Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

Note on randomized filtered experiments in presence of nuisance parameters

K.O. Dzhaparidze

REPORT PNA-R0119 OCTOBER 31, 2001

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

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

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

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

ISSN 1386-3711

Note on Randomized Filtered Experiments in Presence of Nuisance Parameters

K. Dzhaparidze
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
kacha@cwi.nl

ABSTRACT
In this report we define randomized filtered experiments with an abstract parameter space consisting not only parameters of interest but also nuisance parameters. The generalized Hellinger process, introduced in [1] to characterize the experiment, is shown to decompose in two components associated separately with the set of parameters of interest and with the set of nuisance parameters. Explicit expressions are presented for these characteristics in the case of general semimartingale observations and in the case of multivariate counting processes.

2000 Mathematics Subject Classification: 60G07, 60H30, 62B15.
Keywords and Phrases: statistical experiment, randomized filtered experiment, Hellinger process, parameter of interest, nuisance parameter, semimartingale, point process.
Note: Work carried out under the project PNA3.3, 'Stochastic Processes and Applications'.

1 Introduction

This report is a follow-up of the seminal paper [1] where the dynamics of the filtered statistical experiments has been characterized by so-called Hellinger processes indexed by *a priori* probability distributions on an abstract parametric space. The study has been set forth in the recent reports [2] and [3] from the different, information-theoretical point of view, while the present note focuses exclusively on one particular aspect arisen in the situation in which the parametric space consists of two disjoint subsets consisting of parameters of interest and of nuisance parameters. This setup is quite common in statistical literature. We will show how the Hellinger process decomposes in two components associated separately with the set of parameters of interest and with the set of nuisance parameters, see theorem 4.1 and corollary 4.2.

The study of Hellinger integrals and Hellinger processes started in the series of papers [9], [10] and [11]. This theory took a complete form in the book [8] where the notions of Hellinger integrals and Hellinger processes were fully exploited. The focus in these early works was on binary experiments. In the consequent papers [6] and [7] some of the results were generalized to a filtered experiment with a finite number of probability measures; cf. also [4] where some additional aspects of the latter experiment are discussed. As was already pointed out, these results were extended to an arbitrary parameter space in [1] and in the reports [2] and [3]. After the introduction, we reproduce some necessary information from these papers, with further references for more details. The present specification of the parameter space Θ with parameters of interest and nuisance parameters occurs in section 2.3 where some useful formulas are gathered concerning the arithmetic and geometric means in this new situation. In section 3 the conditional Hellinger process is defined completely in parallel to its unconditional counterpart originated in [1]. Section 4 is devoted to main results of this report, theorem 4.1 and corollary 4.2 mentioned above. In section 5 applications are discussed. It is assumed in section 5.1 that the observed process is a certain semimartingale and all representations are given in terms of the associated triplet of predictable characteristics. In the concluding section 5.2 observations are specified to come from a multivariate counting process.

2 Randomized experiments

2.1 Basic conditions

We consider a *statistical experiment* $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P} = \{P_\theta, \theta \in \Theta\}$ 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 $\mathbb{P}$ 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 $B \in \mathcal{F}$

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

Here and elsewhere below we use the expectation sign E indexed by a probability measure.

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.

Observe that in the present setup the probability measure P_θ defined for each $\theta \in \Theta$ by (2.3) (and satisfying $P_\theta(\Omega) = 1$), may be viewed as a regular conditional probability measure, under the condition that the statistical parameter ϑ takes on the particular value θ.

Along with condition (2.1), it is required in paper [1] that the Kullback-Leibler information in P_θ given Q defined as usual by $I(P_\theta|Q) = \mathrm{E}_Q \log\{dQ/dP_\theta\} = -\mathrm{E}_Q \log p_\theta$, is integrable with respect to $\alpha(d\theta)$ and that the *average information* satisfies the condition

$$0 < \mathrm{E}_\alpha I(P_\vartheta|Q) < \infty. \tag{2.4}$$

2.2 Density processes

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 $(\Omega, \mathcal{F}, F, \mathbb{P}, Q)$ so defined, with $\mathbb{P} = \{P_\theta, \theta \in \Theta\}$ as in section 2.1, 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 both 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

$$z_T(P_\theta; Q) = \frac{dP_{\theta,T}}{dQ_T} = \mathrm{E}_Q\{p_\theta | \mathcal{F}_T\}$$

with p_θ as in (2.2). Thus according to [8, Section III.3], for each fixed $\theta \in \Theta$ there is a unique (up to Q-indistinguishability) process $z(P_\theta; Q)$ called the *density process* so that $z_t(P_\theta; Q) = dP_{\theta,t}/dQ_t$ for all $t \geq 0$. For each $\theta \in \Theta$, this density process is in fact a (Q, F)-uniformly integrable martingale with $\mathrm{E}_Q\{z_t(P_\theta; Q)\} = 1$ for all $t \in [0, \infty]$. See [8, Proposition III.3.5] for more details on this and the following two Q-a.s. boundedness properties: $\inf_t z_t(P_\theta; Q) > 0$ and $\sup_t z_t(P_\theta; Q) < \infty$.

