PRAGUE STOCHASTICS 2006

Proceedings of the joint session of

7th Prague Symposium on Asymptotic Statistics

and

15th Prague Conference on Information Theory,
Statistical Decision Functions and Random Processes,

held in Prague from August 21 to 25, 2006

Organised by

Charles University

Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

and

Academy of Sciences of the Czech Republic

Institute of Information Theory and Automation
Department of Stochastic Informatics

Edited by Marie Hušková and Martin Janžura

ISBN 80-86732-75-4

Preface

Prague Stochastics 2006, held in Prague from August 21 to 25, 2006, is an international scientific meeting that continues the tradition of organising Prague conferences on stochastics, established here five decades ago. The first Prague Conference on Information Theory, Statistical Decision Functions and Random Process was initiated by Antonín Špaček in 1956. Prague Symposia on Asymptotic Statistics were founded by Jaroslav Hájek in 1973. This year, we are commemorating the 80th anniversary of the birth date of this untimely deceased outstanding scientist.

Traditionally, the scope of the proceedings, as well as the conference itself, is quite extensive; the topics range from classical to very up-to date ones. It covers both methodological and applied statistics, theoretical and applied probability and, of course, topics from information theory. We hope that all readers will find valuable contributions and a number of papers of their interest in this rich spectrum of scientific ideas.

The printed part contains the plenary and invited papers, and the list of all contributions published in the volume. The CD disc, attached as an official part of the book with the same ISBN code, contains all accepted papers.

The editors would like to express their sincere thanks to the authors for their valuable contributions, to the reviewers for prompt and careful reading of the papers, and to the organisers of the sections for the help with the entire reviewing process.

Our thanks also go to our colleagues, in particular to Pavel Boček and Tomáš Hobza, for their technical editorial work. Without their devotion and diligence, the proceedings would never be completed.

It is our pleasure to acknowledge that Prague Stochastics 2006 is held under the auspices of the Mayor of the City of Prague, the Bernoulli Society for Mathematical Statistics and Probability, and the Czech Statistical Society.

Prague, June 2006 *Marie Hušková, Martin Janžura*

Locally most powerful two-stage tests

Andrey Novikov

Abstract: The problem of testing a simple hypothesis against a composite one-sided alternative is considered. The aim is to find a test which maximizes the slope of the power function at the point of the null-hypothesis over all tests with fixed levels of the first-kind error probability and of the average sample number under the null-hypothesis. For the two-stage tests, the structure of the optimal decision rule and the optimal continuation rule is given (the observations are not supposed to be independent). Numerical results on the efficiency of the optimal two-stage tests, with respect to both the optimal fixed sample-size test and the optimal sequential test by R.Berk, are given.

MSC 2000: 62F03, 62F05, 62L10, 62M02, 62M07
Key words: Statistical hypotheses testing, two-stage test, sequential test, regular experiment, LAN, locally most powerful test, simple hypothesis, composite alternative

1 Introduction

This work is motivated by recent results of the author ([8], [9]) on two-stage hypotheses tests for two simple parametric hypotheses based on regular statistical experiments.

In that case, the optimal two-stage tests perform rather competitively with respect to the optimal sequential tests known as sequential probability ratio tests (SPRT), due to A. Wald [11]. At the same time, the SPRT's are known to be optimal essentially for independent and identically distributed observations (see, for example, [2]), while the two-stage tests have the advantage that they are applicable to any stochastic sequence of observations, and are relatively easy to evaluate (see [8], [9]). In some sense, they are simply "two-step" versions of the well-known Neyman-Pearson test (see, for example, [4]), and with essentially the same way of proof (see [8]), so they are nearly as universal as the Neyman-Pearson test. The real problem of their applicability is the lack of situations in which a simple hypothesis against a *simple* alternative is to be tested.

An approach to sequential testing a simple hypothesis against a *composite* alternative has been proposed by R. Berk in [1]. Again, due to [1], the optimal (called locally most powerful) sequential test exists in the case of independent and

Acknowledgement. The author wishes to thank the anonymous referee for his valuable suggestions on the improvement of the article.

identically distributed observations. There are no known results on optimality of sequential hypotheses testing for more general stochastic sequences in the framework of this approach.

The main aim of this paper is to study the properties of two-stage tests of a simple hypothesis against a composite alternative in the framework of the approach of R. Berk. Because of particular simplicity of two-stage tests, we do not need some of the assumptions made in [1], in particular, we do not suppose the independence of the observations at the two stages of the statistical experiment.

In Section 2, we study the structure of the optimal two-stage test in a rather general context of statistical experiment. To deal with the derivative of the power function we discuss some regularity conditions, which guarantee its existence.

