The Annals of Statistics
1996, Vol. 24, No. 4, 1804–1812

A NOTE ON RITOV'S BAYES APPROACH TO THE MINIMAX PROPERTY OF THE CUSUM PROCEDURE

BY M. BEIBEL

University of Freiburg

We consider, in a Bayesian framework, the model $W_t = B_t + \theta(t-\nu)^+$, where B is a standard Brownian motion, θ is arbitrary but known and ν is the unknown change-point. We transfer the construction of Ritov to this continuous time setup and show that the corresponding Bayes problems can be reduced to generalized parking problems.

1. Introduction. Page (1954) proposed the cusum procedures as a solution to the problem of sequential detection of a change in distribution. Lorden (1971) introduced a notion of minimax optimality for change-point models and showed that the cusum procedures are asymptotically optimal in his sense. Moustakides (1986) then showed the minimax optimality of the cusum procedures, when the initial and final distribution are both known. Ritov (1990) used a Bayes approach and gave an interesting argument for Moustakides' result similar to that of Lehmann (1959) on the optimality of the sequential probability ratio test. We transfer his approach to a continuous time model and discuss the structure of the Bayes risk.

Let θ be an arbitrary but known constant. We will consider a process W with

$$W_t = B_t + \theta(t-\nu)^+, \qquad 0 \le t < \infty,$$

where B is a standard Brownian motion and ν denotes a random change-point. We will specify ν in Section 2. Let $\bar{P}$ denote the corresponding probability measure. For a stopping time T of W we consider the risk function

$$R(T) = \bar{P}(T < \nu) - C_1 E_{\bar{P}}(T \wedge \nu) + C_2 E_{\bar{P}}(T-\nu)^+,$$

where C_1 and C_2 are positive constants. Here stopping with T means deciding that the change has already taken place. To find the minimizing stopping time T^*, let $(\pi_t,\ 0 \le t < \infty)$ denote a continuous version of $\bar{P}\big(\nu \le t | W_s,\ 0 \le s \le t\big)$. We show that for all bounded stopping times T (with respect to the observed process W) the Bayes risk can be written as

$$R(T) = E_{\bar{P}} g(\pi_T)\ ,$$

where g is a twice continuous differentiable function which attains a unique minimum over $[p, 1)$ at some point p^*. This means we can reduce the problem of finding a Bayes solution T^* which minimizes $R(T)$ under all stopping

Received June 1992; revised November 1995.

AMS 1991 *subject classifications*. Primary 62L10; secondary 62L15, 62C10.

Key words and phrases. Change-point, cusum procedures, sequential detection, Wiener process, generalized parking problems.

times T to a generalized parking problem [see Woodroofe, Lerche and Keener (1993)]. Since π_t is continuous in t, one can stop exactly at the minimum and immediately obtain that

$$T^* = \inf\{t > 0 | \pi_t = p^*\} .$$

This approach is especially simple and makes no use of the optimal stopping theory. The method is well adapted to optimality proofs in sequential testing even in the case of composite hypotheses. It works for Shiryaev's disruption problem [see Beibel (1994)] as well as for many other cases [see, e.g., Lerche (1986) and Beibel and Lerche (1995)].

2. The setup. Let $C_0[0, \infty)$ denote the space of continuous, real-valued functions on the interval $[0, \infty)$ which vanish at zero. Let $\mathscr{B}(C_0[0, \infty))$ denote the Borel σ-algebra on $C_0[0, \infty)$ with respect to the topology of uniformly compact convergence. Let $\mathbb{R}_0^+$ denote the set of all nonnegative reals and let $\mathscr{B}(\mathbb{R}_0^+)$ denote the usual Borel σ-algebra on $\mathbb{R}_0^+$. Let μ_W denote the standard Wiener measure on $(C_0[0, \infty), \mathscr{B}(C_0[0, \infty))$. Let $\theta \in \mathbb{R}$ be an arbitrary constant. Let $p \in (0, 1)$. Let ρ denote the probability measure on $\mathbb{R}_0^+$ with

$$\rho(\{0\}) = p,$$
$$\rho((s, \infty]) = (1 - p) \exp\{-ps\} .$$

We shall now consider the product space

$$(\Omega, \mathscr{B}) = \big(C_0[0, \infty) \times \mathbb{R}_0^+, \mathscr{B}(C_0[0, \infty)) \otimes \mathscr{B}(\mathbb{R}_0^+)\big)$$

with product measure $P = \mu_W \otimes \rho$. Let $W(\cdot)$ and V denote the random variables on $(\Omega, \mathscr{B})$ defined by the projection on the first and the second component, respectively. Let

$$\mathscr{F}_t^W := \sigma(W_s;\ s \le t) \quad \text{and} \quad \mathscr{F}_t := \sigma(V, W_s;\ s \le t)$$

for $0 \le t < \infty$. Let $\mathscr{F}_\infty^W := \sigma(W_s,\ s \ge 0)$. Let Λ denote the increasing process on $(\Omega, \mathscr{F}_t)$ given by

$$\Lambda_t := -\min_{0 \le s \le t} \left(\theta W_s - \frac{\theta^2}{2} s\right).$$

