ON TESTING STATISTICAL HYPOTHESES

BY KAMEO MATUSITA, YUKIO SUZUKI and HIROSI HUDIMOTO

(Received Sept. 5, 1954)

1. Introduction

The purpose of this paper is to give some treatment of the problem of testing a composite hypothesis against a simple or composite alternative. In testing a hypothesis, it is very desirable to know bounds of errors, because, even if the most powerful test is adopted, it is often that we cannot obtain the exact value of the power of the test. However, we can easily get them by means of the affinity which was introduced in [1]. Thus far, A. Wald, E. L. Lehmann, C. Stein and other authors treated the problem of testing the composite hypothesis by introducing a measure in the set of probability distributions which makes the composite hypothesis. Especially, Lehmann and Stein [3], [4] reduced the composite hypothesis to a simple one under some conditions. In this paper we also set some conditions on composite hypotheses, which would be natural, and our idea of formulating hypotheses is applicable to non-parametric cases. The bounds of possible errors of our test are easily obtained, although it is not always most powerful. Further, we can do without introducing any measure in the set of probability distributions. Our treatment runs along the line of papers [1], [2].

2. Definitions and fundamental lemmas

First we give the fundamental lemma of Neyman and Pearson for the sake of completeness of description.

LEMMA 1. (Neyman and Pearson) *Let $f(x)$ and $g(x)$ be two probability density functions defined in a space R and let k be a constant and W a region in R such that*

$$(2.1)\qquad \begin{aligned} kf(x) &\leqq g(x) \quad \text{in } W \\ kf(x) &\geqq g(x) \quad \text{in } R-W \end{aligned}$$

Then, if W' is a region such that

$$(2.2)\qquad \int_{W'} f(x)\,dx = \int_W f(x)\,dx\,,$$

we obtain

$$\int_{W'} g(x)\,dx \leqq \int_W g(x)\,dx\,.$$

LEMMA 2. *Let* $\int_R \sqrt{f(x)(x)}\,dx = \rho$ *and* W *the same region as in lemma 1. Then we obtain*

$$\text{(2.3)}\qquad \begin{aligned} &\int_W f(x)\,dx \leqq \frac{\rho}{\sqrt{k}} \\ &\int_W g(x)\,dx \geqq 1-\sqrt{k}\,\rho \end{aligned}$$

PROOF: From (2.1) we have

$$\int_W f(x)\,dx \leqq \int_W \sqrt{f(x)\cdot\frac{g(x)}{k}}\,dx \leqq \frac{1}{\sqrt{k}}\int_R \sqrt{f(x)g(x)}\,dx.$$

Therefore

$$\int_W f(x)\,dx \leqq \frac{\rho}{\sqrt{k}}$$

Similary

$$\begin{aligned} \int_W g(x)\,dx &= 1-\int_{R-W} g(x)\,dx \geqq 1-\int_{R-W} \sqrt{kf(x)g(x)}\,dx \\ &\geqq 1-\sqrt{k}\int_R \sqrt{f(x)g(x)}\,dx \end{aligned}$$

Therefore

$$\int_W g(x)\,dx \geqq 1-\sqrt{k}\,\rho$$

Now, we define a metric in the whole set Ω of absolutely continuous probability distributions as follows:

$$\text{(2.4)}\qquad d(F_1, F_2) = \| F_1 - F_2 \| = \left\{\int_R \left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 dx\right\}^{\frac{1}{2}}$$

where F_1 and F_2 are any two elements with density functions $f_1(x)$, $f_2(x)$, respectively (see [1]). For two sets $\omega_0 = \{F\}$ and $\omega_1 = \{G\}$ of Ω, we define the distance as

$$\text{(2.5)}\qquad d(\omega_0, \omega_1) = \inf_{\substack{F\in\omega_0 \\ G\in\omega_1}} d(F, G)$$