2.3 Parameters of interest and nuisance parameters

As was already mentioned in the introduction, it will be assumed throughout that the parameter space Θ is the union of two disjoint sets Ξ and Υ, i.e. $\Theta = \Xi \cup \Upsilon$, and that the random parameter ϑ consist of two components $\vartheta = (\xi, \eta)$. The first component ξ, taking on the values from Ξ, will be interpreted as a parameter of interest and the second component η, taking on the values from Υ, as a nuisance parameter. Moreover, the prior distribution α is assumed to be defined on the parameter space Θ in the following special manner: for any set $A \cup B \in \Theta$ with $A \in \Xi$ and $B \in \Upsilon$

$$\alpha(A \cup B) = \int_B \chi_y(A) \upsilon(dy)$$

where υ is the marginal prior distribution of the nuisance parameter η (the restriction of the prior α to the subspace Υ), while χ_y is the conditional prior distribution of the parameter of interest ξ under the condition that the nuisance parameter η takes on a particular value $\eta = y \in \Upsilon$.

Let X be a certain function defined on the parametric space Θ possessing all moments required below. In [1] the arithmetic and geometric means with respect to the prior α have been defined as follows:

$$a^\alpha(X) = \mathrm{E}_\alpha X(\vartheta) = \int_\Theta X(\theta)\alpha(d\theta) \text{ and } g^\alpha(X) = e^{\mathrm{E}_\alpha \log X(\vartheta)} = e^{\int_\Theta \log X(\theta)\alpha(d\theta)}$$

(the latter case is restricted exclusively to positive X's) which in the present case means that

$$a^{\alpha}(X) = \mathrm{E}_{\upsilon}\,\mathrm{E}_{\chi_\eta} X(\xi,\eta) = \int_{\Upsilon} \upsilon(dx) \int_{\Xi} X(x,y)\chi_y(dx) \tag{2.5}$$

and

$$g^{\alpha}(X) = e^{\mathrm{E}_{\upsilon}\,\mathrm{E}_{\chi_\eta} \log X(\xi,\eta)} = e^{\int_{\Upsilon} \upsilon(dx) \int_{\Xi} \log X(x,y)\chi_y(dx)} \tag{2.6}$$

where E_{χ_η} denotes the conditional expectation with respect to χ_η. Moreover, the conditional arithmetic and geometric means may be defined in the similar manner by

$$a^{\chi_\eta}(X(\cdot,\eta)) = \mathrm{E}_{\chi_\eta} X(\xi,\eta) = \int_{\Xi} X(x,\eta)\chi_\eta(dx)$$

and

$$g^{\chi_\eta}(X(\cdot,\eta)) = e^{\mathrm{E}_{\chi_\eta} \log X(\xi,\eta)} = e^{\int_{\Xi} \log X(x,\eta)\chi_\eta(dx)}. \tag{2.7}$$

This allows us to rewrite (2.5) and (2.6) in the compact way:

$$a^{\alpha}(X) = a^{\upsilon}(a^{\chi_\cdot}(X)), \quad g^{\alpha}(X) = g^{\upsilon}(g^{\chi_\cdot}(X)). \tag{2.8}$$

Further on we will need the same notations as in [1]:

$$\mathrm{v}^{\alpha}(X) = a^{\alpha}(X^2) - a^{\alpha}(X)^2, \quad \phi^{\alpha}(X) = a^{\alpha}(X) - g^{\alpha}(X) \tag{2.9}$$

and

$$\mathrm{v}^{\chi_\eta}(X(\cdot,\eta)) {=} a^{\chi_\eta}(X(\cdot,\eta)^2) {-} a^{\chi_\eta}(X(\cdot,\eta))^2, \;\; \phi^{\chi_\eta}(X(\cdot,\eta)) {=} a^{\chi_\eta}(X(\cdot,\eta)) {-} g^{\chi_\eta}(X(\cdot,\eta)). \tag{2.10}$$

With these notations at hand one can easily verify that

$$\mathrm{v}^{\alpha}(X) = a^{\upsilon}(\mathrm{v}^{\chi_\cdot}(X)) + \mathrm{v}^{\upsilon}(a^{\chi_\cdot}(X)), \quad \phi^{\alpha}(X) = a^{\upsilon}(\phi^{\chi_\cdot}(X)) + \phi^{\upsilon}(g^{\chi_\cdot}(X)). \tag{2.11}$$

