

## On the Hellinger Type Distances for Filtered Experiments

K. Dzharidze<sup>1</sup> and E. Valkeila<sup>2</sup>

<sup>1</sup> Center for Mathematics and Computer Science, P.O. Box. 4079, NL-1009 AB Amsterdam, The Netherlands

<sup>2</sup> Computing Centre, University of Helsinki, Teollisuuskatu 23, SF-00510 Helsinki, Finland

**Summary.** We study the Hellinger type distances  $\rho_p(P_T, \tilde{P}_T)$  on a filtered space. Here  $p \geq 2$  is an arbitrary number and  $P_T$  and  $\tilde{P}_T$  are two probability measures stopped at a random time  $T$ . We give lower and upper bounds for  $\rho_p(P_T, \tilde{P}_T)$  in predictable terms.

### 1. Introduction

1.1. Let  $(\Omega, F, \mathbf{F})$  be a stochastic basis, i.e. a measurable space with a filtration  $\mathbf{F} = (F_t)_{t \geq 0}$  such that  $\bigvee_{t \geq 0} F_t = F_\infty = F$ . Given two probability measures  $P$  and  $\tilde{P}$ , with  $P_0 = \tilde{P}_0$ , define a probability measure  $Q$  by  $Q = (P + \tilde{P})/2$ . Suppose that  $\mathbf{F}$  satisfies the usual assumptions with respect to  $Q$ . Consider then the optional projections of the measures  $P$ ,  $\tilde{P}$  and  $Q$  with respect to  $\mathbf{F}$ . We will denote these optional valued processes by  $P$ ,  $\tilde{P}$  and  $Q$ , respectively. If  $T$  is a  $\mathbf{F}$ -stopping time, then  $P_T$  is the restriction of the measure  $P$  to the sub- $\sigma$ -field  $F_T$  of  $F$ ; define  $\tilde{P}_T$  and  $Q_T$  similarly. Since the measures  $P_T$  are absolutely continuous with respect to the measure  $Q_T$ , we can define  $(Q, \mathbf{F})$ -martingales  $\xi$  and  $\tilde{\xi}$  by

$$\xi_T = dP_T/dQ_T \quad \text{and} \quad \tilde{\xi}_T = d\tilde{P}_T/dQ_T. \quad (1.1)$$

The collection  $(\Omega, F, \mathbf{F}, P, \tilde{P})$  is called the *binary experiment*.

In the present paper the following distances between stopped measures  $P_T$  and  $\tilde{P}_T$  are studied

$$\rho_p(P_T, \tilde{P}_T) = \{E_Q[|\xi_T^{1/p} - \tilde{\xi}_T^{1/p}|^p]\}^{1/p}, \quad (1.2)$$

where  $p \geq 2$ . Recall that if  $p = 2$  then  $\rho_2(P_T, \tilde{P}_T)$  is called the Hellinger distance. For more details on such kind of distances see Liese and Vajda (1987), where  $\rho_p^2$  is called the  $\chi^{1/p}$ -divergence. Note that the distances  $\rho_p$  are independent of a particular choice of the dominating measure  $Q$ .

1.2. With the binary experiment  $(\Omega, F, \mathbf{F}, P, \tilde{P})$  we associate the Hellinger process by

$$h = (1/2)((\xi - \tilde{\xi})^{-2} \circ \langle \xi^c \rangle + (\sqrt{1 + x/\xi_-} - \sqrt{1 - x/\tilde{\xi}_-})^2 \star v^{\xi, \tilde{\xi}}). \quad (1.3)$$

Here  $\nu^{\xi, Q}$  is the compensator of the jump measure of the process  $\xi$ ,  $\circ$  denotes the ordinary Stieltjes-integral and  $\star$  stands for the double integral  $\nu(ds, dx)$ .

It is known that the Hellinger process controls the Hellinger distance in the sense of Jacod and Shiryaev (1987), Section V.4 (see also Valkeila and Vostrikova (1986)). In particular,

$$\rho_2^2(P_T, \tilde{P}_T) \leq 4\sqrt{E_P h_T}. \quad (1.4)$$

To control  $\rho_p^p$  also for  $p > 2$ , along with the Hellinger process (1.3) we introduce the *p-divergency process* by

$$k(p) = |(1 + x/\xi_-)^{1/p} - (1 - x/\tilde{\xi}_-)^{1/p}|^p \star \nu^{\xi, Q}, \quad (1.5)$$

where  $p \geq 2$ . As is shown in this paper (see Theorem 3.2 below), for each even integer  $p \geq 2$  there is a constant  $C_p > 0$  such that

$$\rho_p^p(P_T, \tilde{P}_T) \leq C_p E_P (h_T^{p/2} + k_T(p)); \quad (1.6)$$

for  $p = 2$ , in particular

$$\rho_2^2(P_T, \tilde{P}_T) \leq 8E_P h_T \quad (1.7)$$

(cf. (1.4)).

1.3. This paper is organized as follows. In Section 2 more details can be found on the quantities introduced above. Besides, in Lemma 2.1 we give a Burkholder type of inequality, expressed in predictable terms, extending the results by Burkholder (1973) and Novikov (1975).

The first of two theorems, presented in Section 3 gives upper and lower bounds for  $\rho_p$  in terms of the expectation with respect to the measure  $Q$ .

In the case where the processes  $h$  and  $k(p)$  are not necessarily deterministic, it is useful to have bounds in terms of the expectation with respect to the measure  $P$ : for an upper bound see Theorem 3.2 below. This upper bound is given in a slightly more general form than (1.6), useful for an application in Section 4, Theorem 4.2.

