B}}p\} \tag{2.4}$$

so that for each $\boldsymbol{\omega} = (\omega, \theta) \in \boldsymbol{\Omega}$ we have $p(\boldsymbol{\omega}) = \frac{d\mathbf{P}}{d\mathbf{Q}}(\boldsymbol{\omega})$. Obviously,

$$p(\boldsymbol{\omega}) = \frac{d\mathbf{P}}{d\mathbf{Q}}(\boldsymbol{\omega}) = \frac{dP_\theta}{dQ}(\omega) = p_\theta(\omega); \tag{2.5}$$

cf (2.2).
Observe that in the present setting the probability measure P_θ defined for each $\theta \in \Theta$ by (2.3) (and satisfying $P_\theta(\Omega) = 1$), can be viewed as a regular conditional probability measure, under the condition that the statistical parameter ϑ takes on the particular value θ. In view of (2.3) we can rewrite (2.4) as follows: for each $\boldsymbol{B} = B \times A \in \boldsymbol{\mathcal{F}}$

$$\mathbf{P}(\boldsymbol{B}) = \int_A p_\theta(B)\alpha(d\theta) = E_\alpha\left\{1_A E_Q\{1_B p\}\right\} = E_Q\left\{1_B E_\alpha\{1_A p\}\right\},$$

since by Loève [20], theorem 8.2B, it is allowed to interchange the integration order.
Throughout the paper the Kullback-Leibler information in $\mathbf{P}$ given $\mathbf{Q}$ is defined by (we follow the notation of [3])

$$I(\mathbf{P}|\mathbf{Q}) = E_{\mathbf{Q}} \log\frac{d\mathbf{Q}}{d\mathbf{P}} = -E_{\mathbf{Q}}\log p \tag{2.6}$$

with the density p as in (2.5). Indeed, in view of our assumptions in section 2.1 this quantity is well-defined, since the measures $\mathbf{P}$ and $\mathbf{Q}$ are equivalent and $I(\mathbf{P}|\mathbf{Q}) = -E_{\mathbf{Q}} \log\{d\mathbf{P}/d\mathbf{Q}\}$. In order to avoid trivialities this information is assumed positive. Often we will also assume that it is finite, i.e.

$$0 < I(\mathbf{P}|\mathbf{Q}) < \infty. \tag{2.7}$$

Notice that we can alternatively write $I(\mathbf{P}|\mathbf{Q}) = E_\alpha I(P_\vartheta|Q)$, showing that we deal with an average information in the experiment given a dominating measure Q.
Already in the next section will also come across the quantity

$$I(\mathbf{Q}|\mathbf{P}) = E_{\mathbf{P}} \log \frac{d\mathbf{P}}{d\mathbf{Q}} = E_{\mathbf{Q}}\, p \log p \tag{2.8}$$

with the density p as in (2.5), cf proposition 2.2. This Kullback-Leibler information in $\mathbf{Q}$ given $\mathbf{P}$ equals to $I(\mathbf{Q}|\mathbf{P}) = E_\alpha I(Q|P_\vartheta)$. Using the terminology of the theory of large deviations we may characterize this quantity as *the average relative entropy in the experiment given a dominating measure* Q (cf e.g. [6], section 1.4; for a different, statistical context, see e.g. [17]). Instead of (2.7) we will sometimes assume that the average relative entropy is positive and finite, i.e.

$$0 < I(\mathbf{Q}|\mathbf{P}) < \infty. \tag{2.9}$$

2.3 Arithmetic mean measure

Usually, one intends to get statements about the experiment that are independent from the choice of a dominating measure Q. But even in that case it is often handy from a technical point of view to make specific computations under a special choice of a dominating measure Q. Following [13] and [3], we will make two different choices, called *the arithmetic mean measure* and *the geodesic measure*, respectively. The motivation will be their extremal properties proved in propositions 2.2 and 2.5.
First, we define the arithmetic mean measure as follows. Consider once more a statistical experiment $(\Omega, \mathcal{F}, \{P_\theta\}_{\theta\in\Theta}, Q)$. With the family of probability measures $\{P_\theta\}_{\theta\in\Theta}$ we associate a new measure defined on the same measurable space $(\Omega, \mathcal{F})$, the so-called *arithmetic mean measure* $\bar{P} = \bar{P}_\alpha$: for each $B \in \mathcal{F}$

$$\bar{P}_\alpha(B) = \mathbf{P}(B \times \Theta) = E_\alpha P_\vartheta(B). \tag{2.10}$$

Let us also introduce the notation $a(\alpha, Q) = E_\alpha\{p_\vartheta\}$ for the arithmetic mean with respect to α of the density p_θ. 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 (2.10) 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 $\bar{P} \ll Q$. □

Dealing with the present statistical experiment one may wish to select a special dominating measure that is most neutral with respect to the given family $\{P_\theta\}_{\theta\in\Theta}$ in the following sense. One may choose for the probability measure that possesses the following extremal property: it minimizes $E_\alpha I(Q|P_\vartheta) = I(\mathbf{Q}|\mathbf{P})$ among all dominating measures Q, which is exactly the average relative entropy in the experiment given a dominating measure Q of (2.8). The arithmetic mean measure achieves this extremal property as we have the following

Proposition 2.2. *For all dominating measures Q*

$$I(\mathbf{Q}|\mathbf{P}) \geq I(\bar{\mathbf{P}}_\alpha|\mathbf{P})$$

where $\bar{\mathbf{P}}_\alpha$ is the product measure $\bar{P}_\alpha \times \alpha$. The equality is attained if and only if $Q = \bar{P}_\alpha$.

Proof. Since $I(\mathbf{Q}|\mathbf{P}) = I(Q|\bar{P}) + I(\bar{\mathbf{P}}|\mathbf{P})$ the assertion is straightforward. □

Remark 2.3. The relation $I(\mathbf{Q}|\mathbf{P}) = I(Q|\bar{P}) + I(\bar{\mathbf{P}}|\mathbf{P})$ is of the same type as a so-called nonsymmetric Pythagorean theorem for orthoprojections obtained in [3, chapter 22].

2.4 Geodesic measure

Now we turn to the geodesic measure, another candidate to serve as a dominating measure. It is denoted by $\breve{C}_\alpha$ and defined for all $B \in \mathcal{F}$ as follows:

$$\breve{C}_\alpha(B) = \frac{\int_B e^{E_\alpha \log p_\vartheta(\omega)} Q(d\omega)}{\int_\Omega e^{E_\alpha \log p_\vartheta(\omega)} Q(d\omega)}. \tag{2.11}$$

Note that the integrand $g(\alpha, Q) = \exp\{E_\alpha \log p_\vartheta\}$ is the geometric mean of the density p_θ with respect to α. In view of Jensen's inequality (applied to the concave function log), we see that $g(\alpha, Q) \leq a(\alpha, Q)$. In the denominator we have $H(\alpha) = E_Q g(\alpha, Q)$ called the *Hellinger integral* of order α, in analogy with [13], section IV.1, formula 1.6. It is now immediate that $H(\alpha) \leq 1$. Note that the Hellinger integral is independent of the choice of the dominating measure Q: if Q' is another dominating measure such that $Q \ll Q'$ and $Z = dQ/dQ'$, then $E_Q g(\alpha, Q) = E_{Q'} g(\alpha, Q')$, since $E_Q g(\alpha, Q) = E_{Q'}\{Z\ g(\alpha, Q)\}$ and by definition of $g(\alpha, Q)$

$$Z\ g(\alpha, Q) = e^{E_\alpha \log\{Z\ p(\vartheta, Q)\}} = e^{E_\alpha \log p(\vartheta, Q')} = g(\alpha, Q'). \tag{2.12}$$

Let then Q and Q_0 be two dominating measures and $Q' = \frac{1}{2}(Q + Q_0)$. A double application of the above result gives $E_Q g(\alpha, Q) = E_{Q'} g(\alpha, Q') = E_{Q_0} g(\alpha, Q_0)$, which establishes the postulated independence of the choice of the dominating measure.
Recall lemma 2.1. The arithmetic mean measure was characterized by having $a(\alpha, Q)$,

the arithmetic mean of the p_θ as its density with respect to Q. Now we are dealing with the measure $\breve{C}_\alpha$ that is absolutely continuous with respect to Q, having the density

$$\frac{d\breve{C}_\alpha}{dQ} = \frac{g(\alpha, Q)}{E_Q g(\alpha, Q)}. \tag{2.13}$$

Moreover, we have

Lemma 2.4. *Assume (2.7). Then the measure $\breve{C}_\alpha$ is equivalent to Q.*

Proof. We only have to prove $Q \ll \breve{C}_\alpha$. From (2.7) and the fact that $\mathbf{P} \sim \mathbf{Q}$ we get that $E_Q E_\alpha \log p_\vartheta > -\infty$. Hence $E_\alpha \log p_\vartheta > -\infty$ with Q-probability one. This is equivalent to $Q(g(\alpha, Q) > 0) = 1$, whence $Q \ll \breve{C}_\alpha$. □

Like the arithmetic mean measure, also the geodesic measure has an extremal property:

Proposition 2.5. *The following relation holds*

$$I(\mathbf{P}|\mathbf{Q}) = I(\breve{C}_\alpha|Q) - \log H(\alpha). \tag{2.14}$$

Moreover, among the dominating measures Q the measure $\breve{C}_\alpha$ is such that $I(\mathbf{P}|\mathbf{Q})$ is minimized with minimum value $-\log H(\alpha)$.

Proof. First, since $\breve{C}_\alpha \sim Q$, we observe that

$$I(\breve{C}_\alpha|Q) = -E_Q \log \frac{d\breve{C}_\alpha}{dQ} = -E_Q \log g(\alpha, Q) + \log H(\alpha)$$

and then $I(\mathbf{P}|\mathbf{Q}) = -E_Q E_\alpha \log p_\vartheta = -E_Q \log g(\alpha, Q)$. Equation (2.14) now follows. Since $I(\breve{C}_\alpha|Q) \geq 0$ we have $I(\mathbf{P}|\mathbf{Q}) \geq -\log H(\alpha)$. Because $H(\alpha)$ is independent of Q, we also have $\inf_Q I(\mathbf{P}|\mathbf{Q}) \geq -\log H(\alpha)$. Clearly we get equality for $Q = \breve{C}_\alpha$. □

Remark 2.6. Equation (2.14) is equivalent to $I(\mathbf{P}|\mathbf{Q}) = I(\breve{C}_\alpha|Q) + I(\mathbf{P}|\breve{C}_\alpha \times \alpha)$. This is another analogue of Pythagoras' theorem. Similar results are obtained in [4] for I-projections.