In Section 3, we apply the results of Section 2 for optimal two-stage tests to testing hypotheses about the drift of a Wiener process with a lineal drift and give some numerical results on the efficiency of the optimal two-stage tests with respect to both optimal one-stage tests, and to the locally most powerful sequential test of R. Berk.

2 The structure of the optimal two-stage tests

In this section, we give a general framework for two-stage hypotheses tests, and describe the structure of the optimal two-stage test.

2.1 General framework. Definitions

Let us assume that we can observe in a statistical experiment a random variable X (the first stage of the experiment), and, depending on its value, either stop at the first stage or get to a second stage, obtaining an additional portion of observations Y. In any case we have to take a final decision about the distribution from which the vector (X, Y) comes. This type of experiment can be thought of as an alternative to fixed-size sampling, as in the Neyman-Pearson test, and to completely sequential tests like the Wald's sequential probability ratio test. For example, the usual fixed sample-size test is a particular case of this scheme, corresponding to *never* going to the second stage, and making the inference on the base of the X-observation.

Let us assume that the vector (X, Y) follows a parametric distribution with a probability density function $f_\theta(x, y)$ with respect to a product-measure $\mu_1 \otimes \mu_2$ on the space of values of (X, Y), where $\theta \in \Theta \subset \mathbb{R}$ is some parameter. Thus, $f_\theta(x) = \int f_\theta(x, y) d\mu_2(y)$ is the marginal density function of the first-stage component X with respect to μ_1.

Let θ_0 be such that there exists θ_1, $\theta_0 < \theta_1 \leqslant \infty$ for which $[\theta_0, \theta_1) \subset \Theta$. In this paper, we deal with testing the simple hypothesis $H_0 : \theta = \theta_0$ against the composite one-sided alternative $H_1 : \theta > \theta_0$.

For a pair of hypotheses H_0 and H_1 let us define a (two-stage) test as a triplet of measurable functions $(\phi_1(x), \phi_2(x, y), \chi(x))$, all of them taking values in $[0, 1]$, with the following interpretation:

- $\phi_1(x)$ being the conditional probability, given a first-stage observation x, to reject H_0,
- $\phi_2(x, y)$ the conditional probability, given observations up to the second stage (x, y), to reject H_0, and
- $\chi(x)$ being the conditional probability, given the first-stage observation x, to get to the second stage (to continue sampling).

The functions $\phi_1(x)$, $\phi_2(x, y)$ can be considered as (randomized) decision rules at the respective stages of the experiment, and $\chi(x)$ as a (randomized) continuation rule. So, for example, the particular case $\chi(x) \equiv 0$ corresponds to a "fixed sample-size" test, with no observations at the second stage (in fact, in this case $\phi_2(x, y)$ has to play no role, although formally we have to give some value to it, for example $\phi_2(x, y) \equiv 0$ or $\phi_2(x, y) \equiv 1$, or whatever. In what follows we will see that it does not have any importance for the performance of the test).

As usual in the context of hypotheses testing we define the power function as the (total) probability to reject H_0 when the true parameter of the distribution of (X, Y) is θ:

$$P(\theta) = E_\theta \left[\phi_1(X)(1 - \chi(X)) + \phi_2(X, Y)\chi(X)\right]. \tag{1}$$

As in [1], we are interested in maximizing $P'(\theta_0)$ and minimizing $P(\theta_0)$ (called the error probability of the first kind). Also, we have to take into account the cost of additional observations, if any. As the first stage is always present, the only variable part is naturally related to

$$C(\theta) = E_\theta \chi(X), \tag{2}$$

the probability of continuing observations up to the second stage, given θ. As in [1], we will only pay attention to the value of $C(\theta)$ under H_0, i.e. $C(\theta_0)$.

2.2 Differentiability of the power function

To deal with the derivative of the power function (1), we have to be sure that it exists. In [1], there are conditions ensuring the differentiability of $P(\theta)$ for the experiment consisting in observing sequentially independent and identically distributed random variables $X_1, X_2, \ldots, X_n, \ldots$ and *any* stopping time τ based on it, for which $E_\theta \tau < \infty$. In our case, in view of (1) the conditions for existence of the derivative might be very mild. A very natural candidate for this is some differentiability condition of the family $\{f_\theta(x, y)\}_{\theta \in \Theta}$ at $\theta = \theta_0$. Essentially, we need the possibility to calculate the derivative of the power function (1) at $\theta = \theta_0$ differentiating under the integral sign.

So, we will suppose that at $\theta = \theta_0$ the following condition holds.

C1. *The power function $P(\theta)$ (1) of any two-stage test is differentiable and there exists a $\mu_1 \otimes \mu_2$-integrable function $\psi_\theta(x,y)$ such that*

$$P'(\theta) = \int \psi_\theta(x,y)\left[\phi_1(x)(1-\chi(x)) + \phi_2(x,y)\chi(x)\right] d\mu_1 \otimes \mu_2(x,y)$$

Typically, one would expect that $\psi_\theta(x,y) = f'_\theta(x,y) = \frac{\partial f_\theta}{\partial \theta}(x,y)$ if this derivative exists. If it does not, but C1 still holds, we will keep using this notation, i.e.

$$f'_\theta(x,y) \equiv \psi_\theta(x,y) \tag{3}$$

by definition.