3 Geometric mean measures and processes

3.1 Multiplicative decomposition

In the present section we shall shortly review the part of the general theory concerning the so-called geometric mean process and its decompositions. For the details and proofs we refer to [1] and [2]. Consider again the parametric family of probability measures $\mathbb{P} = \{P_\theta; \theta \in \Theta\}$ and assume the conditions (2.1) and (2.4). Associate with the parametric family of density processes $\{z(P_\theta; Q), P_\theta \in \mathbb{P}\}$ a new process by taking the geometric mean

$$g^{\alpha}(\mathbb{P}; Q) = e^{E_\alpha \log z(P_\vartheta; Q)} = e^{\int_\Theta \log z(P_\theta;Q)\alpha(d\theta)}.$$

The process $g^{\alpha}(\mathbb{P}; Q)$, called the *geometric mean process*, is a (Q, F)-supermartingale of class (D) with $g_0(\mathbb{P}; Q) = 1$ (see [1, Proposition 4.1] for this and the following two Q-a.s. boundedness properties: $\inf_t g_t^{\alpha}(\mathbb{P}; Q) > 0$ and $\sup_t g_t^{\alpha}(\mathbb{P}; Q) < \infty$). Therefore, the Doob-Meyer decomposition of the geometric mean process is available and there exists a (unique up to Q-indistinguishability) predictable finite-valued increasing process $h^\alpha(\mathbb{P})$ starting from the origin $h_0^\alpha(\mathbb{P}) = 0$, so that the sum $g^\alpha(\mathbb{P};Q) + g_-^\alpha(\mathbb{P};Q) \cdot h^\alpha(\mathbb{P})$ is a (Q,F)-uniformly integrable martingale. Note that the process $h^\alpha(\mathbb{P})$, called the *Hellinger process* of order α, is associated with the family of probability measures $\mathbb{P} = \{P_\theta; \theta \in \Theta\}$ independently of the choice of the dominating measure Q.

Since the geometric mean process is non-negative, one can turn this into the multiplicative decomposition by applying the usual device as described e.g. in [12, Theorem 2.5.1]. This leads us to the conclusion of [2, Theorem 5.13] that the ratio $g^\alpha(\mathbb{P};Q)/\mathcal{E}\{-h^\alpha(\mathbb{P})\}$ is a local (Q,F)-martingale, cf. also [1, Section 5.2] (here and elsewhere below $\mathcal{E}$ means the Doléans-Dade exponential). But we need more. We need additional conditions under which this ratio becomes a (Q,F)-martingale. For discussion on sufficient conditions we refer to [2] and [3], for we prefer here to avoid details and to just assume throughout that the ratio $g^\alpha(\mathbb{P};Q)/\mathcal{E}\{-h^\alpha(\mathbb{P})\}$ is a (Q,F)-martingale. Then we can define on the space $(\Omega,\mathcal{F})$ yet another probability measure G_α, the so-called *geometric mean measure* so that the following multiplicative decomposition holds:

$$g^\alpha(\mathbb{P};Q) = z(G_\alpha;Q)\,\mathcal{E}\{-h^\alpha(\mathbb{P})\} \tag{3.1}$$

with the martingale part that is the density process $z(G_\alpha;Q)$ of the geometric mean measure G_α with respect to the dominating measure Q.

3.2 Conditional Hellinger process

Let us fix a particular value of the nuisance parameter η, say $\eta = y \in \Upsilon$, and consider the subfamily of probability measures $\mathbb{P}_y = \{P_{x,y}, x \in \Xi\}$. The second identity in (2.8) involves the conditional geometric mean process defined by (2.7). Let us apply this definition to the subfamily of density processes $\{z(P_{x,y};Q), x \in \Xi\}$. For each fixed $\eta = y \in \Upsilon$ we get the conditional geometric mean process defined by

$$g^{\chi_y}(\mathbb{P}_y;Q) = e^{\mathrm{E}_{\chi_y}\log z(P_{\xi,y};Q)} = e^{\int_\Xi \log z(P_{x,y};Q)\chi_y(dx)}.$$

For every $y \in \Upsilon$ these processes are again (Q,F)-supermartingales of class (D) with $g_0^{\chi_y}(\mathbb{P}_y;Q) = 1$, so the results parallel to that of the previous section are available. In particular, there exists a (unique up to Q-indistinguishability) predictable finite-valued increasing process $h^{\chi_y}(\mathbb{P}_y)$ starting from the origin $h_0^{\chi_y}(\mathbb{P}_y) = 0$, so that the ratio $g^{\chi_y}(\mathbb{P}_y;Q)/\mathcal{E}\{-h^{\chi_y}(\mathbb{P}_y)\}$ is a local (Q,F)-martingale. Like in the unconditional case, process $h^{\chi_y}(\mathbb{P}_y)$ is called *the conditional Hellinger process* of order χ_y, given $\eta = y \in \Upsilon$. It is associated with the family of probability measures $\mathbb{P}_y = \{P_{x,y}, x \in \Xi\}$ independently of the choice of the dominating measure Q. Assume again that the aforementioned ratio is a (Q,F)-martingale and along with (3.1) get for each $y \in \Upsilon$ the multiplicative decomposition

$$g^{\chi_y}(\mathbb{P}_y;Q) = z(G_{\chi_y};Q)\,\mathcal{E}\{-h^{\chi_y}(\mathbb{P}_y)\} \tag{3.2}$$

where the martingale part $z(G_{\chi_y};Q)$ is the density process of the conditional geometric measure G_{χ_y} with respect to the dominating measure Q.