2.5 Prior and posterior measures

Let us turn for awhile back to the probability measure α that we have defined on the parametric space $(\Theta, \mathcal{A})$ at the beginning of section 2.2. In a Bayesian set up it is called *a priori* probability measure. Along with this one usually defines on the same space the *a posteriori* probability measure β by the following Bayes formula: for all $A \in \mathcal{A}$

$$\beta(A) \doteq \frac{\int_A p(\theta, Q)\alpha(d\theta)}{\int_\Theta p(\theta, Q)\alpha(d\theta)}. \tag{2.15}$$

In other words

$$\frac{d\beta}{d\alpha}(\theta) = \frac{p(\theta, Q)}{\int_\Theta p(\theta, Q)\alpha(d\theta)} \tag{2.16}$$

for each $\theta \in \Theta$. Obviously, the posterior β so defined is free of the choice of a dominating measure Q.
In section 6 we will be interested in the Kullback-Leibler information in the posterior β given the prior α that in virtue of the identity (2.16) is

$$I(\beta|\alpha) = E_\alpha \log \frac{d\alpha}{d\beta}(\vartheta) = \log E_\alpha p(\vartheta, Q) - E_\alpha \log p(\vartheta, Q). \tag{2.17}$$

Note that in terms of section 2.3 the identities (2.16) and (2.17) may be written as follows:

$$\frac{d\beta}{d\alpha}(\theta) = \frac{p(\theta, Q)}{a(\alpha, Q)} = p(\theta, \bar{P}_\alpha) \tag{2.18}$$

and

$$e^{-I(\beta|\alpha)} = g(\alpha, \bar{P}_\alpha). \tag{2.19}$$

The denominator in (2.16) is indeed $a(\alpha, Q)$ so that (2.18) follows from (2.2) and lemma 2.1. The equation (2.19) is obvious, since on the right of (2.17) we have $E_\alpha \log\{(a(\alpha, Q)/p(\vartheta, Q)\} = -E_\alpha \log p(\vartheta, \bar{P}_\alpha)$.
The $\bar{P}_\alpha$-expectation of the Kullback-Leibler information in β given α is then easily seen to be equal to $E_\alpha I(P_\vartheta|\bar{P}_\alpha)$, which in turn is nothing else than $I(\mathbf{P}|\bar{\mathbf{P}}_\alpha)$.
We will also treat the relative entropy of the posterior given a prior that by definition amounts to

$$I(\alpha|\beta) = E_\beta \log \frac{d\beta}{d\alpha}(\vartheta) = E_\alpha p(\vartheta, \bar{P}_\alpha) \log p(\vartheta, \bar{P}_\alpha). \tag{2.20}$$

In Bayesian statistics this quantity is called *information from data* (see [1, Definition 2.26, p.78]). Its expectation $E_{\bar{P}_\alpha} I(\alpha|\beta)$ is an especially important notion. It is called *expected utility from data*. By (2.20) we get the following representation:

$$E_{\bar{P}_\alpha} I(\alpha|\beta) = E_\alpha E_{\bar{P}_\alpha} p(\vartheta, \bar{P}_\alpha) \log p(\vartheta, \bar{P}_\alpha) = E_\alpha I(\bar{P}_\alpha|P_\vartheta) = I(\bar{\mathbf{P}}_\alpha|\mathbf{P}). \tag{2.21}$$

Hence the expected utility from data is finite under assumption (2.9) in view of proposition 2.2.

3 Randomized filtered experiment

Let the measurable space $(\Omega, \mathcal{F})$ be equipped with a filtration $F = \{\mathcal{F}_t\}_{t\geq 0}$, an increasing and right continuous flow of sub-σ-fields of $\mathcal{F}$, so that $\bigvee_{t\geq 0} \mathcal{F}_t = \mathcal{F}_\infty = \mathcal{F}$. Assume that the filtered probability space $(\Omega, \mathcal{F}, F = \{\mathcal{F}_t\}_{t\geq 0}, Q)$ is a stochastic basis: $\mathcal{F}$ is Q-complete and each $\mathcal{F}_t$ contains the Q-null sets of $\mathcal{F}$. We also assume for simplicity that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ Q-a.s. The filtered probability space

$$(\Omega, \mathcal{F}, F, \{P_\theta\}_{\theta\in\Theta}, Q)$$

so defined is called a *filtered statistical experiment.*
Consider now the optional projections of the probability measures Q and P_θ with respect to F, and use the same symbols for resulting optional valued processes: for a F-stopping time T Q_T and $P_{\theta,T}$ are then the restrictions of the measures Q and P_θ to the sub-σ-field $\mathcal{F}_T$. Since $P_{\theta,T}$ is equivalent to Q_T for each $\theta \in \Theta$, we can define the Radon-Nikodym derivatives

$$z_T(\theta) = \frac{dP_{\theta,T}}{dQ_T} = E_Q\{p_\theta|\mathcal{F}_T\}.$$

Thus according to [13], section III.3, for each fixed $\theta \in \Theta$ there is a unique (up to Q-indistinguishability) process $z(\theta) = z(\theta, Q)$ called the *density process* (we usually stress the dependence on a dominating measure Q), so that $z_t(\theta, Q) = dP_{\theta,t}/dQ_t$ for all $t \geq 0$, which possesses the following properties (see [13], 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 $t \in [0, \infty]$.

Later on in this paper we will also encounter various Kullback-Leibler information numbers like $I(P_{\theta,T}|Q_T)$, next to the ones that have been previously introduced.

4 Arithmetic mean process and arithmetic mean measure

4.1 Arithmetic mean process

In section 2.3 we have associated with the parametric family of densities $\{p(\theta, Q)\}_{\theta\in\Theta}$ the arithmetic mean $a(\alpha, Q)$ with respect to the prior α. Likewise, with the parametric family of density processes $\{z(\theta, Q)\}_{\theta\in\Theta}$ the process

$$a(\alpha, Q) = E_\alpha z(\vartheta, Q) \tag{4.1}$$

is associated, that is called the *arithmetic mean process.*
In view of the definition (4.1) and the identity of lemma 2.1 the a-mean process can also be defined as the density process of $\bar{P}_\alpha$ with respect to Q: with the notation like in section 3

$$a(\alpha, Q) = z(\bar{P}_\alpha, Q) \tag{4.2}$$

where $z_t(\bar{P}_\alpha, Q) = E_Q\{\frac{d\bar{P}_\alpha}{dQ}|\mathcal{F}_t\}$ for all $t \geq 0$. Therefore with the choice $\bar{P}_\alpha$ as the dominating measure it becomes particularly simple: identically $a(\alpha, \bar{P}_\alpha) = 1$.
Parallel to the above statements (i) - (iii) on the density processes at the end of section 3, the following properties of the arithmetic mean process can be stated:

Proposition 4.1. *Assume (2.1). The arithmetic mean process $a = a(\alpha, Q)$ possesses the following properties:*
(i) $\inf\limits_t a_t > 0$ *Q-a.s.*
(ii) $\sup\limits_t a_t < \infty$ *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 [13], section III.3, proposition 3.5. □

Recall the Bayesian terminology of section 2.5 where the measure α on $(\Theta, \mathcal{A})$ is called *a priori* probability measure. Along with this one may also define for each stopping time T on the same space the *a posteriori* probability measure α^T by the Bayes formula: for all $A \in \mathcal{A}$

$$\alpha^T(A) \doteq \frac{\int_A z_T(\theta, Q)\alpha(d\theta)}{\int_\Theta z_T(\theta, Q)\alpha(d\theta)}, \tag{4.3}$$

i.e. for each $\theta \in \Theta$

$$\frac{d\alpha^T}{d\alpha}(\theta) = \frac{z_T(\theta, Q)}{\int_\Theta z_T(\theta, Q)\alpha(d\theta)}, \tag{4.4}$$

cf (2.15) and (2.15). Obviously, the posterior α^T so defined is free of the choice of a dominating measure Q. Note that for fixed $A \in \mathcal{A}$ the random variable $\alpha^T(A)$ is $\mathcal{F}_T$-measurable. In virtue of the identity (4.2) we get with $\bar{P}_\alpha$ as the dominating measure that for each $\theta \in \Theta$

$$\frac{d\alpha^T}{d\alpha}(\theta) = \frac{z_T(\theta, Q)}{a_T(\alpha, Q)} = z_T(\theta, \bar{P}_\alpha), \tag{4.5}$$

cf (2.18). In particular, $E_\alpha z_T(\vartheta, \bar{P}_\alpha) = a(\alpha, \bar{P}_\alpha) = 1$. It is now straightforward to define at each stopping time T the information or the relative entropy in the posterior given a prior like in (2.17) or (2.20), respectively. But we postpone this till section 6 where the development in time of the related processes are treated.

4.2 Arithmetic mean process as an exponential

In practice each density process $z(\theta, Q), \theta \in \Theta$, is often given in the form of the Doléans exponential of a certain (Q, F)-local martingale, say $m(\theta, Q)$, that is by definition $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ with

$$\mathcal{E}(m) = e^{m - \frac{1}{2}\langle m^c\rangle} \prod_{s \leq \cdot} (1 + \Delta m_s) e^{-\Delta m_s}. \tag{4.6}$$

Moreover, often the continuous and the discontinuous parts of the (Q, F)-local martingales $m(\theta) = m(\theta, Q), \theta \in \Theta$, are separately modeled as martingale transforms of a certain continuous (Q, F)-local martingale, say M^c, and a certain purely discontinuous (Q, F)-local martingale, say M^d, both free of the parameter θ. The dependence on the parameter is realized via the predictable integrands, say $c(\theta)$ and $d(\theta)$, respectively, so that $m(\theta)$ is given the form

$$m(\theta) = c(\theta) \cdot M^c + d(\theta) \cdot M^d. \tag{4.7}$$

In fact all models of practical interest are of this particular form, see Part II for examples. As we have seen in section 4.1, the arithmetic mean process $a(\alpha, Q)$ is a density process, cf (4.2). Therefore, it may as well be represented as a Doléans exponential. Under the present circumstances this representation is particularly appealing, since it is only required to take the predictable posterior expectations (with respect to the posterior measure defined in (4.3)) of both integrands, namely $\bar{c} \doteq E_{\alpha^-} c(\vartheta)$ and $\bar{d} \doteq E_{\alpha^-} d(\vartheta)$. It suffices then to form the (Q, F)-local martingale

$$\bar{m} \doteq \bar{c} \cdot M^c + \bar{d} \cdot M^d, \tag{4.8}$$

as is shown in theorem 4.2 below.

