stochastic processes and their applications

ELSEVIER Stochastic Processes and their Applications 80 (1999) 177–191

A martingale approach for detecting the drift of a Wiener process

Volkert Paulsen

Mathematisches Seminar, Universität Kiel, D-24098 Kiel, Germany

Received 13 January 1998; received in revised form 20 July 1998; accepted 22 July 1998

Abstract

Lerche (Ann. Statist. 14, 1986b, 1030–1048) considered a sequential Bayes-test problem for the drift of the Wiener process. In the case of a normal prior an o(c)-optimal test could be constructed. In this paper a new martingale approach is presented, which provides an expansion of the Bayes risk for a one-sided SPRT. Relations to the optimal Bayes risk are given, which show the o(c)-optimality for suitable nonnormal priors.

AMS classification: 60G35; 60G40; 60H30; 62L10

Keywords: Sequential probability ratio test; Bayes test; Optimal stopping; Boundary crossing; Stochastic integral; Density process; Wiener process

1. Introduction

We observe a standard Wiener process $(W_t)_{t\geqslant 0}$ with unknown drift $\theta\in\mathbb{R}$ and want to test sequentially

$$\mathrm{H}=\{0\}\quad \text{against } K=\mathbb{R}\backslash\{0\}.$$

Darling and Robbins (1967,1968) introduced the concept of tests of power one. Such a test is given by each stopping time T with the properties

$$P_0(T<\infty)<1, \tag{1.1}$$

$$P_\theta(T<\infty)=1 \quad \text{for all } \theta\in K. \tag{1.2}$$

The purpose of T is to control the drift in the sense that every drift will be detected, if it occurs. As long as we observe, we believe that the hypothesis H is true. Stopping means that due to the observations the hypothesis must be rejected. The power of such a test is determined by the error probability (1.1) and the mean time $E_\theta T$, $\theta\in K$, which is needed to detect the alternative drift θ. Weighting the power function with respect

0304-4149/99/$ – see front matter
PII: S0304-4149(98)00071-4

to a prior probability measure defines the Bayes risk $B(T)$ of T. We consider priors of the form

$$\pi = \gamma_0\delta_0 + (1-\gamma_0)F, \tag{1.3}$$

where $\gamma_0 \in (0,1)$ and F is a probability measure on the real line with $F(\{0\})=0$. Thus $B(T)$ is defined by

$$B(T) = \gamma_0 P_0(T<\infty) + (1-\gamma_0)c\int \theta^2 E_\theta T F(\mathrm{d}\theta). \tag{1.4}$$

Here c is a cost constant, which takes into account the cost of observation. Since the order of the singularity of $E_\theta T$ as function of θ is at least 2, the costs of observation must depend on the parameter to get a finite Bayes risk. The choice of θ^2 is quite natural due to the fact that it is proportional to the Kullback–Leibler information of P_θ with respect to P_0. The problem is to find a test, that minimizes the Bayes risk $B(T)$ among all tests of power one, and to calculate the optimal Bayes risk $B^*(c)=\inf_T B(T)$. As usual this cannot be done in the case of composite hypotheses. Thus, we may ask for an asymptotic analysis of the problem for c tending to zero. An expansion of the Bayes risk up to an o(c)-term should be given and a test T_c^* be determined such that

$$B(T_c^*) - B^*(c) = \mathrm{o}(c) \quad \text{for } c\to 0.$$

One reasonable procedure is to stop, when the posterior probability of the hypothesis falls under a small level. Such a test is given by

$$\begin{aligned} T_{b(c)} &= \inf\left\{t\geqslant 0 : \gamma(W_t,t)\leqslant \frac{2c}{1+2c}\right\} \\ &= \inf\{t\geqslant 0 : f(W_t,t)\geqslant b(c)\} \end{aligned} \tag{1.5}$$

with $b(c)=\gamma_0/(1-\gamma_0)2c$. Here $\gamma(x,t)$ denotes the posterior probability of the hypothesis given $W_t=x$ and can be expressed with

$$f(x,t) = \int \exp\left(\theta x - \frac{1}{2}\theta^2 t\right) F(\mathrm{d}\theta) \tag{1.6}$$

as

$$\gamma(x,t) = \frac{\gamma_0}{\gamma_0 + (1-\gamma_0)f(x,t)} \quad \text{for all } x\in\mathbb{R},\ t>0.$$

An important fact is that $T_{b(c)}$ is a one-sided SPRT, since $(f(W_t,t))_{t\geqslant 0}$ is the density process of $\bar{P}=\int P_\theta F(\mathrm{d}\theta)$ with respect to P_0. Thus properties of one-sided SPRTs can be applied to it. Another way to define T_b for $b>1$ is given by

$$T_b = \inf\{t\geqslant 0 : W_t \notin (\psi_b^-(t), \psi_b^+(t))\},$$

where the functions ψ_b^+, ψ_b^- are the unique solutions of the equation $f(x,t)=b$ in the positive, respectively, negative half-plane. Properties of these functions, especially the asymptotic behaviour at infinity are given in Lai (1976).

Using methods of nonlinear renewal theory, Robbins and Siegmund (1974), Hagwood and Woodroofe (1982), Alsmeyer and Irle (1986) gave expansions for the expectation $E_\theta T_b$ and variance $\mathrm{Var}_\theta T_b$, when the level b tends to infinity. Thus for fixed θ an expansion of the power can be given. Since $\theta^2 E_\theta T_b$ converges to infinity, when θ

tends to zero, compare Farell (1964), it is not obvious, whether the weighted expected sample size $\int \theta^2 E_\theta T_b F(\mathrm{d}\theta)$ has an analogous expansion. In the case of a normal prior F Lerche (1986a, b) obtained that expansion and showed the o(c)-optimality of $T_{b(c)}$, when according to the cost constant c the level $b(c)$ is chosen by $b(c)=\gamma_0/(1-\gamma_0)2c$. The normal case is easier to handle, since posterior distributions are normal again and the weighted likelihood function $f(x,t)$ can be computed.

