Aportaciones Matemáticas
Memorias **35** (2005) 31–37
Artículo de Investigación

ASYMPTOTIC OPTIMALITY OF TWO-STAGE HYPOTHESES TESTS

ANDREY NOVIKOV

Abstract. This paper deals with asymptotically optimal two-stage tests for two close simple hypotheses. The observations are not supposed to be independent, but a local asymptotically normal behavior of the statistical experiment is supposed. The results are applicable, in particular, to hypotheses testing for regular Markov ergodic discrete-time processes.

Resumen. En este artículo, se consideran las pruebas bietápicas asintoticamente óptimas para dos hipótesis simples cercanas. No se supone la independencia de las observaciones, pero se supone un comportamiento localmente asintoticamente normal del experimento estadístico. Los resultados son aplicables, en particular, al caso de pruebas de hipótesis para procesos ergódicos de Markov a tiempo discreto regulares.

1. Introduction: Two-stage hypotheses tests for two simple hypotheses

In this section, we refer to the structure of an optimal two-stage test for two simple hypotheses in a quite general set-up, and to a particular case of this test which will serve as a reference for asymptotic considerations.

1.1. **The form of optimal two-stage tests.** Let us assume that we can observe in a statistical experiment a random "variable" X (the first stage of the experiment), and, depending on it, either stop at the first stage or get to a second stage, obtaining an additional portion of observations Y. In both cases we have to take a final decision about the distribution from which X and Y come. This type of experiment can be thought of as an alternative to fixed-size sampling, as in the Neyman-Pearson test,

---

2000 *Mathematics Subject Classification.* 62F05, 62L10, 62M02, 62M07.

Key words and phrases. Hypothesis testing, two-stage test, optimal continuation rule, optimal decision, dependent observations, time-series, local asymptotic normality, Markov discrete-time process, Wiener process, lineal drift.

This is the final form of the paper.

and to completely sequential tests like the Wald's sequential probability ratio test (SPRT).

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 $\mu_1 \times \mu_2$ on the space of values of (X, Y), so $f_\theta(x) = \int f_\theta(x, y) d\mu_2(y)$ being the marginal density function of the first-stage component X with respect to μ_1.

For two simple hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ let us define a 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).

So the power function of the test is defined as

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

(the total probability to reject H_0 given θ).

We are interested in minimizing $P(\theta_0)$ and $1 - P(\theta_1)$ which are, respectively, the error probabilities of the first and the second kind, and some quantities related to a cost of observations. As the first stage is always present, the only variable part is related to $C(\theta) = E_\theta \chi(x)$, which is the probability of continuing observations up to the second stage, given θ.

As usual in statistical hypotheses testing, we start from a sort of Bayesian set-up: we will be interested in finding tests which minimize the average total loss (ATL):

$$\pi_0 P(\theta_0) + \pi_1(1 - P(\theta_1)) + \pi_0 c_0 C(\theta_0) + \pi_1 c_1 C(\theta_1)$$

where π_0 and π_1 can be interpreted as prior probabilities of H_0 and H_1, respectively, and c_0 and c_1 some constants giving some weight to any of the two average observation costs measured by $C(\theta_0)$ and $C(\theta_1)$.

Let a^- be equal to a, if $a < 0$, and $a^- = 0$ otherwise, and let $I(A)$ be the indicator function of the event A.

The following theorem gives the structure of the test with the minimum ATL (see [4]).

Theorem 1.1. *The minimum average total loss is equal to* $\pi_1 +$

$$\int \left[l_1(x)^- + \left(\int l_2(x, y)^- d\mu_2(y) - l_1(x)^- + \pi_0 c_0 f_{\theta_0}(x) + \pi_1 c_1 f_{\theta_1}(x) \right)^- \right] d\mu_1(x)$$

where $l_1(x) = \pi_0 f_{\theta_0}(x) - \pi_1 f_{\theta_1}(x)$, $l_2(x, y) = \pi_0 f_{\theta_0}(x, y) - \pi_1 f_{\theta_1}(x, y)$, *and this minimum is achieved by a test with*

$$
\begin{aligned}
\phi_1(x) &= I(\{l_1(x) < 0\}) \\
\phi_2(x,y) &= I(\{l_2(x,y) < 0\}) \\
\chi(x) &= I(\{\int l_2(x,y)^- d\mu_2(y) - l_1(x)^- + \pi_0 c_0 f_{\theta_0}(x) + \pi_1 c_1 f_{\theta_1}(x) < 0\})
\end{aligned} \tag{1}
$$

1.2. **Wiener process case.** We will need, as a reference, the following particular case of Theorem 1.1.