Theorem 4.2. *Assume (2.1) and (2.9). Suppose that for each $\theta \in \Theta$ the density process $z(\theta, Q)$ can be represented as the Doléans exponential $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ with the (Q, F)-local martingale $m(\theta, Q)$ of the form (4.7). Then the arithmetic mean process is the Doléans exponential of the (Q, F)-local martingale (4.8), i.e.*

$$a(\alpha, Q) = z(\bar{P}_\alpha, Q) = \mathcal{E}(\bar{m}).$$

Proof. The relation $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ implies that

$$\begin{aligned} z(\theta, Q) =& 1 + z_-(\theta, Q) \cdot m(\theta, Q) \\ =& 1 + z_-(\theta, Q)\, c(\theta) \cdot M^c + z_-(\theta, Q)\, d(\theta) \cdot M^d. \end{aligned}$$

The latter equality follows from the condition (4.7). Replace now θ by ϑ and take the expectation E_α of both sides. On the left we get the arithmetic mean process $a(\alpha, Q)$, cf (4.1). On the right we are allowed to interchange the integration order under assumption (2.9), since we can now use the stochastic Fubini theorem in [2]. The condition in this paper under which the stochastic Fubini theorem holds boils in our case down to $E_Q E_\alpha [z(\vartheta, Q)]_\infty^{1/2} < \infty$. This follows from [2, theorem 3.1] and the martingale case of example 2 of the same paper. Combining inequalities (25.2) in chapter V and (90.2) in chapter VII of [5], we can (using a suitable constant C) bound $E_Q E_\alpha [z(\vartheta, Q)]_\infty^{1/2}$ by $C(1+E_Q E_\alpha z_\infty(\vartheta, Q) \log^+ z_\infty(\vartheta, Q))$ which is finite in view of (2.9) and the trivial relation $x \log^+ x \leq x \log x + \frac{1}{e}$.
Hence we obtain the integrands $a_-(\alpha, Q)\, \bar{c}$ and $a_-(\alpha, Q)\, \bar{d}$, since by (4.5) we have

$$E_\alpha\{z_-(\vartheta, Q)\, c(\vartheta)\} = a_-(\alpha, Q)\, E_\alpha\{z_-(\vartheta, \bar{P}_\alpha)\, c(\vartheta)\} = a_-(\alpha, Q)\, E_{\alpha^-}\, c(\vartheta)$$

that indeed yields the first integrand by the definition of $\bar{c}$. The second one is obtained in the similar way. We thus get by (4.8) $a(\alpha, Q) = 1 + a_-(\alpha, Q) \cdot \bar{m}$ that is equivalent to the desired relation $a(\alpha, Q) = \mathcal{E}(\bar{m})$. □

For more explicit computations of this type in terms of triplets of semimartingales we refer to section 3.2 in Part II. In the set up of that section no reference to a stochastic Fubini theorem is needed. See remark 3.4 in Part II for an explanation.
Theorem 4.2 has important consequence: it allows us to express the density of the posterior with respect to the prior as a Doléans exponential.

Corollary 4.3. *Under the conditions of theorem 4.2 the process $d\alpha^{\cdot}/d\alpha$ is the Doléans exponential*

$$\frac{d\alpha^{\cdot}}{d\alpha}(\theta) = \mathcal{E}(n(\theta, \bar{P}_\alpha)) \tag{4.9}$$

where for each $\theta \in \Theta$ the process $n(\theta, \bar{P}_\alpha)$ is a $(\bar{P}_\alpha, F)$-local martingale given by

$$n(\theta, \bar{P}_\alpha) = (c(\theta) - \bar{c}) \cdot N^c + (d(\theta) - \bar{d}) \cdot N^d \tag{4.10}$$

with $N^c = M^c - \langle \bar{m}, M^c \rangle$ and $N^d = M^d - \frac{1}{1+\Delta\bar{m}} \cdot [\bar{m}, M^d]$.

Proof. By (4.5) and by theorem 4.2 we may write the density $d\alpha^{\cdot}/d\alpha$ as the fraction of two Doléans exponentials:

$$\frac{d\alpha^{\cdot}}{d\alpha}(\theta) = \frac{\mathcal{E}(m(\theta, Q))}{\mathcal{E}(\bar{m})}.$$

Hence we only need to verify the identity $\mathcal{E}(n)\mathcal{E}(\bar{m}) = \mathcal{E}(m)$ with $m = m(\theta, Q)$ and $n = n(\theta, \bar{P}_\alpha)$, by using the multiplication rule for Doléans exponentials (cf [10, proposition (6.4)]). This is a straightforward computation. Indeed, according to this rule we must have $m = n + \bar{m} + [n, \bar{m}]$ but this is satisfied by $n = m - \bar{m} - \frac{1}{1+\Delta\bar{m}} \cdot [\bar{m}, m - \bar{m}]$. This plainly holds always, irrespective the specifications (4.7) and (4.8) that gives n the asserted form (4.10). □

Note the following relation

$$1 + \Delta n(\theta, \bar{P}_\alpha) = \frac{1 + \Delta m(\theta)}{1 + \Delta\bar{m}} > 0. \tag{4.11}$$

5 Geometric mean process and geometric mean measure

5.1 Geometric mean process

Along with the a-mean process (4.1), we associate with the parametric family of density processes $\{z(\theta, Q)\}_{\theta\in\Theta}$ a so-called *geometric mean process*

$$g(\alpha, Q) = e^{E_\alpha\{\log z(\vartheta, Q)\}}, \tag{5.1}$$

conform to the geometric mean of family of densities $\{p(\theta, Q)\}_{\theta\in\Theta}$ of section 2.4. By the same argument as in the latter section (Jensen's inequality) the geometric mean process is dominated by the a-mean process identically, i.e.

$$g(\alpha, Q) \leq a(\alpha, Q) \tag{5.2}$$

so that the geometric mean process also possesses property (ii) of proposition 4.1. As for the lower bound, we have assumed (2.7) in order to guarantee that the geometric mean process has property (i) of proposition 4.1 as well.

Proposition 5.1. *Assume (2.1) and (2.7). 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 (2.7) and Jensen's inequality and (ii) follows from equation (5.2).
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 (5.2)). 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

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

□

Observe that in virtue of (4.2) we have by inequality (5.2) that

$$
\frac{g(\alpha, Q)}{a(\alpha, Q)} = g(\alpha, \bar{P}_\alpha) \leq 1. \tag{5.3}
$$

Surely, this fraction depends only on the prior α and the family $\{P_\theta\}_{\theta\in\Theta}$ but not on the choice of a dominating measure Q.

5.2 Hellinger integrals and Hellinger processes

Let T be a F-stopping time. The Hellinger integral of the family of probability measures $\{P_{\theta,T}\}_{\theta\in\Theta}$, is defined according to [13], section IV.1, as the Q-expectation of the g-mean process evaluated at T:

$$
H_T(\alpha) = E_Q\{g_T(\alpha, Q)\}. \tag{5.4}
$$

This is called the *Hellinger integral of order* α. As is mentioned in section 2.4 the Hellinger integral is independent of the dominating measure Q.
Next, we define the Hellinger process of order α, denoted traditionally by $h(\alpha)$. Like the Hellinger integrals, the Hellinger processes are independent of the choice of the dominating measure Q:

Theorem 5.2. *Assume (2.1) and (2.7). 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*

$$M(\alpha, Q) = g(\alpha, Q) + g_-(\alpha, Q) \cdot h(\alpha) \tag{5.5}$$

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 5.1, 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 (2.12) and (5.5) 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 (2.12) to $g_-(\alpha, Q') \cdot h(\alpha)$. Thus similarly to (5.5)

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

is a Q'-martingale. The proof is complete by the same reasoning as the one after equation (2.12). ☐

Lemma 5.3. *Assume (2.1) and (2.7). Then up to a Q-evanescent set*

$$\Delta h(\alpha) < 1 \tag{5.6}$$

so that the Doléans exponential of $-h(\alpha)$ is well defined:

$$\mathcal{E}(-h(\alpha)) = e^{-h(\alpha)} \prod_{s \le \cdot} (1 - \Delta h_s(\alpha)) e^{\Delta h_s(\alpha)} \tag{5.7}$$

is a positive decreasing finite-valued process.

Proof. It suffices to prove (5.6). But in view of proposition 5.1, property (ii), this follows from the equation

$$E_Q\{g_T | F_{T-}\} - g_{T-}(1 - \Delta h(\alpha)_T) = 0, \tag{5.8}$$

valid on the set $\{T < \infty\}$ with a predictable time T, since by the predictable section theorem I.2.18 in [13] the latter equation implies $1 - \Delta h(\alpha) > 0$ up to a Q-evanescent set. The validity of (5.8) is verified as follows: first take Δ on both sides of (5.5), and then take the conditional Q-expectation given F_{T-}. ☐

Remark 5.4. Notice that equation (5.8) also implies that $\Delta h(\alpha)_T = -E\{\frac{\Delta g_T}{g_{T-}}|\mathcal{F}_{T-}\}$ valid on $\{T < \infty\}$ with a predictable time T.

Remark 5.5. It is easily verified that

$$\mathcal{E}(-h(\alpha))^{-1} = \mathcal{E}((1 - \Delta h(\alpha))^{-1} \cdot h(\alpha)),$$

cf [19], p 199.

Remark 5.6. Note the following relationship between Hellinger integrals and Hellinger processes:

$$H_T(\alpha) = 1 - E_Q\{g_-(\alpha, Q) \cdot h(\alpha)_T\} \tag{5.9}$$

that follows from (5.4) and (5.5). It will be shown below (theorem 5.17) that, under additional assumptions, $H_T(\alpha)$ is in fact the expectation with respect to a certain probability measure of the Doléans exponential (5.7) evaluated at T.
In the special situation in which the Hellinger process is deterministic, equation (5.9) reads $H_T(\alpha) = 1 - H_-(\alpha) \cdot h(\alpha)_T$ and hence $H_T(\alpha) = \mathcal{E}(-h(\alpha))_T$.

Remark 5.7. To draw once more parallels with the basic section 2 (in which the probability space is not yet equipped with the filtration), let us dwell for a moment on the elementary special case called a "one period model", since time $T > 0$ is fixed and $\mathcal{F}_t$ is taken to be trivial if $t < T$, while $\mathcal{F}_t = \mathcal{F}$ for $t \geq T$. In this case we have $z_t(\theta, Q) = 1$ for all θ if $t < T$ and $z_t(\theta, Q) = p_\theta$ for $t \geq T$ with p_θ as in equation (2.2). Consequently, $g_t(\alpha, Q) = 1$ for $t < T$ and for $t \geq T$ we have $g_t(\alpha, Q) = \exp E_\alpha \log p_\vartheta$. In this case it follows from equations (5.4) and (5.5) that $h_t(\alpha) = (1 - H_T(\alpha))1_{\{t \geq T\}}$. Therefore $\mathcal{E}(-h(\alpha))_t$ is equal to 1 if $t < T$ and equals $H_T(\alpha)$ if $t \geq T$. Therefore the density of the geodesic measure of section 2.4 can be alternatively expressed as $g_t(\alpha, Q)/\mathcal{E}(-h(\alpha))_t$ for any $t \geq 0$.