In Section 4 we give some applications to sequences of binary experiments. In Theorem 4.1, in particular, we give necessary and sufficient conditions for the convergence to a limiting Gaussian experiment, alternative to those of Jacod and Shiryaev (1987), Theorems X.1.12 and X.1.64.

Finally, in Section 5, we will demonstrate how to evaluate, based on (1.6), certain modulus of continuity needed in statistical applications. To trace the rôle of the distances  $\rho_p$  in various statistical applications, see e.g. Ibragimov and Has'minskii (1981), Kutoyants (1984), Dzhaparidze (1986), Valkeila and Vostrikova (1987) and Vostrikova (1988). We end this paper by giving some examples of the *p-divergency process*.

## 2. Certain Properties of $\rho_p$ and Related Processes

2.1. We assume that  $(\Omega, \mathcal{F}, \mathbf{F})$  is as described above. Moreover, we assume  $F_0 = \{\emptyset, \Omega\}$   $Q$ -a.s. For unexplained notation below we refer to Jacod (1979), Jacod and Shiryaev (1987) and Liptser and Shiryaev (1989).

Let  $\mathcal{D}$  be the space of right-continuous functions with left-hand limits on  $\mathbf{R}_+ = [0, \infty]$ . We can take such versions of the density processes  $\xi$  and  $\tilde{\xi}$  that their paths are in  $\mathcal{D}$ , and

$$\xi + \tilde{\xi} = 2, \langle \xi^c \rangle = \langle \tilde{\xi}^c \rangle, \Delta \xi = -\Delta \tilde{\xi}, \text{ and } \langle \xi^c, \tilde{\xi}^c \rangle = -\langle \xi^c \rangle \quad (2.1)$$

(here and elsewhere below the angle brackets process is understood as a  $(Q, \mathbf{F})$ -compensator). The facts in (2.1) follow the special choice of the dominating measure  $Q$ .

Note that the jump measure  $\mu^\xi$  of the  $(Q, \mathbf{F})$ -uniformly integrable martingale  $\xi$ , as well as its  $(Q, \mathbf{F})$ -compensator  $\nu^{\xi, Q}$  involved in (1.3) and (1.5), only charges the set  $\{(\omega, t, x): \xi_{t-}(\omega) > 0, \tilde{\xi}_{t-}(\omega) > 0, -\xi_{t-}(\omega) \leq x \leq \tilde{\xi}_{t-}(\omega)\}$ ; see Jacod and Shiryaev (1987), Theorem IV.1.33.

By the easily verified inequality

$$|u^{1/q} - v^{1/q}|^q \leq |u^{1/p} - v^{1/p}|^p, \quad (2.2)$$

valid for each  $u, v \geq 0$  and  $1 < p \leq q$  we get the following facts:

- (i) for  $p \geq 2$  we have  $k(p) \leq k(2) \leq 2h$ ;
- (ii) the process  $k(p)$  decreases as  $p$  increases;
- (iii) as  $p \rightarrow \infty$

$$k(p) \equiv k(p; P, \tilde{P}) \rightarrow h(0; P, \tilde{P}) + h(0; \tilde{P}, P) \quad (Q\text{-a.s.}).$$

Here and below we use the following definition by Jacod and Shiryaev (1987), IV.1.57 and IV.1.36 of the Hellinger processes of order 0 and  $\alpha \in (0, 1)$ :

$$h(0; P, \tilde{P}) = \tilde{\lambda} 1_{\{\lambda=0\}} \star \nu^{\xi, Q}, \quad h(0; \tilde{P}, P) = \lambda 1_{\{\tilde{\lambda}=0\}} \star \nu^{\xi, Q}$$

with

$$\lambda = 1 + x/\xi_-, \quad \tilde{\lambda} = 1 - x/\tilde{\xi}_-, \quad (2.3)$$

and

$$h(\alpha) \equiv h(\alpha; P, \tilde{P}) = \frac{\alpha(1-\alpha)}{2} \left( \frac{1}{\xi_-} + \frac{1}{\tilde{\xi}_-} \right)^2 \circ \langle \xi^c \rangle + \phi_\alpha(\lambda, \tilde{\lambda}) \star \nu^{\xi, Q}$$

with

$$\phi_\alpha(u, v) = \alpha u + (1-\alpha)v - u^\alpha v^{1-\alpha}.$$

Obviously,  $h = h(1/2)$ . Thus, by (i), the  $p$ -divergency processes  $k(p)$ ,  $p \geq 2$ , related to the discontinuous part of  $\xi$  only, exist since  $k(p) \leq 2h$ . Observe also that  $k(p)$ ,  $p \geq 2$ , as well as  $h$ , are independent of the measure  $Q$  (Jacod and Shiryaev (1987), Theorem IV.1.22), and for any even integer  $p > 2$

$$k(p) = - \sum_{k=1}^{p-1} (-1)^k \binom{p}{k} h(k/p)$$

due to the binomial formula and properties of  $h(\alpha)$ .

