Ann. Inst. Statist. Math.
28 (1976), Part A, 371-384

ON A NEW METHOD OF TESTING STATISTICAL HYPOTHESES

G. TRENKLER

(Received Apr. 26, 1973)

In [1] Matusita, Suzuki and Hudimoto have introduced a new concept of testing statistical hypotheses. The purpose of this paper is to improve the quality of their tests by strengthening some inequalities and using a better measure of distance between probability measures. Before stating the major theorem some important notations and lemmas will be needed.

Let Ω be a set, $\mathfrak{A}$ a σ-algebra on Ω and P, Q probability measures on $\mathfrak{A}$ dominated by a σ-finite measure λ on $\mathfrak{A}$. Further let

$$f=\frac{dP}{d\lambda} \tag{1}$$

$$g=\frac{dQ}{d\lambda} \tag{2}$$

denote the $[\lambda]$-uniquely determined Radon-Nikodym-densities. Suppose $E\in\mathfrak{A}$ and $t\in[0, 1]$. We define:

$$I_E(P, Q)(t)=\int_E f(x)^t g(x)^{1-t} d\lambda(x) \tag{3}$$

and in the case $E=\Omega$:

$$I(P, Q)(t)=I_\Omega(P, Q)(t) \tag{4}$$

$I(P, Q)(t)$ will be called the distance-generating-function. Obviously we have

$$\bigwedge_{E\in\mathfrak{A}}^{E} \qquad \bigwedge_{t\in[0, 1]}^{t}$$

(i) $I_E(P, Q)(t)$ exists
(ii) $0\leqq I_E(P, Q)(t)\leqq 1$
(iii) $I_E(P, Q)(t)$ is independent of the choice of the dominating measure λ.
Let $E_i\in\mathfrak{A}$, $i=1,\cdots, n$ and put

$$P^{(n)}\left(\mathop{\times}_{i=1}^{n} E_i\right)=\prod_{i=1}^{n} P(E_i) \tag{5}$$

(5) yields a uniquely determined probability measure on

$$\mathfrak{A}^{(n)}=\sigma\left(\left\{\mathop{\times}_{i=1}^{n} E_i \mid E_i \in \mathfrak{A},\ i=1,\cdots,n\right\}\right).$$

In a similar way we find $Q^{(n)}$ and the dominating σ-finite measure $\lambda^{(n)}$.

Now we can easily see that

$$I(P^{(n)}, Q^{(n)})(t)=(I(P,Q)(t))^n. \tag{6}$$

We proceed now to our first lemma.

LEMMA 1. *Suppose* $\vec{X}=(X_1,\cdots,X_m)^T$, $\vec{Y}=(Y_1,\cdots,Y_m)^T$ *be two multinormally distributed random variables with the properties:*

$$\mathrm{E}(\vec{X})=\vec{\mu} \tag{7}$$

$$\mathrm{E}(\vec{Y})=\vec{\rho}. \tag{8}$$

We assume that the covariance matrices of $\vec{X}$ *and* $\vec{Y}$ *are equal and positive definite. It will be denoted by* V. *The following abbreviations are appropriate:*

$$\vec{X}\sim N(\vec{\mu}, V) \tag{9}$$

$$\vec{Y}\sim N(\vec{\rho}, V). \tag{10}$$

Then

$$f(\vec{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^m \frac{1}{\sqrt{\det V}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})\right) \tag{11}$$

$$g(\vec{y})=\left(\frac{1}{\sqrt{2\pi}}\right)^m \frac{1}{\sqrt{\det V}} \exp\left(-\frac{1}{2}(\vec{y}-\vec{\rho})^T V^{-1}(\vec{y}-\vec{\rho})\right) \tag{12}$$

are the density functions of $\vec{X}$ *and* $\vec{Y}$ *respectively. If* P *and* Q *are the corresponding probability measures (on the* σ*-algebra of Borel-sets* $\mathfrak{B}^m$ *on* $\boldsymbol{R}^m$*), we have:*

$$I(P,Q)(t)=\exp\left(-\frac{1}{2}t(1-t)(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})\right). \tag{13}$$

PROOF. (11) and (12) are well-known facts, (13) follows by simple calculation.