5.3 Geometric mean process of an exponential

The characterization of the Hellinger process $h(\alpha)$, presented in section 5.4, is based on proposition 5.8 below. We use here 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) $g(X) = \exp\{E_\alpha \log X(\vartheta)\}$ denote its arithmetic and geometric mean processes, respectively (cf the special cases (4.1) and (5.1)). Until the end of this section we assume sufficiently strong measurability properties and that the expectation with respect to α is well defined. The results of this section will be applied to the density processes $z(\theta, Q)$ and related processes for which these measurability properties are automatically satisfied.
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

$$\phi(CX) = C\phi(X) \tag{5.10}$$

Proposition 5.8. *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 $a(\sum_{s\leq\cdot}(\Delta X_s - \log(1+\Delta X_s)))$ 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

$$g(\mathcal{E}) = \mathcal{E}\{a(X) - \tfrac{1}{2}\tilde{v}(X^c) - \sum_{s\leq\cdot}\phi_s(1+\Delta X)\} \tag{5.11}$$

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

Proof. Since the Doléans exponential is of the form (4.6), the right hand side of (5.11) equals

$$\begin{aligned} &e^{a(X)-\frac{1}{2}\tilde{v}(X^c)-\frac{1}{2}\langle a(X^c)\rangle}\prod_{s\leq\cdot}(1+\Delta a_s(X) - \phi_s(1+\Delta X))e^{-\Delta a_s(X)} \\ &\qquad = e^{a(X)-\frac{1}{2}a(\langle X^c\rangle)}\prod_{s\leq\cdot} g_s(1+\Delta X)e^{-a_s(\Delta X)} \end{aligned}$$

which in turn is the ordinary exponential of

$$a(X - \tfrac{1}{2}\langle X^c\rangle + \sum_{s\leq\cdot} f(\Delta X_s)) = a\left(\log\mathcal{E}(X)\right)),$$

and thus equals to $g(\mathcal{E})$. This proves (5.11). ☐

Remark 5.9. If the continuous part $X(\vartheta)^c$ possesses the *variance process*

$$v(X^c) \doteq var_\alpha(X(\vartheta)^c) = E_\alpha\{|X(\vartheta)^c|^2\} - |E_\alpha\{X(\vartheta)^c\}|^2 \tag{5.12}$$

that is a (Q, F)-submartingale of class (D), then the compensator is given by $\tilde{v}(X^c)$ that occurred in (5.11).

Remark 5.10. Obviously, the identity (5.11) implies

$$g(\mathcal{E}) = g_-(\mathcal{E})\cdot a(X) - \frac{1}{2}g_-(\mathcal{E})\cdot\tilde{v}(X^c) - \sum_{s\leq\cdot} g_{s-}(\mathcal{E})\phi_s(1+\Delta X)$$

which is reduced in the special binary case to the Itô formula on p.199 in [13].

5.4 The Hellinger process as a compensator

The assertions in section 5.3 for an arbitrary family $\{X(\theta,Q)\}_{\theta\in\Theta}$ are aimed at the application to the special parametric family of processes $\{m(\theta,Q)\}_{\theta\in\Theta}$ with $m(\theta,Q)$ so that each density process $z(\theta,Q)$ is a Doléans exponential $\mathcal{E}(m(\theta,Q))$ of a martingale $m(\theta,Q)$. Then the assumptions made in section 5.3 are satisfied. Write m as a shorthand notation for $m(\vartheta,Q)$.
Below the notations of section 5.3 are used.

Theorem 5.11. *Assume (2.1) and (2.7). Let the process*

$$V = \frac{1}{2}v(m^c) + \sum_{s\leq \cdot} \phi_s(1+\Delta m) \tag{5.13}$$

be a (Q,F)-submartingale of class (D). Then its compensator $\tilde{V}$ and the Hellinger process $h(\alpha)$ are Q-indistinguishable. Furthermore, for a predictable stopping time T on the set $\{T<\infty\}$ it holds that

$$1 - \Delta h_T(\alpha) = E_Q\{g_T(1+\Delta m)|\mathcal{F}_{T-}\}. \tag{5.14}$$

Proof. It holds that $E_\alpha E_Q\{m(\vartheta,Q)^c\}^2 < \infty$, since $\log z(\theta,Q) \leq m(\theta,Q) - \frac{1}{2}\langle m(\theta,Q)^c\rangle$ and the Kullback-Leibler information is finite as postulated in (2.7).
By definition (5.13) and remarks 5.9 and 5.10 in section 5.3, especially equation (5.12), we get with $g(z) = g(\alpha,Q)$ the equation

$$g(z) = g_-(z)\cdot a(m) + \tfrac{1}{2}g_-(z)\cdot\{v(m^c) - \tilde{v}(m^c)\} - g_-(z)\cdot V \tag{5.15}$$

where the first two terms are Q-martingales. In order to see that $\tilde{V}$ and the Hellinger process $h(\alpha)$ are Q-indistinguishable, compare this equation with (5.5).
Next we find from (5.15) that $E_Q\{\frac{\Delta g_T}{g_{T-}}|\mathcal{F}_{T-}\} = -E_Q\{\phi_T(1+\Delta m)|\mathcal{F}_{T-}\}$, which by definition of ϕ is nothing else but $E_Q\{g_T(1+\Delta m)|\mathcal{F}_{T-}\} - 1$. Now use remark 5.4. □

Remark 5.12. It follows from this proof that the martingale $M(\alpha,Q)$ in theorem 5.2 can now be expressed as

$$M = g_-(z)\cdot\{a(m) + \frac{1}{2}\{v(m^c) - \tilde{v}(m^c)\} - (V - \tilde{V})\}. \tag{5.16}$$

5.5 Geometric mean measure

In this section we generalize the approach of Grigelionis [9], who worked with experiments having a finite parameter set. The result is formulated in theorem 5.17 below. In section 4.1 we showed that the density process of the arithmetic mean measure with respect to Q was the arithmetic mean of the density process $z(\theta,Q)$. In this section, that deals with the dynamic version of section 2.4, we define a measure whose density process is based on the geometric mean process. Since this process was shown to be a Q-supermartingale, we need a proper normalization to turn it into a martingale, cf theorem 5.13. To this end the multiplicative decomposition theorem of a nonnegative supermartingale (cf theorem 2.5.1 in [19]) together with the additive decomposition of equation (5.5) is precisely what we need. Notice that we cannot simply normalize the geometric mean process with its expectation to get a martingale (as we did in the static case of section 2.4). However remark 5.7 suggests an alternative.
In the special case that we discuss at the end of this section these two normalizations coincide.

Theorem 5.13. *Assume (2.1) and (2.7). Then the ratio*

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

is a local martingale under Q and, with $M(\alpha, Q)$ as in (5.5), the following relations are valid:

$$\zeta(\alpha, Q) = 1 + \frac{1}{\mathcal{E}(-h(\alpha))} \cdot M(\alpha, Q) \tag{5.17}$$

and

$$\zeta(\alpha, Q) = \mathcal{E}(\frac{1}{(1-\Delta h)g_-} \cdot M(\alpha, Q)). \tag{5.18}$$

Proof. Apply theorem 2.5.1 of [19] to the positive supermartingale $g(\alpha, Q)$ with the Doob-Meyer decomposition as in (5.5). This also yields formula (5.18). The expression (5.17) 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. □

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 [9] 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 [14] 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, see Part II.
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 [9], 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

$$\frac{dG_\alpha}{dQ} = \frac{g_\infty(\alpha, Q)}{\mathcal{E}(-h(\alpha))_\infty}, \tag{5.19}$$

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 5.14. *Assume (2.7). 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 5.1. □

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 Part II when we treat examples. Furthermore, in section 3.5 of part II 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 5.15. *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 the density processes $z(\theta, Q)$ equals one, see section 4.1. 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. □

Remark 5.16. Assume that G_α is a probability measure. Then we can give a rather explicit form of the density process $z(G_\alpha, Q)$ in terms of the martingales $m(\theta, Q)$ of section 5.4. First we get that $z(G_\alpha, Q)$ is nothing else but the $\zeta(\alpha, Q)$ of theorem 5.13. From equation (5.16) we obtain $\zeta(\alpha, Q) = \mathcal{E}(m^{G_\alpha})$, with

$$m^{G_\alpha} = \frac{1}{1 - \Delta h(\alpha)} \cdot \{a(m) + \frac{1}{2}\{v(m^c) - \tilde{v}(m^c)\} - (V - \tilde{V})\}.$$

If the discontinuous part of $a(m)$ is a sum of jumps martingale, we can say more: $m^{G_\alpha} = a(m^c) + \sum_{s \leq \cdot} (\frac{g_s(1+\Delta m)}{1-\Delta h_s(\alpha)} - 1)$. This follows from equation (5.14).

It will be shown next that with the special choice of the geometric mean measure as the dominating probability measure, the geometric mean processes take a particularly simple form, namely that of a predictable decreasing process. As a matter of fact, G_α is the only probability measure that gives this special form to the geometric mean process.

Theorem 5.17. *Assume the conditions (2.1) and (2.7). If G_α is a probability measure and if it is used as a dominating measure, then $g(\alpha, G_\alpha) = \mathcal{E}(-h(\alpha))$ and at each 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. The first assertion immediately follows from remark 5.16. From this the second assertion follows upon substitution of Q in (5.4) by G_α. □

The geometric mean measure has, like the geodesic measure, a minimizing property. The criterion is different from the one in proposition 2.5 however. We need the following notations. For a stopping time T we denote, as in section 3, by Q_T and $P_{\theta,T}$ the restrictions of Q, respectively P_θ, to $\mathcal{F}_T$. By $G_{\alpha,T}$ we denote the measure on $\mathcal{F}_T$ defined by $dG_{\alpha,T}/dQ_T = \zeta_T(\alpha, Q)$. Put

$$K_T(Q_T) = E_\alpha I(P_{\vartheta,T}|Q_T) + E_Q \log \mathcal{E}(-h(\alpha))_T \tag{5.20}$$

as the criterion to be minimized and notice the difference (putting $T = \infty$) with the criterion $E_\alpha I(P_\vartheta|Q)$ of section 2.4.

Proposition 5.18. *(i) Assume that $G_{\alpha,T}$ is a probability measure. Then $K_T(Q_T)$ is equal to $I(G_{\alpha,T}|Q_T)$, hence it is nonnegative and the minimum value zero of $K_T(\cdot)$ is attained at $Q_T = G_{\alpha,T}$.*
(ii) Conversely, if $\inf K_T(Q_T) = 0$, where the infimum is taken over all restrictions Q_T to $\mathcal{F}_T$ of the dominating measures Q, then $G_{\alpha,T}$ is a probability measure on $\mathcal{F}_T$.

