*Proc. Natl. Acad. Sci. USA*
Vol. 83, pp. 1546–1548, March 1986
Statistics

# An optimal property of the repeated significance test

**(sequential Bayes tests/simple Bayes rules)**

HANS RUDOLF LERCHE

Institute for Applied Mathematics, University of Heidelberg, Im Neuenheimer Feld 294, 6900 Heidelberg, Federal Republic of Germany

*Communicated by Herbert Robbins, October 11, 1985*

**ABSTRACT** **It is shown that the repeated significance test is a Bayes test for testing sequentially the sign of the drift of a Brownian motion. Its relation to Wald's sequential probability ratio test is studied.**

## The repeated significance test (RST)

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables, and let $S_n = X_1 + \ldots + X_n$. Assume that the mean value $\theta = E X_1$ is unknown and that the variance $\mathrm{Var}\, X_1 = \sigma^2 < \infty$ is known. In order to make inferences about the size and the sign of the unknown value $\theta$, it is common to use the critical region $|S_n|/(n)^{1/2} \geqq a$ to reject the null hypothesis that $\theta = 0$, when the sample size $n$ is fixed in advance. To accelerate the detection of a certain effect, it seems plausible to apply this significance test repeatedly, which is just the RST. More precisely, the RST stops at the first $n$ for which $|S_n|/(n)^{1/2} \geqq a$ and decides that $\mathrm{sgn}\ \theta = \mathrm{sgn}\ S_n$.

For the following two testing problems, it seems natural to apply the RST. The first is testing $H_0{:}\theta = 0$ versus $H_1{:}\theta \neq 0$; the other is testing $H_0{:}\theta < 0$ versus $H_1{:}\theta > 0$. The second one will be discussed below. For the first problem, Robbins (1) observed that the RST stops almost surely even when $\theta = 0$, by the law of the iterated logarithm.

Procedures without this feature but with operating characteristics similar to the RST are the tests of power one. For more information on this topic, see refs. 2 and 3.

The operating characteristics of the RST were studied by McPherson and Armitage (4) using a Monte Carlo method, and by Siegmund (5) and others, theoretically; they derived refined large deviation results.

The RST is a natural procedure for medical statistics. One can use it, for instance, as a breakoff rule for follow-up studies of survival data. For further discussions of this, see ref. 6.

Here we shall study the optimality properties of the RST, about which nothing is known so far. I shall discuss this topic with the following motivating question in mind: is there a natural counterpart to Wald's sequential probability ratio test (SPRT) for testing composite hypotheses without an indifference zone that has optimality properties similar to the SPRT? To explain the background of this question, I give a short review of the relevant optimality results of the theory.

## Optimality in sequential testing

Let $\{P_\theta;\ \theta \in \mathbb{R}\}$ denote a set of probability measures. Let $\Theta_0 \subset (-\infty, 0)$ and $\Theta_1 \subset (0, \infty)$. Let $X_1, X_2, \ldots$ be independent, identically distributed observations from an unknown $P_\theta$, $\theta \in \Theta_0 \cup \Theta_1$. We consider the hypothesis testing problem $H_0{:}\theta \in \Theta_0$ versus $H_1{:}\theta \in \Theta_1$, with 0–1 loss and cost $c > 0$ per observation. Let $G(d\theta)$ denote a prior on $\Theta_0 \cup \Theta_1$. The Bayes risk for a decision procedure $(T, \delta)$, consisting of a stopping time, $T$, of $X_1, X_2, \ldots$ and a terminal decision rule, $\delta$, is given by

$$R(T, \delta) = \int_{\Theta_0} [P_\theta\{H_0 \text{ rejected } (\delta)\} + cE_\theta T]G(d\theta) + \int_{\Theta_1} [P_\theta\{H_1 \text{ rejected } (\delta)\} + cE_\theta T]G(d\theta). \quad [1]$$

The objective is to find a decision procedure $(T^*, \delta^*)$ with minimal Bayes risk.