Let us assume that we observe a Wiener process with a linear drift: $X(t) = W(t) + \theta t$. Without loss of generality we can assume that $W(t)$ is standard and that we are interested in testing the null hypotheses that $\theta = \theta_0 = 0$, taking as the alternative $\theta_1 = \theta = 1$.

At the first stage of the experiment, we observe the process up to a time t_1, keeping observing, if necessary, a time t_2 more at the second stage.

It can easily be seen (see [4]) that for any given $\pi_0, \pi_1, c_0, c_1, t_1, t_2$ the optimal two-stage test is given by

$$
\begin{aligned}
\phi_1(x) &= I(\{Z_1(x) > \pi_0/\pi_1\}) \\
\phi_2(x,y) &= I(\{Z_1(x)Z_2(y) > \pi_0/\pi_1\}) \\
\chi(x) &= I(\{a < Z_1(x) < b\}),
\end{aligned} \tag{2}
$$

with $Z_1(x) = \exp(x - t_1/2)$, $Z_2(y) = \exp(y - t_2/2)$ applied to $x = X(t_1)$ and $y = X(t_2) - X(t_1)$.

Let us now suppose that t_1 and t_2 are not fixed in advance, but are to be sought in order to minimize the average total loss of the form

$$\pi_0 P(\theta_0) + \pi_1(1 - P(\theta_1)) + \pi_0 c_0 N(\theta_0) + \pi_1 c_1 N(\theta_1), \tag{3}$$

where $N(\theta) = t_1 + t_2 C(\theta)$ is the average "sample number" in the two-stage experiment, given θ.

For any fixed t_1 and t_2 the solution is given by test (2), and the problem turns out to be essentially numerical: to find (t_1, t_2) giving a minimum to (3) using the optimal test in the form of (2). In [4], there is a discussion of the results of evaluation of such tests.

As usual, we can use the solution to the optimality problem (3) to find a minimum of the average "sample size" over all the tests with error probabilities not exceeding some α and β: $P(\theta_0) \le \alpha$ and $1 - P(\theta_1) \le \beta$ for some $0 < \alpha < 1$, $0 < \beta < 1$.

Let us suppose that there exist π_0, π_1, c_0, c_1 such that for the test giving a minimum to (3) the equalities $P(\theta_0) = \alpha$ and $1 - P(\theta_1) = \beta$ hold. Then the "weighted average sample number" $\pi_0 c_0 N(\theta_0) + \pi_1 c_1 N(\theta_1)$ of this test is obviously minimum among all the (two-stage) tests with $P(\theta_0) \le \alpha$ and $1 - P(\theta_1) \le \beta$.

From now on, we will suppose that π_0, π_1, c_0, c_1 are such that there exists a unique test (2) giving a minimum to (3), and that for this test the equalities $P(\theta_0) = \alpha$ and $1 - P(\theta_1) = \beta$ hold.

The purpose of what follows in this article is to make use of the results above for construction of asymptotically optimal two-stage tests for a broad class of discrete-time processes.

2. Asymptotically optimal two-stage tests for LAN experiments

In this section we will show how the results of the preceding section can be applied to construct asymptotically optimal tests for a rather broad class of locally asymptotically normal statistical experiments (LAN).

Let us suppose that we observe a discrete-time stochastic process

$$\{X_1, X_2 \dots X_n, \dots\}$$

with finite-dimensional distributions given by densities

$$f_\theta^{(n)}(X_1, X_2, \dots, X_n), \quad n = 1, 2, \dots$$

The aim of this section is to construct a test of $H_0 : \theta$ vs $H_1 : \theta + \epsilon$, with error probabilities α and β which asymptotically minimizes a weighted average sample number, as $\epsilon(> 0) \to 0$, for some broad class of stochastic processes.