Proof. (i) Notice first that

$$E_\alpha I(P_{\vartheta,T}|Q_T) = -E_Q \log g_T(\alpha, Q) = -E_Q \log \zeta_T(\alpha, Q) - E_Q \log \mathcal{E}(-h(\alpha))_T$$

in view of the definition of $\zeta(\alpha, Q)$, which gives (5.20) with $K_T(Q_T) = -E_Q \log \zeta_T(\alpha, Q) = -E_Q \log \frac{dG_{\alpha,T}}{dQ_T}$. Because we assumed that $G_{\alpha,T}$ is a probability measure, this reduces to $K_T(Q_T) = I(G_{\alpha,T}|Q_T)$. The other statement is now clear.
(ii) From $K_T(Q_T) = -E_Q \log \frac{dG_{\alpha,T}}{dQ_T}$ (see the proof of (i)) we obtain by Jensen's inequality applied to the convex function $-\log$ that $K_T(Q_T) \geq -\log E_Q \frac{dG_{\alpha,T}}{dQ_T} = -\log G_{\alpha,T}(\Omega)$.
If the infimum is zero, then for each $n \in \mathbb{N}$ there exists a probability measure Q_n on $\mathcal{F}_T$ such that $K_T(Q_n) < \frac{1}{n}$. Hence we have $-\log G_{\alpha,T}(\Omega) < \frac{1}{n}$ for all n so that $-\log G_{\alpha,T}(\Omega) \leq 0$, or $G_{\alpha,T}(\Omega) \geq 1$. Since we already know that $G_{\alpha,T}$ is a subprobability measure, it must be a probability measure. □

We close this section with some extra observations concerning the case where the Hellinger process is *deterministic*. In view of remark 5.6 we know that now $H_t(\alpha) = \mathcal{E}(-h(\alpha))_t$ for all $t \geq 0$. Hence $z(G_\alpha, Q) = g(\alpha, Q)/H(\alpha)$, which completely parallels the static case of section 2.4. In particular the restriction of G_α to $\mathcal{F}_T$ coincides with the geodesic measure as it is defined in section 2.4 with $\mathcal{F}$ replaced by $\mathcal{F}_T$. In section 5.6 we investigate some more relations between these measures.
The relation in proposition 5.18 now takes the form $E_\alpha I_T(P_\vartheta|Q) = I_T(G_\alpha|Q) - \log H_T(\alpha)$, precisely as in proposition 2.5 and the restriction of G_α to $\mathcal{F}_T$ minimizes $E_\alpha I_T(P_\vartheta|Q)$ over all dominating measures Q (restricted to $\mathcal{F}_T$).

5.6 Relation between geodesic and geometric mean measure

It is our purpose to study in this subsection some relations between G_α and $\breve{C}_\alpha$, both measures being defined in terms of certain, but different (compare equations (2.13) and (5.19)) normalizations of the geometric mean of densities. It is therefore interesting to compare these two measures. This will also yield a characterization (in theorem 5.19 below) of G_α as a probability measure in terms of properties of $\breve{C}_\alpha$.
Recall from section 2.4 that $\breve{C}_\alpha$ is defined on $\mathcal{F} = \mathcal{F}_\infty$ by its density $g_\infty(\alpha, Q)/H_\infty(\alpha)$ with respect to Q. Call this density $z_\infty(\breve{C}_\alpha, Q)$. Note that $0 < g_\infty(\alpha, Q) < \infty$ Q-a.s. We have seen in section 2.4 that $\breve{C}_\alpha \sim Q$, in other words that $z(\breve{C}_\alpha, Q)$ is Q-a.s. finite and positive. By $z_t(\breve{C}_\alpha, Q)$ we denote the density of the restriction of $\breve{C}_\alpha$ w.r.t. the restriction of Q to $\mathcal{F}_t$. Clearly, $z_t(\breve{C}_\alpha, Q) = E\{z_\infty(\breve{C}_\alpha, Q)|\mathcal{F}_t\} = E_Q\{g_\infty(\alpha, Q)|\mathcal{F}_t\}/H_\infty(\alpha)$.
Now we consider the process $\zeta(\alpha, \breve{C}_\alpha)$. It satisfies

$$\zeta_t(\alpha, \breve{C}_\alpha) = \zeta_t(\alpha, Q) z_t(Q, \breve{C}_\alpha) = \frac{g_t(\alpha, Q)}{\mathcal{E}(-h(\alpha))_t} \frac{H_\infty(\alpha)}{E_Q\{g_\infty(\alpha, Q)|\mathcal{F}_t\}}. \tag{5.21}$$

Recall that in our set up the random variable $h_\infty(\alpha)$, as well as the variable $\frac{1}{\mathcal{E}(-h(\alpha))_\infty}$, is finite a.s. under Q (or any other dominating measure), see [13, page 209] for the binary case. From (5.21) we see, letting $t \to \infty$, that

$$\zeta_\infty(\alpha, \breve{C}_\alpha) = \frac{H_\infty(\alpha)}{\mathcal{E}(-h(\alpha))_\infty}. \tag{5.22}$$

Observe that this random variable has $\breve{C}_\alpha$-expectation less than or equal to 1 and hence $E_{\breve{C}_\alpha} \frac{1}{\mathcal{E}(-h(\alpha))_\infty} \leq \frac{1}{H_\infty(\alpha)}$. We are now in the position to formulate another result that gives necessary and sufficient conditions for G_α to be a probability measure.

Theorem 5.19. *The measure G_α is a probability measure iff*

$$E_{\breve{C}_\alpha} \frac{1}{\mathcal{E}(-h(\alpha))_\infty} = \frac{1}{H_\infty(\alpha)}. \tag{5.23}$$

Moreover, if G_α is a probability measure, then its density process $z(G_\alpha, \breve{C}_\alpha)$ with respect to the geodesic measure has a particularly simple structure. It is uniformly integrable (under $\breve{C}_\alpha$) and given by

$$z_t(G_\alpha, \breve{C}_\alpha) = H_\infty(\alpha) E_{\breve{C}_\alpha} \{\frac{1}{\mathcal{E}(-h(\alpha))_\infty} | \mathcal{F}_t\}. \tag{5.24}$$

Similarly, we have under the same condition that the density process of $\breve{C}_\alpha$ w.r.t. G_α is given by

$$z_t(\breve{C}_\alpha, G_\alpha) = \frac{1}{H_\infty(\alpha)} E_{G_\alpha} \{\mathcal{E}(-h(\alpha))_\infty | \mathcal{F}_t\}. \tag{5.25}$$

Proof. The equivalence of the first claim follows from the identities

$$E_{\breve{C}_\alpha} \zeta_\infty(\alpha, \breve{C}_\alpha) = E_Q \zeta_\infty(\alpha, Q) = G_\alpha(\Omega).$$

and expression (5.22). The validity of (5.24) also follows from (5.22) and (5.25) follows from (5.24). $\square$

Remark 5.20. Equation (5.24) again illustrates the fact that the measures G_α and $\breve{C}_\alpha$ coincide if $h(\alpha)$ is a deterministic process in view of remark 5.6.

6 Information processes

In this section we will treat the Kullback-Leibler information in the posterior probability measure α^T with respect to the prior α, defined according to (2.17). Afterwards, we will use the representation of type (2.19) to relate this information to the Hellinger process and to the density process of the geometric mean measure with respect to the arithmetic mean measure, cf theorem 6.3.
We will treat as well the relative entropy in the posterior probability measure α^T with respect to the prior α, defined according to (2.20). Looking at the development in time, we will see that the related process is a $(\bar{P}_\alpha, F)$-submartingale with a rather complicated martingale part. Therefore, the focus in theorem 6.4 will be on the compensator part that yields the information from data like in (2.21), cf remark 6.5.

6.1 Kullback-Leibler information in a posterior

The definition (2.17) suggests us to 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

$$I(\alpha^T|\alpha) = E_\alpha \log \frac{d\alpha}{d\alpha^T}. \tag{6.1}$$

This is a non-negative quantity by the Jensen inequality. In view of the relation (2.19), we have readily the following statement:

Theorem 6.1. *Let $T > 0$ be a stopping time, let α and α^T be the prior and posterior probability measures on the parametric space $(\Theta, \mathcal{A})$ and let $I(\alpha^T|\alpha)$ be the Kullback-Leibler information in the posterior α^T with respect to the prior α, as defined in (6.1). Then*

$$e^{-I(\alpha^T|\alpha)} = \frac{g_T(\alpha, Q)}{a_T(\alpha, Q)} = g_T(\alpha, \bar{P}_\alpha). \tag{6.2}$$

Proof. Recall the relation (4.5) and made appropriate substitutions in (2.19). □

Observe again that the information $I(\alpha^T|\alpha)$ depends only on the prior α but not on the choice of a dominating measure Q. In view of the propositions 4.1 and 5.1 we have

Proposition 6.2. *Assume (2.1) and (2.7). Let $I(\alpha^\cdot|\alpha)$ be the information process starting from zero, $I(\alpha^0|\alpha) = 0$, and at $t > 0$ defined by (6.1). 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. This is a direct consequence of the propositions 4.1 and 5.1 and theorem 6.1.
□

Departing from the identity (6.2) we obtain the following representation of the information process in terms of the Hellinger process and the geometric mean and arithmetic mean measures:

Theorem 6.3. *Under the conditions (2.1) and (2.7) and the condition that G_α is a probability measure the information $I(\alpha^T|\alpha)$ at a stopping time $T > 0$ can be presented as follows:*

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

Proof. Combine remark 5.16 and theorem 6.1. □

6.2 The information from data

According to (2.20) the relative entropy in the posterior α^T at a stopping time T with respect to a prior α is defined by

$$I(\alpha|\alpha^T) = E_{\alpha^T} \log \frac{d\alpha^T}{d\alpha}(\vartheta) = E_\alpha z_T(\vartheta, \bar{P}_\alpha) \log z_T(\vartheta, \bar{P}_\alpha). \tag{6.3}$$

As was already mentioned, in Bayesian statistics this quantity is called the information from data and its expectation

$$E_{\bar{P}_\alpha} I(\alpha|\alpha^T) = E_\alpha I(\bar{P}_{\alpha,T}|P_{\vartheta,T}) \tag{6.4}$$

is called the expected utility from data, cf (2.21). For $T = \infty$ we can use the notation of section 2 to write the information from data as $I(\alpha|\alpha^\infty) = E_\alpha p(\vartheta, \bar{P}_\alpha) \log p(\vartheta, \bar{P}_\alpha)$ or as $I(\alpha|\beta)$.
For brevity, let us agree to denote $x \log x$ by $\ell(x)$. Then the process of the information from data is simply expressed by $I(\alpha|\alpha^\cdot) = E_\alpha \ell(z(\vartheta, \bar{P}_\alpha))$. We claim that this process is a uniformly integrable submartingale under condition (2.9). Notice first that $|\ell(x)| \leq \ell(x) + 2e^{-1}$. Hence condition (2.9) guarantees that $E_{\bar{P}_\alpha} E_\alpha |\ell(z(\vartheta, \bar{P}_\alpha))| < \infty$. Since the function ℓ on $\mathbb{R}_+$ is convex, we readily conclude from Jensen's inequality and the fact that the $z(\theta, \bar{P}_\alpha)$ are $(\bar{P}_\alpha, F)$-martingales that indeed $I(\alpha|\alpha^\cdot)$ is a $(\bar{P}_\alpha, F)$-submartingale with limit $I(\alpha|\beta)$, satisfying $0 \leq I(\alpha|\alpha^t) \leq E_{\bar{P}_\alpha}\{I(\alpha|\beta)|\mathcal{F}_t\}$ for all $t \geq 0$.
The first step towards the expected utility from data is thus to determine the compensator to it. This will be done in theorem 6.4 under the same special circumstances as in section 4.2: it will be again assumed that each density process $z(\theta, Q), \theta \in \Theta$, is the Doléans exponential of a certain (Q, F)-local martingale $m(\theta, Q)$, i.e. $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ with the exponential (4.6), and that the latter martingale is representable in the special form (4.7). Exactly as in section 4.2 and using (here and elsewhere below) the notations of that section we need to take certain predictable posterior expectations from both integrands c and d. Firstly, the *posterior variance* of the integrand $c(\vartheta)$ is required at each instant $t > 0$ that is defined like in (5.12) as follows: $\bar{v}(c) = E_{\alpha^-}(c(\vartheta) - \bar{c})^2$.
We also use the following notation. Define for a process X depending on θ the process $\bar{\ell}(X) = E_{\alpha^-}\ell(X(\theta))$.