In this paper a new martingale approach is introduced. The representation of the density process $(f(W_t,t))_{t\geqslant 0}$ as exponential martingale together with some facts on one-sided SPRTs provide the key formula

$$\int \theta^2 E_\theta T_b F(\mathrm{d}\theta) = 2\log b + \bar{E}\int_0^{T_b} v(W_s,s)\,\mathrm{d}s,$$

v denoting the variance of the posterior distribution, which allows an asymptotic expansion for b tending to infinity. This will de done in Sections 3 and 4. Finally, in Section 5 this can be used to prove the o(c)-optimality of $T_{b(c)}$.

2. Properties of the one-sided SPRT

We list some basic facts of one-sided SPRTs, which are used in the following sections. Let $(\mathscr{F}_t)_{t\geqslant 0}$ be a right continuous filtration and $\mathscr{F}_\infty=\sigma(\mathscr{F}_t\colon t\geqslant 0)$. We assume that there are orthogonal probability measures P_0 and P_1 on $\mathscr{F}_\infty$, which are locally equivalent. This means that $P_1|_{\mathscr{F}_t}$ is equivalent to $P_0|_{\mathscr{F}_t}$ for each $t\geqslant 0$. Then there exists a density process $(L_t)_{t\geqslant 0}$ such that L_t is the Radon–Nikodym derivative of $P_1|_{\mathscr{F}_t}$ w.r.t. $P_0|_{\mathscr{F}_t}$ for all $t\geqslant 0$. L is up to indistinguishable processes uniquely defined and has right continuous paths. We additionally assume that L has continuous paths. Let

$$M_t \overset{\text{def}}{=} \int_0^t L_s^{-1}\,\mathrm{d}L_s \quad \text{for all } t\geqslant 0. \tag{2.1}$$

M is a local P_0-martingale and L is the exponential martingale of M, i.e.,

$$L_t = \exp(M_t - \tfrac{1}{2}[M]_t), \tag{2.2}$$

see Jacod and Shiryaev (1987), Williams and Rogers (1987). The one-sided SPRT T_b is defined by

$$T_b = \inf\{t\geqslant 0 : L_t\geqslant b\} = \inf\{t\geqslant 0 : M_t - \tfrac{1}{2}[M]_t\geqslant \log b\}. \tag{2.3}$$

The theorem of Girsanov implies:

Proposition 2.1.

$$P_0(T_b<\infty) = \frac{1}{b} \quad \textit{and} \quad E_1[M]_{T_b} = 2\log b \tag{2.4}$$

for all $b\geqslant 1$.

Proof. Due to the continuity of L the one-sided SPRT T_b hits its boundary. The orthogonality yields $P_1(T_b<\infty)=1$. Thus

$$P_0(T_b<\infty) = E_1 L_{T_b}^{-1} = \frac{1}{b}. \tag{2.5}$$

The theorem of Girsanov, see Jacod and Shiryaev (1987), implies that $M - [M]$ is a P_1-local martingale with $[M]$ as quadratic variation process w.r.t. P_1. This provides a sequence $(S_n)_{n\in\mathbb{N}}$ of reducing stopping times such that $S_n \uparrow \infty$ and $(M-[M])^{S_n}$ is a uniformly integrable martingale. Hence the optional sampling theorem yields

$$E_1(M - [M])_{S_n \wedge T_b} = 0 \quad \text{for all } n \in \mathbb{N}. \tag{2.6}$$

Together with $\log b \geqslant (M - [M])_{S_n\wedge T_b} + \frac{1}{2}[M]_{S_n\wedge T_b}$ this provides $E_1[M]_{S_n\wedge T_b} \leqslant 2\log b$. Hence the monotone limit fulfills $E_1[M]_{T_b} \leqslant 2\log b$. This implies, see Weizăcker and Winkler (1990),

$$E_1(M - [M])_{T_b} = 0. \tag{2.7}$$

Hence

$$\tfrac{1}{2}E_1[M]_{T_b} = E_1 \log L_{T_b} = \log b. \qquad \square \tag{2.8}$$

One important poperty of the one-sided SPRT is that it minimizes the expected stopped quadratic variation of the loglikelihood among all tests with no higher error probability.

Proposition 2.2. *Let $b > 1$. Then*

$$E_1[M]_T \geqslant E_1[M]_{T_b} = 2\log b \tag{2.9}$$

for each stopping time T with $P_0(T < \infty) \leqslant 1/b$.

Proof. We may assume $E_1[M]_T < \infty$. Then $E_1 \log L_T = \frac{1}{2}E_1[M]_T$ and

$$E_1 \log L_T \geqslant -\log E_1 L_T^{-1} = -\log P_0(T < \infty) \geqslant \log b.$$

Proposition 2.2 also states that the Kullback Leibler information of $P_1|_{\mathscr{F}_T}$ w.r.t. $P_0|_{\mathscr{F}_T}$ is minimized by T_b among all tests T with $P_0(T<\infty) \leqslant 1/b$. $\square$

Besides these optimality results another useful relation between the stopped quadratic variation of T_b and $\log b$ can be given.

Proposition 2.3.

$$\frac{\log b}{[M]_{T_b}} \xrightarrow{b} \frac{1}{2} \quad P_1\text{-}a.s.$$

Proof. Since T_b hits its boundary,

$$\log b = (M_{T_b} - [M]_{T_b}) + \tfrac{1}{2}[M]_{T_b}.$$

As mentioned before, $M - [M]$ is w.r.t. P_1 a continuous local martingale with $[M]$ as its quadratic variation process. Due to the orthogonality of P_1 w.r.t. P_0 the quadratic variation $[M]_t$ tends to infinity P_1-a.s. Since $M-[M]$ is a time changed Wiener process,

$$\frac{(M-[M])_t}{[M]_t} \xrightarrow{t} 0, \quad P_1\text{-a.s.},$$

see Williams and Rogers (1987) theorem of Dubin, Schwarz. Thus the assertion follows.