2.2. By (2.2)  $\rho_p^p$  decreases too as  $p$  increases.

Besides,

$$\rho_p^p(P, \tilde{P}) \rightarrow \tilde{P}(\xi = 0) + P(\tilde{\xi} = 0) \quad \text{as } p \rightarrow \infty.$$

For the variational distance  $\|P - \tilde{P}\| = \rho_1(P, \tilde{P})$ , in particular, we have (cf. Jacod and Shiryaev (1987), V.4.8, and Liese and Vajda (1987), Ch. 2)

$$\|P - \tilde{P}\| \geq \rho_p^p(P, \tilde{P}), c_p \|P - \tilde{P}\| \leq \rho_p(P, \tilde{P}), p \geq 1$$

where the second inequality is obtained by Jensen's inequality applied to the left-hand side inequality

$$\frac{1}{p} |\xi - \tilde{\xi}| \leq |\xi^{1/p} - \tilde{\xi}^{1/p}| \leq \frac{1}{2^{1-1/p}} |\xi - \tilde{\xi}|, p \geq 1. \quad (2.4)$$

The last relation is easily verified by taking into consideration that  $\xi + \tilde{\xi} = 2$ .

2.3. As the process  $\xi^{1/p} - \tilde{\xi}^{1/p}$  is a martingale if only  $p = 1$ , the relation (2.4) allows us to estimate bounds of  $\rho_p^p$  by applying Burkholder-type inequalities. Namely, there are universal constants  $c_p$  and  $C_p$  such that for a stopping time  $T$

$$c_p E_Q[\xi]_T^{p/2} \leq \rho_p^p(P_T, \tilde{P}_T) \leq C_p E_Q[\xi]_T^{p/2}; \quad (2.5)$$

see, e.g., Liptser and Shiryaev (1989), Section 1.9, Theorem 7.

Furthermore, usual considerations establishing Burkholder-type inequalities allow us to replace (2.5) by

$$c_p E_Q\{\langle \xi \rangle_T^{p/2} + ((\Delta \xi)_T^*)^p\} \leq \rho_p^p(P_T, \tilde{P}_T) \leq C_p E_Q\{\langle \xi \rangle_T^{p/2} + ((\Delta \xi)_T^*)^p\}$$

or by

$$c_p E_Q\{\langle \xi \rangle_T^{p/2} + |x|^p \star v_T^{\xi, Q}\} \leq \rho_p^p(P_T, \tilde{P}_T) \leq C_p E_Q\{\langle \xi \rangle_T^{p/2} + |x|^p \star v_T^{\xi, Q}\} \quad (2.6)$$

with some other constants  $c_p$  and  $C_p$ . The proof of the above inequalities is based on (2.4) and the following general lemma.

**Lemma 2.1.** *Let  $M$  be a  $(P, \mathbb{F})$  martingale with  $M_0 = 0$  and  $T$  a stopping time. Then for  $p \geq 2$*

$$\begin{aligned} c_p E_P\{\langle M \rangle_T^{p/2} + |(\Delta M)_T^*|^p\} &\leq E_P(M_T^*)^p \\ &\leq C_p E_P\{\langle M \rangle_T^{p/2} + |(\Delta M)_T^*|^p\} \end{aligned} \quad (2.7)$$

and with different constants  $c_p$  and  $C_p$

$$\begin{aligned} c_p E_P\{\langle M \rangle_T^{p/2} + |x|^p \star v_T^{M, P}\} &\leq E_P(M_T^*)^p \\ &\leq C_p E_P\{\langle M \rangle_T^{p/2} + |x|^p \star v_T^{M, P}\}. \end{aligned} \quad (2.8)$$

*Proof.* As noted above, we depart from Burkholder inequality (see Lenglart et al. (1980), or Liptser and Shiryaev (1989))

$$c_p E_P[M]_T^{p/2} \leq E_P(M_T^*)^p \leq C_p E_P[M]_T^{p/2}.$$

Since  $|\Delta M|_T^p \leq |x|^p \star \mu_T^M \leq [M]_T^{p/2}$  and  $E_P \langle M \rangle_T^l \leq l^l E_P[M]_T^l$  for  $l \geq 1$  (see Liptser and Shiryaev (1989), Theorem 1.9.7), the left-hand side of (2.7) and (2.8) follows from the left-hand side of Burkholder's inequality.

The right-hand side of (2.8) follows from the right-hand side of (2.7) and from the above mentioned fact  $|\Delta M|_T^p \leq |x|^p \star \mu_T^M$ . To prove the right-hand side of (2.7) introduce the Davies' decomposition of the martingale  $M = K + L$ . Put

$S = (\Delta M)^*$ . Then  $|\Delta L| \leq 4S_-$  and

$$E_p(K^*)_T^p \leq C_p E_p S_T^p, \tag{2.9}$$

(see Liptser and Shiryaev (1989), Lemmas 1.9.4 and 1.9.7). Now, according to the results of Lenglart et al. (1980) we have that

$$E_p(L^*)_T^p \leq C_p E_p (\langle L \rangle_T^{p/2} + S_{T-}^p). \tag{2.10}$$

Furthermore, by the orthogonality of the components  $K$  and  $L$  in Davies decomposition ( $[K, L] = 0$  and hence  $\langle M \rangle = \langle L \rangle + \langle K \rangle$ ) the inequality (2.10) can be extended as follows:

$$E_p(L^*)_T^p \leq C_p E_p (\langle M \rangle_T^{p/2} + S_{T-}^p). \tag{2.11}$$

Finally, combine (2.9) and (2.11) to get the right-hand side inequality in (2.7).  $\square$

*Remark 2.1.* If  $E_p(M_T^*)^p = \infty$ , then we can interpret (2.7) and (2.8) as equalities.