Theorem 6.4. *Under conditions (2.1) and (2.9) the difference between the information from data $I(\alpha|\alpha^\cdot)$ and the nondecreasing finite-valued process*

$$\tfrac{1}{2}\bar{v}(c) \cdot \langle M^c \rangle + \sum_{s \leq \cdot} \bar{\ell}_s(1 + \Delta n_s(\cdot, \bar{P}_\alpha)) \tag{6.5}$$

is a $(\bar{P}_\alpha, F)$-martingale.

Proof. For simplicity, we shall write for the rest of the proof z instead of $z(\vartheta, \bar{P}_\alpha)$ and similarly n instead of $n(\vartheta, \bar{P}_\alpha)$. The latter comes from corollary 4.3 where its explicit representation can be found, together with $z = \mathcal{E}(n)$, the consequence of

(4.5) and (4.9). This will cause no ambiguity. With these notations (6.3) reduces to $I(\alpha|\alpha^{\cdot}) = E_\alpha \ell(z)$. Integration by parts gives

$$\ell(z) = z \log z = z_- \cdot \log z + \log z_- \cdot z + [z, \log z]. \tag{6.6}$$

It will be seen in a moment that the $(\bar{P}_\alpha, F)$-local martingale mentioned in the assertion stems from the second term and equals to

$$E_\alpha[\log z_- \cdot z] = E_\alpha[\ell(z_-) \cdot n]. \tag{6.7}$$

Let us examine the first and third terms. In view of the general formula (4.6) we have

$$z_- \cdot \log z = z_- \cdot n - \tfrac{1}{2} z_- \cdot \langle n^c \rangle + \sum_{s \leq \cdot} z_{s-} \left\{ \log(1 + \Delta n_s) - \Delta n_s \right\}$$

and

$$[z, \log z] = z_- \cdot \langle n^c \rangle + \sum_{s \leq \cdot} z_{s-} \Delta n_s \log(1 + \Delta n_s).$$

Take now the expectation of both sides with respect to the prior α and take into consideration that $E_\alpha z_- \cdot n = 0$ and $E_\alpha \sum_{s \leq \cdot} z_{s-} \Delta n_s = 0$ (this follows e.g. from $E_\alpha z = 1$, see (4.5), and $z = 1 + z_- \cdot n$). Substituting the explicit expression for n, see corollary 4.3, we get by Fubini's theorem

$$E_\alpha z_- \cdot \log z = -\tfrac{1}{2} \bar{v}(c) \cdot \langle M^c \rangle + \sum_{s \leq \cdot} E_{\alpha^{s-}} \log(1 + \Delta n_s)$$

and

$$E_\alpha[z, \log z] = \bar{v}(c) \cdot \langle M^c \rangle + \sum_{s \leq \cdot} E_{\alpha^{s-}} \Delta n_s \log(1 + \Delta n_s),$$

where according to equation (4.11) $1 + \Delta n_s$ can be written as in (4.11), i.e.

$$\frac{1 + \Delta m(\vartheta)}{1 + \Delta \bar{m}} = 1 + (d(\vartheta) - \bar{d}) \Delta N^d.$$

These two formulas allow us to find the expectation with respect to α of (6.6). On the right we get the information from data and on the left the local martingale (6.7) plus the expression (6.5). Thus the local martingale mentioned in the assertion is indeed given by (6.7). The proof is complete. □

Remark 6.5. Notice that the decomposition in theorem 6.4 of the information from data process is in general not its Doob-Meyer decomposition, since the discontinuous part (if it is not vanishing) of the increasing process of (6.5) is not predictable.
Since the expectation with respect to the measure $\bar{P}_\alpha$ of the martingale in theorem 6.4 is zero (under the conditions of the theorem), the expected utility from data at the stopping time T equals to

$$\begin{aligned} E_{\bar{P}_\alpha} I(\alpha|\alpha^T) =& E_\alpha I(\bar{P}_{\alpha,T} | P_{\vartheta,T}) \\ =& E_{\bar{P}_\alpha} \{ \tfrac{1}{2} \bar{v}(c) \cdot \langle M^c \rangle + \sum_{s \leq \cdot} \bar{\ell}_s(\Delta N^d_s). \end{aligned}$$

The first identity is already known, see (6.4). The second one follows from (6.5).

Acknowledgements

Dzhaparidze and Spreij were partially supported by the Academy of Finland Project No. 37689 and Valkeila was supported by the Research Grants Committee of the University of Helsinki.

References

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

[2] K. Bichteler and S.J. Lin (1995), On the stochastic Fubini theorem, *Stochastics and Stochastics Reports* **54**, pp 271-279.

[3] N.N. Čencov (1980), *Statistical Decision Rules and Optimal Inference*, Translations of Mathematical Monographs, American Mathematical Society.

[4] I. Csiszár (1975), I-divergence Geometry of Probability Distributions and Minimization Problems, *Ann. Prob.* **3**, 146-158.

[5] C. Dellacherie and P.A. Meyer (1982), *Probabilities and potential B*, North Holland.

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

[7] 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. (eds. Yu. Kabanov, B. Rozovskii and A. Shiryaev) World Scientific, Singapore, 41-62.

[8] K. Dzhaparidze, P.J.C. Spreij and E. Valkeila (2000), Information processes in filtered experiments, Part II: explicit representations and examples.

[9] 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.

[10] J. Jacod (1979), *Calcul Stochastique et Problèmes de Martingales*, Springer.

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

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

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

[14] 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.

[15] 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).

[16] 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.

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

[18] 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.

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

[20] M. Loève (1977,1978), *Probability Theory*, Springer.

Information processes in filtered experiments
Part II: explicit representations and examples

Kacha Dzhaparidze[1], Peter Spreij[2] and Esko Valkeila[3]

January 31, 2001

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.

keywords Kullback-Leibler information, Hellinger process, posterior distribution, semimartingales, predictable characteristics, stochastic calculus, fractional Brownian motions

AMS-classification 60G07, 60H30, 62B15, 94A17

[1]Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam
[2]Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam
[3]Department of Mathematics, University of Helsinki, P.O. Box 4, 00014 Helsinki

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. In a first part [5], often referred to as 'Part I', we derived some results of a general nature, whereas in the present Part II we assume that we deal with observations either coming to us in the form of a semimartingale or by a process that is driven by 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 geometric mean measure and arithmetic mean measure as they were developed in Part I we are now able 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 [10]), 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 paper is organized as follows. In section 2 we repeat and summarize some notions and results from Part I. In section 3 we present explicit version of results of Part I by assuming that we observe a semimartingale. In particular we compute the Hellinger process as it is defined in Part I 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 [11] 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 $\bigvee_t \mathcal{F}_t = \mathcal{F}$.
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

$$0 < I(\mathbf{P}|\mathbf{Q}) < \infty. \tag{2.2}$$

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

$$\alpha^T(A) = \frac{\int_A z_T(\theta, Q)\alpha(d\theta)}{\int_\Theta z_T(\theta, Q)\alpha(d\theta)}. \tag{2.3}$$

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 introduced in [4] and in Part I, where more details can be found. Both measures are defined on the aforementioned filtered space $(\Omega, \mathcal{F}, F)$. For $B \in \mathcal{F}$ we set

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

Under the assumption (2.1) the measure $\bar{P}_\alpha$ is equivalent to the measure Q, see lemma 2.1 in Part I. In [4] and Part I 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 (2.3) is equivalent to

$$\frac{d\alpha^T}{d\alpha}(\theta) = z_T(\theta, \bar{P}_\alpha). \tag{2.4}$$

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

$$g(\alpha, Q) = e^{E_\alpha \log z(\vartheta, Q)}. \tag{2.5}$$

As is shown in proposition 5.1 of Part I, under the present conditions this geometric mean process is a (Q, F)-supermartingale of class (D). Moreover, we have $\inf_t g_t > 0$ Q-a.s. and $\sup_t g_t < \infty$ Q-a.s. These properties of $g(\alpha, Q)$ allow us to characterize its compensator as follows. As in theorem 5.2 of Part I define the *Hellinger process* of order α as the unique (up to Q-indistinguishability) predictable finite-valued increasing process $h(\alpha)$ starting from the origin $h_0(\alpha) = 0$ so that

$$M(\alpha, Q) = g(\alpha, Q) + g_-(\alpha, Q) \cdot h(\alpha) \tag{2.6}$$

is a (Q, F)-uniformly integrable martingale. We are now in the position to define the *geometric mean measure* G_α with the help of the ratio

$$\zeta(\alpha, Q) = \frac{g(\alpha, Q)}{\mathcal{E}(-h(\alpha))} \tag{2.7}$$

as a density process, where $\mathcal{E}(-h(\alpha))$ is the Doléans-Dade exponential of $-h(\alpha)$. The ratio $\zeta(\alpha, Q)$ is a nonnegative local martingale, with terminal value $\zeta_\infty(\alpha, Q) \geq 0$. By definition the geometric mean measure is defined as the measure on $(\Omega, \mathcal{F})$ whose density with respect to Q is given by $\zeta_\infty(\alpha, Q)$. Under additional assumptions, see section 5.5 in Part I, $\zeta(\alpha, Q)$ is a uniformly integrable martingale. Then, at each stopping time T the geometric mean measure G_α possesses the restriction to $\mathcal{F}_T$ defined for each $B \in \mathcal{F}_T$ by

$$G_{\alpha,T}(B) = \int_B \zeta_T(\alpha, Q)dQ.$$

If the experiment satisfies the properties (2.1) and (2.2), then the geometric mean measure G_α is equivalent to Q, see lemma 5.14 in Part I. In section 5 of Part I one can also find a discussion concerning additional assumptions under which the measure G_α is a probability measure. Some concrete examples will be worked out in the forthcoming section 5.
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

$$\phi(CX) = C\phi(X). \tag{2.8}$$

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

$$v(X^c) \doteq var_\alpha X(\vartheta)^c = E_\alpha |X(\vartheta)^c|^2 - |E_\alpha X(\vartheta)^c|^2 \tag{2.9}$$

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

$$\tilde{v}(X^c) \doteq a(\langle X^c \rangle) - \langle a(X^c) \rangle. \tag{2.10}$$

Throughout we well use common notions and facts of the general theory of stochastic processes as developed e.g. in [6] or [10]. 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 [6, Theorem III.3.24] or [10, 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 [10, 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 [4, Section 3.3] (a generalization of a result by Kolomiets [8] that also can be found in [6, Theorem III.3.40] or [10, 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 [4, 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 (2.4) 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 [10, 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 (2.2). For each $\theta \in \Theta$ let $m(\theta, Q)$ be a (Q, F)-local martingale given by

$$m(\theta, Q) = z_-(\theta, Q)^{-1} \cdot z(\theta, Q) \tag{3.5}$$

so that the density process is represented as the Doléans exponential $z(\theta, Q) = \mathcal{E}(m(\theta, Q))$ of this martingale.
Upon further specification of the randomized experiment in question, one can assign to the (Q, F)-local martingale (3.5) 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 (3.5) gets the form

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

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 4.1 of Part I. Departing from the representation property (3.6), 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 (3.6). The rest of this section parallels section 4.2 of Part I. The treatment is even more transparent due to references to the prior section 3.1.