This optimality problem reduces to an optimal stopping problem by the following consideration. We assume that $S_n$ is a sufficient statistic for the first $n$ observations. Let $G_{x,n}$ denote the posterior distribution of $\theta$ with respect to $G$, given that $S_n = x$. Let $T$ be an arbitrary stopping time. Let $\delta^*$ denote the terminal decision rule (after stopping at $T$) which rejects the hypothesis $H_0$ if and only if $G_{S_T,T}(\Theta_0) < G_{S_T,T}(\Theta_1)$. It is well known that $\rho(T, \delta^*) \leqq \rho(T, \delta)$. Since

$$\rho(T, \delta^*) = \int h(S_T, T)d\overline{Q},$$

with $h(x, n) = G_{x,n}(\Theta_0) \wedge G_{x,n}(\Theta_1) + cn$ and $\overline{Q} = \int P_\theta G(d\theta)$, one has only to find the optimal stopping rule.

Two types of optimality results are known: (*i*) For the case that there is an indifference zone in the parameter space (i.e., a positive distance between $\Theta_0$ and $\Theta_1$), it is known that certain simple Bayes rules are optimal or almost optimal for the Bayes risk (Eq. **1**) (7–10). The simple Bayes rules stop sampling when the posterior probability of $\Theta_0$ or $\Theta_1$ is too small. The SPRT is exactly optimal. (*ii*) For the case that there is no indifference zone in the parameter space, the optimal stopping rules are not simple Bayes rules (11–13).

In the following section, I show that if we let the cost $c$ depend on the parameter $\theta$ in a natural way, then a simple Bayes rule is optimal for the testing problem without an indifference zone. For related results about tests of power one see ref. 14.

## The RST and optimality

For simplicity, consider the continuous time problem of testing the sign of the drift $\theta$ of Brownian motion $W(t)$. The parameter sets of $H_0$ and $H_1$ are given by $\Theta_0 = \{\theta < 0\}$ and $\Theta_1 = \{\theta > 0\}$. The observation cost is taken to be $c\theta^2$, where $c$ is a positive constant. On the parameter space $\Theta_0 \cup \Theta_1$ we put the normal prior $G(d\theta) = \phi[r^{1/2}(\theta - \mu)]\, r^{1/2}d\theta$ with $\phi(y) = (2\pi)^{-1/2}e^{-y^2/2}$. The Bayes risk for a decision procedure $(T, \delta)$ is given by

$$R(T, \delta) = \int_{-\infty}^{0} [P_\theta\{H_0 \text{ rejected } (\delta)\} + c\theta^2 E_\theta T]G(d\theta)$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RST, repeated significance test; SPRT, sequential probability ratio test.

$$+ \int_0^\infty [P_\theta\{H_1 \text{ rejected } (\delta)\} + c\theta^2 E_\theta T]G(d\theta). \quad [2]$$

The objective is to find a decision procedure $(T^*, \delta^*)$ that minimizes $R$.

The assumption about the observation cost is somewhat unusual, but its meaning becomes apparent from the following consideration. Let us consider the two testing problems (*i*) $H_0{:}\theta = 0$ versus $H_1{:}\theta = \theta_1$ and (*ii*) $H_0{:}\theta = 0$ versus $H_1{:}\theta = \theta_2$, with $\theta_i > 0$ $(i = 1, 2)$. Let $t_i$ $(i = 1, 2)$ denote the sample sizes. Then the level-$\alpha$ Neyman–Pearson tests for both problems have the same power if and only if $\theta_1^2 t_1 = \theta_2^2 t_2$. [This follows from the form of the power function of a Neyman–Pearson test of level $\alpha$: $\Phi(-c_\alpha + \theta t^{1/2})$.] Thus, the factor $\theta^2$ standardizes the sample sizes in such a way that the testing problems are of equal difficulty.

There exists also some nonmathematical motivation arising from medical statistics for letting the cost of an observation depend on the unknown parameter. In a medical trial, the "cost" of an observation is more an ethical than an economic quantity, a measure of the regret for giving a subject an inferior treatment. There it seems quite reasonable to formulate the cost as a function of the parameter, since the regret for giving a subject a slightly inferior treatment will be less than that for giving a markedly inferior one. For references to literature relating to this, see ref. 3.

