Теор. ймовірност. та матем. статист.
Вип. 67, 2002, стор. 38–56

Teor. Ĭmovirnost. ta Matem. Statyst.
No. 67, 2002, pp. 38–56

INFORMATION CONCEPTS IN FILTERED EXPERIMENTS

UDC 519.21

K. DZHAPARIDZE, P. SPREIJ, AND E. VALKEILA

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.

1. Introduction

In this paper we study the Kullback–Leibler information between a posterior and a prior distribution on an abstract parameter space. First we consider static randomized experiments. Then we analyze the dynamics of Kullback–Leibler information in experiments equipped with a filtration.

For a proper analysis of the dynamical information 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. We give some extremal properties of the arithmetic and geometric mean measures and the related geodesic measure. To understand the dynamical behavior 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. 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] and [8].

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 we 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. The results of the present paper are of a general nature. For applications to semimartingale experiments we refer to [8].

1991 *AMS Mathematics Subject Classification.* Primary 60G07, 60H30, 62B15, 94A17.

Key words and phrases. Kullback–Leibler information, relative entropy, information from data, expected utility from data, Hellinger processes and Hellinger integrals, posterior and prior distributions.

The first and the second authors were partially supported by the Academy of Finland Project No. 37689, and the third author was supported by the Research Grants Committee of the University of Helsinki.

2. Randomized experiments

2.1. Statistical experiment. We consider a *statistical experiment* $(\Omega, \mathcal{F}, \{\mathsf{P}_\theta\}_{\theta\in\Theta})$, where $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$ 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 $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$ is equivalent to a certain probability measure Q, i.e.,

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

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

$$p_\theta = \frac{d\mathsf{P}_\theta}{dQ}. \tag{2.2}$$

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

$$\mathsf{P}_\theta(B) = \int_B p_\theta(\omega)\, Q(d\omega) = \mathsf{E}_Q\{\mathbb{I}_B p_\theta\}. \tag{2.3}$$

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 $(\boldsymbol{\Omega}, \mathfrak{F}, \mathbf{Q})$ of two probability spaces $(\Omega, \mathcal{F}, Q)$ and $(\Theta, \mathcal{A}, \alpha)$, where $\boldsymbol{\Omega} = \Omega \times \Theta$, $\mathfrak{F} = \mathcal{F} \otimes \mathcal{A}$, and $\mathbf{Q} = Q \times \alpha$. Along with $\mathbf{Q}$ define on $(\boldsymbol{\Omega}, \mathfrak{F})$ another probability measure $\mathbf{P}$ as follows: for each $\mathbf{B} \in \mathfrak{F}$

$$\mathbf{P}(\mathbf{B}) = \int_{\mathbf{B}} p_\theta(\omega)\, Q(d\omega)\, \alpha(d\theta) \doteq \mathsf{E}_{\mathbf{Q}}\{\mathbb{I}_{\mathbf{B}} p\} \tag{2.4}$$

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

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

(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 $\mathsf{P}_\theta(\Omega) = 1$), 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 $\mathbf{B} = B \times A \in \mathfrak{F}$

$$\mathbf{P}(\mathbf{B}) = \int_A p_\theta(B)\, \alpha(d\theta) = \mathsf{E}_\alpha\left\{\mathbb{I}_A\, \mathsf{E}_Q\{\mathbb{I}_B p\}\right\} = \mathsf{E}_Q\left\{\mathbb{I}_B\, \mathsf{E}_\alpha\{\mathbb{I}_A p\}\right\},$$

since by the Fubini theorem for positive functions (e.g. Loève [15, 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])

$$I(\mathbf{P}|\mathbf{Q}) = \mathsf{E}_{\mathbf{Q}} \log \frac{d\mathbf{Q}}{d\mathbf{P}} = -\,\mathsf{E}_{\mathbf{Q}} \log p \tag{2.6}$$

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

$$d\mathbf{Q}/d\mathbf{P} = (d\mathbf{P}/d\mathbf{Q})^{-1} \quad \mathbf{Q}\text{-a.s.},$$

so that $I(\mathbf{P}|\mathbf{Q}) = -\mathsf{E}_{\mathbf{Q}} \log\{d\mathbf{P}/d\mathbf{Q}\}$. 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 $I(\mathbf{P}|\mathbf{Q}) = \mathsf{E}_\alpha\, I(\mathsf{P}_\vartheta|Q)$, 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