3. The martingale approach

At first some notations are given. The observed process $(W_t)_{t\geqslant 0}$ is adapted to a right continuous filtration $(\mathscr{F}_t)_{t\geqslant 0}$. On $\mathscr{F}_\infty = \sigma(\mathscr{F}_t\colon t\geqslant 0)$ probability measures $(P_\theta)_{\theta\in\mathbb{R}}$ are defined such that $(W_t)_{t\geqslant 0}$ is a standard Wiener process with drift θ according to P_θ. Let F be the part of the prior on the alternative. This means that F is a probability measure on the real line with $F(\{0\})=0$. We assume throughout this paper that F has a finite second moment. The weighted-likelihood function f is defined by

$$f(x,t) = \int \exp\left(\theta x - \frac{1}{2}\theta^2 t\right) F(\mathrm{d}\theta) \quad \text{for all } x\in\mathbb{R},\ t\geqslant 0. \tag{3.1}$$

Obviously, f is on $\mathbb{R}\times(0,\infty)$ a real-valued function, which solves the heat equation

$$(\partial_t + \tfrac{1}{2}\partial_x^2) f = 0. \tag{3.2}$$

The mixture $\bar{P}$ of $(P_\theta)_{\theta\in\mathbb{R}}$ with respect to F is defined by

$$\bar{P}(A) = \int P_\theta(A) F(\mathrm{d}\theta) \quad \text{for all } A\in\mathscr{F}_\infty.$$

The posterior distribution $F_{x,t}$ w.r.t. F given $W_t = x$ is defined by

$$F_{x,t}(\mathrm{d}\theta) = \frac{\exp(\theta x - 1/2\theta^2 t)}{f(x,t)} F(\mathrm{d}\theta).$$

The mean, second moment and variance of the posterior distribution $F_{x,t}$ are denoted by $\mu(x,t), \rho(x,t), v(x,t)$. Since the second moment of F is finite, they are well defined. If h is a function of θ such that $\int |h(\theta)|^p F(\mathrm{d}\theta) < \infty$ for some $p\geqslant 1$, it is easy to check that the posterior expectation

$$M_h(W_t,t) = \int h(\theta) F_{W_t,t}(\mathrm{d}\theta) \quad \text{for all } t\geqslant 0$$

defines a L_p-bounded $\bar{P}$-martingale. In view of this we can ensure that $(\mu(W_t,t))_{t\geqslant 0}$ and $(\rho(W_t,t))_{t\geqslant 0}$ are $\bar{P}$-martingales.

The dependence of the density process on the posterior mean process can be given by

Lemma 3.1. *$(f(W_t,t))_{t\geqslant 0}$ is the density process of $\bar{P}$ with respect to P_0 and has the martingale representation*

$$f(W_t,t) = \exp(M_t - \tfrac{1}{2}[M]_t) \tag{3.3}$$

with

$$M_t = \int_0^t \mu(W_s,s)\,\mathrm{d}W_s, \quad t\geqslant 0. \tag{3.4}$$

$[M] = \int_0^\cdot \mu(W_s,s)^2\,\mathrm{d}s$ denotes the quadratic variation process of M.

Proof. $(\exp(\theta W_t - \frac{1}{2}\theta^2 t))_{t\geqslant 0}$ is the density process of P_θ with respect to P_0 for each $\theta \in \mathbb{R}$. Integrating over θ provides the first part of the assertion.

Since f solves the heat equation, Ito's formula yields

$$f(W_t,t) = 1 + \int_0^t \partial_x f(W_s,s)\,\mathrm{d}W_s = 1 + \int_0^t f(W_s,s)\frac{\partial_x f(W_s,s)}{f(W_s,s)}\,\mathrm{d}W_s$$

$$= 1 + \int_0^t f(W_s,s)\mu(W_s,s)\,\mathrm{d}W_s.$$

Hence $(f(W_t,t))_{t\geqslant 0}$ is a solution of the stochastic differential equation

$$\mathrm{d}L_t = L_t\,\mathrm{d}M_t, \quad L_0 = 1,$$

which implies representation (3.3), see Williams and Rogers (1987).

Let

$$\bar{W}_t = W_t - \int_0^t \mu(W_s,s)\,\mathrm{d}s \quad \text{for all } t\geqslant 0.$$

Then in view of Lemma 3.1 Girsanov's theorem implies that $\bar{W}$ is a standard Wiener-process with respect to $\bar{P}$. It is natural to search for a stochastic integral representation for a given $\bar{P}$-semimartingale. $\mu(W_t,t)$ and $t\mu(W_t,t) - W_t$ fulfill a representation which is very useful in the following section.

Lemma 3.2. *The $\bar{P}$-semimartingales $\mu(W_t,t)$ and $t\mu(W_t,t) - W_t$ admit the following representation as stochastic integral processes*:

$$\mu(W_t,t) = \mu(W_0,0) + \int_0^t v(W_s,s)\,\mathrm{d}\bar{W}_s, \tag{3.5}$$

$$t\mu(W_t,t) - W_t = \int_0^t (v(W_s,s)s - 1)\,\mathrm{d}\bar{W}_s. \tag{3.6}$$

Proof. f solves the heat equation and $\partial_x f(x,t)/f(x,t) = \mu(x,t)$. Hence

$$\partial_x \mu(x,t) = v(x,t),$$

$$\tfrac{1}{2}\partial_x^2 \mu(x,t) + \partial_t \mu(x,t) = -\mu(x,t)\partial_x \mu(x,t).$$

Thus Ito's formula implies

$$\mathrm{d}\mu(W_t,t) = \partial_x \mu(W_t,t)\,\mathrm{d}W_t + (\partial_t \mu(W_t,t) + \tfrac{1}{2}\partial_x^2 \mu(W_t,t))\,\mathrm{d}t$$

$$= v(W_t,t)\,\mathrm{d}W_t - v(W_t,t)\mu(W_t,t)\,\mathrm{d}t$$

$$= v(W_t,t)\,\mathrm{d}\bar{W}_t$$

and

$$\mathrm{d}(t\mu(W_t,t)) = t\,\mathrm{d}\mu(W_t,t) + \mu(W_t,t)\,\mathrm{d}t$$

$$= tv(W_t,t)\,\mathrm{d}\bar{W}_t + \mathrm{d}W_t - \mathrm{d}\bar{W}_t. \qquad \square$$

This provides the above stochastic integral representations.