*Remark 2.2.* We did not specify the constants  $c_p$  and  $C_p$ . We note only that

$$c_p \rightarrow 0 \quad \text{and} \quad C_p \rightarrow \infty \quad \text{as} \quad p \rightarrow \infty.$$

*Remark 2.3.* If  $\xi$  is a continuous martingale we have

$$[\xi] = \langle \xi \rangle = 2X^2 \circ h, \quad X = \xi \tilde{\xi} \tag{2.12}$$

(cf. (1.3)), so in this special case (2.5) implies directly the following inequalities

$$c_p E_Q (X^2 \circ h)_T^{p/2} \leq \rho_p^p(P_T, \tilde{P}_T) \leq C_p E_Q (X^2 \circ h)_T^{p/2}. \tag{2.13}$$

The inequalities of this type, valid in the general case, not necessarily for continuous  $\xi$ , are derived from (2.7) (instead of (2.5)) in the next section.

### 3. Main Results

3.1. The inequalities (1.6) and (1.7) follow easily from the corresponding statements of Theorem 3.2 below. The proof of this theorem is based on the following statements of independent interest (note that here, in contrast with Theorem 3.2 below,  $p$  is not necessarily positive even integer):

**Theorem 3.1.** *Let  $S$  and  $T$  be stopping times,  $S \leq T$ . For  $p \geq 2$  there are constants  $c_p$  and  $C_p$  such that*

$$\rho_p^p(P_T, \tilde{P}_T) \geq c_p E_Q \{ (X_-^2 \circ h)_T^{p/2} + (X_-^p \circ k(p))_T \} \tag{3.1}$$

and

$$\rho_p^p(P_T, \tilde{P}_T) \leq C_p E_Q \{ (X_- \circ h)_S^{p/2} + (X_- \circ k(p))_S \} + 2Q(S < T). \tag{3.2}$$

with  $X = \xi \tilde{\xi}$ .

*Remark 3.1.* Note that in the special case of a continuous martingale  $\xi$  (and  $S = T$ ) the right-hand side inequality in (2.13) is sharper than (3.2). This follows from the

fact that in this particular case we can use (2.12) instead of the right-hand side inequality in Lemma 3.1 below, valid in the general case.

If the measures  $P$  and  $\tilde{P}$  correspond to processes with independent increments, then the processes  $h$  and  $k(p)$  can be assumed to be deterministic (see the last section of this paper). In this particular case we have

**Corollary 3.1.** *Suppose that the processes  $h$  and  $k(p)$  and the stopping time  $T$  are deterministic. Then we can replace (3.1) and (3.2) with the following inequalities:*

$$\rho_p^p(P_T, \tilde{P}_T) \geq c(p; T, h)(h_T^{p/2} + k_T(p)) \quad (i)$$

and

$$\rho_p^p(P_T, \tilde{P}_T) \leq C_p(h_T^{p/2} + k_T(p)). \quad (ii)$$

*Proof of Theorem 3.1.* In view of (2.6) it suffices to apply the following lemmas, the first two of which give the corresponding estimates of the expectations of two terms involved in (2.6), and the third one leads to the upper bound of form (3.2).

**Lemma 3.1.** *Let  $X = \xi \tilde{\xi}$  and let  $h$  be given by (1.3). Then*

$$\frac{1}{2} X_-^2 \circ h \leq \langle \xi \rangle \leq 2 X_- \circ h$$

(cf. Jacod and Shiryaev (1987), Lemma V.4.26).

*Proof.* In view of (1.3) and the easily verified facts that  $X \leq 1$  and

$$\langle \xi \rangle = \langle \xi^c \rangle + \left( X_- \left( \frac{\lambda - \tilde{\lambda}}{2} \right) \right)^2 \star \nu^{\xi, \mathcal{Q}}$$

by (2.3), it suffices to verify only that

$$X_- (\sqrt{\lambda} - \sqrt{\tilde{\lambda}})^2 \leq X_- (\lambda - \tilde{\lambda})^2 \leq 4 (\sqrt{\lambda} - \sqrt{\tilde{\lambda}})^2, \quad (3.3)$$

by taking into consideration that  $\sqrt{\lambda} + \sqrt{\tilde{\lambda}} \geq 1$ , and that  $X_- (\sqrt{\lambda} + \sqrt{\tilde{\lambda}})^2 \leq 4$  due to Shwartz' inequality and the identities:  $\xi + \tilde{\xi} = 2$  and  $\xi_- \lambda + \tilde{\xi}_- \tilde{\lambda} = 2$ .  $\square$

The inequalities (3.3) can easily be extended to the case  $p \geq 2$ :

$$X_-^{p-1} (\lambda^{1/p} - \tilde{\lambda}^{1/p})^p \leq X_-^{p-1} (\lambda - \tilde{\lambda})^p \leq 4^{p-1} (\lambda^{1/p} - \tilde{\lambda}^{1/p})^p,$$

and this gives

**Lemma 3.2.** *For  $p \geq 2$*

$$(1/2)^p X_-^p \circ k(p) \leq |x|^p \star \nu^{\xi, \mathcal{Q}} \leq 2^{p-2} X_- \circ k(p)$$

**Lemma 3.3.** *For two stopping times  $S \leq T$ , and  $p \geq 1$*

$$\rho_p^p(P_T, \tilde{P}_T) \leq \rho_p^p(P_S, \tilde{P}_S) + 2Q(S < T)$$

*Proof.* For  $p = 1$  see Jacod and Shiryaev (1987), p. 280. The general case is treated analogously.  $\square$

*Proof of Corollary 3.1.* By  $X \leq 1$  (ii) is obvious. To prove (i) observe that the function  $f_t = E_Q \sqrt{X_t}$  is decreasing, in fact  $f_t = \mathcal{E}_t(-h)$  where  $\mathcal{E}$  is Dolean-Dade's exponential (as it satisfies  $f_t = 1 - (f_- \circ h)_t$  in accordance with Jacod and Shiryaev