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

(cf. proposition 2.2) with the density p as in (2.5), where we assume that $p \log p$ is integrable with respect to $\mathbf{Q}$. This Kullback–Leibler information in $\mathbf{Q}$ given $\mathbf{P}$ equals to $I(\mathbf{Q}|\mathbf{P}) = \mathsf{E}_\alpha\, I(Q|\mathsf{P}_\vartheta)$. 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. [13]). Instead of (2.7) we will sometimes assume that the average relative entropy is well-defined, 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 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 $(\Omega, \mathcal{F}, \{\mathsf{P}_\theta\}_{\theta\in\Theta}, Q)$. With the family of probability measures $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$ we associate a new measure defined on the same measurable space $(\Omega, \mathcal{F})$, the so-called *arithmetic mean measure* $\bar{P}_\alpha$: for each $B \in \mathcal{F}$

$$\bar{P}_\alpha(B) = \mathbf{P}(B \times \Theta) = \mathsf{E}_\alpha\, \mathsf{P}_\vartheta(B). \tag{2.10}$$

Let us also introduce the notation $a(\alpha, Q) = \mathsf{E}_\alpha\{p_\vartheta\}$ 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 $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$:

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. See [8, Lemma 2.1]. □

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 $\{\mathsf{P}_\theta\}_{\theta\in\Theta}$ in the following sense. One may choose for the probability measure that possesses the following extremal property: it minimizes $\mathsf{E}_\alpha\, I(Q|\mathsf{P}_\vartheta) = I(\mathbf{Q}|\mathbf{P})$ 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.

Proposition 2.2. *For all dominating measures Q verifying (2.1)*

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

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

Remark 2.3. The relation $I(\mathbf{Q}|\mathbf{P}) = I(Q|\bar{P}_\alpha) + I(\bar{\mathbf{P}}_\alpha|\mathbf{P})$ is of the same type as a so-called nonsymmetric Pythagorean theorem for orthoprojections obtained in [3, Chapter 22]. The first term equals $I(Q|\bar{P}_\alpha) = \mathsf{E}_{\bar{P}_\alpha} \ln\{d\bar{P}_\alpha/dQ\}$ which is well-defined by Jensen's inequality for the convex function $\ell(x) = x\ln x$ and by the equivalence $\bar{P}_\alpha \sim Q$.

2.4. Geodesic measure. In this section another candidate will be defined to serve as a dominating measure. This new measure, called below the *geodesic measure*, is closely related to the notion of the *Hellinger integral* of order α, which in analogy with [11, Section IV.1, formula 1.6], is denoted by $H(\alpha)$ and defined as the Q-expectation of the geometric mean of the density p_θ with respect to α. That is, $H(\alpha) = \mathsf{E}_Q\, g(\alpha, Q)$ where $g(\alpha, Q) = \exp\{\mathsf{E}_\alpha \log p_\vartheta\}$ is this geometric mean. In view of Jensen's inequality (applied to the concave function log), we see that $g(\alpha, Q) \le a(\alpha, Q)$, therefore $0 \le H(\alpha) \le 1$. 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 $\mathsf{E}_Q\, g(\alpha, Q) = \mathsf{E}_{Q'}\, g(\alpha, Q')$, since $\mathsf{E}_Q\, g(\alpha, Q) = \mathsf{E}_{Q'}\{Zg(\alpha, Q)\}$ and by definition of $g(\alpha, Q)$

$$Zg(\alpha, Q) = \exp\{\mathsf{E}_\alpha \log\{Zp(\vartheta, Q)\}\} = \exp\{\mathsf{E}_\alpha \log p(\vartheta, Q')\} = g(\alpha, Q'). \tag{2.11}$$

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 $\mathsf{E}_Q\, g(\alpha, Q) = \mathsf{E}_{Q'}\, g(\alpha, Q') = \mathsf{E}_{Q_0}\, g(\alpha, Q_0)$, which establishes the postulated independence of the choice of the dominating measure.

The geodesic measure mentioned at the beginning of this section, is denoted by $\check{C}_\alpha$ and defined for all $B \in \mathcal{F}$ as follows:

$$\check{C}_\alpha(B) = \frac{\int_B \exp\{\mathsf{E}_\alpha \log p_\vartheta(\omega)\}\, Q(d\omega)}{\int_\Omega \exp\{\mathsf{E}_\alpha \log p_\vartheta(\omega)\}\, Q(d\omega)} = \frac{\mathsf{E}_Q\{\mathbb{I}_B g(\alpha, Q)\}}{H(\alpha)}. \tag{2.12}$$

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 $\check{C}_\alpha$ that is absolutely continuous with respect to Q, having the density, proportional to the geometric mean,

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

Moreover, we have the following lemma.