Then, we obtain the following

LEMMA 3. *Let* H_0 *be the hypothesis that the random variable* $(X_1, \cdots, X_n)$ *is distributed according to a distribution function* F *in* ω_0, *and let H_1 be the alternative that $(X_1, \cdots, X_n)$ is distributed according to a distribution function G in ω_1. Let $d(\omega_0, \omega_1)=d_0>0$. If there exist F_0 in ω_0 and G in ω_1 such that*

$$d(F_0, G_0)=d_0 \tag{2.6}$$

$$F_0(W) \geqq F(W), \qquad G_0(W) \leqq G(W) \tag{2.7}$$

for some $W=\{(x_1, \cdots, x_n);\ kf_0(x_1, \cdots, x_n) \leqq g_0(x_1, \cdots, x_n)\}$, and for any element F in ω_0 and any element G in ω_1, where $f_0(x_1, \cdots, x_n)$ and $g_0(x_1, \cdots, x_n)$ are density functions of F_0 and G_0, respectively, then we can use W as a critical region for testing H_0 against H_1 and we obtain

$$F(W) \leqq \frac{1}{\sqrt{k}}\rho_0 \tag{2.8}$$

$$G(W) \leqq 1-\sqrt{k}\,\rho_0 \tag{2.9}$$

for any F in ω_0 and G in ω_1, where ρ_0 is the affinity between F_0 and G_0.

This lemma is simply showed by lemmas 1 and 2.

Generally, for testing efficiently a hypothesis H_0 that the true distribution is contained in ω_0 against a hypothesis that the true distribution is contained in ω_1 on the basis of a finite number of observations, it is necessary that ω_0 and ω_1 are discriminated from each other so that $d(\omega_0, \omega_1)>0$. To discriminate ω_0, ω_1 from each other, we want to employ F_0, G_0 which satisfy (2.6) and (2.7) as discriminating distributions. Then, the above lemma serves as fundamental. In the continuous case, we can take Gaussian distributions, for instance, as F_0 and G_0. This is why we give examples concerning Gaussian distributions in the following section.

3. Examples

PROBLEM 1. *Let $(X_1, \cdots, X_n)$ be a sample from a Gaussian population with unknown mean ξ and variance 1. Test the hypothesis $H_0: \{\xi \leqq -\varepsilon\}$ against $H_1: \{\xi \geqq \varepsilon\}$.*

Under hypothesis H_0 the joint probability density function of $X_1, \cdots, X_n$ is

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\xi_0)^2}{2}} \quad \text{where} \quad \xi_0 \leqq -\varepsilon. \tag{3.1}$$

and under hypothesis H_1 it is

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\xi_1)^2}{2}} \quad \text{where} \quad \xi_1 \geqq \varepsilon. \tag{3.2}$$

Then the affinity $\rho(\xi_0, \xi_1)$ between (3.1) and (3.2) is

$$\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\int_{R_n} e^{-\frac{\sum\{(x_i-\xi_0)^2+(x_i-\xi_1)^2\}}{4}}dx_1\cdots dx_n$$

$$=\prod_{i=1}^{n}\left(\frac{1}{2\pi}\right)^{\frac{1}{2}}\int_{-\infty}^{\infty} e^{-\frac{\left(x_i-\frac{\xi_0+\xi_1}{2}\right)^2+\left(\frac{\xi_0-\xi_1}{4}\right)^2}{2}}dx_i$$

$$=e^{-\frac{n(\xi_0-\xi_1)^2}{8}}.$$

Therefore, when $\xi_0=-\varepsilon$ and $\xi_1=\varepsilon$, $\rho(\xi_0, \xi_1)$ assumes its maximum value and consequently the distance $d(\xi_0, \xi_1)$ is minimum.

The most powerful critical region for testing $N(-\varepsilon, 1)$ against $N(+\varepsilon, 1)$ is given by

$$k\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i+\varepsilon)^2}{2}}\leqq\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\varepsilon)^2}{2}}$$

that is,

$$e^{\frac{\sum_{i=1}^{n}(x_i+\varepsilon)^2-(x_i-\varepsilon)^2}{2}}\geqq k$$

or

$$2\varepsilon\sum_{i=1}^{n}x_i\geqq\log k.$$

When we denote this critical region by W, it is obvious that the condition (2.7) is satisfied for a suitable k. Therefore, we can use W for such k as the critical region for testing H_0 against H_1 and the errors are bounded as follows:

$$\text{size of } W\leqq\frac{\rho(-\varepsilon, \varepsilon)}{\sqrt{k}}=\frac{1}{\sqrt{k}}e^{-\frac{n\varepsilon^2}{2}},$$

$$1-\text{power of } W\leqq\sqrt{k}\,\rho(-\varepsilon, \varepsilon)=\sqrt{k}\,e^{-\frac{n\varepsilon^2}{2}}.$$

Now, we want to remark that we can treat quite in the same way H_0' and H_1' which consist of all distribution functions satisfying the condition (2.7), respectively.

