COMPSTAT

Physica-Verlag
A Springer Company

COMPSTAT

Proceedings
in Computational Statistics

16th Symposium Held in Prague,
Czech Republic, 2004

Edited by
Jaromir Antoch

With 151 Figures
and 38 Tables

Physica-Verlag
A Springer Company

Prof. Dr. Jaromir Antoch
Charles University Prague
Faculty of Mathematics and Physics
Department of Statistics and Probability
Sokolovska 83
18675 Prague 8 – Karlin
Czech Republic
antoch@karlin.mff.cuni.cz

ISBN 3-7908-1554-3 Physica-Verlag Heidelberg

Cataloging-in-Publication Data
Library of Congress Control Number: 2004108446

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Softcover-Design: Erich Kirchner, Heidelberg

SPIN 11015154 88/3130-5 4 3 2 1 0 – Printed on acid-free paper

OPTIMALITY OF TWO-STAGE HYPOTHESIS TESTS

Andrei Novikov

Key words: Hypothesis testing, two-stage test, optimal decision, local asymptotic normality.

COMPSTAT 2004 section: Optimization.

Abstract: This paper deals with optimal two-stage tests for two simple hypotheses. The structure of both the optimal decision rule and the optimal continuation rule is given. The results are applied to the optimal two-stage tests for a Wiener process with a linear drift, and to obtain an asymptotically optimal test for two close hypotheses in the case of locally asymptotically normal statistical experiment. The numerical results of comparison between the optimal Neyman-Pearson test, Wald's SPRT and the proposed optimal two-stage test are given.

1 The structure of an optimal two-stage test

In this section, we give the structure of an optimal two-stage test for two simple hypotheses.

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, 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 $[0, 1]$, interpreting them as follows: $\phi_1(x)$ 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 will be 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.

Theorem 1. *The minimum average total loss is equal to* π_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*

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

Proof. For any test $(\phi_1(x),\phi_2(x,y),\chi(x))$ let us represent the ATL$-\pi_1$ as

$$\begin{aligned}&\int l_1(x)\phi_1(x)(1-\chi(x))d\mu_1(x)\\ &+\int\left(\int l_2(x,y)\phi_2(x,y)\chi(x)d\mu_2(y)\right)d\mu_1(x)\\ &+\int(\pi_0 c_0 f_{\theta_0}(x)+\pi_1 c_1 f_{\theta_1}(x))\chi(x)d\mu_1(x)\end{aligned}\tag{2}$$

The first term in (2) is greater or equal than

$$\int l_1(x)I(\{l_1(x)<0\})(1-\chi(x))d\mu_1(x)\tag{3}$$

because $l_1(x)(\phi_1(x)-I(\{l_1(x)<0\}))(1-\chi(x))\geq 0$ for any $0\leq\phi_1(x)\leq 1$ (this is an almost literal repetition of the proof of the Neyman-Pearson's theorem).

The second term in (2) is greater or equal than

$$\int \left(\int l_2(x,y) I(\{l_2(x,y) < 0\}) \chi(x) d\mu_2(y) \right) d\mu_1(x), \tag{4}$$

because $l_2(x,y)(\phi_2(x,y) - I(\{l_2(x,y) < 0\}))\chi(x) \geq 0$ for any $0 \leq \phi_2(x,y) \leq 1$.

So from (2-4) we have that the (2) is greater or equal than

$$\begin{aligned} &\int l_1(x) I(\{l_1(x) < 0\})(1 - \chi(x)) d\mu_1(x) \\ &+ \int \left(\int l_2(x,y) I(\{l_2(x,y) < 0\}) \chi(x) d\mu_2(y) \right) d\mu_1(x) \\ &+ \int (\pi_0 c_0 f_{\theta_0}(x) + \pi_1 c_1 f_{\theta_1}(x)) \chi(x) d\mu_1(x) \end{aligned} \tag{5}$$

which is equal to

$$\begin{aligned} &\int l_1(x)^- d\mu_1(x) \\ &+ \int \left(\int l_2(x,y)^- d\mu_2(y) - l_1(x)^- + \pi_0 c_0 f_{\theta_0}(x) + \pi_1 c_1 f_{\theta_1}(x) \right) \\ &\chi(x) d\mu_1(x) \end{aligned}$$

and in the same way as above this is greater or equal than

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

which proves the first affirmation of Theorem 1.