There are different ways to guarantee C1. A very closely related discussion can be found in [5], where some references to earlier papers are given.

In particular, it is easy to see that condition C1 is satisfied if $f_\theta(x,y)$ is L_1-differentiable in the following sense (cf., e.g., [5]).

C2. *There exists a function $\psi_\theta(x,y)$ such that $\int |\psi_\theta(x,y)| d\mu_1 \otimes \mu_2(x,y) < \infty$ and*

$$\int |f_{\theta+u}(x,y) - f_\theta(x,y) - \psi_\theta(x,y)u| d\mu_1 \otimes \mu_2(x,y) = o(u),$$

as $u \to 0$.

In a rather standard way, in turn, this condition holds if $\sqrt{f_\theta(x,y)}$ is L_2-differentiable in the following sense (see, for example, [3]).

C3. *There exists a function $\psi_\theta(x,y)$ such that $\int \psi^2_\theta(x,y) d\mu_1 \otimes \mu_2(x,y) < \infty$ and*

$$\int (\sqrt{f_{\theta+u}(x,y)} - \sqrt{f_\theta(x,y)} - \psi_\theta(x,y)u)^2 d\mu_1 \otimes \mu_2(x,y) = o(u^2),$$

as $u \to 0$.

Although C2 seems to be more natural in the context of hypotheses testing, C3 may be preferable dealing with regular statistical experiments and/or locally asymptotically normal (LAN) experiments (see, for example, [3]).

In what follows, we will only use condition C1, seemingly close to the weakest possible one.

Note that in all conditions C1-C3 we need in effect only the right-differentiability at $\theta = \theta_0$ due to the essence of our testing problem.

Concluding this section let us note that condition C1 implies that for any one-stage test $\phi_1(x)$ (with $\chi(x) \equiv 0$) by Fubini's theorem

$$P'(\theta) = \int \phi_1(x) \left[\int \psi_\theta(x,y) d\mu_2(y)\right] d\mu_1(x),$$

justifying the notation

$$f'_\theta(x) \equiv \int \psi_\theta(x,y) d\mu_2(y). \tag{4}$$

2.3 Optimal two-stage tests

To study the structure of the optimal test let us start with a Lagrange-multiplier-type optimization. Let λ and c be two positive constants. Then our Lagrange function is

$$P'(\theta_0) - \lambda P(\theta_0) - cC(\theta_0) \tag{5}$$

with $P(\theta)$ defined by (1) and $C(\theta)$ defined by (2).

In what follows we use the following notation:

$$a^+ = \frac{a + |a|}{2}$$

and

$$I(A) = \begin{cases} 1, & \text{if } A \text{ occurs,} \\ 0, & \text{if not.} \end{cases}$$

Theorem 1. *Let condition C1 at $\theta = \theta_0$ be fulfilled. Then the maximum value of (5) over all two-stage tests is equal to*

$$\int \left(l_1(x)^+ + \rho(x)^+\right) d\mu_1(x) \tag{6}$$

with

$$\rho(x) = \int l_2(x,y)^+ d\mu_2(y) - l_1(x)^+ - cf_{\theta_0}(x), \tag{7}$$

$$l_1(x) = f'_{\theta_0}(x) - \lambda f_{\theta_0}(x),$$

$$l_2(x,y) = f'_{\theta_0}(x,y) - \lambda f_{\theta_0}(x,y).$$

The maximum value (6) is achieved at any two-stage test of the form:

$$\begin{aligned} \phi_1(x) &= I(\{l_1(x) > 0\}) + \gamma_1(x) I(\{l_1(x) = 0\}), & (8)\\ \phi_2(x,y) &= I(\{l_2(x,y) > 0\}) + \gamma_2(x,y) I(\{l_2(x,y) = 0\}), & (9)\\ \chi(x) &= I(\{\rho(x) > 0\}) + \gamma_3(x) I(\{\rho(x) = 0\}), & (10) \end{aligned}$$

where $\gamma_1(x)$, $\gamma_2(x,y)$ and $\gamma_3(x)$ (randomization constants) are some measurable functions taking values in [0,1].

Proof. In what follows $\theta = \theta_0$.