LEMMA 2 (Neyman-Pearson). *Let* $E \in \mathfrak{A}$ *and* $k \in \boldsymbol{R}^+$ *have the following properties:*

$$\bigwedge_{\substack{x \\ x \in E}} \quad k\cdot f(x) \leqq g(x) \tag{14}$$

$$\bigwedge_{\substack{x \\ x \in \bar{E}}} \quad k \cdot f(x) > g(x) . \tag{15}$$

Then we can conclude that

$$\bigwedge_{\substack{E^* \\ E^* \in \mathfrak{A} \\ P(E^*) = P(E)}} \quad Q(E^*) \leqq Q(E) . \tag{16}$$

PROOF. See for example Hogg-Craig [2], p. 262. The set E is often called critical region.

LEMMA 3. *Suppose* $E \in \mathfrak{A}$, $k \in \boldsymbol{R}^+$ *be as in Lemma* 2. *Then we have*:

$$\bigwedge_{\substack{t \\ t \in [0, 1]}}$$

$$\text{(i)} \quad P(E) \leqq \left(\frac{1}{k}\right)^{1-t} I(P, Q)(t) \tag{17}$$

$$\text{(ii)} \quad 1 - Q(E) \leqq k^t I(P, Q)(t) . \tag{18}$$

PROOF. Let $t \in [0, 1]$ be arbitrary. First we have

$$\bigwedge_{\substack{x \\ x \in E}} \quad f(x) \leqq f(x)^t \left(\frac{g(x)}{k}\right)^{1-t} \tag{19}$$

which gives

$$\int_E f(x) d\lambda(x) \leqq \int_E f(x)^t \left(\frac{g(x)}{k}\right)^{1-t} d\lambda(x) \tag{20}$$

or

$$P(E) \leqq \left(\frac{1}{k}\right)^{1-t} I_E(P, Q)(t) . \tag{21}$$

Since $I_E(P, Q)(t) \leqq I(P, Q)(t)$ it follows that

$$P(E) \leqq \left(\frac{1}{k}\right)^{1-t} I(P, Q)(t) . \tag{22}$$

From (15) we conclude that

$$\bigwedge_{\substack{x \\ x \in \bar{E}}} \quad g(x) \leqq k^t f(x)^t g(x)^{1-t} \tag{23}$$

which yields

$$\int_{\bar{E}} g(x) d\lambda(x) \leqq k^t \int_{\bar{E}} f(x)^t g(x)^{1-t} d\lambda(x) \tag{24}$$

or

$$Q(\bar{E}) \leqq k^t I_{\bar{E}}(P, Q)(t) \tag{25}$$

and hence

$$1 - Q(E) \leqq k^t I(P, Q)(t) . \tag{26}$$

Suppose now we have two nonempty families

$$\mathfrak{P} = \{P_i\}_{i \in I} , \qquad \mathfrak{Q} = \{Q_j\}_{j \in J}$$

of probability measures on $\mathfrak{A}$ dominated by the σ-finite measure λ. For the following considerations we propose a useful distance between these two families: Let $t \in [0, 1]$. Then

$$d_t(\mathfrak{P}, \mathfrak{Q}) = \sup_{(i,j) \in I \times J} I(P_i, Q_j)(t) \tag{27}$$

will be called the t-distance between $\mathfrak{P}$ and $\mathfrak{Q}$.

Remark. (27) is a generalization of the distance of Matusita, Suzuki and Hudimoto [1].

Clearly we have

$$\bigwedge_{\substack{t \\ t \in [0,1]}} \qquad 0 \leqq d_t(\mathfrak{P}, \mathfrak{Q}) \leqq 1 . \tag{28}$$

The preceding lemmas and definitions enable us to state the

THEOREM. *Suppose X be a random variable. On the basis of n independent observations $X_1, \cdots, X_n$ on X we now test the hypotheses:*
H: X has a probability distribution belonging to $\mathfrak{P}$
G: X has a probability distribution belonging to $\mathfrak{Q}$.
Suppose the following conditions are fulfilled:

$$\text{(i)} \quad \bigvee_{\substack{t_0 \\ t_0 \in [0,1]}} \qquad d_{t_0}(\mathfrak{P}, \mathfrak{Q}) > 0 \tag{29}$$

$$\text{(ii)} \quad \bigvee_{\substack{P_0 \\ P_0 \in \mathfrak{P}}} \ \bigvee_{\substack{Q_0 \\ Q_0 \in \mathfrak{Q}}} \qquad d_{t_0}(\mathfrak{P}, \mathfrak{Q}) = I(P_0, Q_0)(t_0) \tag{30}$$