(1987), IV.1.20), and this and Jensen's inequality entail

$$\inf_{s \leq t} E_Q(X_{s-})^p \geq \inf_{s \leq t} (E_Q \sqrt{X_{s-}})^{2p} \geq f_t^{2p}$$

Hence (i) takes place with  $c(p; T, h) = c_p(\mathcal{E}_T(-h))^{2p}$ .  $\square$

*Remark 3.2.* In the simplest case  $p = 2$  we have the following representation

$$\rho_2^2(P_T, \tilde{P}_T) = 2E_Q(X_-^{1/2} \circ h)_T \tag{3.4}$$

(from the definition of the Hellinger process, for example, Valkeila and Vostrikova (1986)). Comparing (3.4) and (3.1) for  $p = 2$ , with  $2E_Q(X_-^2 \circ h)$  on the right-hand side (constants here and in the next paragraph are defined by (2.4) with  $p = 2$ ) we see that the lower bound obtained is quite crude; cf. also Corollary 3.1, Assertion (i).

As for the upper bound (3.2) for  $p = 2$  and  $S = T$ , with  $4E_Q(X_- \circ h)$  on the right-hand side, it is simply derived from (3.4) by the following considerations:

$$\begin{aligned} \rho_2^2 &\geq \frac{1}{2} E_Q |\xi^{1/2} - \tilde{\xi}^{1/2}|^4 \\ &\geq E_Q |\xi^{1/2} - \tilde{\xi}^{1/2}|^2 \circ |\xi^{1/2} - \tilde{\xi}^{1/2}|^2 \\ &= 2\rho_2^2 - 2E_Q X_-^{1/2} \circ |\xi^{1/2} - \tilde{\xi}^{1/2}|^2 \\ &= 2\rho_2^2 - 4E_Q X_- \circ h \end{aligned}$$

Here we have first used the inequality  $|\xi^{1/2} - \tilde{\xi}^{1/2}|^2 \leq 2$ , then Ito's formula and, finally, (3.4).

*Remark 3.3.* By Jacod and Shiryaev (1987), Lemma I.3.12, we have

$$E_Q(X_- \circ h)_T \leq 2E_Q(\xi_- \circ h)_T = 2E_Q \xi_T h_T = 2E_P h_T,$$

since  $\tilde{\xi} \leq 2$ , and this gives (1.7). Thus the upper bound here can be given in terms of the expectation with respect to the measure  $P$ . For the general result see the following theorem.

**Theorem 3.2.** *Let  $S$  and  $T$  be stopping times,  $S \leq T$ . For a positive even integer  $p$  there are constants  $C_p$  and  $B_p$  such that*

$$\rho_p^p(P_T, \tilde{P}_T) \leq C_p E_P(h_s^{p/2} + k(p)_S) + B_p P^{1/p}(S < T)$$

*Proof.* In view of (3.2) it suffices to show that

$$E_Q(X_- \circ k(p)_T) \leq 2E_P k(p)_T, \tag{3.5}$$

$$E_Q(X_- \circ h)_T^{p/2} \leq p E_P h_T^{p/2} \tag{3.6}$$

and

$$\rho_p^p(P_T, \tilde{P}_T) - \rho_p^p(P_S, \tilde{P}_S) \leq B_p P^{1/p}(S < T). \tag{3.7}$$

Since  $\tilde{\xi} \leq 2$ , (3.5) follows from Jacod and Shiryaev (1987), Lemma I.3.12.

To prove (3.6) apply the same lemma, along with the considerations of Liptser and Shiryaev (1989), Lemma I.9.6: for  $A = X_- \circ h$  we have

$$\begin{aligned} E_Q A_T^{p/2} &\leq \frac{p}{2} E_Q \int_0^T A_s^{(p/2)-1} dA_s \\ &\leq \frac{p}{2} E_Q \int_0^T X_{s-} h_s^{(p/2)-1} dh_s \\ &\leq p E_Q \int_0^T \zeta_{s-} h_s^{(p/2)-1} dh_s = p E_P \int_0^T h_s^{(p/2)-1} dh_s \leq p E_P h_T^{p/2}. \end{aligned}$$

For (3.7) we use the approach of Vostrikova (1987), Theorem 2.2. We note first that

$$\rho_p^p(P_T, \tilde{P}_T) - \rho_p^p(P_S, \tilde{P}_S) = E_P \left( 1_{\{S < T\}} \sum_{k=1}^{p-1} \binom{n}{k} (-1)^k (Z_S^{k/p} - Z_T^{k/p}) \right)$$

with  $Z = \tilde{\xi}/\xi$ . Then we apply Hölder's inequality with exponents  $1/p$  and  $p - 1/p$  to the right-hand side of the above equation. Finally, evaluating the expression

$$\left( E_P \sum_{k=1}^{p-1} (-1)^k (Z_T^{k/p} - Z_S^{k/p})^{p/(p-1)} \right)^{(p-1)/p}$$

we use the fact that  $E_P Z_T^\alpha \leq 1$  for  $0 < \alpha \leq 1$ . This shows that one can in (3.7) take  $B_p = p^{1/p}(2^p + 1 - 4)^{(p-1)/p}$  which gives, in particular,  $B_2 = 2\sqrt{2}$ .  $\square$

#### 4. Sequences of Binary Experiments

4.1. In the present section we consider certain applications to sequences of binary experiments

$$(\Omega^n, \mathbf{F}^n, P^n, \tilde{P}^n), \quad n = 1, 2, \dots \quad (4.1)$$

with the associated density processes  $\xi^n$  and  $\tilde{\xi}^n$  as in (1.1), and the corresponding Hellinger process  $h^n$  and  $p$ -divergency processes  $k^n(p)$ ,  $p \geq 2$  defined as in (1.3) and (1.5).