Let us start with a fixed continuation rule. For any $\chi(x)$ fixed let us find the maximum value of (5). As $C(\theta)$ depends only on $\chi(x)$, it suffices to find a maximum of

$$P'(\theta) - \lambda P(\theta) = \int\int (f'_\theta(x,y) - \lambda f_\theta(x,y))\phi_1(x)(1 - \chi(x)) d\mu_1(x) d\mu_2(y) +$$

$$+ \int\int (f'_\theta(x,y) - \lambda f_\theta(x,y))\phi_2(x,y)\chi(x) d\mu_1(x) d\mu_2(y) =$$

$$= \int (f'_\theta(x) - \lambda f_\theta(x))\phi_1(x)(1-\chi(x))d\mu_1(x) +$$

$$+ \int\int (f'_\theta(x,y) - \lambda f_\theta(x,y))\phi_2(x,y)\chi(x)d\mu_1(x)d\mu_2(y), \tag{11}$$

were we used condition C1 to calculate $P'(\theta)$ and definitions (3) and (4).

The first summand on the right-hand side of (11) does not exceed

$$\int (f'_\theta(x) - \lambda f_\theta(x))^+(1-\chi(x))d\mu_1(x),$$

because their difference is equal to

$$\int (f'_\theta(x) - \lambda f_\theta(x))(I(\{f'_\theta(x) - \lambda f_\theta(x) \geqslant 0\}) - \phi_1(x))(1-\chi(x))d\mu_1(x), \tag{12}$$

which is non-negative due to

$$(f'_\theta(x) - \lambda f_\theta(x))(I(\{f'_\theta(x) - \lambda f_\theta(x) \geqslant 0\}) - \phi_1(x)) \geqslant 0,$$

because $0 \leqslant \phi_1(x) \leqslant 1$.

At the same time we see that the difference (12) is equal to 0 if $\phi_1(x)$ has the form (8).

In the same way we see that the second summand on the right-hand side of (11) does not exceed

$$\int\int (f'_\theta(x,y) - \lambda f_\theta(x,y))^+\chi(x)d\mu_1(x)d\mu_2(y),$$

and, again, this maximum is achieved if $\phi_2(x,y)$ has the form (9).

Now we have that for any $\chi(x)$

$$P'(\theta) - \lambda P(\theta) - cC(\theta) \leqslant$$

$$\int (f'_\theta(x) - \lambda f_\theta(x))^+(1-\chi(x))d\mu_1(x) + \int\int (f'_\theta(x,y) - \lambda f_\theta(x,y))^+\chi(x)d\mu_1(x)d\mu_2(y)$$

$$-c\int \chi(x)f_\theta(x)d\mu_1(x)$$

$$= \int (f'_\theta(x) - \lambda f_\theta(x))^+ d\mu_1(x)$$

$$+ \int \left(\int (f'_\theta(x,y) - \lambda f_\theta(x,y))^+ d\mu_2(y) - (f'_\theta(x) - \lambda f_\theta(x))^+ - cf_\theta(x) \right) \chi(x)d\mu_1(x)$$

$$= \int l_1(x)^+ d\mu_1(x) + \int \rho(x)\chi(x)d\mu_1(x). \tag{13}$$

In the same way as above we see that the second term on the right-hand side of (13) does not exceed

$$\int \rho(x)^+ d\mu_1(x), \tag{14}$$

and that it coincides with (14) if $\chi(x)$ has the form (10).

From this fact and (13) we conclude that

$$P'(\theta)-\lambda P(\theta)-cC(\theta) \leqslant \int l_1(x)^+ d\mu_1(x)+\int \rho(x)^+ d\mu_1(x) = \int (l_1(x)^+ + \rho(x)^+) d\mu_1(x),$$

with the equality if the test has the form (8)-(10). □

Note. From the proof it is obvious that, more generally, to reach the maximum value (6) the relations (8)–(10) may be satisfied almost everywhere.

More than that, it is not difficult to see that if the maximum (6) is reached, then

$$\phi_1(x)(1-\chi(x)) = (I(\{l_1(x)>0\}) + \gamma_1(x)I(\{l_1(x)=0\}))\,(1-\chi(x)),$$

$$\phi_2(x,y)\chi(x) = (I(\{l_2(x,y)>0\}) + \gamma_2(x,y)I(\{l_2(x,y)=0\}))\,\chi(x),$$

$$\chi(x) = I(\{\rho(x)>0\}) + \gamma_3(x)I(\{\rho(x)=0\}),$$

almost everywhere, so this is *the necessary and sufficient condition for reaching the maximum value (6).*

2.4 Locally most powerful two-stage tests

Let us show now how the result of the preceding section can be applied to finding locally most powerful two-stage tests.

Suppose first that we observe, in two stages, a discrete-time stochastic process $X_1, X_2, \ldots, X_n, \ldots$. In terms of the preceding section we have: $X = (X_1, X_2, \ldots, X_{n_1})$ and $Y = (X_{n_1+1}, X_{n_2+2}, \ldots, X_{n_1+n_2})$, where n_1 (n_2) is the number of observations taken at the first (second) stage of the experiment.