(iii) *If*

$$f_0(x_1, \cdots, x_n) = \prod_{i=1}^{n} \frac{dP_0}{d\lambda}(x_i) \tag{31}$$

$$g_0(x_1, \cdots, x_n) = \prod_{i=1}^{n} \frac{dG_0}{d\lambda}(x_i) \tag{32}$$

(33) $$W=\{(x_1,\cdots,x_n)\mid k\cdot f_0(x_1,\cdots,x_n)\leqq g_0(x_1,\cdots,x_n)\}$$

and

(34) $$\bigvee_{\substack{k\\ k\in R^+}}\ \bigwedge_{\substack{i\\ i\in I}}\ \bigwedge_{\substack{j\\ j\in J}}\qquad P_i^{(n)}(W)\leqq P_0^{(n)}(W)\qquad Q_j^{(n)}(W)\geqq Q_0^{(n)}(W)$$

we can conclude:

(a) *W is a critical region (for the test H against G)*

(35) (b) $$\bigwedge_{\substack{i\\ i\in I}}\qquad P_i^{(n)}(W)\leqq\left(\frac{1}{k}\right)^{1-t_0}(I(P_0,Q_0)(t_0))^n$$

(36) (c) $$\bigwedge_{\substack{j\\ j\in J}}\qquad 1-Q_j^{(n)}(W)\leqq k^{t_0}(I(P_0,Q_0)(t_0))^n\,.$$

PROOF. It follows immediately from (i)–(iii) and Lemmas 3 and 4.

Remark. This theorem is a generalization of Lemma 3 [1], p. 134.

With the preceding theorem bounds for the error probabilities are available. We have:

(37) $$\Pr(\text{Error of the I. kind})\leqq P_0^{(n)}(W)\leqq\left(\frac{1}{k}\right)^{1-t_0}(I(P_0,Q_0)(t_0))^n$$

(38) $$\Pr(\text{Error of the II. kind})\leqq 1-Q_0^{(n)}(W)\leqq k^{t_0}(I(P_0,Q_0)(t_0))^n\,.$$

For practical purposes it is often required to construct a critical region W which satisfies the following conditions:

(39) (i) $$\bigwedge_{\substack{i\\ i\in I}}\qquad P_i^{(n)}(W)\leqq\alpha$$

(40) (ii) $$\bigwedge_{\substack{j\\ j\in J}}\qquad 1-Q_j^{(n)}(W)=\inf_{W'}(1-Q_j^{(n)}(W'))\qquad W'\text{ critical region}$$

where $0<\alpha<1$ is a given bound for Pr (Error of the I. kind).

Modifying these so called uniformly most powerful tests in view of the considerations from above we compute k (and hence W) from the following conditions (assuming the theorem is valid):

(41) (a) $k\cdot f_0(x_1,\cdots,x_n)\leqq g_0(x_1,\cdots,x_n)$

(42) (b) t_0 is chosen so that $t_0\in T_k$ and
$$k^{t_0}(I(P_0,Q_0)(t_0))^n=\inf_{t\in T_k}k^t(I(P_0,Q_0)(t))^n$$

where

$$T_k=\left\{t \mid t\in[0,1]\wedge\left(\frac{1}{k}\right)^{1-t}(I(P_0,Q_0)(t))^n=\alpha\right\}. \tag{43}$$

In the following two examples it will be seen that k is uniquely determined. Unfortunately there exist considerable difficulties in computing k if the random variable X is not normally distributed. This fact is due to the rather complicated form of $k^t I(P_0, Q_0)(t)$ in those cases. Even the normal case is troublesome if the variance is not known.

If one compares the testing principles developed in [1] with those just proposed one must admit that in the new proposal considerably smaller bounds for Pr (Error of the II. kind) are obtained.