The second one is immediate in view of the above proof (any step in it does not take to an inequality if the functions ϕ_1, ϕ_2 and χ are defined as in (1)), so test (1) is the optimal one.

2 Testing hypotheses about a drift of a Wiener process

2.1 The structure of the optimal two-stage test

Let us assume that we observe a Wiener process with a linear drift $W(t) + \theta t$. 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 some $\theta \neq 0$.

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.

So, in terms of the above section, denoting by $\varphi(x)$ the standard normal probability density function we have:

$$f_\theta(x) = f_\theta^1(x) = \frac{1}{\sqrt{t_1}}\varphi(\frac{x-\theta t_1}{\sqrt{t_1}}), \quad f_\theta(x,y) = f_\theta^1(x)f_\theta^2(y)$$
$$f_\theta^2(y) = \frac{1}{\sqrt{t_2}}\varphi(\frac{x-\theta t_2}{\sqrt{t_2}})$$

the two components (X, Y) being independent.

By Theorem 1, for any given $\pi_0, \pi_1, c_0, c_1, t_1, t_2$ the optimal two-stage test is given by

$$\begin{aligned}
\phi_1(x) &= I(\{Z_1(x) > \pi_0/\pi_1\}) \\
\phi_2(x,y) &= I(\{Z_1(x)Z_2(y) > \pi_0/\pi_1\}) \\
\chi(x) &= I(\{E\{(\pi_0 - \pi_1 Z_1(X)Z_2(Y))^-|X = x\} \\
&\quad -(\pi_0 - \pi_1 Z_1(x))^- + \pi_0 c_0 + \pi_1 c_1 Z_1(x) < 0\})
\end{aligned}$$

where $Z_1(x) = \exp(x\theta - \theta^2 t_1/2)$, $Z_2(y) = \exp(y\theta - \theta^2 t_2/2)$ are so-called likelihood ratios:

$$Z_1(x) = \frac{f_\theta^1(x)}{f_0^1(x)}, \quad Z_2(x) = \frac{f_\theta^2(y)}{f_0^2(y)}.$$

It is easy to see that the continuation rule is based on the function

$$g(z) = E_0(\pi_0 - \pi_1 z Z_2(Y))^-$$

and is equivalent to proceed to the second stage if and only if $z = Z_1(X)$ is such that

$$g(z) < (\pi_0 - \pi_1 z)^- - \pi_0 c_0 - \pi_1 c_1 z. \tag{6}$$

It is easy to observe that the function $g(z)$ is concave, and the right-hand side of (6) is piece-wise linear and concave, too. So if there are z satisfying (6), it is equivalent to $a < z < b$ with some a, b such that $a < \pi_0/\pi_1 < b$.

Due to this fact, it is obvious that the optimal two-stage test has the form:

$$\begin{aligned}
\phi_1(x) &= I(\{Z_1(x) > \pi_0/\pi_1\}) \\
\phi_2(x,y) &= I(\{Z_1(x)Z_2(y) > \pi_0/\pi_1\}) \\
\chi(x) &= I(\{a < Z_1(x) < b\}),
\end{aligned} \tag{7}$$

or, in other words, the optimal rule says:

1. Observe X. Stop observations at this stage if $Z_1(X) < a$ (accepting H_0), or if $Z_1(X) > b$ (rejecting H_0); continue observing otherwise.

2. At the second stage, obtain Y. Accept H_0 if

   $$Z_1(X)Z_2(Y) < \pi_0/\pi_1$$

   and reject it otherwise.