We remark first that in view of the properties of the distances  $\rho_p$  indicated in Subsection 2.2, the limiting (as  $n \rightarrow \infty$ ) behaviour of  $\rho_p(P_{T_n}^n, \tilde{P}_{T_n}^n)$ , defined by (1.2) with a sequence of stopping times  $T_n$ ,  $n = 1, 2, \dots$ , is controlled under the circumstances

$$h_{T_n}^n \xrightarrow{P^n} 0 \quad \text{or} \quad h_{T_n}^n \xrightarrow{P^n} \infty \quad (4.2)$$

in the exactly same way as that of the variational distance  $\|P_{T_n}^n - \tilde{P}_{T_n}^n\|$  (see Jacod and Shiryaev (1987), Theorem 4.32).

Contrary to (4.2), in the next subsection we consider the situation in which a sequence of the Hellinger possesses a certain limit in  $P^n$ -probability.

4.2. Let  $t \mapsto C_t$  be a non-decreasing continuous function with  $C_0 = 0$ . Let  $M$  be a continuous martingale with  $M_0 = 0$  and  $\langle M, M \rangle_t = C_t$ , on some stochastic basis  $(\Omega, \mathbf{F}, P)$  (so  $M$  is Gaussian).



Let  $\mathcal{D}$  be a dense subset in  $\mathbf{R}_+$ . Consider the following conditions:

$$h_t^n \xrightarrow{P^n} \frac{1}{8} C_t \quad \text{for all } t \in \mathcal{D} \quad (\text{a})$$

(that is Condition  $[H - D]$  in Jacod and Shiryaev (1987), Theorem X.1.12) and

$$k_t^n(p) \xrightarrow{P^n} 0 \quad \text{for some } p > 2. \quad (\text{b})$$

Along with the processes  $h^n$  and  $k^n(p)$ ,  $p \geq 2$ , we will associate with (4.1) a new process  $I_t^n(a)$  for  $a > 1$ :

$$I^n(a) = 1_{\{1/a < \lambda^n/\tilde{\lambda}^n < a\}^c} |\lambda^n - \tilde{\lambda}^n| \star \nu^{\xi^n, Q^n}$$

where  $\tilde{\lambda}^n$  and  $\lambda^n$  are defined as in (2.2) and  $Q^n = (P^n + \tilde{P}^n)/2$  obviously, and we consider Condition  $[L - D]$  in Jacod and Shiryaev (1987), Theorem X.1.12:

$$I^n(1 + \varepsilon)_t \xrightarrow{P^n} 0 \quad \text{for all } t \in \mathcal{D}, \varepsilon > 0. \quad (\text{c})$$

Set  $Z^n = \tilde{\xi}^n/\xi^n$ , and consider the following statement:

$$Z^n \xrightarrow{\mathcal{L}(P^n)} Z = e^{M - C/2}, \quad (\text{i})$$

with  $M$  and  $\langle M \rangle = C$  defined above.

The following extension of Theorem X.1.12 by Jacod and Shiryaev (1987) takes place:

**Theorem 4.1.** *The statement (i) is equivalent to the following two statements:*

- (ii) *Conditions (a) and (c) hold;*
- (iii) *Conditions (a) and (b) hold;*

*Proof.* For (i)  $\Leftrightarrow$  (ii) see Jacod and Shiryaev (1987), Theorem X.1.12. To show (ii)  $\Leftrightarrow$  (iii) denote by  $A_a(\lambda, \tilde{\lambda})$  the set the indicator of which is involved in (4.3) (we suppress the index  $n$ , as it is superfluous here).

It is easily verified that the validity of the following two statements is enough for the proof:

1) for each  $\varepsilon$ ,  $0 < \varepsilon < 1$  and  $p \geq 2$

$$k(p) \leq \left( \frac{2\varepsilon}{1 - \varepsilon} \right)^{p-2} k(2) + I\left( \frac{1 + \varepsilon}{1 - \varepsilon} \right);$$

2) for each  $p > 2$  and  $a > 1$  there is a constant  $C_{a,p} > 0$  such that

$$1_{A_a(\lambda, \tilde{\lambda})} |\lambda^{1/p} - \tilde{\lambda}^{1/p}|^p \star \nu^{\xi, Q} \leq I(a) \leq C_{a,p} 1_{A_a(\lambda, \tilde{\lambda})} |\lambda^{1/p} - \tilde{\lambda}^{1/p}|^p \star \nu^{\xi, Q}$$

Statement 1) follows from the simply verified inequalities

$$(u^{1/p} - 1)^p \leq \begin{cases} (u^{1/2} - 1)^2 (2\varepsilon/(1 - \varepsilon))^{p-2} & \text{if } 1 \leq u \leq \frac{1 + \varepsilon}{1 - \varepsilon}, \\ u - 1 & \text{if } u > \frac{1 + \varepsilon}{1 - \varepsilon}, \end{cases}$$

and Statement 2) from (2.2) and the fact that the continuous function  $|u^{1/p} - 1|^p / |u - 1|$  vanishes as  $u \rightarrow 1$  and tends to one as  $u \rightarrow \infty$ .  $\square$

*Remark 4.1.* The relation between  $k(p)$  and the Hellinger processes  $h(\alpha)$  of order  $\alpha \in (0, 1)$  indicated at the end of Subsection 2.2, allows one to trace directly the equivalence of (iii) above and (ii) or (iii) in Jacod and Shiryaev (1987), Theorem X.1.64.