Example 1. Let $\vec{X}=(X_1,\cdots,X_m)^T$ be a vector-valued multinormally distributed random variable with unknown mean $\vec{\theta}$ and known positive definite covariance matrix V. We will consider n independent observations $\vec{X}_1,\cdots,\vec{X}_n$ on $\vec{X}$ and pose the simple testing problem:

$$\begin{aligned} H:&\quad \vec{\theta}=\vec{\mu} \qquad \text{with } \vec{\mu}=(\mu_1,\cdots,\mu_m)^T \\ G:&\quad \vec{\theta}=\vec{\rho} \qquad \text{with } \vec{\rho}=(\rho_1,\cdots,\rho_m)^T. \end{aligned} \tag{44}$$

In this case we have

$$\mathfrak{P}=\{P\} \qquad \mathfrak{Q}=\{Q\} \tag{45}$$

$$\frac{dP}{d\lambda^{(m)}}(\vec{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^m\frac{1}{\sqrt{\det V}}\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})\right) \tag{46}$$

$$\frac{dQ}{d\lambda^{(m)}}(\vec{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^m\frac{1}{\sqrt{\det V}}\exp\left(-\frac{1}{2}(\vec{x}-\vec{\rho})^T V^{-1}(\vec{x}-\vec{\rho})\right) \tag{47}$$

where $\lambda^{(m)}$ is the Lebesgue-measure on the Borel-σ-algebra $\mathfrak{B}^{(m)}$. From Lemma 1 we conclude:

$$I(P^{(n)},Q^{(n)})(t)=\exp\left(-\frac{n}{2}t(1-t)(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})\right) \tag{48}$$

with the notations:

$$\frac{dP^{(n)}}{d\lambda^{(mn)}}(\vec{x}_1,\cdots,\vec{x}_n)=f(\vec{x}_1,\cdots,\vec{x}_n)=\prod_{i=1}^n\frac{dP}{d\lambda^{(m)}}(\vec{x}_i) \tag{49}$$

$$\frac{dQ^{(n)}}{d\lambda^{(mn)}}(\vec{x}_1,\cdots,\vec{x}_n)=f(\vec{x}_1,\cdots,\vec{x}_n)=\prod_{i=1}^n\frac{dQ}{d\lambda^{(m)}}(\vec{x}_i) \tag{50}$$

and

$$d_t(\mathfrak{P},\mathfrak{Q})=\exp\left(-\frac{t}{2}(1-t)(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})\right) \tag{51}$$

so that

(52) $$\bigwedge_{t \in [0,1]} d_t(\mathfrak{P}, \mathfrak{Q}) > 0 .$$

If we put

(53) $$P_0 = P , \qquad Q_0 = Q$$

(54) $$f_0 = f , \qquad g_0 = g$$

we have

(55) $$\bigwedge_{t \in [0,1]} d_t(\mathfrak{P}, \mathfrak{Q}) = I(P_0, Q_0)(t) .$$

To find a suitable t_0 and a suitable k we consider the inequality

(56) $$k \cdot f_0(\vec{x}_1, \cdots, \vec{x}_n) \leqq g_0(\vec{x}_1, \cdots, \vec{x}_n)$$

which is equivalent to (verified by straightforward computation):

(57) $$(\vec{\rho} - \vec{\mu})^T V^{-1} \sum_{i=1}^{n} \vec{x}_i \geqq \log k + \frac{n}{2} (\vec{\rho}^T V^{-1} \vec{\rho} - \vec{\mu}^T V^{-1} \vec{\mu}) .$$

The number

(58) $$c = \log k + \frac{n}{2} (\vec{\rho}^T V^{-1} \vec{\rho} - \vec{\mu}^T V^{-1} \vec{\mu})$$

will be called critical value for the test H against G. (58) is equivalent to

(59) $$k = \exp \left[c - \frac{n}{2} (\vec{\rho}^T V^{-1} \vec{\rho} - \vec{\mu}^T V^{-1} \vec{\mu}) \right] .$$

Thus we have

(60) $$\left(\frac{1}{k} \right)^{1-t} I(P^{(n)}, Q^{(n)})(t) = \exp \left\{ (1-t) \cdot \left[\frac{n}{2} (\vec{\rho}^T V^{-1} \vec{\rho} - \vec{\mu}^T V^{-1} \vec{\mu}) - c \right.\right.$$
$$\left.\left. - \frac{n}{2} t (\vec{\mu} - \vec{\rho})^T V^{-1} (\vec{\mu} - \vec{\rho}) \right] \right\} .$$