PROBLEM 2. *Let* $(X_1, \cdots, X_n)$ *be a sample from a Gaussian population with unknown mean* ξ *and variance 1. Test the hypothesis* H_0: $\{|\xi|\leqq\eta_0\}$ *against* H_1: $\{|\xi|\geqq\eta_1\}$, *where* $\eta_1-\eta_0=\varepsilon>0$.

Under hypotheses H_0 and H_1, the joint probability density functions of $X_1, \cdots, X_n$ are

(3.3) $$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\xi_0)^2}{2}} \qquad (|\xi_0| \leqq \eta_0)$$

(3.4) $$\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\xi_1)^2}{2}} \qquad (|\xi_0| \geqq \eta_1)$$

respectively. The affinity $\rho(\xi_0, \xi_1)$ between (3.3) and (3.4) is $e^{-\frac{n(\xi_0-\xi_1)^2}{8}}$. Therefore, it attains its maximum value ρ_0 at $\xi_0=\eta_0$, $\xi_1=\eta_1$ and $\xi_0=-\eta_0$, $\xi_1=-\eta_1$. The most powerful critical region W_1 for testing $N(\eta_0, 1)$ against $N(\eta_1, 1)$ is easily obtained by lemma 1 as follows:

$$e^{-\frac{\sum(x_i-\eta_1)^2}{2}} \geqq k\, e^{-\frac{\sum(x_i-\eta_0)^2}{2}}$$

$$\sum(\eta_1-\eta_0)(2x_i-\eta_0-\eta_1) \geqq 2\log k$$

$$2\sum x_i \geqq \frac{2\log k}{\varepsilon}+n(\eta_0-\eta_1)$$

Similarily we get the most powerful critical region W_2 for testing $N(-\eta_0, 1)$ against $N(-\eta_1, 1)$, that is,

$$2\sum x_i \leqq -\left[\frac{2\log k}{\varepsilon}+n(\eta_0+\eta_1)\right]$$

Therefore, we may take the set-theoretical sum W of W_1 and W_2 for the critical region for testing H_0 against H_1. Then

$$\text{size of } W \leqq \frac{2}{\sqrt{k}}\rho_0 = \frac{2}{\sqrt{k}} e^{-\frac{n\varepsilon^2}{8}}$$

$$\text{power of } W \geqq 2(1-\sqrt{k}\,\rho_0) = 2\left(1-\sqrt{k}\, e^{-\frac{n\varepsilon^2}{8}}\right)$$

PROBLEM 3. *When the population is Gaussian $N(0, \sigma^2)$ with mean 0 and unknown variance σ^2, test the hypothesis H_0: $\{0<\sigma^2 \leqq a^2\}$ against the alternative H_0; $\{b^2 \leqq \sigma^2 < c^2\}$ where a and b are given numbers such that $0 < a < b$.*

According to hypotheses H_0 and H_1, the joint probability density functions of sample $(X_1, \cdots, X_n)$ are respectively.

(3.5) $$\left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{\sum x_i^2}{2\sigma_0^2}} \qquad (0<\sigma_0^2 \leqq a^2)$$

(3.6) $$\left(\frac{1}{2\pi\sigma_1^2}\right)^{\frac{n}{2}} e^{-\frac{\sum x_i^2}{2\sigma_1^2}} \qquad (b^2 \leqq \sigma_1^2)$$

The affinity $\rho(\sigma_0, \sigma_1)$ between (3.5) and (3.6) is $\left(\frac{2\sigma_0\sigma_1}{\sigma_0^2+\sigma_1^2}\right)^{\frac{n}{2}}$ as is easily calculated. Also we can easily verify that $\rho(\sigma_0, \sigma_1)$ attains its maximum value at $\sigma_0^2=a^2$, $\sigma_1^2=b^2$.

The most powerful critical region W for testing $N(0, a^2)$ against $N(0, b^2)$ is given by

$$k\left(\frac{1}{2\pi a^2}\right)^{\frac{n}{2}} e^{-\frac{\sum x_i^2}{a^2}} \leqq \left(\frac{1}{2\pi b^2}\right)^{\frac{n}{2}} e^{-\frac{\sum x_i^2}{b^2}}$$

that is,

$$\sum x_i^2 \geqq \frac{2a^2b^2(\log k + n\log b - n\log a)}{b^2-a^2}.$$

Since we can easily verify that the condition (2.7) in lemma 3 is satisfied by taking k suitably, we can use W as the critical region for testing H_0 against H_1. Then

$$\text{size of } W \leqq \frac{1}{\sqrt{k}}\left(\frac{2ab}{a^2+b^2}\right)^{\frac{n}{2}}$$

$$\text{power of } W \geqq 1-\sqrt{k}\left(\frac{2ab}{a^2+b^2}\right)^{\frac{n}{2}}$$