Let $G_{x,t}$ denote the posterior distribution of $\theta$, given that the process $[W(s), s]$ has reached $(x, t)$; $G_{x,t} = N(x + r\mu/t + r, 1/t + r)$, where $N(\rho, \sigma^2)$ denotes the normal distribution with mean $\rho$ and variance $\sigma^2$. For $\lambda > 0$, the simple Bayes rule is defined as

$$T_\lambda = \inf\{t > 0 | \min_{i=0,1} G_{W(t),t}(\Theta_i) \leqq \Phi(-\lambda)\},$$

where $\Phi$ denotes the standard normal distribution function. It can also be expressed as

$$T_\lambda = \inf\{t > 0 | \frac{|W(t) + r\mu|}{(t + r)^{1/2}} \geqq \lambda\}.$$

The following result states that a simple Bayes rule is optimal for the risk (2). The corresponding stopping boundary defines a repeated significance test ($\mu = 0$ is the usual case). To my knowledge, this is the only exact optimality result in sequential testing besides those about the SPRT.

THEOREM: *Let* $0 < c < \infty$. *Let* $\lambda(c)$ *denote the solution of the equation* $\phi(\lambda)/\lambda = 2c$, *and let*

$$T^* = \inf\{t > 0: \frac{|W(t) + r\mu|}{(t + r)^{1/2}} \geqq \lambda(c)\}.$$

*Let* $\delta^* = 1_{\{W(T^*) > -r\mu\}}$ *and* $|\mu| r^{1/2} \leqq \lambda(c)$. *Then the procedure* $(T^*, \delta^*)$ *minimizes the Bayes risk (Eq. 2).*

*Proof*: Let $\overline{Q} = \int_{-\infty}^{\infty} P_\theta \, \phi[r^{1/2}(\theta - \mu)] \, r^{1/2} d\theta$. Then the Bayes formula $P_{\theta,t}(dW)\phi[r^{1/2}(\theta - \mu)]r^{1/2}d\theta = G_{W(t),t}(d\theta)\overline{Q}(dW)$ holds. A well-known argument yields that $R(T, \delta^*) \leqq R(T, \delta)$ for every stopping time $T$. Let $r(T, \delta)$ denote the part of the Bayes risk (2) consisting of the error probabilities. Then

$$r(T, \delta^*) = \int \min_{i=0,1} G_{W(T),T}(\Theta_i) d\overline{Q}$$

$$= \int \Phi\left(-\frac{|W(T) + r\mu|}{(T + r)^{1/2}}\right) d\overline{Q}. \quad [3]$$

On the other hand, the Bayes formula and Fubini's theorem yield

$$\int_{-\infty}^{\infty} \theta^2 E_\theta T \phi[r^{1/2}(\theta - \mu)] r^{1/2} d\theta$$

$$= \int_{-\infty}^{\infty} \theta^2 \left[\int (T + r) dP_\theta\right] \phi[r^{1/2}(\theta - \mu)] r^{1/2} d\theta - (r\mu^2 + 1)$$

$$= \int (T + r)\left(\int_{-\infty}^{\infty} \theta^2 N\left[\frac{W(T) + r\mu}{T + r}, \frac{1}{T + r}\right](d\theta)\right) d\overline{Q} - (r\mu^2 + 1)$$

$$= \int (T + r)\left(\left[\frac{W(T) + r\mu}{T + r}\right]^2 + \frac{1}{T + r}\right) d\overline{Q} - (r\mu^2 + 1)$$

$$= \int \frac{[W(T) + r\mu]^2}{T + r} d\overline{Q} - r\mu^2. \quad [4]$$

Thus, Eqs. **3** and **4** yield the representation of the Bayes risk for $(T, \delta^*)$:

$$\rho(T, \delta^*) = \int f\left(\frac{|W(T) + r\mu|}{(T + r)^{1/2}}\right) d\overline{Q}, \quad [5]$$

with $f(\lambda) = \Phi(-\lambda) + c\lambda^2 - cr\mu^2$. For $\lambda > 0$, $f$ has a unique minimum at $\lambda(c)$ which is defined as the solution of the equation $\phi(\lambda)/2\lambda = c$. Since $|\mu| r^{1/2} \leqq \lambda(c)$, the stopping time $T^* = \inf\{t > 0: |W(t) + r\mu|/(t + r)^{1/2} \geqq \lambda(c)\}$ satisfies $\overline{Q}\{T^* < \infty\} = 1$. Thus,

$$\rho(T, \delta^*) = \int f\left(\frac{|W(T) + r\mu|}{(T + r)^{1/2}}\right) d\overline{Q} \geqq f(\lambda(c)) = \rho(T^*, \delta^*),$$

which completes the proof.