Demanding $t \in T_k$ we obtain

(61) $$c = -\frac{\log \alpha}{1-t} + \frac{n}{2} (\vec{\rho}^T V^{-1} \vec{\rho} - \vec{\mu}^T V^{-1} \vec{\mu}) - \frac{n}{2} t (\vec{\mu} - \vec{\rho})^T V^{-1} (\vec{\mu} - \vec{\rho})$$

which gives

(62) $$k^t I(P^{(n)}, Q^{(n)})(t) = \exp \left\{ t \cdot \left[-\frac{\log \alpha}{1-t} - \frac{n}{2} (\vec{\mu} - \vec{\rho})^T V^{-1} (\vec{\mu} - \vec{\rho}) \right] \right\} .$$

To determine the value of t_0 which minimizes $k^t I(P^{(n)}, Q^{(n)})(t)$ we minimize:

(63) $$\bar{H}(t) = -\frac{t}{1-t}\log\alpha - \frac{nt}{2}(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho}) .$$

Differentiating $\bar{H}(t)$ we obtain

(64) $$\frac{d\bar{H}(t)}{dt} = -\frac{1}{(1-t)^2}\log\alpha - \frac{n}{2}(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})$$

(65) $$\frac{d^2\bar{H}(t)}{dt^2} = -\frac{2}{(1-t)^3}\log\alpha .$$

As $t \in [0, 1]$, (64) and (65) confirm that

(66) $$t_0 = 1 - \sqrt{\frac{2\log(1/\alpha)}{n(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})}}$$

is a minimum of $\bar{H}(\cdot)$. (Of course n is to be chosen so that $t \in [0, 1)$). From (61) we derive that

(67) $$c = \sqrt{2n\log(1/\alpha)(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})} - n\vec{\mu}^T V^{-1}(\vec{\mu}-\vec{\rho}) .$$

Thus we accept G (reject H) if

(68) $$(\vec{\rho}-\vec{\mu})^T V^{-1}\sum_{i=1}^{n}\vec{x}_i \geq c$$

or equivalently

(69) $$\sum_{i=1}^{n}(\vec{\rho}-\vec{\mu})^T V^{-1}(\vec{x}_i-\vec{\mu}) \geq \sqrt{2n\log(1/\alpha)(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})} .$$

To size up the probability of the error of the II. kind we calculate the upper bound

$$k^{t_0} I(P^{(n)}, Q^{(n)})(t_0)$$

which yields

(70) $$1 - Q^{(n)}(W) \leq \exp\left(-\frac{1}{2}\left(\sqrt{n(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})} - \sqrt{2\log(1/\alpha)}\right)^2\right) .$$

We consider now two

Special cases:

(71) 1) $$V = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{pmatrix}$$

Omitting some steps of computation we have: Accept G if

$$\sum_{j=1}^{m}\frac{(\rho_j-\mu_j)(\bar{x}_{\cdot j}-\mu_j)}{\sigma_j^2}\geqq\sqrt{\frac{2}{n}\log\frac{1}{\alpha}\sum_{j=1}^{m}\left(\frac{\mu_j-\rho_j}{\sigma_j}\right)^2} \tag{72}$$

where

$$\bar{x}_{\cdot j}=\frac{1}{n}\sum_{i=1}^{n}x_{ij}\qquad j=1,\cdots,m \tag{73}$$

is the arithmetic mean taken over the columns of the observation matrix

$$((x_{ij}))_{\substack{i=1,\cdots,n\\ j=1,\cdots,m}}$$

and

$$1-Q^{(n)}(W)\leqq\exp\left(-\frac{1}{2}\left(\sqrt{n\sum_{j=1}^{m}\left(\frac{\mu_j-\rho_j}{\sigma_j}\right)^2}-\sqrt{2\log\frac{1}{\alpha}}\right)^2\right). \tag{74}$$