PROBLEM 4. *When the population is Gaussian* $N(\xi, \sigma^2)$, *test the hypothesis* H_0: $\{\xi \leqq c,\ 0 < \sigma^2 \leqq a^2\}$ *against the alternative* H_1: $\{\xi \geqq d,\ \sigma^2 = b^2\}$, *where* a, b, c, *and* d *are given numbers and* $b > a$, $d-c=\varepsilon > 0$.

Under the hypothesis H_0 and the alternative H_1 we have the joint probability density functions of $X_1, \cdots, X_n$

$$\left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-\xi_0)^2}{2\sigma_0^2}} \qquad (0 < \sigma_0 \leqq a^2,\ \xi \leqq c) \tag{3.7}$$

and

$$\left(\frac{1}{2\pi\sigma_1^2}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-\xi_1)^2}{2\sigma_1^2}} \qquad (\sigma_1^2 = b^2,\ \xi \geqq d)\ . \tag{3.8}$$

The affinity $\rho(\xi_0, \sigma_0; \xi_1, b)$ between (3.7) and (3.8) is

$$\left(\frac{2\sigma_0 b}{\sigma_0^2+b^2}\right)^{\frac{n}{2}} e^{-\frac{n(\xi_0-\xi_1)^2}{4(\sigma_0^2+b^2)}}\ .$$

As can easily be seen, $\rho(\xi_0, \sigma_0; \xi_1\ b)$ attains its maximum value at $\xi_0=c$, $\xi_1=d$, and $\sigma_0^2=a^2$.

The most powerful critical region W for testing $N(c, a^2)$ against $N(d, b^2)$ is obtained in the following form

$$\left(\frac{1}{2\pi b^2}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-d)^2}{2b^2}} \geqq k\left(\frac{1}{2\pi a^2}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-c)^2}{2a^2}}$$

that is,

$$\sum_{i=1}^{n}\left\{\frac{(x_i-c)^2}{2a^2}-\frac{(x_i-d)^2}{2b^2}\right\} \geqq \log k + n\log\frac{b}{a}\ .$$

After a simple calculation, this becomes

$$\sum_{i=1}^{n}\left(x_i-\frac{b^2c-a^2d}{b^2-a^2}\right)^2 \geqq \frac{a^2b^2}{(b^2-a^2)^2}\left\{n(c-d)^2+2(b^2-a^2)\left(\log k+n\log\frac{b}{a}\right)\right\}.$$

and we get

$$\text{size of } W \leqq \frac{1}{\sqrt{k}}\rho(c,\ a^2;\ d,\ b^2)$$

$$\text{power of } W \geqq 1-\sqrt{k}\ \rho(c,\ a^2;\ d,\ b^2)$$

where $\rho(c,\ a^2;\ d,\ b^2)=\left(\frac{2ab}{a^2+b^2}\right)^{\frac{n}{2}} e^{-\frac{n\varepsilon^2}{4(a^2+b^2)}}$.

The condition (2.7) is satisfied for W, H_0 and H_1 when taking a proper value of k.

PROBLEM 5. *Let H_0 be the hypothesis that two independent random variables X_1, X_2 have the same Gaussian distribution, i.e., the distribution of $(X_1,\ X_2)$ has the density function of the form*

$$f(x_1,\ x_2)=\prod_{i=1}^{2} f(x_i)=\frac{1}{2\pi\sigma^2}\, e^{-\frac{\sum_{i=1}^{2}(x_i-m_0)^2}{2\sigma^2}}$$

Let H_1 be that $(X_1,\ X_2)$ the alternative hypothesis has any distribution function with density

$$g(x_1,\ x_2)=\frac{1}{2\pi\sigma^2}\, e^{-\frac{\sum_{i=1}^{2}(x_i-m_i)^2}{2\sigma^2}}$$

where the point $(m_1,\ m_2)$ lies more than δ_0 distant from any point $(m_0,\ m_0)$. Then, test H_0 against H_1 on the basis of a random sample of size n.