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

Proof. We only have to prove $Q \ll \check{C}_\alpha$. If $\check{C}_\alpha(B) = 0$ for a $B \in \mathcal{F}$, then $g(\alpha, Q) = 0$ and $\mathsf{E}_\alpha \log p_\vartheta = -\infty$ on this B Q-a.s. But $\log p_\vartheta$ is integrable with respect to $\mathbf{Q}$ by assumption. Hence $Q(B) = 0$ by the Fubini theorem. The proof is complete. □

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

Proposition 2.5. *The following relation holds*

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

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

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

$$I(\check{C}_\alpha|Q) = -\mathsf{E}_Q \log \frac{d\check{C}_\alpha}{dQ} = -\mathsf{E}_Q \log g(\alpha, Q) + \log H(\alpha)$$

and then $I(\mathbf{P}|\mathbf{Q}) = -\mathsf{E}_Q\,\mathsf{E}_\alpha \log p_\vartheta = -\mathsf{E}_Q \log g(\alpha, Q)$. Equation (2.14) now follows.

Since $I(\check{C}_\alpha|Q) \geq 0$ 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 $Q = \check{C}_\alpha$. □

Remark 2.6. Equation (2.14) is equivalent to $I(\mathbf{P}|\mathbf{Q}) = I(\check{C}_\alpha|Q) + I(\mathbf{P}|\check{C}_\alpha \times \alpha)$. 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

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

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

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

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

$$\mathsf{E}_\alpha \log\{a(\alpha, Q)/p(\vartheta, Q)\} = -\,\mathsf{E}_\alpha \log p(\vartheta, \bar{P}_\alpha).$$

The $\bar{P}_\alpha$-expectation of the Kullback–Leibler information in β given α is then easily seen to be equal to $\mathsf{E}_\alpha\, I(\mathsf{P}_\vartheta|\bar{P}_\alpha)$, 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

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

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

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

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 $\mathcal{F}_0 = \{\varnothing, \Omega\}$ Q-a.s. The filtered probability space

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

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 $\mathsf{P}_{\theta,T}$ are then the restrictions of the measures Q and P_θ to the sub-σ-field $\mathcal{F}_T$. Since $\mathsf{P}_{\theta,T}$ is equivalent to Q_T for each $\theta \in \Theta$, we can define the Radon–Nikodym derivatives

$$z_T(\theta) = \frac{d\mathsf{P}_{\theta,T}}{dQ_T} = \mathsf{E}_Q\{p_\theta|\mathcal{F}_T\}.$$

Thus according to [11, 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 $z_t(\theta, Q) = d\mathsf{P}_{\theta,t}/dQ_t$ for all $t \geq 0$, which possesses the following properties (see [11, 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 $\mathsf{E}_Q\{z_t(\theta, Q)\} = 1$, for all $t \in [0, \infty]$.

Later on in this paper we will also encounter various Kullback–Leibler information numbers like $I(\mathsf{P}_{\theta,T}|Q_T)$, 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

$$a(\alpha, Q) = \mathsf{E}_\alpha\, z(\vartheta, Q) \tag{4.1}$$

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

$$z_t(\bar{P}_\alpha, Q) = \mathsf{E}_Q\left\{\frac{d\bar{P}_\alpha}{dQ}\Big|\mathcal{F}_t\right\}$$

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) $\inf_t a_t > 0$ *Q-a.s.*
(ii) $\sup_t a_t < \infty$ *Q-a.s.*
(iii) *a is a (Q, F)-uniformly integrable martingale with $\mathsf{E}_Q\, a_t = 1$ for all $t \geq 0$.*

Proof. In view of lemma 2.1 it suffices to refer again to [11, 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 β^T by the Bayes formula: for all $A \in \mathcal{A}$

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

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

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

cf. (2.15) and (2.15). Obviously, the posterior β^T so defined is free of the choice of a dominating measure Q. Note that for fixed $A \in \mathcal{A}$ the random variable $\beta^T(A)$ 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 $\theta \in \Theta$