We now define random variables τ_u by

$$\tau_u := \inf\{0 \le t |\ \Lambda_t = u\}$$

for $u \in [0, \infty)$. The τ_u are stopping times with respect to the filtration $(\mathscr{F}_t;\ 0 \le t < \infty)$. Further we define the random variable ν by $\nu := \tau_V$. The random time ν is a stopping time with respect to the filtration $(\mathscr{F}_t;\ 0 \le t < \infty)$ and

$$P(\nu > t) = (1 - p) E \exp\{-p\Lambda_t\} .$$

Note that

$$V = \Lambda_\nu = -\theta W_\nu + \frac{\theta^2}{2}\nu.$$

The expression above for $P(\nu > t)$ illustrates the analogy between the priors which we consider and the exponential priors of Shiryayev [(1973), page 158]. Loosely speaking, we replace linear deterministic time by the random time scale Λ. We will consider the stopping time ν as a random change-point. Such a change-point depends on the observations before the change occurs. We will now construct a probability measure $\bar{P}$ on $(\Omega, \mathscr{B})$ such that

$$B_t := W_t - \theta(t-\nu)^+$$

is a standard Brownian motion. For this purpose we apply Girsanov's formula to the exponential supermartingale Z_t given by

$$Z_t := \exp\left\{\theta \int_0^t 1_{\{\nu \le s\}}\, dW_s - \frac{\theta^2}{2}\int_0^t 1_{\{\nu \le s\}}\, ds\right\}.$$

Note that

$$Z_t = \exp\left\{\theta W_t - \frac{\theta^2}{2}t + V\right\} 1_{\{\nu \le t\}} + 1_{\{\nu > t\}}.$$

Obviously, for all $t \ge 0$,

$$E \exp\left\{\frac{\theta^2}{2}(t-\nu)^+\right\} < +\infty.$$

This means that Novikov's condition is fulfilled and there exists a probability measure $\bar{P}$ on $(\Omega, \mathscr{B})$ such that B_t $(0 \le t < \infty)$ is a standard Brownian motion on $(\Omega, \mathscr{B}, \bar{P})$ [see Karatzas and Shreve (1988), page 192]. For simplicity of notation, we shall write $\bar{E}$ instead of $E_{\bar{P}}$.

In the next two sections the parameter p will be arbitrary but fixed. Therefore we do not indicate explicitly the dependence of P and $\bar{P}$ on p.

3. The posterior probability. We now calculate the posterior probability under $\bar{P}$ that a change has taken place up to time t given the observation W up to time t, $\bar{P}(\nu \le t | \mathscr{F}_t^W)$, which we denote by π_t.

LEMMA 1. *For* $A \in \mathscr{F}_t^W$,

(i) $$\bar{E}1_A 1_{\{\nu > t\}} = (1-p)\; E1_A \exp\{-p\Lambda_t\},$$

(ii) $$\bar{E}1_A 1_{\{\nu \le t\}} = p\; E1_A \exp\left\{(1-p)\Lambda_t + \theta W_t - \frac{\theta^2}{2}t\right\}.$$

PROOF. Equality (i) follows by straightforward calculations. For $A \in \mathscr{F}_t^W$,

$$\begin{aligned}\bar{E}1_A 1_{\{\nu \le t\}} &= EZ_t 1_A 1_{\{\nu \le t\}} \\ &= E\left[1_A 1_{\{\Lambda_t \ge V\}} \exp\left\{\theta W_t - \frac{\theta^2}{2}t + V\right\}\right] \\ &= E\left[1_A \exp\left\{\theta W_t - \frac{\theta^2}{2}t\right\}\left(\int_0^{\Lambda_t} e^v \rho(dv)\right)\right] \\ &= pE\left(1_A \exp\left\{\theta W_t - \frac{\theta^2}{2}t + (1-p)\Lambda_t\right\}\right) .\end{aligned}$$

This proves (ii). □

From Lemma 1 we immediately come to the following conclusion:

LEMMA 2.