Theorem 3.3. *Assume (2.1), (2.2) and the representation property (3.6). Then the arithmetic mean process is the Doléans exponential $a(\alpha, Q) = \mathcal{E}(\bar{m})$ of a (Q, F)-local martingale*

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

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 4.1 of Part I, 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 (3.7). 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. □

Remark 3.4. Alternatively, we could argue as in section 4.2 of Part I by observing first that the relation (3.5) and the condition (3.6) imply

$$\begin{aligned} z(\vartheta, Q) =& 1 + z_-(\vartheta, Q) \cdot m(\vartheta, Q) \\ =& 1 + z_-(\vartheta, Q)\, \beta(\vartheta) \cdot X^c + z_-(\vartheta, Q)\{Y(\vartheta) - 1 + \frac{\hat{Y}(\vartheta) - \hat{1}}{1 - \hat{1}}\} * (\mu^X - \nu). \end{aligned}$$

By taking then the expectation E_α we get on the left the arithmetic mean process $a(\alpha, Q)$ whose explicit expression was given in theorem 3.3. Compare this to what we get on the right to conclude that

$$E_\alpha\{z_-(\vartheta, Q)\,\beta(\vartheta)\cdot X^c\} = \{E_\alpha z_-(\vartheta, Q)\,\beta(\vartheta)\}\cdot X^c,$$

i.e. that the integration order here is interchangeable (this holds for the discontinuous term too, of course). No stochastic Fubini theorem is needed, unlike in the proof of theorem 4.2 in Part I.

Exactly as in section 4.2 of Part I we obtain the following corollary for the density (2.4) of the posterior with respect to the prior (we provide the details of the proof in the spirit of Part I, though due references to section 3.1 should easily bring the reader to the assertion). The notations that we use are those of section 3.1.

Corollary 3.5. *Assume (2.1), (2.2) and assume the representation property (3.6). 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

$$m(\theta, \bar{P}_\alpha) = \big(\beta(\theta) - \bar{\beta}\big)\cdot X^{c,\bar{P}_\alpha} + \{\frac{Y(\theta)}{\bar{Y}} - 1 + \frac{\hat{Y}(\theta) - \hat{\bar{Y}}}{1 - \hat{\bar{Y}}}\} * \big(\mu^X - \bar{\nu}\big) \tag{3.8}$$

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 (2.4) 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

$$1 + \Delta m(\theta, \bar{P}_\alpha) = \frac{1 + \Delta m(\theta, Q)}{1 + \Delta\bar{m}}, \tag{3.9}$$

cf formula (4.11) in Part I. To this end, observe first that

$$\begin{aligned} 1 + \Delta m(\theta, Q) =& 1 + \{Y(\theta; \cdot, \Delta X) - 1\}I_{\{\Delta X \neq 0\}} - \frac{\hat{Y}(\theta) - \hat{1}}{1 - \hat{1}} I_{\{\Delta X = 0\}} \\ =& Y(\theta; \cdot, \Delta X) I_{\{\Delta X \neq 0\}} + \frac{1 - \hat{Y}(\theta)}{1 - \hat{1}} I_{\{\Delta X = 0\}} \end{aligned} \tag{3.10}$$

(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 [6, definition II.1.27] or [10, theorem 3.5.1]). Substitute then θ by ϑ and take on both sides the expectation with respect to the posterior α^- to get

$$1 + \Delta\bar{m} = \bar{Y}(\cdot;\cdot,\Delta X) I_{\{\Delta X \neq 0\}} + \frac{1-\hat{\bar{Y}}}{1-\hat{1}} I_{\{\Delta X = 0\}} \tag{3.11}$$

(one may derive this directly from (3.7), of course). Finally, apply this device to (3.8). 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 (3.9). The proof is complete. □

3.3 Representation of Hellinger processes

The main result of this section is based upon a characterization of the Hellinger processes which was theorem 5.11 in Part I. Therefore we begin with a short description of the assertion of this theorem. Assume again (2.1), (2.2) and the representation (3.6) 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

$$V = \frac{1}{2} v(m^c) + \sum_{s \leq \cdot} \phi_s(1 + \Delta m) \tag{3.12}$$

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. This is exactly the assertion of theorem 5.11 in Part I. We are going to present in the next theorem this compensator in terms of the triplet of predictable characteristics of the observations (cf [4], theorem 5.3; the proof is reproduced below, since the basic arguments are needed anew in the subsequent sections).

Theorem 3.6. *Along with the conditions (2.1) and (2.2) assume the representation property (3.6). Then*

$$h(\alpha) = \frac{1}{2} v(\beta) \cdot C + \phi(Y) \cdot \nu + \sum_{s \leq \cdot} \phi_s(1 - \hat{Y}). \tag{3.13}$$

Proof. The first term in (3.12) 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 (2.9) and (2.10) to $m(\theta,Q)^c = \beta(\theta) \cdot X^c$. Next, we have to show that the second term in (3.12) is compensated by the sum of the last two terms in (3.13), i.e. that

$$\sum_{s \leq \cdot} \phi_s(1 + \Delta m) - \{\phi(Y) \cdot \nu + \sum_{s \leq \cdot} \phi_s(1 - \hat{Y})\} \tag{3.14}$$

is a (Q,F)-local martingale. But this claim holds true, in view of lemma 3.7 below upon noting that ϕ is homogeneous (see equation (2.8)). □

Now we formulate a lemma (related to [10, theorem 3.5.1]) with the computational tool that we needed in the course of proving theorem 3.6 and that we will also use in the proof of theorem 4.3.

Lemma 3.7. *Let $m = m(\theta, Q)$ be given by (3.6) 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*

$$f(Y)\cdot\nu + \sum_{s\leq\cdot} f(\frac{1-\hat{Y}_s}{1-\hat{1}_s})(1-\hat{1}_s) \tag{3.15}$$

and the local martingale in its semimartingale decomposition takes the form

$$\{f(Y) - f(\frac{1-\hat{Y}}{1-\hat{1}})\} * (\mu^X - \nu). \tag{3.16}$$

Proof. The jumps $\Delta m(\theta, Q)$ satisfy equation (3.10), as we have seen. Apply the function f to both sides of this equation and take the sum to obtain

$$\sum_{s\leq\cdot} f(1+\Delta m_s) = \sum_{s\leq\cdot} f(Y(\theta; s, \Delta X_s))I_{\{\Delta X_s\neq 0\}} + \sum_{s\leq\cdot} f(\frac{1-\hat{Y}_s(\theta)}{1-\hat{1}_s})I_{\{\Delta X_s = 0\}}. \tag{3.17}$$

Therefore, by the same considerations as in [6, Lemma IV.3.22], the compensator of $\sum_{s\leq\cdot} f(1+\Delta m_s)$ has the asserted form. To check that the local martingale part can be written as the given stochastic integral with respect to $\mu^X - \nu$ of (3.16) one proceeds along the same lines that were followed to arrive at (3.10). $\square$

Remark 3.8. The explicit expression for the (Q, F)-local martingale (3.14) is then (use again lemma 3.7 and the fact that ϕ is homogeneous)

$$\{\phi(Y) - \frac{\phi(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu). \tag{3.18}$$

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 3.9. *Assume (2.1), (2.2) and the representation property (3.6). 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*

$$z(G_\alpha, Q) = \mathcal{E}(\frac{1}{1-\Delta h(\alpha)} \cdot N(\alpha, Q)) \tag{3.19}$$

where

$$N(\alpha, Q) = a(\beta) \cdot X^c + \{g(Y) - \frac{g(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu) \tag{3.20}$$

is a (Q, F)-local martingale that is simply related to $M(\alpha, Q)$ defined by (2.6):

$$M(\alpha, Q) = g_-(\alpha, Q) \cdot N(\alpha, Q). \tag{3.21}$$

Proof. Relation (3.19) follows from theorem 5.13 in Part I and equation (3.21). Next we verify that $N(\alpha, Q)$ has the representation (3.20). From remark 5.12 and equation (5.13) in Part I 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 (3.14) that is according to remark 3.8 given by $\{\phi(Y) - \frac{\phi(1-\hat{Y})}{1-\hat{1}}\} * (\mu^X - \nu)$, cf (3.18). 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 (3.20). □

Remark 3.10. Of course the decomposition (3.19) is also valid for the process $\zeta(\alpha, Q)$ of equation (2.7) without condition (3.6). We only needed it to specify the martingale in (3.20).

Theorem 3.11. *Assume (2.1), (2.2) and the representation property (3.6). 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*

$$\begin{cases} B^{G_\alpha} &= a(B) + (Y^{G_\alpha} - a(Y))\hbar \cdot \nu \\ C^{G_\alpha} &= C \\ \nu^{G_\alpha} &= Y^{G_\alpha} \cdot \nu \ \ \textit{with} \ \ Y^{G_\alpha} = \frac{g(Y)}{1-\Delta h(\alpha)}. \end{cases} \tag{3.22}$$

Proof. We use theorem 3.9 and focus first on the third characteristic ν^{G_α}. By Girsanov's theorem for random measures (see [6, 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 [6] 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 (3.13) and by $\phi = a - g$ we have

$$1 - \Delta h(\alpha) = \hat{g}(Y) + g(1 - \hat{Y}). \tag{3.23}$$

Observe that the latter identity yields $1 - \Delta h(\alpha) + \Delta N(\alpha, Q) = g(Y(\cdot, \cdot, \Delta X))$, hence by (3.19) 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 (3.22) is trivial, we finally prove the first one. According to Girsanov's theorem III.3.24 in [6] we have

$$B^{G_\alpha} = B + a(\beta) \cdot C + (Y^{G_\alpha} - 1)\hbar \cdot \nu, \tag{3.24}$$

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. $\square$

Remark 3.12. 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 (3.24), 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 (3.19), the multiplicative decomposition of the geometric mean process $g(\alpha,Q)$ resulting from (2.7), can be given a specific form.