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

cf. (2.18). In particular, $\mathsf{E}_\alpha\, z_T(\vartheta, \bar{P}_\alpha) = a(\alpha, \bar{P}_\alpha) = 1$. 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

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

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 many models of practical interest are of this particular form, the examples can be found in [8] (the specification (4.7), however, is superfluous throughout most of the present paper, except in the present and in the concluding section 6.2). 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 $\bar{c} \doteq \mathsf{E}_{\beta-}\, c(\vartheta)$ and $\bar{d} \doteq \mathsf{E}_{\beta-}\, d(\vartheta)$. 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 $\mathsf{E}_Q\, \mathsf{E}_\alpha[z(\vartheta, Q)]_\infty^{1/2} < \infty$. 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 $\mathsf{E}_Q\, \mathsf{E}_\alpha[z(\vartheta, Q)]_\infty^{1/2}$ by $C(1 + \mathsf{E}_Q\, \mathsf{E}_\alpha\, z_\infty(\vartheta, Q) \log^+ z_\infty(\vartheta, Q))$ which is finite in view of (2.9) and the trivial relation $x \log^+ x \le x \log x + e^{-1}$.

Hence we obtain the integrands $a_-(\alpha, Q)\, \bar{c}$ and $a_-(\alpha, Q)\, \bar{d}$, since by (4.5) we have

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

that indeed yields the first integrand by the definition of $\bar{c}$. The second one is obtained in a 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 $a(\alpha, Q) = \mathcal{E}(\bar{m})$. $\square$

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 $d\beta^{\cdot}/d\alpha$ is the Doléans exponential*

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

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 $N^d = M^d - (1 + \Delta\bar{m})^{-1} \cdot [\bar{m}, M^d]$.

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

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

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 $n = m - \bar{m} - (1 + \Delta\bar{m})^{-1}\, [\bar{m}, m - \bar{m}]$. This plainly holds always, irrespective the specifications (4.7) and (4.8) that gives n the asserted form (4.10). $\square$

Note the following relation

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

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*

$$g(\alpha, Q) = \exp\{\mathrm{E}_\alpha\{\log z(\vartheta, Q)\}\}, \tag{5.1}$$

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 $\mathrm{E}_Q\{g_t|\mathcal{F}_s\} \leq g_s$ 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

$$\begin{aligned}\mathrm{E}_Q\left\{\frac{g_t}{g_s}\Big|\mathcal{F}_s\right\} &= \mathrm{E}_Q\left\{\exp\left\{\mathrm{E}_\alpha \log\frac{z_t(\vartheta, Q)}{z_s(\vartheta, Q)}\right\}\Big|\mathcal{F}_s\right\} \leq \mathrm{E}_Q\left\{\mathrm{E}_\alpha\frac{z_t(\vartheta, Q)}{z_s(\vartheta, Q)}\Big|\mathcal{F}_s\right\}\\ &= \mathrm{E}_\alpha\left\{\mathrm{E}_Q\frac{z_t(\vartheta, Q)}{z_s(\vartheta, Q)}\Big|\mathcal{F}_s\right\} = 1. \qquad \square\end{aligned}$$

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 $\{\mathrm{P}_\theta\}_{\theta\in\Theta}$ 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 $\{\mathrm{P}_{\theta,T}\}_{\theta\in\Theta}$, is defined according to [11, Section IV.1], as the Q-expectation of the g-mean process evaluated at T:

$$H_T(\alpha) = \mathrm{E}_Q\{g_T(\alpha, Q)\}. \tag{5.4}$$

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. See [8, Theorem 5.2]. □

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:

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

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

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

valid on the set $\{T < \infty\}$ with a predictable time T, since by the predictable section theorem I.2.18 in [11] 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

$$\Delta h(\alpha)_T = -\mathsf{E}\left\{\frac{\Delta g_T}{g_{T-}} \Big| \mathcal{F}_{T-}\right\}$$

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. [14, p. 199].

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

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

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 $g_t(\alpha, Q) = \exp \mathsf{E}_\alpha \log p_\vartheta$. In this case it follows from equations (5.4) and (5.5) that $h_t(\alpha) = (1 - H_T(\alpha))\mathbb{I}_{\{t \geq T\}}$. 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 $a(X) = \mathsf{E}_\alpha X(\vartheta)$ and (for a nonnegative family) $g(X) = \exp\{\mathsf{E}_\alpha \log X(\vartheta)\}$ 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 $a(X) = \mathsf{E}_\alpha X(\vartheta)$ be a (Q,F)-semimartingale and $a_-(X) = \mathsf{E}_\alpha X_-(\vartheta)$. Suppose that the increasing processes $a(\langle X^c\rangle)$ and*

$$a\left(\sum_{s\le\cdot}(\Delta X_s - \log(1+\Delta X_s))\right)$$

are finite-valued.