$$\pi_t = \frac{p\exp\{\Lambda_t + \theta W_t - (\theta^2/2)t\}}{p\exp\{\Lambda_t + \theta W_t - (\theta^2/2)t\} + (1-p)}.$$

According to Lemma 2 the process π_t is an isotonic function of the cusum statistic Y_t given by

$$\begin{aligned}Y_t &:= \max_{0\le s\le t}\left\{\theta W_t - \theta W_s - \frac{\theta^2}{2}t + \frac{\theta^2}{2}s\right\} \\ &= \theta W_t - \frac{\theta^2}{2}t + \Lambda_t .\end{aligned}$$

Let $\overline{W}$ denote the innovation process defined by

$$\overline{W}_t = W_t - \theta\int_0^t \pi_s\, ds . \tag{1}$$

The innovation process $\overline{W}$ is adapted to the filtration $(\mathscr{F}_t^W;\ 0 \le t < \infty)$ and is a standard Brownian motion under $\bar{P}$ [see Kallianpur (1980), pages 192–199]. Equation (1) yields

$$Y_t = \theta^2\int_0^t \pi_s\, ds + \theta\overline{W}_t - \frac{\theta^2}{2}t + \Lambda_t.$$

By application of Itô's formula for continuous semimartingales [see, e.g., Karatzas and Shreve (1988), page 149] we arrive at

$$d\pi_t = \theta\pi_t(1-\pi_t)\, d\overline{W}_t + \pi_t(1-\pi_t)\, d\Lambda_t, \tag{2}$$

with $\pi_0 = p$. Now Λ is flat off $\{s|\ Y_s = 0\}$ [see, e.g., Karatzas and Shreve (1988), page 210], so

$$\int_0^\infty 1_{(0,\infty)}(Y_s)\, d\Lambda_s = 0$$

and further, for all measurable functions h with $h(0) = 0$,

$$\int_0^\infty h(Y_s)\, d\Lambda_s = 0.$$

Therefore, equation (2) yields, for all functions f which are twice continuous differentiable in an open set containing $[p, 1)$ with $f'(p) = 0$,

$$f(\pi_t) - f(p) = \theta \int_0^t f'(\pi_s)\pi_s(1-\pi_s)\, d\overline{W}_s + \frac{\theta^2}{2}\int_0^t f''(\pi_s)\pi_s^2(1-\pi_s)^2\, ds. \tag{3}$$

4. Bayes optimality. Throughout this and the next section the expression "stopping time" always refers to a stopping time with respect to the filtration $(\mathscr{F}_t^W;\ 0 \le t < \infty)$. Let us introduce the risk function

$$R(T) = \overline{P}(T < \nu) - C_1\overline{E}(T \wedge \nu) + C_2\overline{E}(T - \nu)^+,$$

where C_1 and C_2 are positive constants. Note that $\overline{E}\nu < +\infty$. Thus $R(T)$ is defined for all stopping times T. Further, we have $R(T) = +\infty$ if $\overline{P}(T = \infty) > 0$. Let us now define two functions f_1 and f_2 on $(0, 1)$ by

$$f_1(x) := \frac{2}{\theta^2}\int_p^x \int_p^y z^{-2}(1-z)^{-1}\, dz\, dy$$

and

$$f_2(x) := \frac{2}{\theta^2}\int_p^x \int_p^y z^{-1}(1-z)^{-2}\, dz\, dy.$$

Let g denote the function on $(0, 1)$ defined by

$$g(x) := (1-x) - C_1 f_1(x) + C_2 f_2(x).$$

For $A > 0$, let N_A denote the stopping time defined by

$$N_A := \inf\{t \mid Y_t \ge A\}.$$

This is the usual cusum procedure.

THEOREM 1. *Let* $A^* = \ln(p^*/(1-p^*)) - \ln(p/(1-p))$, *where* p^* *is the unique solution in* $(p, 1)$ *of* $g'(x) = 0$. *Then, for the cusum procedure* N_{A^*},

$$N_{A^*} = \inf\{t \mid Y_t \ge A^*\},$$
$$R(N_{A^*}) = \inf_T R(T),$$

where the infimum on the right-hand side is taken over all stopping times T.

The main idea of the proof of Theorem 1 is to reformulate our problem. We rewrite the risk $R(T)$ for bounded stopping times T as the expectation of $g(\pi_T)$, where g is a function with a unique minimum on $[p, 1)$. Lemmas 3 and 4 below establish this alternative formulation of $R(T)$. Lemma 5 states the necessary property of the function g just mentioned.