To explain the relation between Wald's SPRT and the RST, we consider the problem of testing the sign of the drift of Brownian motion for the simple hypotheses $-\theta$ versus $+\theta$, with $\theta > 0$. We take loss and cost as above (0–1 loss and cost $c\theta^2$) and restrict our considerations to a symmetric prior, $G = 1/2\delta_{-\theta} + 1/2\delta_{+\theta}$, where $\delta_a$ denotes the point mass at $a$. Because here the cost is constant, it is well known from the theorem on page 197 of ref. 15 that the SPRT minimizes the Bayes risk (Eq. **1**) with cost $c\theta^2$ and with respect to the prior $G$. Calculations similar to those in the proof above show that the Bayes risk can be expressed as $R(T, \delta^*) = \int g(\theta | W(T)|) d\overline{Q}$, with $g(x) = [e^{-2x}/(1 + e^{-2x})] + cx(1 - e^{-2x})/(1 + e^{-2x})$ and $\overline{Q} = \frac{1}{2}P_{-\theta} + \frac{1}{2}P_\theta$. For $x \geqq 0$, $g(x)$ has a unique minimum, say at $b(c)$. Let

$$T^* = \inf\{t > 0: \theta | W(t)| \geqq b(c)\}. \quad [6]$$

Then $(T^*, \delta^*)$ minimizes the Bayes risk (Eq. **1**) with respect to the prior $G$.

Now we consider the testing problem for composite hypotheses $H_0{:}\theta < 0$ versus $H_1{:}\theta > 0$. In ignorance of the parameter $|\theta|$, we estimate it for instance by $|\hat{\theta}_t| = |W(t)|/(t + r)$. Then

$$|\hat{\theta}_t| |W(t)| = \frac{W(t)^2}{t + r},$$

which, together with Eq. **6**, shows that the RST is an adapted version of Wald's SPRT.

The related cases of normal random walks with known and unknown variance can be treated similarly. Since for those cases the overshoot has to be taken into account, exact results no longer hold. The details will be discussed elsewhere.

I thank T. Sellke for useful discussions and the referees for helpful suggestions. This work was supported by the Deutsche Forschungsgemeinschaft at SFB 123 and by National Science Foundation Grant 8120790.

1. Robbins, H. (1952) *Bull. Am. Math. Soc.* **58,** 527–534.
2. Robbins, H. (1970) *Ann. Math. Stat.* **41,** 1397–1409.
3. Lerche, H. R. (1985) *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, eds. Le Cam, L. & Olshen, R. (Wadsworth), Vol. 2, 579–597.
4. McPherson, C. K. & Armitage, P. (1971) *J. R. Stat. Soc. A* **134,** 15–26.
5. Siegmund, D. (1977) *Biometrika* **64,** 177–189.
6. Armitage, P. (1975) *Sequential Medical Trials* (Blackwell, Oxford).
7. Wald, A. & Wolfowitz, J. (1948) *Ann. Math. Stat.* **19,** 326–339.
8. Schwarz, G. (1962) *Ann. Math. Stat.* **33,** 224–236.
9. Kiefer, J. & Sacks, J. (1963) *Ann. Math. Stat.* **34,** 705–750.
10. Lorden, G. (1967) *Ann. Math. Stat.* **38,** 1399–1422.
11. Chernoff, H. (1961) *Proc. Fourth Berkeley Symp. Math. Stat. Probab.* **1,** 79–91.
12. Bather, J. A. (1962) *Proc. Cambridge Philos. Soc.* **58,** 599–620.
13. Chernoff, H. (1972) *Sequential Analysis and Optimal Design* (SIAM, Philadelphia).
14. Lerche, H. R. (1986) *Ann. Stat.*, in press.
15. Wald, A. (1947) *Sequential Analysis* (Dover, New York).