Then the g-mean process $g(\mathcal{E}) = \exp \mathsf{E}_\alpha\{\log \mathcal{E}(X(\vartheta))\}$ of the family of the Doléans exponentials $\{\mathcal{E}(X(\theta))\}_{\theta\in\Theta}$ is well-defined and

$$g(\mathcal{E}) = \mathcal{E}\left\{a(X) - \frac{1}{2}\tilde{v}(X^c) - \sum_{s\le\cdot}\phi_s(1+\Delta X)\right\}, \tag{5.11}$$

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

For the proof we refer to [7, Proposition 4.5].

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

$$v(X^c) \doteq \operatorname*{var}_\alpha(X(\vartheta)^c) = \mathsf{E}_\alpha\left\{|X(\vartheta)^c|^2\right\} - \left|\mathsf{E}_\alpha\left\{X(\vartheta)^c\right\}\right|^2 \tag{5.12}$$

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

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

which is reduced in the special binary case to the Itô formula (cf. [11, p. 199]).

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*

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

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

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

Proof. It holds that $\mathsf{E}_\alpha\,\mathsf{E}_Q\{m(\vartheta,Q)^c\}^2<\infty$, 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

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

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

$$\mathsf{E}_Q\left\{\frac{\Delta g_T}{g_{T-}}\Big|\mathcal{F}_{T-}\right\}=-\,\mathsf{E}_Q\{\phi_T(1+\Delta m)|\mathcal{F}_{T-}\},$$

which by definition of ϕ is nothing else but $\mathsf{E}_Q\{g_T(1+\Delta m)|\mathcal{F}_{T-}\}-1$. 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

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

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. To this end the multiplicative decomposition theorem of a nonnegative supermartingale (cf. [14, Theorem 2.5.1]) 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, cf the right-hand side of (2.13) where the ratio $g(\alpha,Q)/\,\mathsf{E}_Q\,g(\alpha,Q)$ occurs. However remark 5.7 suggests an alternative, namely, the normalization with $\mathcal{E}(-h(\alpha))$. These two normalizations coincide only in the case of a deterministic Hellinger process when $\mathsf{E}_Q\,g(\alpha,Q)=H(\alpha)=\mathcal{E}(-h(\alpha))$, cf. remark 5.6.

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

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

Proof. Apply theorem 2.5.1 of [14] 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 $\mathsf{E}_Q\, \zeta(\alpha, Q) \equiv 1$. 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 [12] that a nonnegative local martingale is in fact a martingale. So working in discrete time one obtains $\mathsf{E}_Q\, \zeta(\alpha, Q) \equiv 1$, see the examples in [8] for other possibilities.

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

$$\mathsf{E}_Q\{\zeta_\infty(\alpha, Q)|\mathcal{F}_t\} = \zeta_t(\alpha, Q).$$

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 $G_\alpha(B) = \mathsf{E}_Q\, \mathbb{I}_B \zeta_\infty(\alpha, Q)$. 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 and we put $\frac{0}{0} = 0$.

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 is aimed at the applications in the spirit of [8]. 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 $\mathsf{E}_{\bar{P}_\alpha}\{1/\mathcal{E}(-h(\alpha))_\infty\} < \infty$. 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 the equations (5.16) and (5.18) we obtain $\zeta(\alpha, Q) = \mathcal{E}(m^{G_\alpha})$, with

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

If the discontinuous part of $a(m)$ is a sum of jumps martingale, we can say more:

$$m^{G_\alpha} = a(m^c) + \sum_{s \leq \cdot} \left(\frac{g_s(1+\Delta m)}{1-\Delta h_s(\alpha)} - 1 \right).$$

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. *Along with the conditions (2.1) and (2.7), assume that G_α is a probability measure and that it is used as a dominating measure. Then the following multiplicative decomposition holds for the geometric mean process $g(\alpha, Q)$:*

$$g(\alpha, Q) = z(G_\alpha, Q)\,\mathcal{E}(-h(\alpha)), \tag{5.20}$$

in particular, $g(\alpha, G_\alpha) = \mathcal{E}(-h(\alpha))$. Besides, at each stopping time T the Hellinger integral and the Hellinger process are related as follows:

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

Proof. The representation (5.20) immediately follows from the remark 5.16 that $z(G_\alpha, Q)$ is nothing else but the $\zeta(\alpha, Q)$ of theorem 5.13. The rest follows upon substitution of Q in (5.20) and 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 $\mathsf{P}_{\theta,T}$ 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

$$K_T(Q_T) = \mathsf{E}_\alpha\, I(\mathsf{P}_{\vartheta,T}|Q_T) + \mathsf{E}_Q \log \mathcal{E}(-h(\alpha))_T \tag{5.21}$$

as the criterion to be minimized and notice the difference (putting $T = \infty$) with the criterion $\mathsf{E}_\alpha\, I(\mathsf{P}_\vartheta|Q)$ 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

$$\mathsf{E}_\alpha\, I(\mathsf{P}_{\vartheta,T}|Q_T) = -\,\mathsf{E}_Q \log g_T(\alpha, Q) = -\,\mathsf{E}_Q \log \zeta_T(\alpha, Q) - \mathsf{E}_Q \log \mathcal{E}(-h(\alpha))_T$$

in view of the definition of $\zeta(\alpha, Q)$, which gives (5.21) with

$$K_T(Q_T) = -\,\mathsf{E}_Q \log \zeta_T(\alpha, Q) = -\,\mathsf{E}_Q \log \frac{dG_{\alpha,T}}{dQ_T}.$$

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

$$K_T(Q_T) = -\,\mathsf{E}_Q \log \frac{dG_{\alpha,T}}{dQ_T}$$

(see the proof of (i)) we obtain by Jensen's inequality applied to the convex function $-\log$ that

$$K_T(Q_T) \geq -\log \mathsf{E}_Q \frac{dG_{\alpha,T}}{dQ_T} = -\log G_{\alpha,T}(\Omega).$$

If the infimum is zero, then for each $n \in \mathbf{N}$ there exists a probability measure Q_n on $\mathcal{F}_T$ such that $K_T(Q_n) < n^{-1}$. Hence we have $-\log G_{\alpha,T}(\Omega) < n^{-1}$ 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*. Note once again that now $H(\alpha) = \mathcal{E}(-h(\alpha))$. 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 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

$$\mathsf{E}_\alpha\, I_T(\mathsf{P}_\vartheta|Q) = I_T(G_\alpha|Q) - \log H_T(\alpha),$$

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 $\check{C}_\alpha$, both measures being defined in terms of certain normalizations of the geometric mean of densities (to see the difference, compare equations (2.13) and (5.19)). It is therefore interesting to relate 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 $\check{C}_\alpha$.

Recall from section 2.4 that $\check{C}_\alpha$ 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 $z_\infty(\check{C}_\alpha, Q)$. Note that $0 < g_\infty(\alpha, Q) < \infty$ Q-a.s.

We have seen in section 2.4 that $\check{C}_\alpha \sim Q$, in other words that $z(\check{C}_\alpha, Q)$ is Q-a.s. finite and positive. By $z_t(\check{C}_\alpha, Q)$ we denote the density of the restriction of $\check{C}_\alpha$ w.r.t. the restriction of Q to $\mathcal{F}_t$. Clearly,

$$z_t(\check{C}_\alpha, Q) = \mathsf{E}\{z_\infty(\check{C}_\alpha, Q)|\mathcal{F}_t\} = \mathsf{E}_Q\{g_\infty(\alpha, Q)|\mathcal{F}_t\}/H_\infty(\alpha).$$

Now we consider the process $\zeta(\alpha, \check{C}_\alpha)$. It satisfies

$$\zeta(\alpha, \check{C}_\alpha) = \zeta(\alpha, Q) z_t(Q, \check{C}_\alpha) = \frac{g(\alpha, Q)}{\mathcal{E}(-h(\alpha))} \frac{H_\infty(\alpha)}{\mathsf{E}_Q\{g_\infty(\alpha, Q)|\mathcal{F}.\}}. \tag{5.22}$$

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 [11, p. 209] for the binary case. From (5.22) we see, letting $t \to \infty$, that

$$\zeta_\infty(\alpha, \check{C}_\alpha) = \frac{H_\infty(\alpha)}{\mathcal{E}(-h(\alpha))_\infty}. \tag{5.23}$$