2) $m=1$

In this case we have the classical simple testing problem:

$$H:\ \mu=\mu_0\qquad\text{against}\qquad G:\ \mu=\mu_1\,. \tag{75}$$

If $\mu_1>\mu_0$, we accept G if

$$\bar{x}\geqq\mu_0+\sigma\sqrt{\frac{2}{n}\log\frac{1}{\alpha}}\,. \tag{76}$$

If $\mu_1<\mu_0$, we accept G if

$$\bar{x}\leqq\mu_0-\sigma\sqrt{\frac{2}{n}\log\frac{1}{\alpha}}\,. \tag{77}$$

($\bar{x}=(1/n)\sum_{i=1}^{n}x_i$ is the arithmetic mean of the observation.)
In both cases ($\mu_1<\mu_0$, $\mu_1>\mu_0$) we obtain

$$1-Q^{(n)}(W)\leqq\exp\left[-\frac{1}{2}\left(\sqrt{n}\,\frac{|\mu_0-\mu_1|}{\sigma}-\sqrt{2\log(1/\alpha)}\right)^2\right]. \tag{78}$$

Example 2. We shall now study the testing problem

$$H:\ \sum_{j=1}^{m}\mu_j=a_1\qquad\text{against}\qquad G:\ \sum_{j=1}^{m}\mu_j=a_2 \tag{79}$$

with $a_2-a_1>0$. The same assumptions as in Example 1 will be made on X. It is obvious that both $\mathfrak{P}$ and $\mathfrak{Q}$ are infinite for

$$\mathfrak{P}=\left\{P\,\middle|\,\frac{dP}{d\lambda^{(m)}}(\vec{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^m\frac{1}{\sqrt{\det V}}\times\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})\right)\wedge\vec{\mu}^T\vec{e}=a_1\right\} \tag{80}$$

$$\mathfrak{Q}=\left\{Q\,\middle|\,\frac{dQ}{d\lambda^{(m)}}(\vec{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^m\frac{1}{\sqrt{\det V}}\times\exp\left(-\frac{1}{2}(\vec{x}-\vec{\rho})^T V^{-1}(\vec{x}-\vec{\rho})\right)\wedge\vec{\rho}^T\vec{e}=a_2\right\} \tag{81}$$

where

$$\vec{e}=(1,1,\cdots,1)^T\,. \tag{82}$$

Clearly $I(P^{(n)},Q^{(n)})(t)$, $f(\vec{x}_1,\cdots,\vec{x}_n)$, $g(\vec{x}_1,\cdots,\vec{x}_n)$ are the same as in Example 1. To obtain $d_t(\mathfrak{P},\mathfrak{Q})$ we consider $I(P,Q)(t)$ with arbitrarily chosen $P\in\mathfrak{P}$, $Q\in\mathfrak{Q}$ and fixed $t\in[0,1]$. To maximize $I(P,Q)(t)$ subject to the given conditions on $\vec{\mu}$ and $\vec{\rho}$ means to minimize

$$\Phi(\vec{\mu},\vec{\rho})=(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho}) \tag{83}$$

subject to

$$\vec{\mu}^T\vec{e}=a_1 \qquad \vec{\rho}^T\vec{e}=a_2\,. \tag{84}$$

Using the method of Lagrange multipliers we consider the function

$$L(\vec{\mu},\vec{\rho},z_1,z_2)=\Phi(\vec{\mu},\vec{\rho})+z_1(\vec{\mu}^T\vec{e}-a_1)+z_2(\vec{\rho}^T\vec{e}-a_2) \tag{85}$$

which gives

$$\frac{\partial L}{\partial\vec{\mu}}=2V^{-1}\vec{\mu}-2V^{-1}\vec{\rho}+z_1\vec{e} \tag{86}$$

$$\frac{\partial L}{\partial\vec{\rho}}=2V^{-1}\vec{\rho}-2V^{-1}\vec{\mu}+z_2\vec{e} \tag{87}$$

$$\frac{\partial L}{\partial z_1}=\vec{\mu}^T\vec{e}-a_1 \tag{88}$$

$$\frac{\partial L}{\partial z_2}=\vec{\rho}^T\vec{e}-a_2\,. \tag{89}$$

For minimum values all the partial derivatives (86)–(89) must be zero. All values which satisfy these necessary conditions will be indexed by the sign $_0$. From (86) and (87) we obtain

$$z_{10}=-z_{20}\,. \tag{90}$$

Multiplying (86) and (87) from the left with V yields

(91) $$\vec{\mu}_0-\vec{\rho}_0=-\frac{z_{10}}{2}V\vec{e}\ .$$

Using the facts $\vec{\mu}^T\vec{e}=a_1$ and $\vec{\rho}^T\vec{e}=a_2$ we can see that