For the expansion of the Bayes risk for one-sided SPRTs in the following section we need sufficient conditions to check, whether the local martingale $t\mu((W_t, t)) - W_t$ is in fact a L_p bounded martingale. This leads to

Proposition 3.3. *Let F have a density g with the properties:*
(A1) g is absolutely continuous with $\int((g'(\theta)/g(\theta))1_{\{g>0\}})^p F(\mathrm{d}\theta) < \infty$ for some $p \geqslant 1$,
(A2) $g(\theta) \leqslant C\exp(|\theta|^\alpha)$ for some $C > 0$ and $\alpha < 2$.
Then $(t\mu(W_t, t) - W_t)_{t\geqslant 0}$ is a L_p-bounded $\bar{P}$-martingale.

Proof. It suffices to show that

$$t\mu(W_t, t) - W_t = \int \frac{g'(\theta)}{g(\theta)} 1_{\{g>0\}} F_{(W_t,t)}(\mathrm{d}\theta), \tag{3.7}$$

since the right-hand side of Eq. (3.7) is the posterior expectation of the function $h(\theta) = (g'(\theta)/g(\theta))1_{\{g>0\}}$. The integrability condition in (A1) implies then the L_p-bounded martingale property. Eq. (3.7) follows with partial integration:

$$\begin{aligned} t\mu(W_t, t) - W_t &= \int (t\theta - W_t) F_{(W_t,t)}(\mathrm{d}\theta) \\ &= \frac{1}{f(W_t, t)} \int (t\theta - W_t) \exp\left(\theta W_t - \frac{1}{2}\theta^2 t\right) g(\theta)\,\mathrm{d}\theta \\ &= \frac{1}{f(W_t, t)} \int \exp\left(\theta W_t - \frac{1}{2}\theta^2 t\right) g'(\theta)\,\mathrm{d}\theta \\ &= \int \frac{g'(\theta)}{g(\theta)} 1_{\{g>0\}}(\theta) F_{(W_t,t)}(\mathrm{d}\theta). \end{aligned}$$

The boundary terms in the partial integration step vanish since due to (A2) $g(\theta)\exp(\theta x - \frac{1}{2}\theta^2 t) \to 0$ for all x, t. Hence Eq. (3.7) is valid and the proof is finished. □

Finally the relation of the Bayes risk to the posterior second moment process $(\rho(W_t, t))_{t\geqslant 0}$ is given. This can be exploited in the one-sided SPRT case.

Proposition 3.4. *For each $(\mathscr{F})_{t\geqslant 0}$ stopping time T it holds*

$$\int \theta^2 E_\theta T F(\mathrm{d}\theta) = \bar{E} \int_0^T \rho(W_s, s)\,\mathrm{d}s. \tag{3.8}$$

In particular, this provides for $T_b = \inf\{t \geqslant 0\colon f(W_t, t) \geqslant b\}$, $b > 1$,

$$B(T_b) = \gamma_0 \frac{1}{b} + (1 - \gamma_0)c\left(2\log b + \bar{E}\int_0^{T_b} v(W_s, s)\,\mathrm{d}s\right). \tag{3.9}$$

Proof. Let us first assume that T is a bounded stopping time. Recall that F has a finite second moment. Thus $(\rho(W_t, t))_{t\geqslant 0}$ is a $\bar{P}$-martingale as was pointed out at the beginning of this section. Hence,

$$\bar{E}\rho(W_T, T)1_{\{T>s\}} = \bar{E}\rho(W_s, s)1_{\{T>s\}} \quad \text{for all } s \geqslant 0.$$

This leads to

$$\int \theta^2 E_\theta T F(\mathrm{d}\theta) = \int\int \theta^2 T dP_\theta F(\mathrm{d}\theta) = \int T \int \theta^2 F_{(W_T,T)}(\mathrm{d}\theta)\,\mathrm{d}\bar{P}$$
$$= \bar{E} T \rho(W_T, T) = \bar{E}\int_0^\infty \rho(W_T, T) 1_{\{T>s\}}\,\mathrm{d}s$$
$$= \int_0^\infty \bar{E}\rho(W_s, s) 1_{\{T>s\}}\,\mathrm{d}s = \bar{E}\int_0^T \rho(W_s, s)\,\mathrm{d}s.$$

Thus formula (3.8) is valid for bounded stopping times. In the unbounded case we approximate T by $T\wedge t$ and use the monotone convergence on both sides of Eq. (3.8). To get Eq. (3.9), notice that

$$\bar{E}\int_0^{T_b} \rho(W_s, s)\,\mathrm{d}s = \bar{E}\int_0^{T_b} \mu(W_s, s)^2\,\mathrm{d}s + \bar{E}\int_0^{T_b} v(W_s, s)\,\mathrm{d}s$$

and

$$\bar{E}\int_0^{T_b} \mu(W_s, s)^2\,\mathrm{d}s = \bar{E}[M]_{T_b} = 2\log b, \quad P_0(T_b < \infty) = \frac{1}{b},$$

see Proposition 2.1.

4. Asymptotic expansion of $\int \theta^2 E_\theta T_b F(\mathrm{d}\theta)$

Starting point for the analysis of the integrated expected sample size of a one-sided SPRT is the fact that

$$\int \theta^2 E_\theta T_b F(\mathrm{d}\theta) = 2\log b + \bar{E}\int_0^{T_b} v(W_s, s)\,\mathrm{d}s.$$

The main idea is that we can replace the posterior variance $v(W_t, t)$ by $1/(t+r)$, the posterior variance with respect to the prior $F = \mathrm{N}(0, 1/r)$. Thus the problem is reduced to find an expansion for $\bar{E}\log((T_b + r)/r)$. But as we will see

$$\frac{T_b}{2\log b}\theta^2 \stackrel{b\to\infty}{\longrightarrow} 1, \quad P_\theta\text{-a.s.}$$

Hence we guess

$$\bar{E}\log\left(\frac{T_b + r}{r}\right) = \log(2\log b) - \int \log\theta^2 F(\mathrm{d}\theta) - \log r + \mathrm{o}(1).$$

In the following we become more precise and verify each step.