Let us say that a statistical experiment $\{X_1, X_2 \dots X_n\}$ is locally asymptotically normal if for any $\epsilon > 0$ there exists $n = n(\epsilon)$ such that the likelihood ratio for two simple hypotheses θ and $\theta + \epsilon$

$$Z_\epsilon^n = \frac{f_{\theta+\epsilon}^n(X_1, X_2, \dots, X_n)}{f_\theta^n(X_1, X_2, \dots, X_n)}$$

converges weakly, when $X_1, \dots X_n$ follow the distribution with the parameter θ, to that of two normal distributions:

$$Z = \exp\{\xi - 1/2\},$$

where ξ is a standard normal random variable (cf., e.g., [1]).

We will need a slightly stronger condition which we will call *two-dimensional LAN*:

For any $t_1 > 0$ and $t_2 > 0$ the two-dimensional distribution of

$$(Z_\epsilon^{[t_1 n]}, Z_\epsilon^{[t_2 n]})$$

converges weakly to that of the random vector

$$(Z(t_1), Z(t_2)),$$

where $Z(t) = \exp\{W(t) - t/2\}$, where $W(t)$ a standard Wiener process (here and in what follows $[a]$ is the integer part of the number a).

An example of two-dimensional (and finite-dimensional of any order) LAN stochastic process can be found in [2].

From now on, we will suppose that the stochastic process under consideration is two-dimensionally LAN.

For any two-stage test

$$< \phi_1(X_1, X_2, \ldots, X_{n_1}), \phi_2(X_1, X_2, \ldots, X_{n_1+n_2}), \chi(X_1, X_2, \ldots, X_{n_1}) >$$

let us define as

$$\alpha(\phi_1, \phi_2, \chi) = E_\theta(\phi_1(1-\chi) + \phi_2\chi)$$

and

$$\beta(\phi_1, \phi_2, \chi) = 1 - E_{\theta+\epsilon}(\phi_1(1-\chi) + \phi_2\chi)$$

the probabilities of error of the first and the second kind, respectively, and also the average sample number given θ as

$$N_\theta(\phi_1, \phi_2, \chi) = n_1 + n_2 E_\theta \chi.$$

Theorem 2.1. *Let π_0, π_1, c_0, c_1 be such numbers that there exists a two-stage test (2) minimizing (3) for which the equalities $P(\theta_0) = \alpha$ and $1 - P(\theta_1) = \beta$ hold. Then the two-stage test taking $n_1 = [t_1 n(\epsilon)]$ observations at the first stage, and additional $n_2 = [t_2 n(\epsilon)]$ observations at the second stage and defined as*

$$\begin{aligned} \phi_1^\epsilon &= I(\{Z_\epsilon^{n_1} > \pi_0/\pi_1\}) \\ \phi_2^\epsilon &= I(\{Z_\epsilon^{n_1+n_2} > \pi_0/\pi_1\}) \\ \chi^\epsilon &= I(\{a < Z_\epsilon^{n_1} < b\}) \end{aligned} \tag{4}$$

has the following asymptotic properties:

$$\lim_{\epsilon \to 0} \alpha(\phi_1^\epsilon, \phi_2^\epsilon, \chi^\epsilon) = \alpha \tag{5}$$

$$\lim_{\epsilon \to 0} \beta(\phi_1^\epsilon, \phi_2^\epsilon, \chi^\epsilon) = \beta, \tag{6}$$

and

$$\lim_{\epsilon \to 0} N_\theta(\phi_1^\epsilon, \phi_2^\epsilon, \chi^\epsilon)/n(\epsilon) = t_1 + t_2 P(a < Z(t_1) < b)) \tag{7}$$

$$\lim_{\epsilon \to 0} N_{\theta+\epsilon}(\phi_1^\epsilon, \phi_2^\epsilon, \chi^\epsilon)/n(\epsilon) = t_1 + t_2 P(a < \frac{1}{Z(t_1)} < b)) \tag{8}$$