Obviously, for any n_1 and n_2 fixed, Theorem 1 gives us the form of the optimal test, which maximizes

$$P'(\theta_0) - \lambda P(\theta_0) - cN(\theta_0) \tag{15}$$

over all tests with n_1 observations at the first and n_2 at the second stage of the experiment, where $N(\theta) = n_1 + n_2 C(\theta)$ is the average sample number.

Let us denote by Δ the class of all two-stage tests of the form (8)–(10) with $x = (x_1, x_2, \ldots, x_{n_1})$ and $y = (x_{n_1+1}, x_{n_1+2}, \ldots, x_{n_1+n_2})$, corresponding to any combination of $n_1 \geqslant 1$, $n_2 \geqslant 1$, $\lambda > 0$ and $c > 0$.

By Theorem 1, for any fixed $\lambda > 0$ and $c > 0$ *any* two-stage test has its corresponding test in Δ with a greater (or equal) value of the Lagrange function (15).

Let us start with the locally most powerful two-stage tests now.

From now on, for a two-stage test $\phi =< \phi_1, \phi_2, \chi >$ let us use $P(\theta;\phi)$ and $N(\theta;\phi)$ for its power function and average sample number, respectively.

As in [1], we are interested in finding a test maximizing $P'(\theta_0;\phi)$ over all two-stage tests ϕ with
$$P(\theta_0;\phi) \leqslant \alpha, \tag{16}$$
and
$$N(\theta_0;\phi) \leqslant \nu, \tag{17}$$
where $\alpha \in (0,1)$ and $\nu > 0$ are some fixed numbers (locally most powerful test at $\theta = \theta_0$).

Let
$$L(\phi;\lambda,c) = P'(\theta_0;\phi) - \lambda P(\theta_0;\phi) - cN(\theta_0;\phi)\,. \tag{18}$$

Let us suppose now that for some $\lambda > 0$ and $c > 0$ there is a test $\phi^* \in \Delta$ such that
$$\sup_{\phi\in\Delta} L(\phi;\lambda,c) = L(\phi^*;\lambda,c), \tag{19}$$
and let $\alpha = P(\theta_0;\phi^*)$ and $\nu = N(\theta_0;\phi^*)$.

It is easy to see that in this case ϕ^* is the locally most powerful test among all two-stage tests satisfying (16) and (17).

Indeed, if ϕ is any such test, then
$$L(\phi;\lambda,c) \leqslant L(\phi^*;\lambda,c)$$

$$= P'(\theta_0;\phi^*) - \lambda P(\theta_0;\phi^*) - cN(\theta_0;\phi^*) = P'(\theta_0;\phi^*) - \lambda\alpha - c\nu \tag{20}$$
because of Theorem 1 and (19).

On the other hand, because of (16) and (17),
$$L(\phi;\lambda,c) = P'(\theta_0;\phi) - \lambda P(\theta_0;\phi) - cN(\theta_0;\phi) \geqslant P'(\theta_0;\phi) - \lambda\alpha - c\nu, \tag{21}$$

Combining (20) and (21) we have
$$P'(\theta_0;\phi) \leqslant P'(\theta_0;\phi^*),$$
which proves that ϕ^* is locally most powerful.

It is quite obvious that in the same way we can apply the result of the preceding section for construction of the locally most powerful two-stage test for a continuous-time stochastic process.

In this case, the observations will be taken from a stochastic process $X(t), t \geqslant 0$, and, in terms of the preceding section, the observations X and Y at the two stages, of the respective duration t_1 and t_2, can be taken as $X(t_1)$ and $X(t_1+t_2)$, respectively.

Again, by Theorem 1, the optimal two-stage test has the form (8)–(10) with $x = x(t_1)$ and $y = x(t_1+t_2)$, where $x(t_1)$ and $x(t_1+t_2)$ are the observed values of

$X(t_1)$ and $X(t_1+t_2)$, respectively, and $f_\theta(x,y)$ corresponds to the two-dimensional distribution of $(X(t_1), X(t_1+t_2))$. Because of this, all the elements of the optimal two-stage test (8)–(10) are relatively easy to calculate, nearly as easy as in the case of the discrete-time stochastic process above.

Acting in the same way as above in this section (see (19) and what follows), we can find the locally most powerful two-stage test in this case as well.

It is worth mentioning that, generally speaking, there is no guarantee that this way we can find the locally most powerful two-stage test *for any* given α and/or ν (and even for *some* of them), but neither is it there in the case of [1], even for independent and identically distributed observations.

As a promising fact let us note that in any case the optimization problem (19) is essentially numerical (two-dimensional optimization), so there is a hope that, in any concrete case, it can be solved with more or less difficulty, at least numerically.