Lemma 4.1. *Assume that the $\bar{P}$-local martingale $((t+r)\mu(W_t, t) - W_t)_{t\geqslant 0}$ is in fact an L_2-bounded martingale. Then*

$$\bar{E}\int_0^\infty \left| v(W_s, s) - \frac{1}{s+r}\right|\mathrm{d}s < \infty \tag{4.1}$$

and

$$\bar{E}\int_0^{T_b} v(W_s, s)\,\mathrm{d}s = \bar{E}\log\left(\frac{T_b + r}{r}\right) + \chi(r) + \mathrm{o}(1) \tag{4.2}$$

for all $r > 0$ with $\chi(r) = \bar{E}\int_0^\infty v(W_s, s) - 1/(s+r)\,\mathrm{d}s$.

Proof. Let μ denote the first moment of F. It follows from Eq. (3.6) that the martingale $X_t = (t+r)\mu(W_t,t) - W_t - \mu r$ has the representation

$$X_t = \int_0^t (s+r)v(W_s,s) - 1\,\mathrm{d}\bar{W}_s.$$

Hence

$$\bar{E}\int_0^\infty ((s+r)v(W_s,s)-1)^2\,\mathrm{d}s = \bar{E}[X]_\infty < \infty.$$

Hölder's inequality provides

$$\int_0^\infty \left| v(W_s,s) - \frac{1}{s+r}\right| \mathrm{d}s \leqslant \left(\int_0^\infty \left(\frac{1}{s+r}\right)^2\right)^{1/2} \left(\int_0^\infty ((s+r)v(W_s,s)-1)^2\,\mathrm{d}s\right)^{1/2}.$$

This implies Eq. (4.1). The second assertion follows from the dominated convergence theorem, since

$$\bar{E}\int_0^{T_b} v(W_s,s)\,\mathrm{d}s - \bar{E}\log\left(\frac{T_b+r}{r}\right) = \bar{E}\int_0^{T_b} v(W_s,s) - \frac{1}{s+r}\,\mathrm{d}s.$$

Proposition 4.2. *Let Θ denote the limiting random variable of the uniformly integrable martingale $(\mu(W_t,t))_{t\geqslant 0}$.*

Then

$$\frac{T_b}{2\log b} \overset{b\to\infty}{\longrightarrow} \frac{1}{\Theta^2}, \quad \bar{P}\text{-a.s.}$$

and $\Theta = \theta$ P_θ-a.s. for F-almost every θ.

Proof. It holds

$$\frac{1}{T_b}\int_0^{T_b} \mu(W_s,s)^2\,\mathrm{d}s \overset{b\to\infty}{\longrightarrow} \Theta^2, \quad \bar{P}\text{-a.s.},$$

$$\frac{[M]_{T_b}}{2\log b} \overset{b\to\infty}{\longrightarrow} 1, \quad \bar{P}\text{-a.s.},$$

see Proposition 2.3. Hence the first part of the assertion follows. To prove the additional remark, recall that $W_t - \int_0^t \mu(W_s,s)\,\mathrm{d}s$ is a standard Wiener process w.r.t. $\bar{P}$. Hence

$$\frac{W_t}{t} \overset{t\to\infty}{\longrightarrow} \Theta, \quad \bar{P}\text{-a.s.},$$

due to $\lim_{t\to\infty} 1/t\int_0^t \mu(W_s,s) = \Theta$. But W_t/t tends to θ P_θ-a.s. for all $\theta\in\mathbb{R}$. Thus $\Theta=\theta$ P_θ-a.s. for F-almost every θ.

In view of the limiting random variable Θ it holds

$$\int \theta^2 E_\theta T_b F(\mathrm{d}\theta) = \bar{E}\Theta^2 T_b.$$

From Eq. (3.9) or Proposition 2.2 we obtain the trivial lower bound

$$\int \theta^2 E_\theta T_b F(\mathrm{d}\theta) \geqslant 2\log b.$$

A first upper bound is given in the following:

Lemma 4.3. *Let* $((t+r)\mu(W_t,t)-W_t)_{t\geqslant 0}$ *be an* L_2*-bounded martingale and* $\int|\log(\theta^2)|F(\mathrm{d}\theta)<\infty$. *Then*

$$\int \theta^2 E_\theta T_b F(\mathrm{d}\theta)\leqslant 4\log b+\mathrm{o}(1).$$

Proof. From Lemma 4.1 and Proposition 4.2 it follows:

$$\begin{aligned}\bar{E}\Theta^2(T_b+r)&=r\bar{E}\Theta^2+\bar{E}\Theta^2T_b\\&=r\bar{E}\Theta^2+2\log b+\bar{E}\int_0^{T_b} v(W_s,s)\,\mathrm{d}s\\&=r\bar{E}\Theta^2+2\log b+\bar{E}\log\left(\frac{(T_b+r)}{r}\right)+\bar{E}\int_0^{T_b} v(W_s,s)-\frac{1}{s+r}\,\mathrm{d}s\\&=2\log b+\bar{E}\log(\Theta^2(T_b+r))+\mathrm{O}(1).\end{aligned}$$

Using Jensen's inequality and $\log x\leqslant\frac{1}{2}x$ for all $x\geqslant 2$, we obtain f.a. $b\geqslant e$,

$$\begin{aligned}\bar{E}\Theta^2(T_b+r)&\leqslant 2\log b+\log\bar{E}(\Theta^2(T_b+r))+\mathrm{O}(1)\\&\leqslant 2\log b+\tfrac{1}{2}\bar{E}\Theta^2(T_b+r)+\mathrm{O}(1),\end{aligned}$$

since $\bar{E}\Theta^2(T_b+r)\geqslant 2\log b\geqslant 2$. Hence the assertion follows. □

To get the desired expansion for $\bar{E}\log((T_b+r)/r)$, notice that

$$P_\theta\left(T_b\leqslant\frac{(2-\varepsilon)\log b}{\theta^2}\right)=\mathrm{o}((\log b)^{-1})\tag{4.3}$$

uniformly in $\theta\in\mathbb{R}\setminus\{0\}$ for all $0<\varepsilon<2$. This can be shown by nearly the same arguments as in Woodroofe (1982), (p. 69).