LEMMA 3. *For all stopping times* T,

$$R(T) = \bar{E}(1-\pi_T)1_{\{T<\infty\}} - C_1\bar{E}\int_0^T (1-\pi_t)\,dt + C_2\bar{E}\int_0^T \pi_t\,dt.$$

Lemma 3 can be proved in the same way as the corresponding statement in Shiryayev [(1973), page 161].

The functions f_1 and f_2 defined above are solutions of the differential equations

$$\frac{\theta^2}{2}x^2(1-x)^2 f_1''(x) = (1-x)$$

and

$$\frac{\theta^2}{2}x^2(1-x)^2 f_2''(x) = x.$$

The properties of f_1 and f_2 yield, together with (3),

$$\int_0^t (1-\pi_s)\,ds = f_1(\pi_t) - \theta\int_0^t f_1'(\pi_s)\pi_s(1-\pi_s)\,d\overline{W}_s$$

and

$$\int_0^t \pi_s\,ds = f_2(\pi_t) - \theta\int_0^t f_2'(\pi_s)\pi_s(1-\pi_s)\,d\overline{W}_s,$$

where both equalities hold $\bar{P}$-a.s.. The stochastic integrals on the right-hand side are martingales, because [see Karatzas and Shreve (1988), page 139]

$$\bar{E}\int_0^t f_i'(\pi_s)^2\pi_s^2(1-\pi_s)^2\,ds < \infty$$

for $i = 1, 2$ and all $t \geq 0$. The optional stopping theorem thus implies

$$\bar{E}\int_0^T f_i'(\pi_s)\pi_s(1-\pi_s)\,d\overline{W}_s = 0$$

for $i = 1, 2$ and for all bounded stopping times T. This yields the following lemma:

LEMMA 4. *For all bounded stopping times* T,

$$R(T) = \bar{E}\{(1-\pi_T) - C_1 f_1(\pi_T) + C_2 f_2(\pi_T)\} = \bar{E}g(\pi_T)\,.$$

LEMMA 5. *There exists a unique* $p^* \in [p, 1)$ *with* $g(p^*) \leq g(x)$ *for all* $x \in [p, 1)$.

PROOF. For $x \in (0, 1)$,

$$\frac{\theta^2}{2} g''(x) = \frac{1}{x^2(1-x)^2} \{x(C_1 + C_2) - C_1\}.$$

For $C_1 = 0$, the function g is strictly convex and from $\lim_{x \to 1} g(x) = +\infty$ and $g'(p) = -1$ the assertion follows.

Now let $C_1 > 0$. Then

$$g''(x) < 0 \quad \text{for } x \in \left(0, \frac{C_1}{C_1 + C_2}\right),$$
$$g''(x) > 0 \quad \text{for } x \in \left(\frac{C_1}{C_1 + C_2}, 1\right)$$

and $g''(C_1/(C_1+C_2)) = 0$. Rolle's theorem therefore yields that g' has at most two zeros in (0, 1). Now $g'(p) = -1$ and $\lim_{x \to 1} g'(x) = \lim_{x \to 0} g'(x) = +\infty$ imply that there exist $p_1 \in (0, p)$ and $p_2 \in (p, 1)$ with $g'(p_1) = g'(p_2) = 0$. Because g is continuous and $\lim_{x \to 1} g(x) = +\infty$, g assumes its minimum over $[p, 1)$ at least at one point. By $g'(p) < 0$, such points lie in the interior. Hence the function g assumes its minimum over $[p, 1)$ uniquely at $p^* = p_2$. □

We now proceed with the proof of Theorem 1. From Lemmas 4 and 5, we immediately infer that

$$R(T) \geq g(p^*)$$

holds for all bounded stopping times T. This inequality extends to arbitrary stopping times. Let

$$T^* := \inf\{t \mid \pi_t \geq p^*\}.$$

In order to complete the proof of Theorem 1, it is now only left to show that

$$R(T^*) = g(p^*).$$

We have $\bar{P}(T^* < \infty) = 1$. This follows immediately from the fact that π_t converges $\bar{P}$-a.s. to 1 for $t \to \infty$, which can be shown in the same way as the corresponding statement of Shiryayev [(1973), page 153]. Now

$$\lim_{n \to \infty} g(\pi_{T^* \wedge n}) = g(\pi_{T^*}) = g(p^*) \qquad \bar{P}\text{-a.s.}$$

The function g is bounded on $[p, p^*]$ and, therefore,

$$R(T^*) = \lim_{n \to \infty} R(T^* \wedge n) = g(p^*).$$

The stopping time T^* is obviously a cusum procedure N_A with threshold $A = A^*$ given by

$$A^* = \ln(p^*/(1-p^*)) - \ln(p/(1-p)).$$

The last identity follows immediately from Lemma 2. This completes the proof of Theorem 1.