To treat this problem we employ the set

$$W=\left\{(x_{11},\ \cdots,\ x_{1n};\ x_{21},\ \cdots,\ x_{2n});\ k\leqq e^{\frac{1}{2\sigma^2}\sum_{i}^{2}\sum_{j}^{n}\{(x_{ij}-m_0)^2-(x_{ij}-m_i)^2\}}\right\}$$

as the critical region. The point of this set satisfies

$$\frac{\delta_0}{\sqrt{2}}+\frac{\sqrt{2}\,\sigma^2}{n\delta_0}\log k \leqq (\bar{x}_2-\bar{x}_1)$$

and vice versa, where

$$\bar{x}_1=\frac{x_{11}+\cdots+x_{1n}}{n}, \quad \bar{x}_2=\frac{x_{21}+\cdots+x_{2n}}{n}$$

As to the affinity ρ between two distributions with $f(x)$ and $g(x)$, the size ε of W, and the power η of W we have

$$\rho_{\max.}=e^{-\frac{n\delta_0^2}{8\sigma^2}}$$

$$(*) \qquad \varepsilon \leqq \frac{\rho_{\max.}}{\sqrt{k}}=\frac{1}{\sqrt{k}}e^{-\frac{n\delta^2}{8\sigma^2}}$$

$$1-\eta \leqq \sqrt{k}\,\rho_{\max.}=\sqrt{k}\,e^{-\frac{n\delta^2}{8\sigma^2}}$$

If we put $\delta_0=\frac{h\sigma}{\sqrt{n}}$, where h is any number, then (*) is rewritten as

$$\varepsilon \leqq \frac{1}{\sqrt{k}}e^{-\frac{h^2}{8}}$$

and

$$1-\eta \leqq \sqrt{k}\,e^{-\frac{h^2}{8}}$$

Therefore, when we wish to test the hypothesis of the same distribution against the alternative which is defined by the distribution functions more than $\delta_0=\frac{5\sigma}{\sqrt{n}}$ distant from any distribution function with density $\frac{1}{2\pi\sigma^2}e^{-\frac{\sum_{i=1}^{2}(x_i-m_0)^2}{2\sigma^2}}$ and moreover, when we wish $\varepsilon \leqq 0.05$, then putting $k=0.764$, $3.5\frac{\sigma}{\sqrt{n}} \leqq \bar{x}_2-\bar{x}_1$ defines a wanted critical region. In this case $1-\eta \leqq 0.0384$.

The following examples are taken from our experiment on effectiveness of the warm-shades made of the agricultural vinyle over Japanese cedars and Japanese cypresses. In autumn, 1953 we got the following data of the length of their young plants.

〈Japanese cedars〉

the sample mean *for the vinyle treatment* $\bar{x}_2=33.27$

the sample mean *for the control treatment* $\bar{x}_1=28.06$

$\bar{x}_2-\bar{x}_1=$ 5.21

$\frac{3.5s}{\sqrt{n}}=2.17$ unit: cm

where s is an unbiased estimate of σ

〈Japanese cypresses〉

the sample mean *for the vinyle treatment* $\bar{x}_2=23.62$

the sample mean *for the control treatment* $\bar{x}_1=19.17$

$\bar{x}_2-\bar{x}_1=$ 4.45

$\frac{3.5s}{\sqrt{n}}=1.79$ unit: cm

These results seem satisfactory to reject our hypothesis of the same distribution. Still more, this test is applicable to the case where the tail of any distribution in H_t which is nearer to the alternative distribution is covered by a Gaussian distribution for some suitable h, or any devised $\sigma_i(i=1, 2)$.

〈Case: $\varepsilon \leqq 0.05$〉

h	4	5	6	7	8	9
$\sqrt{k}$	2.707	0.874	0.222	0.044	0.068	0.0008
$1-\eta\leqq$	0.365	0.0385	0.0024	0.001	0.000002	0.00000
$\frac{h^2+2\log k}{\sqrt{2h}}$	3.536	3.465	3.536	3.687	3.885	4.125

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Matusita, K.: On the theory of statistical decision functions, *Annals of the Institute of Statistical Mathematics*, Vol. 3, 1951.

[2] Matusita, K. and H. Akaike: Note on the decision problem, *Annals of the Institute of Statistical Mathematics*, Vol. IV, 1952.

[3] Lehmann, E. L. and C. Stein: Most powerful tests of composite hypotheses. I. Normal distributions, Annals of Mathematical Statistics, Vol. 19, 1948.

[4] Lehmann, E. L. and C. Stein: On the theory of some non-parametric hypotheses, *Annals of Mathematical Statistics*, Vol. 20, 1949.