4 Relationship between the Hellinger processes

Consider the family of probability measures $\mathbb{G} = \{G_{\chi_y}, y \in \Upsilon\}$ that are conditional geometric mean measures. With the family of density processes $\{z(G_{\chi_y}; Q), y \in \Upsilon\}$ we associate in the usual way the geometric mean process

$$g^\upsilon(\mathbb{G}; Q) = e^{\mathrm{E}_\upsilon \log z(G_{\chi_\eta}; Q)} = e^{\int_\Upsilon \log z(G_{\chi_y}; Q)\upsilon(dy)}.$$

This geometric mean process possesses the following properties:

Theorem 4.1. *Along with the basic conditions (2.1) end (2.4), assume that all geometric mean measures are probability measures. Then*
(i) The geometric mean process $g^\upsilon(\mathbb{G}; Q)$ is a (Q, F)-supermartingale of class (D) with $g_0^\upsilon(\mathbb{G}; Q) = 1$ and with the following Q-a.s. boundedness properties: $\inf_t g_t^\upsilon(\mathbb{G}; Q) > 0$ and $\sup_t g_t^\upsilon(\mathbb{G}; Q) < \infty$.
(ii) There exists a (unique up to Q-indistinguishability) predictable finite-valued increasing process $h^\upsilon(\mathbb{G})$ starting from the origin $h_0^\upsilon(\mathbb{G}) = 0$, so that $g^\upsilon(\mathbb{G}; Q) + g_-^\upsilon(\mathbb{G}; Q) \cdot h^\upsilon(\mathbb{G})$ is a uniformly integrable (Q, F)-martingale. The process $h^\upsilon(\mathbb{G})$, called the Hellinger process of order υ, is associated with the family of probability measures $\mathbb{G} = \{G_{\chi_y}, y \in \Upsilon\}$ independently of the choice of the dominating measure Q.
(iii) The Hellinger process $h^\upsilon(\mathbb{G})$ of order υ is related to the original Hellinger process $h^\alpha(\mathbb{P})$ and to the conditional Hellinger process $h^{\chi_\eta}(\mathbb{P}_\eta)$ by the following relations:

$$\begin{aligned} h^\alpha(\mathbb{P})^c &= h^\upsilon(\mathbb{G})^c + \mathrm{E}_\upsilon h^{\chi_\eta}(\mathbb{P}_\eta)^c \\ \log(1 - \Delta h^\alpha(\mathbb{P})) &= \log(1 - \Delta h^\upsilon(\mathbb{G})) + \mathrm{E}_\upsilon \log(1 - \Delta h^{\chi_\eta}(\mathbb{P}_\eta)). \end{aligned}$$

(iv) The geometric mean process $g^\upsilon(\mathbb{G}; Q)$ has the following multiplicative decomposition:

$$g^\upsilon(\mathbb{G}; Q) = z(G_\alpha; Q)\, \mathcal{E}\{-h^\upsilon(\mathbb{G})\}.$$

Proof. The poof of assertion (i) follows arguments used in the course of proving [1, Proposition 4.1]. The existence and uniqueness of the Hellinger process of order υ, as well as the Doob-Mayer decomposition asserted in (ii), will become evident from the forthcoming considerations. The notion defined in [1, Lemma 4.3] of independence of the choice of the dominating measure, does easily extend to the present Hellinger process of order υ. Let us focus now on the remainning assertions (iii) and (iv).

Take g^υ from both sides in (3.2). Apply to the left hand side the second identity in (2.8) to get $g^\alpha(\mathbb{P}; Q)$. On the right hand side take into consideration the obvious property that the geometric mean of a product is the product of geometric means. This results in

$$g^\alpha(\mathbb{P}; Q) = g^\upsilon(\mathbb{G}; Q)\, g^\upsilon(\mathcal{E}\{-h^{\chi_\cdot}(\mathbb{P}_\cdot)\}).$$

By (3.1) we have

$$g^\upsilon(\mathbb{G}; Q) = z(G_\alpha; Q)\, \mathcal{E}\{-h^\alpha(\mathbb{P})\}\, g^\upsilon(\mathcal{E}\{(1 - \Delta h^{\chi_\cdot}(\mathbb{P}_\cdot)) \cdot h^{\chi_\cdot}(\mathbb{P}_\cdot)\}), \tag{4.1}$$