5. Minimax optimality. For $0 \leq t < \infty$, let P_t denote the probability measure on $\mathscr{F}_\infty^W$ which corresponds to a change in the drift of W from 0 to θ at time t. From now on we will write P_∞ instead of μ_W. For simplicity of notation, the expectation with respect to P_t is denoted by E_t and the expectation with respect to P_∞ is denoted by E_∞. For arbitrary $\gamma \geq 0$, let $\mathscr{S}(\gamma)$ denote the class of all stopping times N with $E_\infty N \geq \gamma$. We shall now consider various values of p, C_1 and C_2 and thus like to indicate the dependence of A^* on p, C_1 and C_2 as well as the dependence of $\bar{P}$ on p. Therefore, we write $A^*(p, C_1, C_2)$ and $\bar{P}_p$.

LEMMA 6. *Given* $p \in (0,1)$ *and* $A > 0$, *there exist constants* $C_1(p, A) > 0$ *and* $C_2(p, A) > 0$ *such that*

(i) $$A^*(p, C_1(p, A), C_2(p, A)) = A,$$

(ii) $$1 - C_1(p, A)E_\infty N_A - C_2(p, A)E_0 N_A = 0.$$

PROOF. The assertion is equivalent to solving a system of two linear equations in C_1 and C_2. The assertion therefore follows from Cramer's rule. □

Choosing $C_1(p, A)$ and $C_2(p, A)$ according to Lemma 6, we can follow the arguments of Ritov and therefore obtain the minimax optimality of the cusum procedure N_A in the sense of Lorden (1971).

THEOREM 2. *Let* $A > 0$. *Then, for all* $N \in \mathscr{S}(E_\infty(N_A))$, *the following statement holds:*

$$\sup_{t \geq 0} \text{ess-sup } E_t\left((N - t)^+ | \mathscr{F}_t^W\right) \geq \sup_{t \geq 0} \text{ess-sup } E_t\left((N_A - t)^+ | \mathscr{F}_t^W\right) .$$

REMARK. Our method to construct the measures $\bar{P}_p$ works for arbitrary random variables τ, which are randomized stopping times with respect to the process W. By the use of Girsanov's theorem one can find a probability measure P_τ under which $W_t - \theta(t - \tau)^+$ is a standard Brownian motion. So under P_τ the observed process W is a Brownian motion process which has drift zero during the random time interval $[0, \tau)$ and drift θ during $[\tau, \infty)$. We can now consider the same decision theoretic problem as in Ritov (1990). Nature chooses τ and thus determines P_τ, whereas the statistician chooses any stopping time T with respect to W. The loss structure is taken as

$$P_\infty(T < \tau) - C_1 E_\infty T \wedge \tau + C_2 E_\tau (T - \tau)^+.$$

Equalities (i) and (ii) of Lemma 6 then imply that the pair of strategies $(N_A, \bar{P}_p)$ is a saddle point for this game if C_1 and C_2 are taken as $C_1(p, A)$ and $C_2(p, A)$. This can be shown in a similar way as the corresponding statement in Ritov (1990).

Acknowledgment. The author thanks Professor H. R. Lerche for posing the problem and for helpful discussions.

REFERENCES

BEIBEL, M. (1994). Bayes problems in change-point models for the Wiener process. In *Change-Point Problems*, 1–6. IMS, Hayward, CA.

BEIBEL, M. and LERCHE, H. R. (1995). A new look at optimal stopping problems related to mathematical finance. *Statist. Sinica.* To appear.

KALLIANPUR, G. (1980). *Stochastic Filtering Theory*. Springer, Berlin.

KARATZAS, I. and SHREVE, S. E. (1988). *Brownian Motion and Stochastic Calculus*. Springer, Berlin.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LERCHE, H. R. (1986). An optimality property of the repeated significance test. *Proc. Nat. Acad. Sci. U.S.A.* **83** 1546–1548.

LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.

MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387.

PAGE, E. L. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

RITOV, Y. (1990). Decision theoretic optimality of the cusum procedure. *Ann. Statist.* **18** 1464–1469.

SHIRYAYEV, A. N. (1973). Statistical sequential analysis. *Transl. Math. Monographs* **8**.

WOODROOFE, M., LERCHE, R. and KEENER, R. (1993). A generalized parking problem. In *Statistical Decision Theory and Related Topics V* (S. S. Gupta and J. O. Berger, eds.) 523–532. Springer, Berlin.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT FREIBURG
ECKERSTR. 1
79104 FREIBURG
GERMANY