4.3. Under the circumstances of the previous subsection we have

**Theorem 4.2.** *Statement (i) implies*

$$\limsup_{n \rightarrow \infty} \rho_p^n(P_t^n, \tilde{P}_t^n) \leq K_p C_t^{p/2}$$

with a certain constant  $K_p$ .

*Proof.* Let  $S_n = \inf\{s | h_s^n \geq C_t + 1\}$ . Then

$$k_{S_n \wedge t}^n(p) \leq 2h_{S_n \wedge t}^n \leq 2C_t + 3,$$

since  $\Delta h \leq 1$ , and  $\{S_n < t\} \subset \{h_t^n \geq C_t + 1\}$ . Hence  $P^n(S_n < t) \rightarrow 0$  under (i), and this implies in turn that

$$h_{S_n \wedge t}^n \xrightarrow{P^n} \frac{1}{8} C_t$$

and

$$k_{S_n \wedge t}^n(p) \xrightarrow{P^n} 0.$$

But the sequences  $k_{S_n \wedge t}^n(p)$  and  $h_{S_n \wedge t}^n$  are bounded and hence under (i)

$$E_{P^n}(h_{S_n \wedge t}^n)^{p/2} \rightarrow \left(\frac{1}{8} C_t\right)^{p/2}$$

and

$$E_{P^n}(k_{S_n \wedge t}^n(p)) \rightarrow 0.$$

This, in view of Theorem 3.2, gives the result.  $\square$

## 5. Parametric Families of Experiments and Examples

5.1. We consider here an application to a parametric family of experiments

$$(\Omega, \mathcal{F}, \mathbb{F}, \{P^\theta, \theta \in \Theta\}, Q)$$

where  $\Theta$  is a closed subset of the Euclidean space  $R^d$ , and  $Q$  is a measure dominating the family  $\{P^\theta, \theta \in \Theta\}$  of probability measures depending continuously on a parameter  $\theta$ .

We retain here the assumptions and notations of Introduction (with a general dominating measure  $Q$ , however) writing specifically (for  $\theta, \theta + u \in \Theta$ )

$$\rho_p(P_T^{\theta+u}, P_T^\theta) = \{E_Q |\xi_T(\theta + u)^{1/p} - \xi_T(\theta)^{1/p}|^p\}^{1/p} \quad (5.1)$$

with  $p \geq 2$  and

$$\xi_T(\theta) = dP_T^\theta / dQ_T. \quad (5.2)$$

Analogously, we define the processes  $h(\theta + u, \theta)$  and  $k(p; \theta + u, \theta)$  by the formulas (1.3) and (1.5) respectively, with  $\xi = \xi(\theta)$  and  $\tilde{\xi} = \xi(\theta + u)$  this time.

5.2. We wish to evaluate the expectation  $E_Q$  with respect to the dominating measure  $Q$  of the following modulus of continuity (for a certain  $p > d$ )

$$\omega_p(\delta, L; P_T^\theta, P_T^{\theta+u}) = \sup |\xi_T(\theta + u)^{1/p} - \xi_T(\theta)^{1/p}|^p, \quad (5.3)$$

where sup is taken over  $\theta, \theta + u \in \Theta$  with  $|\theta| \leq L, |\theta + u| \leq L$  and  $|u| \leq \delta$ .

**Theorem 5.1.** *Let the following Lipschitz type conditions be satisfied: there is a bounded function  $B_\theta$  of  $\theta$  such that for each  $\theta, \theta + u \in \Theta$*

$$E^\theta h_T^{p/2}(\theta, \theta + u) \leq B_\theta |u|^p, \quad E^\theta k_T(p; \theta, \theta + u) \leq B_\theta |u|^p \quad (5.4)$$

with the expectation relative to the measure  $P^\theta$ .

Then for an even  $p > d$

$$E_Q(\omega_p(\delta, L; P_T^\theta, P_T^{\theta+u}) \leq B_0 \sup_{|\theta| < L} B_\theta L^{dp} \delta^{p-d}$$

where the constant  $B_0$  depends on  $d$  and  $p$  only.

*Proof.* We apply here Theorem 19 in Ibragimov and Has'minskii (1981), Appendix I. All of its conditions are satisfied: the first one in (7), p. 372 by  $E_Q \xi_T(\theta) \leq 1$  and the second one by Theorem 3.2 above which implies

$$\rho_p(P_T^{\theta+u}, P_T^\theta) \leq C_p E^\theta \{h_T^{p/2}(\theta, \theta + u) + k_T(p; \theta, \theta + u)\} \leq C_p B_\theta |u|^p$$

in view of (5.1), (5.2) and (5.4).

5.3. We give some examples of the  $p$ -divergence process, where  $p \geq 2$  and  $p$  is even. Because  $p$  is even we use the following equality to construct our examples from the corresponding examples for the Hellinger processes of order  $\alpha$ :

$$k(p) = - \sum_{k=1}^{p-1} (-1)^k \binom{p}{k} h(k/p)$$

(see Section 2.1). We take  $(\Omega, \mathcal{F})$  to be the canonical space of all càdlàg functions  $\omega: \mathbf{R}_+ \rightarrow \mathbf{R}^1$ ;  $X$  is the "canonical process"  $X_t(\omega) = \omega_t$  and  $F_t = \sigma\{X_s | s \leq t\}$  is generated by  $X$  (and completed with respect to  $P$  and  $\tilde{P}$ ). We refer to Jacod and Shiryaev (1987) for more details.