3 Example: A Wiener process with a lineal drift

In this section, we apply the results of the preceding section to the case of testing hypotheses about the drift of a Wiener process with a lineal drift.

We observe the process $\xi(t) = W(t) + \theta t$, where $W(t)$ is a standard Wiener process. At the first stage, we observe $\xi(t)$ up to the time t_1, then, if necessary, at the second stage we observe $\xi(t)$ for t_2 time units more.

We are interested in testing $H_0 : \theta = \theta_0 = 0$ vs $H_1 : \theta > 0$ using a two-stage test.

Because $(\xi(t_1), \xi(t_1+t_2))$ is a sufficient statistics, we can restrict our attention to the distribution of the vector (X, Y), where $X = \xi(t_1)$ and $Y = \xi(t_1+t_2) - \xi(t_1)$. So, in the terms of the preceding section

$$f_\theta(x) = f_\theta^1(x) = \frac{1}{\sqrt{t_1}}\phi\left(\frac{x-\theta t_1}{\sqrt{t_1}}\right), \quad f_\theta(x,y) = f_\theta^1(x) f_\theta^2(y),$$

$$f_\theta^2(y) = \frac{1}{\sqrt{t_2}}\phi\left(\frac{x-\theta t_2}{\sqrt{t_2}}\right),$$

$\phi(x)$ being the probability distribution function of the standard normal distribution.

In this case ($\theta_0 = 0$) it is not difficult to calculate:

$$f'_{\theta_0}(x) = x f_{\theta_0}(x)$$

$$f'_{\theta_0}(x,y) = (x+y) f_{\theta_0}(x,y)$$

$$\rho(x) = (E_{\theta_0}(x+Y-\lambda)^+ - (x-\lambda)^+ - ct_2) f_{\theta_0}(x)$$

$$= (\sqrt{t_2}\phi((x-\lambda)/\sqrt{t_2}) - |x-\lambda|\Phi(-|x-\lambda|/\sqrt{t_2}) - ct_2) f_{\theta_0}(x),$$

where $\Phi(x)$ is the standard normal distribution function.

To define the optimal continuation rule (see Theorem 1) let us note that $\rho(x) > 0$ is equivalent to $\phi(u) - |u|\Phi(-|u|) > c\sqrt{t_2}$, where $u = (x-\lambda)/\sqrt{t_2}$. Thus, the optimal continuation rule is

$$\chi(x) = I(\{|x - \lambda|/\sqrt{t_2} < a\}),$$

where $a = a(c\sqrt{t_2})$ is the positive solution of the equation

$$\phi(u) - |u|\Phi(-|u|) = c\sqrt{t_2}.$$

Jointly with $\phi_1(x) = I(\{x > \lambda\})$ and $\phi_2(x) = I(\{x + y > \lambda\})$ this gives a complete description of the optimal test (8)-(10) for any fixed t_1 and t_2.

Let us denote $G(u) = \phi(u) - |u|\Phi(-|u|)$ for $u \in \mathbb{R}$. Then for any $\lambda > 0$ and $c > 0$ and $t_1 > 0$, $t_2 > 0$ fixed the value (18) for the above test is equal to

$$\sqrt{t_1}\left(G\left(\frac{\lambda}{\sqrt{t_1}}\right) - c\sqrt{t_1}\right) + \sqrt{t_2}E_{\theta_0}\left(G\left(\frac{\xi(t_1) - \lambda}{\sqrt{t_2}}\right) - c\sqrt{t_2}\right)^+ \tag{22}$$

To find the locally most powerful test following the plan of Section 2.4, we need to find the supremum of (22) over all $t_1 > 0$ and $t_2 > 0$ (see (19)) for any $\lambda > 0$ and $c > 0$ fixed.

Surprisingly, for some $\lambda > 0$ and $c > 0$ the maximum value of (22) is equal to 0 and is achieved at $t_1 = t_2 = 0$, so the procedure of Section 2.4 fails. The worst of all is that this happens for large values of λ which are necessary to hold the error probability (16) at a reasonably low level of α. Our numerical estimations show that for $\alpha \approx 0.1$ or less the maximum value of (22) over all $t_1 > 0$ and $t_2 > 0$ is 0. Nevertheless, for greater values of α the direct maximization of (22) gives us a definite level-α two-stage test, which turns out to be the locally most powerful level-α two-stage test, by the results of Section 2.4. Some numerical results for larger α we show below in this Section.

To treat lower levels of α in this example we propose another plan.

The idea is to start with a fixed $t_1 > 0$ in maximization of (22), say, $t_1 = 1$. It is easy to see that the maximum of (22) with $t_1 = 1$ over $t_2 > 0$ is achieved at some $t_2 = r^2$. Starting from the pair $(1, r^2)$ it is not difficult to construct the two-stage test giving the maximum to (22) (over $t_2 > 0$) for any fixed t_1.