It is interesting to note that of c_0 and/or c_1 (the costs of additional observations) are sufficiently large, there are no solutions to (6), so the optimal test will stop at the first stage.

2.2 Comparison between optimal tests

Let us now pose a more realistic problem in relation with the two-stage tests.

Let us 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

$$\pi_0 P(\theta_0) + \pi_1(1 - P(\theta_1)) + \pi_0 c_0 N(\theta_0) + \pi_1 c_1 N(\theta_1), \tag{8}$$

say, 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 the above test, and the problem turns to be essentially numerical: to find (t_1, t_2) giving a minimum to (8) using the optimal test in the form of (7), say. It is obvious that test (7) has four parameters (t_1, t_2, a, b), and the minimum ATL in (8) can be calculated minimizing over all of them. Properly saying, parameters a and b are uniquely defined by (6) for any t_1, t_2, but there is no explicit way to calculate them, so we prefer to optimize over them as well, which is an equivalent procedure due to the results above.

We developed a program module (unit, in terms of Borland Pascal 6.0) for numerical optimization of (8), given any π_0,π_1,c_0,c_1. This module is available from the author.

Below we present some results of evaluation of optimal tests.

The most appropriate context of such an evaluation seems to be a comparison between different competing tests including the optimal two-stage test above.

So we compare the optimal two-stage test with the classic Neyman - Pearson and Wald's test, similar to [1]. Obviously, an optimal test (7) with $P(\theta_0) = \alpha$ and $1 - P(\theta_1) = \beta$, would minimize

$$\pi_0 c_0 N(\theta_0) + \pi_1 c_1 N(\theta_1)$$

among all the (two-stage) tests with error probabilities of the first and second kind not exceeding α and β, respectively. So it is interesting to compare its average sample number(ASN) $N(\theta_0)$ and $N(\theta_1)$ with the ASN of the Neyman-Pearson and Wald's test (see [1], see also [4]) .

The following table contains the respective characteristics of the three competing tests for a series of $\alpha = \beta$ evaluated for the null hypothesis $\theta = 0$ against the alternative $\theta = 1$. The numbers in the respective columns are the ASN of the three tests which correspond to the same level of $\alpha = \beta$ indicated in column "α".

The results above give a very clear evidence that two-stage tests have rather competitive properties concerning the average sample number.

The following table gives an idea about the parameters of the respective optimal two-stage test. Because $Z_1(x)$ and $Z_2(y)$ in (7) are monotone functions of x and y respectively, the parameters of the two-stage test are given in terms of x and y rather than in terms of Z_i. For example, to

α	**Wald**	**Neyman-Pearson**	**Two-Stage**
0.0072	9.71	23.94	15.17
0.0150	8.12	18.83	12.48
0.0231	7.15	15.91	10.84
0.0313	6.44	13.87	9.66
0.0396	5.87	12.32	8.73
0.0480	5.40	11.09	7.97
0.0563	5.00	10.07	7.33
0.0646	4.65	9.20	6.77
0.0729	4.34	8.46	6.28
0.0811	4.07	7.81	5.85
0.1563	2.32	4.08	3.22
0.2161	1.46	2.47	2.00

Table 1: Average sample number.

achieve $\alpha = \beta = 0.0072$, you have first to observe the process up to the time 11.54 (t_1), then if the value x of the process at that time is less than 2.58 (a), then accept H_0 and stop observing. If x is greater than 8.96 (b) then stop observing as well, and reject H_0. Otherwise keep observing for 16.53 time units more (t_2), obtaining the value y of the process at the end of this period. Based on this, accept H_0 if $y < 14.03$ (c) and reject H_0 otherwise.