since it follows from [2, Remark 5.5] that $g^{\upsilon}(\mathcal{E}\{-h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})\})\, g^{\upsilon}(\mathcal{E}\{(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))\cdot h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})\}) = 1$. Compare (4.1) with the multiplicative decomposition in assertion (iv). We see that the martingale parts on the right hand side coincide. It remains therefore to prove that the relations in assertion (iii) are equivalent to $\mathcal{E}\{-h^{\upsilon}(\mathbb{G})\} = \mathcal{E}\{-h^{\alpha}(\mathbb{P})\}\, g^{\upsilon}(\mathcal{E}\{(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))\cdot h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})\})$ which in turn is equivalent to $\mathcal{E}\{-h^{\alpha}(\mathbb{P})\} = \mathcal{E}\{-h^{\upsilon}(\mathbb{G})\}\, g^{\upsilon}(\mathcal{E}\{-h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})\})$. Let us now apply [2, Proposition 5.8]. We get $g^{\upsilon}(\mathcal{E}\{-h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})\}) = \mathcal{E}\{-a^{\upsilon}(h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})) - \sum_{s\leq\cdot}\phi^{\upsilon}_{s}(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))\}$. Hence by the well-known formula for the product of two stochastic exponentials and by elementary algebra taking into consideration $\phi = a - g$, it is easily seen that

$$\begin{aligned}\mathcal{E}\{-h^{\alpha}(\mathbb{P})\} =&\mathcal{E}\{-h^{\upsilon}(\mathbb{G})\}\,\mathcal{E}\{-a^{\upsilon}(h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})) - \sum_{s\leq\cdot}\phi^{\upsilon}_{s}(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))\}\\ =&\mathcal{E}\{-h^{\upsilon}(\mathbb{G}) - a^{\upsilon}(h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})^{c}) - \sum_{s\leq\cdot}(1-\Delta h^{\upsilon}_{s}(\mathbb{G}))(1-g^{\upsilon}_{s}(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))\}\end{aligned}$$

which implies $h^{\alpha}(\mathbb{P}) = h^{\upsilon}(\mathbb{G})^{c} + a^{\upsilon}(h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})^{c}) + \sum_{s\leq\cdot}\{1-(1-\Delta h^{\upsilon}_{s}(\mathbb{G}))g^{\upsilon}_{s}(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))\}$. This is indeed equivalent to the identities in assertion (iii). The case of continuous parts is clear. Take jumps from both sides of the latter display to get the identity $1-\Delta h^{\alpha}(\mathbb{P}) = (1-\Delta h^{\upsilon}_{s}(\mathbb{G}))\, g^{\upsilon}_{s}(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot}))$, equivalent to the second of identities in assertion (iii). □

It has been shown in the course of the proof that the following corollary holds:

Corollary 4.2. *Under the conditions of theorem 4.1*

$$h^{\alpha}(\mathbb{P}) = h^{\upsilon}(\mathbb{G}) + a^{\upsilon}(h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})) - \sum_{s\leq\cdot}\{a^{\upsilon}_{s}(\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})) - (1-\Delta h^{\upsilon}_{s}(\mathbb{G}))(1-g^{\upsilon}_{s}(1-\Delta h^{\chi_{\cdot}}(\mathbb{P}_{\cdot})))\}. \quad (4.2)$$

5 Applications

5.1 Semimartingale observations

Throughout we well use common notions and facts of the general theory of stochastic processes as developed e.g. in [5], [8] or [12]. 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 next display, together with $\nu(\theta)$ the compensator with respect to the measure $P_{\theta}, \theta\in\Theta$. We will also deal with the compensators $\nu^{G_{\alpha}}$ and $\nu^{G_{\chi_y}}$ with respect to the geometric mean measures G_{α} and G_{χ_y}, respectively.

Suppose that we observe a semi-martingale 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 [8, Theorem III.3.24] or [12, 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:

$$\begin{cases} B(\theta) & = B + \beta(\theta) \cdot C + (Y(\theta) - 1)\, \hbar \cdot \nu \\ C(\theta) & = C \\ \nu(\theta) & = Y(\theta) \cdot \nu \end{cases} \tag{5.1}$$

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 [12, Lemma IV.5.6] one can find the relationship of these processes to the density process $z(P_\theta; Q)$.

Assume (2.1) and (2.4). For each $\theta \in \Theta$ let $m(P_\theta; Q)$ be a (Q, F)-local martingale given by

$$m(P_\theta; Q) = z_-(P_\theta; Q)^{-1} \cdot z(P_\theta; Q) \tag{5.2}$$

so that the density process is represented as the Doléans-Dade exponential $z(P_\theta; Q) = \mathcal{E}(m(P_\theta; Q))$ of this martingale. Upon further specification of the randomized experiment in question, one can assign to the (Q, F)-local martingale (5.2) explicit form in terms of the triplet of predictable characteristics $T = (B, C, \nu)$ of the observed (Q, F)-semimartingale X. In addition to (2.1) and (2.4), 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 (5.2) gets the form

$$m(P_\theta; Q) = \beta(\theta) \cdot X^c + \{Y(\theta) - 1 + \frac{\hat{Y}(\theta) - \hat{1}}{1 - \hat{1}}\} * (\mu^X - \nu) \tag{5.3}$$

where $\beta(\theta) = \beta(\theta, Q)$ and $Y(\theta) = Y(\theta, Q)$ are the same as (5.1). 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 $\nu(\theta)$ 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 at the beginning of this section.