Let us denote by $P_1(\theta; \lambda, c)$ the power function of the two-stage test based on $t_1 = 1$ and $t_2 = r^2$ giving the maximum (over t_2) to (22) with $t_1 = 1$, and let $P_1'(\theta_0; \lambda, c)$ be its derivative at $\theta = \theta_0$. Let, finally, be $N_1(\theta_0; \lambda, c)$ its average sample number.

For $t_1 \neq 1$ we use, respectively, $P_{t_1}(\theta; \lambda, c)$, $P_{t_1}'(\theta_0; \lambda, c)$ and $N_{t_1}(\theta_0; \lambda, c)$ for the corresponding characteristics of the test giving the maximum (over t_2) to (22) when t_1 is held fixed.

It is not difficult to see that

$$P_{t_1}(\theta_0; \lambda, c) = P_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1}), \tag{23}$$

$$P'_{t_1}(\theta_0; \lambda, c) = \sqrt{t_1} P'_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1}), \tag{24}$$

$$N_{t_1}(\theta_0; \lambda, c) = t_1 N_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1}). \tag{25}$$

If now

$$P_{t_1}(\theta_0; \lambda, c) \approx \alpha \tag{26}$$

and

$$N_{t_1}(\theta_0; \lambda, c) \approx \nu, \tag{27}$$

from (23) and (25) we have that

$$P_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1}) \approx \alpha \tag{28}$$

and

$$t_1 \approx \nu / N_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1}),$$

and, by virtue of (24),

$$P'_{t_1}(\theta_0; \lambda, c) \approx \sqrt{\nu} \frac{P'_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1})}{\sqrt{N_1(\theta_0; \frac{\lambda}{\sqrt{t_1}}, c\sqrt{t_1})}}. \tag{29}$$

Thus, to maximize the left-hand side of (29) over all tests subject to (26) and (27), it suffices to maximize the right-hand side of (29) subject to (28).

So, our candidate for the locally most powerful two-stage test is the test giving the maximum to

$$\frac{P'_1(\theta_0; \lambda, c)}{\sqrt{N_1(\theta_0; \lambda, c)}} \tag{30}$$

subject to

$$P_1(\theta_0; \lambda, c) = \alpha. \tag{31}$$

The value of (30) is natural to interpret as the efficiency of the two-stage test, representing the "specific slope", per square root of the average sample number unit. Because of that, let us denote its maximum value by $E_2 = E_2(\alpha)$ (here 2 stands for "two-stage"). Some numerical results on the evaluation of E_2 can be found below.

It is very interesting to calculate the relative efficiency of the optimal two-stage tests with respect to the fixed sample-size test, and to the optimal sequential test.

Let us start with the "fixed sample-size" tests.

In terms of Section 2.1 it is a "one-stage" test with no continuation region ($\chi(x) \equiv 0$). In the context of this example, this means that it is defined by a fixed time t_1 of the observation at the first stage, with no continuation.

Because the form of the decision rule is fixed by Theorem 1, we have as the optimal one-stage test $\phi_1(x) = I(\{x > \lambda\})$ with $x = \xi(t_1)$. Obviously, the power function of such a test is $P(\theta_0) = P_{\theta_0}(\xi(t_1) > \lambda) = 1 - \Phi(\lambda/\sqrt{t_1})$ and $P'(\theta_0) = E_{\theta_0}\xi(t_1)I(\{\xi(t_1) > \lambda\}) = \sqrt{t_1}\exp\{-\lambda^2/(2t_1)\}/\sqrt{2\pi}$ with $N(\theta) = t_1$.

Defining λ in such a way that $P(\theta_0) = \alpha$ we have:

$$\lambda = \sqrt{t_1}\Phi^{-1}(1-\alpha),$$

where $\Phi^{-1}(1-\alpha)$ it the $(1-\alpha)$-quantile of the standard normal distribution, and hence for the efficiency $E_1 = E_1(\alpha)$:

$$E_1 = \phi(\Phi^{-1}(1-\alpha)). \tag{32}$$

Again, 1 in E_1 stands for "one-stage".

Now, let us evaluate the efficiency of the optimal sequential test. Formally, the case of continuous-time stochastic processes is not covered in [1], so we make use of the results of [10] extending the locally most powerful tests of R. Berk to processes with stationary independent increments. Because, as stated in [10], for exponential families the locally most powerful sequential test is a Wald's SPRT for a pair of conjugate values of θ, for which θ_0 is an "exceptional point", we see that for the case we are considering the optimal sequential test is an SPRT for two symmetrical values of θ. Being so, it is easy to calculate the characteristics of the test under H_0, because in this case the SPRT is defined by two constants $-A < 0$ and $B > 0$ and stops when $\xi(t)$ for the first time hits any one of the two boundaries, that is, its stopping time is $\tau = \sup\{t : -A < \xi(t) < B\}$. Because under H_0 there is no drift $(\xi(t) \equiv W(t))$, all the characteristics are easy to calculate using the well-known formulas for the ruin probability. This way, we come to the efficiency E_∞ of the optimal sequential test:

$$E_\infty = \sqrt{\alpha(1-\alpha)}.$$

Obviously, it should be $E_1 < E_2 < E_\infty$. In the table below we give the values of E_1, E_2 and E_∞ for some usual (or interesting) values of α. For convenience, we also show their relative values $RE_2 = E_2/E_1$ and $RE_\infty = E_\infty/E_2$ to make visible the increase in the specific slope from using "respectively more" stages of the experiment. It should be noted that the increase E_2/E_1 is really due to one additional stage of the experiment, while E_∞/E_2 corresponds to going to an "infinitely much" richer experiment, with a continuum of "stages" in place of two stages as in E_2.

α	E_1	E_2	E_∞	RE_2	RE_∞
50%	0.39894	0.43484	0.50000	1.090	1.150
20%	0.28067	0.32867	0.40063	1.171	1.219
10%	0.17550	0.23374	0.30000	1.332	1.283
5.0%	0.10313	0.16143	0.21794	1.565	1.350
2.5%	0.05844	0.11054	0.15612	1.891	1.412
1.0%	0.02665	0.06733	0.09950	2.526	1.478
0.5%	0.01446	0.04630	0.07053	3.202	1.523

We note that the two-stage hypotheses tests are good competitors to the fully sequential tests. Taking into account that they do not require the independence of the observations, and thus are more applicable, they are a good prospect to study in more details.

Examples of the two-stage tests for dependent observations will be given somewhere else. The example we consider here deals with *independent* observations for two reasons. We need this case as a "reference point" for efficiency evaluation, because there are no known optimality results of purely sequential tests for reasonably general model with dependent observations, and, consequently, the efficiency comparison of the two-stage tests with purely sequential tests would not be feasible. Now, we can state that two-stage perform well *even* in the case of *independent* observations. The second reason is that the model considered here serve as the "limiting" case for a rather broad class of locally asymptotically normal (LAN) experiments. In particular, we can mention, besides the well-known LAN experiments with independent identically distributed observations (see, e.g., [3]), regular discrete-time Markov ergodic stochastic processes (see [6], cf. also [7]). We expect that the optimality results for the Wiener process will have consequences in *asymptotic* optimality for a larger class of LAN experiments.

Other promising application of two-stage tests seems to be the problem of testing a simple hypothesis versus a *two-sided* alternative, in which case unbiased tests are needed. As stated in [5] the form of the locally most powerful unbiased sequential test is difficult to find. At the same time, unbiased two-stage tests are easy to find, at least in the example we considered here. So the locally most powerful unbiased two-stage tests are waiting for being studied.

References

[1] Berk R.H. Locally most powerful sequential tests. *Ann. Statist.* V.3, No. 2, 373–381, 1975

[2] DeGroot M. H. *Optimal statistical decisions.* McGraw-Hill Book Co., New York-London-Sydney, 1970.

[3] Le Cam, L. *Asymptotic methods in statistical decision theory.* Springer Series in Statistics. Springer-Verlag, New York–Berlin, 1986.

[4] Lehmann E.L. *Testing statistical hypotheses.* John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1959.

[5] Müller-Funk U., Pukelsheim F., Witting H. Locally most powerful tests for two-sided hypotheses. *Probability and statistical decision theory, Vol. A (Bad Tatzmannsdorf, 1983)*, 31–56, Reidel, Dordrecht, 1985.

[6] Novikov A. Uniform asymptotic expansion of likelihood ratio for Markov dependent observations. *Ann. Inst. Statist. Math.*, V. 53, No. 4, 799–809, 2001.

[7] Novikov A. Efficiency of sequential hypotheses testing. *Aportaciones matemáticas. Serie Comunicaciones*, **30**:71–79, 2002.

[8] Novikov A. Optimality of two-stage hypothesis tests *COMPSTAT Proceedings in Computational Statistics, 16th Symposium held in Prague, Czech Republic*, Physica-Verlag, pp. 1601–1608, 2004.

[9] Novikov A. Asymptotic optimality of two-stage hypotheses tests. Aportaciones Matemáticas. Serie Comunicaciones, 35:37–43, 2005.

[10] Roters M. Locally most powerful sequential tests for processes of the exponential class with stationary and independent increments. *Metrika*, 39, no. 3-4, 177–183, 1992.

[11] Wald A. *Sequential analysis.* John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1947

Andrey Novikov: UAM-Iztapalapa, Depto. Matemáticas, San Rafael Atlixco #186, col. Vicentina, México, 09340, México, *an@xanum.uam.mx*