Lemma 4.4. *If the assumptions of Lemma* 4.3 *are valid, then*

$$\bar{E}\log\left(\frac{T_b+r}{r}\right)=\log(2\log b)-\int\log\theta^2F(\mathrm{d}\theta)-\log r+\mathrm{o}(1).\tag{4.4}$$

Proof. Due to Proposition 4.2 it holds

$$\log\left(\frac{(T_b+r)\Theta^2}{2\log b}\right)\overset{b\to\infty}{\longrightarrow}0,\quad \bar{P}\text{-a.s.}$$

Let $\xi_b=((T_b+r)\Theta^2)/2\log b$ and $A_b=\{T_b\geqslant(\log b)/\Theta^2\}$. $|\log(\xi_b)|$ is bounded on $A_b\cap\{\xi_b\leqslant M\}$ by $\log M$ for each $M>2$. Hence we may conclude by dominated convergence

$$\bar{E}\log\xi_b 1_{\{\xi_b\leqslant M\}\cap A_b}\overset{b\to\infty}{\longrightarrow}0\quad\text{for all } M>2.\tag{4.5}$$

On $A_b^c\cap\{\xi_b\leqslant M\}$ it holds

$$\frac{r\Theta^2}{2\log b}\leqslant\xi_b\leqslant M.$$

Furthermore,

$$\bar{E}\log\left(\frac{r\Theta^2}{2\log b}\right)1_{A_b^c\cap\{\xi_b\leqslant M\}} = \bar{E}\log r\Theta^2 1_{A_b^c\cap\{\xi_b\leqslant M\}} - \log(2\log b)\bar{P}(A_b^c\cap\{\xi_b\leqslant M\}).$$

Both terms on the right tend to zero due to Eq. (4.3). Thus it follows

$$\lim_{b\to\infty}\bar{E}1_{A_b^c\cap\{\xi_b\leqslant M\}}\log\xi_b = 0. \tag{4.6}$$

The last step is

$$\bar{E}1_{\{\xi_b\geqslant M\}}\log\xi_b \leqslant \sup_{x\geqslant M}\frac{\log x}{x}\bar{E}\xi_b.$$

The first factor tends to zero for M to infinity, the second remains bounded in b due to Lemma 4.3. This together with Eqs. (4.5) and (4.6) yields

$$\lim_{b\to\infty}\bar{E}\log\xi_b = 0$$

from which the desired expansion follows.

The preceding considerations run into the following:

Theorem 4.5. *Under the assumptions of Lemma* 4.3 *it holds*

$$\int\theta^2 E_\theta T_b F(\mathrm{d}\theta) = 2\log b + \log(2\log b) - \int\log\theta^2 F(\mathrm{d}\theta) - \log r + \chi(r) + \mathrm{o}(1) \tag{4.7}$$

and

$$B(T_{b(c)}) = \frac{\gamma_0}{b(c)} + (1-\gamma_0)c(2\log b(c) + \log(2\log b(c)) - A - \log r + \chi(r) + \mathrm{o}(1))$$

with $\chi(r) = \bar{E}\int_0^\infty v(W_s,s) - 1/(s+r)\mathrm{d}s,\ b(c) = \gamma_0/(1-\gamma_0)2c, A = \int\log(\theta^2)F(\mathrm{d}\theta)$.

Proof. Eq. (4.7) follows immediatly from Eqs. (3.9), (4.2) and (4.4). Due to $P_0(T_b<\infty) = 1/b$ the expansion for $B(T_{b(c)})$ holds.

5. o(c)-optimality of the one-sided SPRT $T_{b(c)}$

In this section we show that the expansion for the Bayes risk of the one-sided SPRT $T_{b(c)}$ in Theorem 4.5 is an expansion for the optimal Bayes risk $B^*(c)$ too. Throughout this section let F be a probability measure on the real line with finite second moment and Lebesgue-density g such that

(A1) g is absolutely continuous with $\int((g'(\theta)/g(\theta))1_{\{g>0\}})^2\,F(\mathrm{d}\theta)<\infty$

(A2) $g(\theta)\leqslant C\exp(|\theta|^\alpha)$ for some $C>0$ and $\alpha<2$.

Note that then the assumptions of Theorem 4.5 are fulfilled. Furthermore, let T_c^* be a o(c)-optimal Bayes test. This means that

$$B(T_c^*) - B^*(c) = \mathrm{o}(c).$$

At first the Bayes test problem is expressed as an optimal stopping problem.

Lemma 5.1. *Let* $h(x) = \gamma_0 \exp(-x) + (1-\gamma_0)2cx$ *for all* $x \in \mathbb{R}$ *and* $D_t = \log f(W_t t)$. *Then*

$$B(T) = \bar{E}\left(h(D_T) + (1-\gamma_0)c\int_0^T v(W_s, s)\,\mathrm{d}s\right)$$
$$\geqslant h(\log(b(c)) + (1-\gamma_0)c\bar{E}\int_0^T v(W_s, s)\,\mathrm{d}s \tag{5.1}$$

holds for each stopping time T *with finite Bayes risk.*

Proof. Since T has finite Bayes risk it is $\bar{P}$-a.s. finite. Then due to the assumptions on F Eq. (3.8) provides

$$B(T) = \gamma_0 P_0(T < \infty) + (1-\gamma_0)c\left(\bar{E}[M]_T + \bar{E}\int_0^T v(W_s, s)\,\mathrm{d}s\right).$$

$\bar{E}D_T = \frac{1}{2}\bar{E}[M]_T$, since $\bar{E}[M]_T < \infty$ and $P_0(T < \infty) = \bar{E}\exp(-D_T)$. Hence the first equation of Eq. (5.1) follows. The inequality holds due to the fact that h is a convex function with minimum at $\log b(c)$.