Proof. The theorem is rather straightforward if we note that due to the two-dimensional LAN condition the distributions of $(Z_\epsilon^{n_1}, Z_\epsilon^{n_1+n_2})$ defining test (4) converge weakly to the distribution $(Z_1, Z_1 Z_2)$ in test (2). Therefore its error probabilities converge to that of test (2) (which gives (5) and (6)), and so does the continuation probability. Because of that the average sample number $N_\theta(\phi_1^\epsilon, \phi_2^\epsilon, \chi^\epsilon)$ of the test, normalized by $n(\epsilon)$, tends to $N(\theta_0)$ of test (2), which is equal to the right-hand side of (7). Equality (8) is proven using the usual arguments of contiguity (see [3], for example). □

Theorem 2.1 does not state any kind of optimality with respect to the average sample number but its parameters are defined to minimize a weighted average sample number for the limit test (2), so there is a hope that the performance of this test is near to the optimal.

In [4], it is stated that, applied to independent and identically distributed LAN observations, the test (4) is asymptotically optimal.

We will now state that the same optimality holds in the case of Markov ergodic observations (under conditions of [2]).

Theorem 2.2. *Let the conditions* A1 – A4 *of* [2] *be fulfilled. Then test (4) defined as in Theorem 2.1 has the following asymptotic property:*

For any test $<\varphi_1, \varphi_2, \kappa>$ *for which*

$$\limsup_{\epsilon \to 0} \alpha(\varphi_1, \varphi_2, \kappa) \leq \alpha \tag{9}$$

$$\limsup_{\epsilon \to 0} \beta(\varphi_1, \varphi_2, \kappa) \leq \beta \tag{10}$$

it holds

$$\liminf_{\epsilon \to 0} \frac{(\pi_0 c_0 N_\theta(\varphi_1, \varphi_2, \kappa) + \pi_1 c_1 N_{\theta+\epsilon}(\varphi_1, \varphi_2, \kappa))}{(\pi_0 c_0 N_\theta(\phi_1, \phi_2, \chi) + \pi_1 c_1 N_{\theta+\epsilon}(\phi_1, \phi_2, \chi))} \geq 1 \tag{11}$$

Proof. The result can easily be obtained observing the structure of the optimal continuation rule.

Let us consider a two-stage test based on n_1 and n_2 observation on the first and the second stage, respectively. From Theorem 1.1 we have, that the optimal continuation rule is based on

$$g_n(z, X_1) = E_{\theta_0}\{(\pi_0 - \pi_1 z Z)^- | X_1\}$$

where

$$Z = Z(X_1, X_2, \dots, X_{n+1}) = \frac{\prod_{i=1}^n f_{\theta+\epsilon}(X_{i+1}|X_i)}{\prod_{i=1}^n f_\theta(X_{i+1}|X_i)}.$$

More precicely, the optimal test continues observations to the second stage if and only if

$$g_{n_2}(z, X_{n_1}) < (\pi_0 - \pi_1 z)^- - \pi_0 c_0 - \pi_1 c_1 z$$

where $z = Z_\epsilon^{n_1}$.

Now, the proof can be completed in a similar way to that of independent observations by noting that under the conditions of [2] $g_{n_2}(z, X_1)$ converges in distribution to $E(\pi_0 - \pi_1 z Z_2(W(t_2)))^-$ which is exactly the same as in the case of independent observations. □

The results of Theorem 2.1 and Theorem 2.2 can be considered as an advance towards optimality of sequential tests of close hypotheses for dependent observations, discussed in [3], in which case the main problem is to work around the uniform integrability of the stopping time. Theorem 2.2, in particular, gives some sort of optimality, which is an open, and seemingly very complex, problem for fully sequential tests, in any case which is different from independent identically distributed observations.

References

[1] Le Cam, L., Asymptotic methods in statistical decision theory, Springer Series in Statistics, Springer-Verlag, New York-Berlin (1986). .

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

[3] Novikov A., *Efficiency of sequential hypotheses testing*, Aportaciones Mat. Serie Comun. Soc. Mat. Mexicana **30** (2002), 71–79.

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

Departamento de Matemáticas, Universisdad Autónoma Metropolitana-Iztapalapa, San Rafael Atlixco #186, Col. Vicentina, 09340, México D.F., México

E-mail address: an@xanum.uam.mx