Observe that this random variable has $\check{C}_\alpha$-expectation less than or equal to 1 and hence

$$\mathsf{E}_{\check{C}_\alpha} \frac{1}{\mathcal{E}(-h(\alpha))_\infty} \leq \frac{1}{H_\infty(\alpha)}.$$

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 if and only if*

$$\mathsf{E}_{\check{C}_\alpha} \frac{1}{\mathcal{E}(-h(\alpha))_\infty} = \frac{1}{H_\infty(\alpha)}. \tag{5.24}$$

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

$$z_t(G_\alpha, \check{C}_\alpha) = H_\infty(\alpha)\, \mathsf{E}_{\check{C}_\alpha} \left\{ \frac{1}{\mathcal{E}(-h(\alpha))_\infty} \Big| \mathcal{F}_t \right\}. \tag{5.25}$$

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

$$z_t(\check{C}_\alpha, G_\alpha) = \frac{1}{H_\infty(\alpha)}\, \mathsf{E}_{G_\alpha}\{\mathcal{E}(-h(\alpha))_\infty | \mathcal{F}_t\}. \tag{5.26}$$

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

$$\mathsf{E}_{\check{C}_\alpha}\, \zeta_\infty(\alpha, \check{C}_\alpha) = \mathsf{E}_Q\, \zeta_\infty(\alpha, Q) = G_\alpha(\Omega)$$

and expression (5.23). The validity of (5.25) also follows from (5.23) and (5.26) follows from (5.25). □

Remark 5.20. Equation (5.25) again illustrates the fact that the measures G_α and $\check{C}_\alpha$ 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 β^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 β^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 β^T with respect to the prior α by

$$I(\beta^T | \alpha) = \mathsf{E}_\alpha \log \frac{d\alpha}{d\beta^T}. \tag{6.1}$$

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

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

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

Observe again that the information $I(\beta^T|\alpha)$ 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 the following proposition.

Proposition 6.2. *Assume (2.1) and (2.7). Let $I(\beta^\cdot|\alpha)$ be the information process starting from zero, $I(\beta^0|\alpha) = 0$, and at $t > 0$ defined by (6.1). Then it possesses the following properties:*

(i) $\inf_t I(\beta^t|\alpha) > 0$ *Q-a.s.*
(ii) $\sup_t I(\beta^t|\alpha) < \infty$ *Q-a.s.*
(iii) $\exp\{-I(\beta^\cdot|\alpha)\}$ *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 $I(\beta^T|\alpha)$ at a stopping time $T > 0$ can be presented as follows:*

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

Proof. Combine (5.20) and (6.2). □

6.2. The information from data. According to (2.20) the relative entropy in the posterior β^T at a stopping time T with respect to a prior α is defined by

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

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

$$\mathsf{E}_{\bar{P}_\alpha}\, I\left(\alpha|\beta^T\right) = \mathsf{E}_\alpha\, I(\bar{P}_{\alpha,T}|P_{\vartheta,T}) \tag{6.4}$$

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 $I(\alpha|\beta^\infty) = \mathsf{E}_\alpha\, p(\vartheta, \bar{P}_\alpha) \log p(\vartheta, \bar{P}_\alpha)$ 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 $I(\alpha|\beta^\cdot) = \mathsf{E}_\alpha \ell(z(\vartheta, \bar{P}_\alpha))$. 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 $\mathsf{E}_{\bar{P}_\alpha}\, \mathsf{E}_\alpha\, |\ell(z(\vartheta, \bar{P}_\alpha))| < \infty$. Since the function ℓ on $\mathbf{R}_+$ 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 $I(\alpha|\beta^\cdot)$ is a $(\bar{P}_\alpha, F)$-submartingale with limit $I(\alpha|\beta)$, satisfying $0 \leq I(\alpha|\beta^t) \leq \mathsf{E}_{\bar{P}_\alpha}\{I(\alpha|\beta)|\mathcal{F}_t\}$ 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: $\bar{v}(c) = \mathsf{E}_{\beta\text{-}}(c(\vartheta) - \bar{c})^2$. Secondly, we require the posterior moment $\mathsf{E}_{\beta\text{-}}\,\ell(1 + (d(\vartheta) - \bar{d})\Delta N)$ with N as in corollary 4.3.