Corollary 3.13. *Assume (2.1), (2.2) and the representation property (3.6). Then the geometric mean process possesses the multiplicative decomposition*

$$g(\alpha,Q) = \mathcal{E}(\frac{1}{1-\Delta h(\alpha)}\cdot N(\alpha,Q))\,\mathcal{E}(-h(\alpha)), \tag{3.25}$$

with $N(\alpha,Q)$ as in (3.20).
If G_α is taken as the dominating measure, then the above identity can be replaced with

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

Proof. Combine equation (2.7) and theorem 3.9 to get (3.25), whereas (3.26) immediately follows from (2.7). $\square$

Another important consequence is the following useful representation of the Hellinger integral that is defined at the beginning of section 5.2 in Part I as

$$H_T(\alpha) = E_Q g_T(\alpha,Q) \tag{3.27}$$

where T is a stopping time. Note that the Hellinger integral so defined is independent of the choice of the dominating measure Q, see Part I for details.

Corollary 3.14. *Assume (2.1), (2.2) and the representation property (3.6). 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 (3.27) by G_α and apply (3.26). $\square$

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. [16].
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, see Part I, section 5.5). As we will show, this identity holds iff $\{P_\theta\}_{\theta\in\Theta}$ is an exponential family. In the book [9] (in particular chapter 11, that deals with semimartingales) exponential families are defined in a number of specific cases, see also [15, chapter 2]. Exponential families of a general nature in a static context have been studied in [13]. The definition below is much in spirit of [9, 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 3.15. 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 $x \in E$

$$Y(\cdot; t, x) = \frac{l(\cdot; t, x)}{1_{\{\hat{1}<1\}} + \hat{l}_t(\cdot)} \tag{3.28}$$

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

Remark 3.16. Note that if $\hat{1} < 1$, then (3.28) 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 3.17. *(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 3.15, at $\theta_\alpha^\circ = E_\alpha \vartheta$ we have $\beta(\theta_\alpha^\circ) = a(\beta) = \beta^{G_\alpha}$ and (excluding the trivial case of vanishing Y) $Y(\theta_\alpha^\circ) = \frac{l(\theta_\alpha^\circ)}{1_{\{\hat{1}<1\}}+\hat{l}(\theta_\alpha^\circ)} = \frac{g(l)}{1_{\{\hat{1}<1\}}+\hat{g}(l)}$. Thus to show that $Y(\theta_\alpha^\circ) = Y^{G_\alpha}$ as well, it suffices to verify

$$\frac{g(Y)}{\hat{g}(Y) + g(1-\hat{Y})} = \frac{g(l)}{1_{\{\hat{1}<1\}} + \hat{g}(l)}, \tag{3.29}$$

cf definition of Y^{G_α} in (3.22) in conjunction with (3.23). But due to (3.28) 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 (3.29).
(ii) Since now $G_\alpha = P_{\theta_\alpha^\circ}$ for all α with finite support, we have $a(\beta) = \beta(\theta_\alpha^\circ)$ and $Y^{G_\alpha} = Y(\theta_\alpha^\circ)$. 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 [6, theorem III.3.17] then also all $\hat{Y}_t(\theta) = 1$. Hence we have $\frac{g(Y)}{\hat{g}(Y)} = Y(\theta_\alpha^\circ)$. Choosing x_0 such that $Y(\theta_\alpha^\circ; t, x_0) > 0$, we obtain that (we now and henceforth in this part of the proof suppress the fixed time instant t)

$$g\left(\frac{Y(\cdot;x)}{Y(\cdot;x_0)}\right) = \frac{g(Y(\cdot;x))}{g(Y(\cdot;x_0))} = \frac{Y(\theta_\alpha^\circ;x)}{Y(\theta_\alpha^\circ;x_0)}.$$

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 (3.28)).
In the second case we assume that $\hat{1}_t < 1$. In view of remark 3.16 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 $Y^{G_\alpha} = Y(\theta_\alpha^\circ)$ we deduce that

$$g\left(\frac{Y}{1-\hat{Y}}\right) = \frac{Y(\theta_\alpha^\circ)}{1-\hat{Y}(\theta_\alpha^\circ)}.$$

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. $\square$

Remark 3.18. The statements of theorem 3.17 can be summarized by saying that the geometric mean measures G_α are equal to $P_{\theta_\alpha^\circ}$ for all α such that $E_\alpha\vartheta$ exists iff they are equal to $P_{\theta_\alpha^\circ}$ for all α with finite support.

We recall the multiplicative decomposition of the geometric mean process given by (3.25) in conjunction with (3.19):

$$g(\alpha, Q) = z(G_\alpha, Q)\mathcal{E}(-h(\alpha)). \tag{3.30}$$

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

Proposition 3.19. *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 3.17 and equation (3.30) 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 (2.3). Recall that, using the arithmetic mean measure $\bar{P}_\alpha$ as a dominating measure, we may present this formula as the following identity (2.4) 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 related to the arithmetic and geometric mean processes as follows (see theorem 6.1 in Part I):

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

Observe again that the information $I(\alpha^T|\alpha)$ depends only on the prior α but not on the choice of a dominating measure Q. By (2.4) we have

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

Assume (2.1) and (2.2). 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 $\inf_t I(\alpha^t|\alpha) > 0$ Q-a.s., $\sup_t I(\alpha^t|\alpha) < \infty$ Q-a.s. and $e^{-I(\alpha^\cdot|\alpha)}$ is a $(\bar{P}_\alpha, F)$-supermartingale of class (D) (see section 6.1 in Part I.

4.2 Information in the prior given the posterior

The previous considerations rely on the condition (2.2) 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, e.g. in Part I, the *relative entropy* in $\mathbf{P}$ given $\mathbf{Q}$. Contrary to (2.2), 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 according to proposition 2.2 in Part I).

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 (2.4) 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 4.1. *Assume is (2.1), (2.2), the representation property (3.6) 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

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

with

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

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 (3.8).

Proof. Equation (4.7) has already been obtained in Part I as theorem 6.3. Then, in view of identity (4.2) it suffices to substitute Q in (3.20) 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 3.5. 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 (3.9). To this end use (3.11) and (3.23) and determine $\Delta N(\alpha, Q)/(1 - \Delta h(\alpha))$ and $\Delta N(\alpha, \bar{P}_\alpha)/(1 - \Delta h(\alpha))$ from (3.20) and (4.8), respectively, by following the same device as in the course of proving corollary 3.5. $\square$

Remark 4.2. Under the conditions of theorem 4.1 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 (2.9) as follows: $\bar{v}(\beta) = E_{\alpha^-}(\beta(\vartheta) - \bar{\beta})^2$. In the present circumstances we can make the assertion of theorem 6.4 in Part I more precise. In fact, we get the Doob-Meyer decomposition of the of the information from data process.

Theorem 4.3. *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*

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

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

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

By applying theorem 6.4 in Part I to the present situation we 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 $\sum_{s\le\cdot} E_{\alpha^{s-}} \ell(1+\Delta m_s\left(\vartheta, \bar{P}_\alpha\right))$ with $\ell(x) = x\log x$ (the latter comes from the second term in formula (6.5) of Part I). 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 3.7 – substitute f in its assertion by ℓ to see that this compensator is given by $E_{\alpha^-}\ell(Y(\vartheta)/\bar{Y}) \cdot \bar{\nu} + \sum_{s\le\cdot} E_{\alpha^{s-}}\ell\left((1-\hat{Y}_s(\vartheta))/(1-\hat{\bar{Y}}_s)\right)(1-\hat{\bar{Y}}_s)$, 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 4.4. We obtain from theorem 4.3 that the expected utility from data at the stopping time T equals to

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

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 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 (3.5), 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 [6, 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 (2.2) 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 (3.25) 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 $E_Q h_\infty(\alpha) \leq -E_Q \sum_{n=1}^\infty \log \gamma_n(\alpha, Q) < \infty$ 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

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

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 [7] 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 (2.4). 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 (2.2) 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 (3.27). 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_θ+Q-a.s. for all $\theta \in \Theta$ (see [6, 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 (2.2) 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 5.15 in Part I. According to (3.22) 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 (2.4) 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 4.4, the expected utility from the present data equals to

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

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 [11]. In this special case the above expression for the expected utility from data reduces to equation (19.132) in [11]. The latter expression was derived in [11] 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 $t \geq 0$

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

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 $\int_0^\infty \beta_s^2(\theta) ds < \infty \;\; P_\theta + Q$-a.s. for all $\theta \in \Theta$ (see [6, 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 (2.2) 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 4.3 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 5.15 in Part I) 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 3.14 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 (2.4) 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 4.3 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 [14] and [11]. In fact, [11, equation (16.65)] is nothing else but $E_{\bar P_\alpha} I(\alpha|\alpha^T)$, although the context is different. In [11] 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 [12]:

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 [12, 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 [12] 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

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

Remark 5.2. Consider for each t the equation

$$\int_0^t s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H}\beta(s)ds = \int_0^t \rho(s)s^{1-2H}ds. \tag{5.11}$$

So we look at (5.10), but suppress dependence on θ and have the constant in front of the integral absorbed by ρ. In general there is no closed form solution to this equation, but an integral representation is available. Rewrite the equation as

$$t^{2-2H}\int_0^1 u^{\frac{1}{2}-H}(1-u)^{\frac{1}{2}-H}\beta(tu)du = \int_0^t \rho(s)s^{1-2H}ds.$$

As an intermediate step (assuming for a while that β is differentiable and ρ is continuous), differentiate both sides of (5.11) with respect to t to get after rearrangement of terms a solution for ρ as the sum of the integral terms

$$\rho(t) = (2-2H)\int_0^1 u^{\frac{1}{2}-H}(1-u)^{\frac{1}{2}-H}\beta(tu)du + t\int_0^1 u^{\frac{3}{2}-H}(1-u)^{\frac{1}{2}-H}\dot\beta(tu)du.$$

For $H < \frac{1}{2}$ we use integration by parts to arrive at the simpler expression

$$\rho(t) = (\tfrac{1}{2}-H)\int_0^1 u^{\frac{1}{2}-H}(1-u)^{-\frac{1}{2}-H}\beta(ut)du.$$

This expression for ρ holds in general if $H < \frac{1}{2}$, provided that the last integral is well defined. The solution ρ has in this case the following simple interpretation: if U has a $B(\frac{3}{2}-H, \frac{1}{2}-H)$ distribution, we have $\rho(t) = (\frac{1}{2}-H)B(\frac{3}{2}-H, \frac{1}{2}-H)E\beta(tU)$. An explicit solution can be obtained if β is a linear combination of functions of the form $t \mapsto t^\gamma$. Substituting this directly into (5.11), one sees that the solution is then the same linear combination of $t \mapsto (2-2H+\gamma)B(\frac{3}{2}-H+\gamma, \frac{3}{2}-H)t^\gamma$. In particular, an analytic function β results in an analytic solution ρ.

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 [6, Theorem IV.2.1]). Switching to $\bar P_\alpha$ as the dominating measure, we likewise obtain

$$z(\theta, \bar P_\alpha) = \mathcal{E}((\rho(\theta)-\bar\rho)\cdot \bar M), \tag{5.12}$$

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 (2.2) 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 5.15 in Part I.

According to corollary 3.14 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 (2.4), the equation (5.12) 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 4.3 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] 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.

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

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

[7] 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.

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

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

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

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

[12] 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.

[13] 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.

[14] 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.

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

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