Under the additional assumption – the representation property (5.3) – theorem 3.6 from [3] is applicable and we have the following explicit representation for the Hellinger process $h^\alpha(\mathbb{P})$ of order α: in terms of the triplet of predictable characteristics

$$h^\alpha(\mathbb{P}) = \frac{1}{2}\mathrm{v}^\alpha(\beta) \cdot C + \phi^\alpha(Y) \cdot \nu + \sum_{s \leq \cdot} \phi^\alpha_s(1 - \hat{Y}) \tag{5.4}$$

where the notations (2.9) are used.

According to [3, Theorem 3.11], the geometric mean measure G_α of order α prescribes to the observations the triplet of predictable characteristics $T^{G_\alpha} = (B^{G_\alpha}, C^{G_\alpha}, \nu^{G_\alpha})$ given by

$$\begin{cases} B^{G_\alpha} & = a^\alpha(B) + (Y^{G_\alpha} - a^\alpha(Y))\hbar \cdot \nu \\ C^{G_\alpha} & = C \\ \nu^{G_\alpha} & = Y^{G_\alpha} \cdot \nu \end{cases} \tag{5.5}$$

with

$$Y^{G_\alpha} \doteq \frac{g^\alpha(Y)}{1-\Delta h^\alpha(\mathbb{P})} = \frac{g^\alpha(Y)}{\hat{g}^\alpha(Y)+g^\alpha(1-\hat{Y})},$$

see [3, formula (3.23)] for the identity

$$1-\Delta h^\alpha(\mathbb{P}) = \hat{g}^\alpha(Y)+g^\alpha(1-\hat{Y}). \tag{5.6}$$

The expression similar to (5.4) is available for the conditional Hellinger process $h^{\chi_y}(\mathbb{P}_y)$ of order χ_y in terms of the conditional variance process and conditional difference process: with the notations (2.10) we have

$$h^{\chi_y}(\mathbb{P}_y) = \frac{1}{2}\mathrm{v}^{\chi_y}(\beta(\cdot,y))\cdot C + \phi^{\chi_y}(Y(\cdot,y))\cdot\nu + \sum_{s\le\cdot}\phi_s^{\chi_y}(1-\hat{Y}(\cdot,y)). \tag{5.7}$$

Similarly to (5.5), the conditional geometric mean measure G_{χ_y} of order χ_y prescribes to the observations the triplet of predictable characteristics $T^{G_{\chi_y}} = (B^{G_{\chi_y}}, C^{G_{\chi_y}}, \nu^{G_{\chi_y}})$ given by

$$\begin{cases} B^{G_{\chi_y}} &= a^{\chi_y}(B(\cdot,y)) + (Y^y - a^{\chi_y}(Y(\cdot,y)))\hbar\cdot\nu \\ C^{G_{\chi_y}} &= C \\ \nu^{G_{\chi_y}} &= Y^{G_{\chi_y}}\cdot\nu \end{cases} \tag{5.8}$$

with

$$Y^{G_{\chi_y}} = \frac{g^{\chi_y}(Y(\cdot,y))}{1-\Delta h^{\chi_y}(\mathbb{P}_y)} = \frac{g^{\chi_y}(Y(\cdot,y))}{\hat{g}^{\chi_y}(Y(\cdot,y))+g^{\chi_y}(1-\hat{Y}(\cdot,y))}. \tag{5.9}$$

In the latter equality the same device is used as above for calculating jumps of Hellinger processes according to which the jump of process (5.7) satisfies

$$1-\Delta h^{\chi_y}(\mathbb{P}_y) = \hat{g}^{\chi_y}(Y(\cdot,y))+g^{\chi_y}(1-\hat{Y}(\cdot,y)). \tag{5.10}$$

From (5.9) it follows that

$$1-\hat{Y}^{G_{\chi_\cdot}} = \frac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}(\mathbb{P}_\cdot)}. \tag{5.11}$$

The considerations leading us to the expressions (5.7) and (2.10) for the Hellinger process $h^\alpha(\mathbb{P})$ of order α and for the conditional Hellinger process $h^{\chi_y}(\mathbb{P}_y)$ of order χ_y, given $y\in\Upsilon$, allows us to assert that also the Hellinger process $h^\upsilon(\mathbb{G})$ of order υ, introduced in 4.1, is expressible in terms of the prior moments of the local characteristics β and Y as follows:

$$h^\upsilon(\mathbb{G}) = \frac{1}{2}\mathrm{v}^\upsilon(a^{\chi_\cdot}(\beta))\cdot C + \phi^\upsilon(Y^{G_{\chi_\cdot}})\cdot\nu + \sum_{s\le\cdot}\phi_s^\upsilon(1-\hat{Y}^{G_{\chi_\cdot}})$$