We also use the following notation. Define for a process X depending on θ the process $\bar{\ell}(X) = \mathsf{E}_{\beta\text{-}}\,\ell(X(\theta))$. With this notation the latter central posterior moment is expressed as $\bar{\ell}(X) = \mathsf{E}_{\beta\text{-}}\,\ell(1 + \Delta n(\cdot, \bar{P}_\alpha))$, cf. (4.11). From lemma 2.1 we know that the argument of ℓ is strictly positive. Moreover, invoking once more Jensen's inequality one sees that this posterior moment is nonnegative.

Theorem 6.4. *Along with the conditions (2.1), and (2.9), assume the representation property (4.7). Then the difference between the information from data $I(\alpha|\beta^{\cdot})$ and the nondecreasing finite-valued process*

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

is a $(\bar{P}_\alpha, F)$-martingale (cf. the expression (4.11) for $1 + \Delta n_s(\cdot, \bar{P}_\alpha)$).

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 $I(\alpha|\beta^{\cdot}) = \mathsf{E}_\alpha\,\ell(z)$. 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

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

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

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

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 $\mathsf{E}_\alpha\, z_- \cdot n = 0$ and $\mathsf{E}_\alpha \sum_{s\leq \cdot} z_{s-} \Delta n_s = 0$ (this follows e.g. from $\mathsf{E}_\alpha\, z = 1$, 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

$$\mathsf{E}_\alpha\, z_- \cdot \log z = -\frac{1}{2}\bar{v}(c) \cdot \langle M^c \rangle + \sum_{s\leq \cdot} \mathsf{E}_{\beta^{s}\text{-}} \log(1 + \Delta n_s)$$

and

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

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

$$\begin{aligned} \mathsf{E}_{\bar{P}_\alpha} I(\alpha|\beta^T) &= \mathsf{E}_\alpha I(\bar{P}_{\alpha,T}|P_{\vartheta,T}) \\ &= \mathsf{E}_{\bar{P}_\alpha} \left\{ \frac{1}{2}\bar{v}(c)\cdot\langle M^c\rangle + \sum_{s\le\cdot} \bar{\ell}_s\left(\Delta N_s^d\right) \right\}. \end{aligned}$$

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

Bibliography

1. J. M. Bernardo and A. F. M. Smith, *Bayesian theory*, Wiley, Chichester, 1994.
2. K. Bichteler and S. J. Lin, *On the stochastic Fubini theorem*, Stochastics and Stochastics Reports **54** (1995), 271–279.
3. N. N. Čencov, *Statistical decision rules and optimal inference*, Translations of Mathematical Monographs, American Mathematical Society, Providence, 1980.
4. I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Prob. **3** (1975), 146–158.
5. C. Dellacherie and P. A. Meyer, *Probabilities and potential B*, North Holland, Amsterdam, 1982.
6. P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, Wiley, New York, 1997.
7. K. Dzhaparidze, P. J. C. Spreij, and E. Valkeila, *On Hellinger processes for parametric families of experiments*, Statistics and control of stochastic processes. The Liptser festschrift (Yu. Kabanov, B. Rozovskii, and A. Shiryaev, eds.), World Scientific, Singapore, 1997, pp. 41–62.
8. ———, *Information processes for semimartingale experiments*, Ann. Probab. (2002) (to appear).
9. B. Grigelionis, *Hellinger integrals and Hellinger processes for solutions of martingale problems*, Prob. Theory and Math. Stat. (B. Grigelionis et al., eds.), vol. 1, VSP/Mokslas, Utrecht/Vilnius, 1990, pp. 446–454.
10. J. Jacod, *Calcul stochastique et problèmes de martingales*, Springer, Berlin–Heidelberg–New York, 1979.
11. J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, Springer, Berlin–Heidelberg–New York, 1987.
12. ———, *Local martingales and the fundamental asset pricing theorems in the discrete-time case*, Finance and Stochastics **2** (1998), № 3, 259–273.
13. S. Kullback, *Information theory and statistics*, Wiley, New York, 1959.
14. R. Sh. Liptser and A. N. Shiryayev, *Theory of martingales*, Kluwer, Dordrecht, 1986.
15. M. Loève, *Probability theory*, Springer, New York–Heidelberg, 1977, 1978.

Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam
E-mail address: kacha@cwi.nl

Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam
E-mail address: spreij@science.uva.nl

Department of Mathematics, University of Helsinki, Yliopistonkatu 5, Helsinki
E-mail address: Esko.Valkeila@Helsinki.Fi

Received 28/01/2002