α	a	b	c	t_1	t_2
0.0072	2.58	8.96	14.03	11.54	16.53
0.0150	2.01	7.29	11.08	9.30	12.87
0.0231	1.66	6.31	9.38	7.97	10.79
0.0313	1.41	5.62	8.20	7.03	9.36
0.0396	1.22	5.09	7.29	6.31	8.28
0.0480	1.06	4.66	6.57	5.72	7.42
0.0563	0.93	4.30	5.97	5.23	6.72
0.0646	0.82	3.99	5.47	4.81	6.12
0.0729	0.72	3.72	5.03	4.44	5.62
0.0811	0.64	3.49	4.65	4.12	5.18
0.1563	0.16	2.06	2.45	2.22	2.67
0.2161	-0.03	1.39	1.49	1.36	1.61

Table 2: The optimal two-stage test.

3 Asymptotically optimal two-stage tests for LAN experiments

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

Let us say that a statistical experiment $\{X_1, X_2 \dots X_n\}$ with independent and identically distributed observations 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 $\theta + \epsilon$

$$Z_\epsilon^n = \prod_{i=1}^n f_{\theta+\epsilon}(X_i)/f_\theta(X_i)$$

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., [2]).

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 $\epsilon \to 0$.

Theorem 2. *Let π_0, π_1, c_0,c_1 be such numbers that there exists a two-stage test (7) minimizing (8) with $P(\theta_0) = \alpha$ and $1 - P(\theta_1) = \beta$. 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*

$$\begin{aligned} \phi_1 &= I(\{Z_\epsilon^{n_1} > \pi_0/\pi_1\}) \\ \phi_2 &= I(\{Z_\epsilon^{n_1+n_2} > \pi_0/\pi_1\}) \\ \chi &= I(\{a < Z_\epsilon^{n_1} < b\}) \end{aligned} \tag{9}$$

is asymptotically optimal in the sense that it minimizes

$$\lim_{\epsilon \to 0} (\pi_0 c_0 N_\epsilon(\theta) + \pi_1 c_1 N_\epsilon(\theta + \epsilon))/n(\epsilon)$$

in the class of all two-stage tests whose error probabilities of the first and the second kind asymptotically do not exceed α and β, respectively.

Proof is rather straightforward if we note that due to the LAN condition and independence of the observations the distributions of $Z_\epsilon^{n_1}$ and $Z_\epsilon^{n_2}$ defining test (9) converge weakly to the distribution of Z_1 and Z_2 in test (7), so its error probabilities converge to that of test (7), and so the continuation probability, and thus the average sample number $N_\epsilon(\theta)$ of the test in Theorem 2 normalized by $n(\epsilon)$ tends to $N(\theta_0)$ of test (7). The rest of the proof is due to the optimality of (7).

Another promising application of two-stage tests seems to be the construction of a test similar to that of Theorem 2 for statistical experiments with Markov dependent observations (see [3]), in which case the likelihood ratio behaves exactly the same way as in the case of independent observations. Unfortunately, a proof as in Theorem 2 does not proceed, because for dependent observations the structure of continuation rule is not as simple as in (7) any more. So the problem of finding an optimal sequential test for non-independent observations is still open even in the simplest case of two-stage tests.

References

[1] Aivazjan S.A. (1959). *A comparison of the optimal properties of the Neyman-Pearson and the Wald sequential probability ratio test.* Theory Prob. Appl. **4**, 86 – 93.

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

[3] Novikov A. (2001). *Uniform asymptotic expansion of likelihood ratio for Markov dependent observations.* Ann. Inst. Statist. Math. **53** (4), 799 – 809.

[4] Novikov A. (2002). *Efficiency of sequential hypotheses testing.* Aportaciones matemáticas. Serie Comunicaciones **30** 71 – 79.

Acknowledgement: My deep gratitude to Mexico and its wonderful people.

Address: A. Novikov, Universidad Autonoma Metropolitana - Unidad Iztapalapa, Departamento de Matematicas, San Rafael Atlixco #186, col. Vicentina, C.P. 09340, Mexico D.F., Mexico

E-mail: an@xanum.uam.mx