*Point processes.* Our first example deals with *point processes* (or *counting processes*). Let  $A$  and  $\tilde{A}$  be the compensators of  $X$  with respect to the measures  $P$  and  $\tilde{P}$ . Let  $B$  be any increasing predictable càdlàg process such that  $dA \ll dB$  and  $d\tilde{A} \ll dB$ . There are two predictable non-negative processes  $\lambda$  and  $\tilde{\lambda}$  such that

$$A = \lambda \circ B, \quad \tilde{A} = \tilde{\lambda} \circ B \quad (P + \tilde{P})\text{-a.s.}$$

Then (cf. Jacod and Shiryaev Theorem IV.4.2) we have

$$k(p; P, \tilde{P}) = |\lambda^{1/p} - \tilde{\lambda}^{1/p}|^p \circ B + \sum_{s \leq t} |(1 - \Delta A_s)^{1/p} - (1 - \Delta \tilde{A}_s)^{1/p}|^p. \quad (5.5)$$

*Independent random variables.* Our second example deals with *discrete case and independent random variables* (cf. Jacod and Shiryaev, IV.1). Here  $P$  (resp.  $\tilde{P}$ ) is the law of a sequence of independent real valued random variables. Suppose that the law of  $n^{\text{th}}$  coordinate is given by the density  $f_n$  (resp.  $\tilde{f}_n$ ) with respect to some common measure  $\nu$ . Then we have

$$k(p)_n = \sum_{t \leq n} \int_{\mathbf{R}} |f_t^{1/p}(x) - \tilde{f}_t^{1/p}(x)|^p \nu(dx). \quad (5.6)$$

*Processes with independent increments.* We follow again Jacod and Shiryaev (1987) IV.4. We assume that the canonical process  $X$  is a process with independent increments under  $P$  (resp.  $\tilde{P}$ ) and hence its canonical triplet  $T = (B, C, \nu)$  (resp.  $\tilde{T} = (\tilde{B}, \tilde{C}, \tilde{\nu})$ ) is deterministic. Suppose now that  $C = \tilde{C}$ , put  $\pi = \nu + \tilde{\nu}$  and define

$$U = d\nu/d\pi \quad \text{and} \quad \tilde{U} = d\tilde{\nu}/d\pi$$

and

$$a_t = \nu(\{t\} \times \mathbf{R}) \quad \text{and} \quad \tilde{a}_t = \tilde{\nu}(\{t\} \times \mathbf{R}).$$

Then the  $p$ -divergency process has the following version

$$k(p) = |U^{1/p} - \tilde{U}^{1/p}|^p \star \pi + \sum_{s \leq \cdot} |(1 - a_s)^{1/p} - (1 - \tilde{a}_s)^{1/p}|^p. \quad (5.7)$$

*Remark 5.1.* We note that one can construct examples of the  $p$ -divergency process also for *general semimartingales* using the connection between the  $p$ -divergency process and Hellinger-processes of order  $\alpha$ . This can be done with the help of the corresponding examples in Jacod and Shiryaev (1987). We leave this for the interested reader.

*Acknowledgements.* The authors thank the Centre for Mathematics and Computer Science, the Finnish Academy and the Ella and Georg Ehrnrooth Foundation for the financial support. This paper has appeared as a preprint (CWI Report MS-R8818) in a slightly different form.

## References

1. Burkholder, D.L.: Distribution function inequalities for martingales, *Ann. Probab.* **1**, 19–42 (1973)
2. Dzharidze, K.: Parameter estimation and hypothesis testing in spectral analysis of stationary time series. Berlin Heidelberg New York: Springer 1986
3. Ibragimov, I.A., Has'minskii, R.Z.: Statistical estimation – asymptotic theory. Berlin Heidelberg New York: Springer 1981
4. Jacod, J.: Calcul stochastique et problèmes de Martingales, (Lect. Notes Math., vol. 714) Berlin Heidelberg New York: Springer 1979
5. Jacod, J., Shiryaev, A.N.: Limit theorems for stochastic processes. Berlin Heidelberg New York: Springer 1987
6. Kutoyants, Yu. A.: Parameter estimation for stochastic processes. Berlin: Heldermann 1984
7. Lenglar, E., Lepingle, D., Pratelli, M.: Presentation unifiée de certaines inegalites de la theorie des Martingales. (Lect. Notes Math., vol. 784, pp. 26–48). Berlin Heidelberg New York: Springer 1980
8. Liese, F., Vajda, I.: Convex statistical distances. Leipzig: Teubner 1987
9. Liptser, R.Sh., Shiryaev, A.N.: Theory of Martingales. Dordrecht: Kluwer 1989
10. Novikov, A.A.: On discontinuous martingales. *Theory Probab. Appl.* **20**, 11–26 (1975)
11. Valkeila, E., Vostrikova, L.: An integral representation for the Hellinger distance. *Math. Scand.* **58**, 239–254 (1986)

12. Valkeila, E., Vostrikova, L.: On predictable criteria of  $(c_n)$ -consistency of estimates. *Theory Probab. Appl.* **32**, 477–489 (1987)
13. Vostrikova, L.: On  $F$ -processes and their applications. In: Kalashnikov, V.V. Penkov, B., Zolotarev, V.M. (eds.) *Stability problems for stochastic models.* (Lect. Notes Math., vol. 1233, pp. 190–203) Berlin Heidelberg New York: Springer 1987
14. Vostrikova, L.: On the weak convergence of likelihood ratio processes of general statistical parametric models. *Stochastics* **23**, 277–298 (1988)

Received January 31, 1989; in revised form October 2, 1989