where $Y^{G_{\chi_\cdot}}$ is defined by (5.9) and $1-\hat{Y}^{G_{\chi_\cdot}}$ by (5.11). Obviously, the asserted expression for $h^\upsilon(\mathbb{G})$ may also be written as

$$h^\upsilon(\mathbb{G}) = \frac{1}{2}\mathrm{v}^\upsilon(a^{\chi_\cdot}(\beta))\cdot C + \phi^\upsilon\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right)\cdot\nu + \sum_{s\le\cdot}\phi_s^\upsilon\left(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}(\mathbb{P}_\cdot)}\right) \tag{5.12}$$

so that

$$1 - \Delta h^{\upsilon}(\mathbb{G}) = \hat{g}^{\upsilon}\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right) + g^{\upsilon}\left(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}\right). \tag{5.13}$$

We are now going to verify that the Hellinger processes (5.4), (5.7) and (5.12) satisfy the relationship (4.2) which will imply that our assertion (5.12) indeed holds true. Note first that by the identities (2.11)

$$h^{\alpha}(\mathbb{P}) - \mathrm{E}_{\upsilon} h^{\chi_y}(\mathbb{P}_y) = \frac{1}{2}\mathrm{v}^{\upsilon}(a^{\chi_\cdot}(\beta)) \cdot C + \phi^{\upsilon}\left(g^{\chi_\cdot}(Y)\right) \cdot \nu + \sum_{s\leq \cdot} \phi^{\upsilon}_s\left(g^{\chi_\cdot}\right).$$

Therefore it suffices to examine the difference

$$\{\phi^{\upsilon}\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right) - \phi^{\upsilon}\left(g^{\chi_\cdot}(Y)\right)\} \cdot \nu + \sum_{s\leq\cdot}\{\phi^{\upsilon}_s\left(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}\right) - \phi^{\upsilon}_s(g^{\chi_\cdot}(1-\hat{Y}))\}$$

and to show that firstly

$$\begin{aligned}
&\{a^{\upsilon}\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right) - a^{\upsilon}\left(g^{\chi_\cdot}(Y)\right)\} \cdot \nu + \sum_{s\leq\cdot}\{a^{\upsilon}_s(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}) - a^{\upsilon}_s(g^{\chi_\cdot}(1-\hat{Y}))\}\\
=&\sum_{s\leq\cdot} a^{\upsilon}_s\left(\tfrac{\Delta h^{\chi_\cdot}\hat{g}^{\chi_y}(Y)}{1-\Delta h^{\chi_\cdot}}\right) + \sum_{s\leq\cdot} a^{\upsilon}_s\left(\tfrac{\Delta h^{\chi_\cdot} g^{\chi_y}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}\right) = \sum_{s\leq\cdot} a^{\upsilon}_s(\Delta h^{\chi_\cdot})
\end{aligned}$$

and secondly

$$\begin{aligned}
&\{g^{\upsilon}\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right) - g^{\upsilon}\left(g^{\chi_\cdot}(Y)\right)\} \cdot \nu + \sum_{s\leq\cdot}\{g^{\upsilon}_s(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}) - g^{\upsilon}_s(g^{\chi_\cdot}(1-\hat{Y}))\}\\
=&g^{\upsilon}\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right)(1-g^{\upsilon}(1-\Delta h^{\chi_\cdot})) \cdot \nu + \sum_{s\leq\cdot} g^{\upsilon}_s\left(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}\right)(1-g^{\upsilon}_s(1-\Delta h^{\chi_\cdot}))\\
=&\sum_{s\leq\cdot}(\hat{g}^{\upsilon}_s\left(\tfrac{g^{\chi_\cdot}(Y)}{1-\Delta h^{\chi_\cdot}}\right) + g^{\upsilon}_s\left(\tfrac{g^{\chi_\cdot}(1-\hat{Y})}{1-\Delta h^{\chi_\cdot}}\right))(1 - g^{\upsilon}_s(1-\Delta h^{\chi_\cdot}(\mathbb{P}_\cdot)))\\
=&\sum_{s\leq\cdot}(1-\Delta h^{\upsilon}_s(\mathbb{G}))(1 - g^{\upsilon}_s(1-\Delta h^{\chi_\cdot}(\mathbb{P}_\cdot)))
\end{aligned}$$

Both of these equations are obtained by repeatedly evoking the formulas for jumps of Hellinger processes, namely (5.6), (5.10) and (5.13).