Lemma 5.1 provides a lower bound for the optimal Bayes risk with the leading term $h(b(c))$ from the expansion of $B(T_{b(c)})$. The remainder term $\bar{E}\int_0^T v(W_s, s)\,\mathrm{d}s$ can be expanded by the methods of Section 4. Eq. (4.2) yields

$$\bar{E}\int_0^{T_c^*} v(W_s, s)\,\mathrm{d}s = \bar{E}\left(\log\left(\frac{T_c^* + r}{r}\right)\right) + \chi(r) + \mathrm{o}(1)$$

for all $r > 0$. To get an expansion for the first term on the right, we have to analyse the error probability $\alpha(c) = P_0(T_c^* < \infty)$ of the o(c)-optimal test T_c^*. Eq. (5.1) together with the expansion of $B(T_{b(c)})$ can be used to get an upper bound for $\alpha(c)$. Especially it tends to zero fast enough.

Lemma 5.2. *The error probability* $\alpha(c)$ *of* T_c^* *is a function less than* $\mathrm{O}(\log b(c)/b(c))$.

Proof. We compare T_c^* to $T_{\alpha(c)^{-1}}$, the one-sided SPRT with the same error probability. $T_{\alpha(c)^{-1}}$ minimizes the expected quadratic variation $\bar{E}[M]_T$ among all tests T with $P_0(T < \infty) \leqslant \alpha$, see Proposition 2.2. Hence

$$\bar{E}[M]_{T_c^*} \geqslant \bar{E}[M]_{T_{\alpha(c)^{-1}}} = 2\log\alpha(c)^{-1}$$

and in view of Eq. (5.1)

$$B(T_c^*) \geqslant \gamma_0\alpha(c) + (1-\gamma_0)c\left(2\log\alpha(c)^{-1} + \bar{E}\int_0^{T_c^*} v(W_s, s)\,\mathrm{d}s\right)$$
$$\geqslant h(\log\alpha(c)^{-1}).$$

Together with the expansion of $B(T_{b(c)})$ and $b(c)^{-1} = O(c)$ the estimation

$$h(\log\alpha(c)^{-1}) \leqslant B(T_c^*) \leqslant B(T_{b(c)}) + \mathrm{o}(c) \leqslant h(\log b(c)) + \mathrm{O}\left(c\log\log\frac{1}{c}\right)$$

follows. The convex function h has a unique minimum at $\log b(c)$. Hence

$$0 \leqslant \gamma_0 \left(\alpha(c) - \frac{1}{b(c)} \right) - (1-\gamma_0) 2c \log(\alpha(c) b(c)) \leqslant \mathrm{O}\left(c \log\log \frac{1}{c} \right).$$

Thus

$$\alpha(c) \leqslant \frac{1}{b(c)} + \frac{1}{b(c)} \log(\alpha(c) b(c)) + \mathrm{O}\left(c \log\log \frac{1}{c} \right) \leqslant \mathrm{O}\left(\frac{\log b(c)}{b(c)} \right).$$

In particular, this lemma yields $b(c)^{\lambda}\alpha(c) \to 0$ for $0 < \lambda < 1$. Thus nearly the same arguments as in Woodroofe (1982) p. 69 provide

$$P_\theta \left(T_c^* \leqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) = \mathrm{o}((\log b(c))^{-1}) \tag{5.2}$$

for all $0 < \varepsilon < 2$ uniformly in $\theta \in \mathbb{R}\backslash\{0\}$. This is the key to:

Lemma 5.3. *For each $r > 0$ and $0 < \varepsilon < 2$*

$$\bar{E} \log\left(\frac{T_c^* + r}{r} \right) \geqslant \log((2-\varepsilon)\log b(c)) - \int \log \theta^2 F(\mathrm{d}\theta) - \log r + \mathrm{o}(1).$$

Proof. Let

$$R(c) = \int \log \left(\frac{(2-\varepsilon)\log b(c)}{\theta^2 r} \right) P_\theta \left(T_c^* \leqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) F(\mathrm{d}\theta)$$

Eq. (5.2) yields

$$\begin{aligned} |R(c)| &\leqslant \int \left| \log\left(\frac{(2-\varepsilon)\log b(c)}{\theta^2 r} \right) \right| P_\theta \left(T_c^* \leqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) F(\mathrm{d}\theta) \\ &\leqslant \log((2-\varepsilon)\log b(c)) \int P_\theta \left(T_c^* \leqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) F(\mathrm{d}\theta) \\ &\quad + \int \left| \log \frac{1}{\theta^2 r} \right| P_\theta \left(T_c^* \leqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) F(\mathrm{d}\theta) \\ &= \mathrm{o}(1). \end{aligned}$$

Hence for all $0 < \varepsilon < 2$ it holds

$$\begin{aligned} \bar{E} \log\left(\frac{T_c^* + r}{r} \right) &= \int E_\theta \log\left(\frac{T_c^* + r}{r} \right) F(\mathrm{d}\theta) \\ &\geqslant \int E_\theta \log\left(\frac{T_c^* + r}{r} \right) 1_{\left\{ T_c^* \geqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right\}} F(\mathrm{d}\theta) \\ &\geqslant \int \log\left(\frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) P_\theta \left(T_c^* \geqslant \frac{(2-\varepsilon)\log b(c)}{\theta^2} \right) F(\mathrm{d}\theta) \\ &= \log((2-\varepsilon)\log b(c)) - \int \log \theta^2 F(\mathrm{d}\theta) - \log r - R(c). \end{aligned}$$

Thus the claim holds.

We can use Lemma 5.3 to give a lower bound for the Bayes risk of T_c^*. This differs from the Bayes risk of $T_{b(c)}$ only up to a $\mathrm{o}(c)$ term. Hence $T_{b(c)}$ is a $\mathrm{o}(c)$-optimal test. This is summarized in the following.

Theorem 5.4. *Let F have a density g with properties* (A1) *and* (A2). *Then for all $r > 0$ and $0 < \varepsilon < 2$*

$$B(T_c^*) \geqslant \frac{\gamma_0}{b(c)} + (1-\gamma_0)c(2\log b(c) + \log((2-\varepsilon)\log b(c)) \\ -A - \log r + \chi(r) + \mathrm{o}(1))$$

with $A = \int \log\theta^2 F(\mathrm{d}\theta)$, $\chi(r) = \bar{E}\int_0^\infty v(W_s,s) - 1/(s+r)\mathrm{d}s, b(c) = \gamma_0/(1-\gamma_0)2c$. $T_{b(c)}$ is an $\mathrm{o}(c)$*-optimal test and the expansion*

$$B^*(c) = \frac{\gamma_0}{b(c)} + (1-\gamma_0)c(2\log b(c) + \log(2\log b(c)) - A - \log r + \chi(r) + \mathrm{o}(1)).$$

holds for the Bayes risk.