5.2 Point processes

The setup in the present section is the same as in [3, Section 5.2]. Suppose that observed is a d-dimensional counting process $(N^1, \ldots, N^d)$. 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 are positive increasing processes so that the densities $d\Lambda^i(\theta)/dA^i = Y^i(\theta)$ exist for all $i = 1, \ldots, d$ and $\theta \in \Theta$. The expression for the corresponding density process is well-known:

$$z(P_\theta; Q) = e^{-\Lambda(\theta)^c + A^c} \prod_{s\leq\cdot}\left(\frac{1-\Delta\Lambda_s(\theta)}{1-\Delta A_s}\right)^{1-\Delta N_s} \prod_{i=1}^{d} Y^i_s(\theta)^{\Delta N^i_s}$$

with $N = N^1 + \cdots + N^d$, $\Lambda = \Lambda^1 + \cdots + \Lambda^d$ and $A = A^1 + \cdots + A^d$. Since $\Lambda(\theta) = \int_0^{\cdot} \sum_{i=1}^d Y^i(\theta) dA^i$, we have $\Delta\Lambda(\theta) = \sum_{i=1}^d Y^i(\theta)\Delta A^i$. Assume in addition the following moment condition

$$\mathrm{E}_\alpha \log \frac{Y_s^i(\vartheta)}{1 - \Delta\Lambda_s(\vartheta)} > -\infty$$

for all $s > 0$ and all densities $Y^i(\theta), i = 1, \ldots, d$. Then the Hellinger process of order α is given by

$$h^\alpha(\mathbb{P}) = \int_0^{\cdot} \sum_{i=1}^d \phi_s^\alpha(Y^i) dA_s^i + \sum_{s \le \cdot} \phi_s^\alpha(1 - \Delta\Lambda) \tag{5.14}$$

so that $1 - \Delta h^\alpha(\mathbb{P}) = \sum_{i=1}^d g^\alpha(Y^i \Delta A^i) + g^\alpha(1 - \Delta\Lambda)$. According to (5.5), the geometric mean measure G_α (that is again supposed to be a probability measure) assigns to the component N^i the intensity density with respect to the same A^i given by

$$Y^{i\,G_\alpha} = \frac{g^\alpha(Y^i)}{\sum\limits_{i=1}^d g^\alpha(Y^i \Delta A^i) + g^\alpha(1 - \Delta\Lambda)}.$$

Therefore the density process of the geometric mean measure G_α with respect to the dominating measure Q is given by

$$z(G^\alpha; Q) = \prod_{i=1}^d e^{-\int_0^{\cdot} (g^\alpha(Y^i) - 1) dA^{i\,c}} \prod_{s \le \cdot} \frac{g_s^\alpha(\frac{1-\Delta\Lambda}{1-\Delta A})^{1-\Delta N_s} \prod\limits_{i=1}^d g_s^\alpha(Y^i)^{\Delta N_s^i}}{\sum\limits_{i=1}^d g_s^\alpha(Y^i \Delta A^i) + g_s^\alpha(1 - \Delta\Lambda)}.$$

The conditional Hellinger process is defined similarly to (5.14) with the obvious substitutions leading to

$$h^{\chi_\eta}(\mathbb{P}_\eta) = \int_0^{\cdot} \sum_{i=1}^d \phi_s^{\chi_\eta}(Y^i(\cdot, \eta)) dA_s^i + \sum_{s \le \cdot} \phi_s^{\chi_\eta}(1 - \Delta\Lambda(\cdot, \eta))$$

and $1 - \Delta h^{\chi_\eta}(\mathbb{P}_\eta) = \sum_{i=1}^d g^{\chi_\eta}\left(Y^i(\cdot, \eta)\Delta A^i\right) + g^{\chi_\eta}(1 - \Delta\Lambda(\cdot, \eta))$. According to the general formula (5.12), the expression for the Hellinger process $h^\upsilon(\mathbb{G})$ in the present case is as follows:

$$h^\upsilon(\mathbb{G}) = \int_0^{\cdot} \sum_{i=1}^d \phi_s^\upsilon \left(\frac{g^{\chi_\cdot}(Y^i)}{1 - \Delta h^{\chi_\eta}(\mathbb{P}_\eta)} \right) dA_s^i + \sum_{s \le \cdot} \phi_s^\upsilon \left(\frac{g^{\chi_\cdot}(1 - \Delta\Lambda)}{1 - \Delta h^{\chi_\cdot}(\mathbb{P}_\cdot)} \right).$$

References

[1] K. Dzhaparidze, P.J.C. Spreij and E. Valkeila (1997), On Hellinger Processes for Parametric Families of Experiments, in *Statistics and Control of Stochastic Processes, The Liptser Festschrift* (eds. Yu. Kabanov, B. Rozovskii and A. Shiryaev), World Scientific, 41-62.

[2] K. Dzhaparidze, P. Spreij and E. Valkeila (2000), Information processes in filtered experiments, Part I: general concepts, *Report of the Department of Mathematics, University of Helsinki,* Preprint 264.

[3] K. Dzhaparidze, P. Spreij and E. Valkeila (2000), Information processes in filtered experiments, Part II: explicit representations and examples, *Report of the Department of Mathematics, University of Helsinki,* Preprint 265.

[4] 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.

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

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

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

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

[9] 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).

[10] 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.

[11] 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.

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