Proof. Theorem 4.5 and Lemma 5.3 provide for each $\varepsilon > 0$ a function $R_\varepsilon(c)$ with $R_\varepsilon(c) = \mathrm{o}(1)$ such that

$$B(T_{b(c)}) - B(T_c^*) \leqslant (1-\gamma_0)2c\left(\log\left(\frac{2}{2-\varepsilon}\right) + R_\varepsilon(c)\right).$$

Let $\delta > 0$. We can choose a ε_0 and a c_0 such that

$$2(1-\gamma_0)\log\left(\frac{2}{2-\varepsilon_0}\right) < \frac{\delta}{2},\ 2(1-\gamma_0)R_{\varepsilon_0}(c) < \frac{\delta}{2} \quad \text{for all } c \leqslant c_0.$$

Hence

$$\frac{B(T_{b(c)}) - B(T_c^*)}{c} \leqslant \delta \quad \text{for all } c < c_0$$

and the claim follows. □

6. Examples of priors

In this section some examples for the choice of a prior probability measure F on the alternative are given. One interesting class was introduced by Diaconis and Ylvisaker (1979). Let μ be a measure on the real line with the Borel σ-field and $\Lambda = \{(\lambda_1,\lambda_2) \in \mathbb{R}^2 : \int \exp(\theta\lambda_1 - \frac{1}{2}\lambda_2\theta^2)\mu(\mathrm{d}\theta) < \infty\}$. For $\lambda \in \Lambda$ a probability measure F^λ is defined by the μ-density

$$g^\lambda(\theta) = \exp(\lambda_1\theta - \tfrac{1}{2}\lambda_2\theta^2 - C(\lambda_1,\lambda_2))$$

with

$$C(\lambda_1,\lambda_2) = \log\left(\int \exp(\lambda_1\theta - \tfrac{1}{2}\lambda_2\theta^2)\mu(\mathrm{d}\theta)\right).$$

$(F^\lambda)_{\lambda\in\Lambda}$ defines a two parameter exponential family that is closed under posterior distribution. Starting with a prior F^λ, the posterior distribution given $W_t = x$ is $F^{(\lambda_1+x,\lambda_2+t)}$ for all $x \in \mathbb{R},\ t > 0$. The moments can be calculated by the partial derivatives of the function C. Hence the posterior mean- and second moment process are given by

$$\mu(W_t,t) = \partial_1 C(\lambda_1 + W_t, \lambda_2 + t), \quad \rho(W_t,t) = -2\partial_2 C(\lambda_1 + W_t, \lambda_2 + t).$$

If μ is the Lebesgue measure, $F^\lambda = N(\lambda_1/\lambda_2, 1/\lambda_2)$. Thus the normal case is included.

We can verify the assumptions on the prior in Theorem 5.4 if the measure μ has a Lebesgue density h with the following properties:
(i) h is absolutely continuous with $\int (h'(\theta)/h(\theta))^2 \exp(\theta\lambda_1 - \frac{1}{2}\lambda_2\theta^2)h(\theta)\,\mathrm{d}\theta < \infty$,
(ii) $h(0) > 0$.
The following choices of h are possible:
(a) $h \equiv 1$ leads to a normal prior,
(b) $h(\theta) = (1 + |\theta|)^p$, $p > 0$,
(c) $h(\theta) = 1_{[0,\infty)}(\theta) + (\theta + \varepsilon)^p \varepsilon^{-p} 1_{(-\varepsilon,0)}(\theta) \quad \varepsilon > 0, p > 1$.

Condition (i) is not valid for $p = 1$. For $\lambda_1 < 0$, $\lambda_2 = 0$ and small ε this is nearly an exponential distribution. The second term has to smooth the jump at zero.

References

Alsmeyer, G., Irle, A., 1986. Asymptotic expansions for the variance of stopping times in nonlinear renewal theory. Stochastic Proc. Appl. 23, 235–258.

Darling, D., Robbins, H., 1967. Iterated logarithm inequalities. Proc. Nat. Acad. Sci. USA 57, 1188–1192.

Darling, D., Robbins, H., 1968. Some further remarks on inequalities for partial sums. Proc. Nat. Sci. USA 60, 1175–1182.

Diaconis, P., Ylvisaker, D., 1979. Conjugate priors for exponential families. Ann. Statist. 7, 269–281.

Farell, R., 1964. Asymptotic behaviour of expected sample size in certain one-sided tests. Ann. Math. Statist. 35, 36–72.

Hagwood, C., Woodroofe, M., 1982. On the expansion for expected sample size in nonlinear renewal theory. Ann. Probab. 10, 844–848.

Jacod, J., Shiryaev, A.N., 1987. Limit Theorems for Stochastic Processes. Springer, Berlin.

Lai, T.L., 1976. Boundary crossing probabilities for sample sums and confidence sequences. Ann. Probab. 4, 299–312.

Lerche, H.R., 1986a. Boundary Crossing of Brownian Motion. Lecture Notes in Statistics, vol. 40. Springer, Berlin.

Lerche, H.R., 1986b. The shape of Bayes tests of power one. Ann. Statist 14, 1030–1048.

Robbins, H., Siegmund, D., 1974. The expected sample size of some tests of power one. Ann. Statist 2, 415–436.

Weizäcker, H.v., Winkler, F., 1990. Stochastic Integrals. Vieweg Braunschweig.

Williams, D., Rogers, C., 1987. Diffusions, Markov Processes and Martingales, vol II. Wiley, New York.

Woodroofe, M., 1982. Nonlinear Renewal Theory in Sequential Analysis, CBMS-NSF Regional Conf. Series in Applied Mathematics, vol. 39. SIAM, Philadelphia.