(92) $$z_{10}=2\frac{a_2-a_1}{\vec{e}^T V\vec{e}}$$

(93) $$\vec{\mu}_0-\vec{\rho}_0=\frac{a_1-a_2}{\vec{e}^T V\vec{e}}V\vec{e}$$

(94) $$(\vec{\mu}_0-\vec{\rho}_0)^T V^{-1}(\vec{\mu}_0-\vec{\rho}_0)=\frac{(a_2-a_1)^2}{\vec{e}^T V\vec{e}}\ .$$

Next we will show that $(\vec{\mu}-\vec{\rho})^T V^{-1}(\vec{\mu}-\vec{\rho})$ must attain its minimum. As V^{-1} is positive definite we can find a regular matrix B with

(95) $$B^T B=V^{-1}\ .$$

We must prove the existence of

(96) $$\min_{(\vec{\mu},\vec{\rho})}\|B(\vec{\mu}-\vec{\rho})\|^2$$

subject to $\vec{\mu}^T\vec{e}=a_1$ $\vec{\rho}^T\vec{e}=a_2$ ($\|\cdot\|$ the Euclidean norm) which is the problem of finding

(97) $$\min_{\substack{\vec{y}\\ \vec{y}\in M}}\|B\vec{y}\|$$

where

(98) $$M=\{\vec{y}\,|\,\vec{y}=\vec{\mu}-\vec{\rho},\ \vec{\mu}^T\vec{e}=a_1,\ \vec{\rho}^T\vec{e}=a_2\}\ .$$

Clearly M is a convex and closed set in $|\boldsymbol{R}^m|$ and so is $B(M)$ (B understood as a linear mapping). Because

(99) $$\min_{\substack{\vec{y}\\ \vec{y}\in M}}\|B\vec{y}\|^2=\min_{\substack{\vec{z}\\ \vec{z}\in B(M)}}\|z\|^2$$

we can confirm the existence of a minimum using a well-known result from the theory of topological linear spaces [3] p. 347. $\vec{\mu}_0$ and $\vec{\rho}_0$ are, of course, not uniquely determined. However, we obtain:

(100) $$d_t(\mathfrak{P},\mathfrak{Q})=\exp\left[-\frac{t}{2}(1-t)\frac{(a_2-a_1)^2}{\vec{e}^T V\vec{e}}\right]\ .$$

We now test

(101) $$H:\ \vec{\mu}=\vec{\mu}_0 \qquad \text{against} \qquad G:\ \vec{\mu}=\vec{\rho}_0$$

which is equivalent to the testing problem:

$$(102)\qquad H_0:\ \vec{\mu}=\vec{\mu}_0 \quad\text{against}\quad G_0:\ \vec{\mu}=\vec{\mu}_0+\frac{a_2-a_1}{\vec{e}^T V\vec{e}}V\vec{e}$$

where $\vec{\mu}_0$ is fixed for the moment. (102) is a special case of Example 1. We accept G_0, if

$$(103)\qquad \sum_{i=1}^{n}(\vec{x}_i-\vec{\mu}_0)^T\frac{a_2-a_1}{\vec{e}^T V\vec{e}}\vec{e}\geqq\sqrt{\frac{(a_2-a_1)^2 2n\log(1/\alpha)}{\vec{e}^T V\vec{e}}}$$

or equivalently

$$(104)\qquad \sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij}\geqq na_1+\sqrt{\vec{e}^T V\vec{e}\cdot 2n\log(1/\alpha)}\ .$$

We now return to the more general problem of testing:

$$(105)\qquad H:\ \vec{\mu}^T\vec{e}=a_1 \qquad G:\ \vec{\rho}^T\vec{e}=a_2\ .$$

As critical region for the test H against G we naturally choose:

$$(106)\qquad W=\left\{(\vec{x}_1,\cdots,\vec{x}_n)^T\ \middle|\ \sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij}\geqq na_1+\sqrt{\vec{e}^T V\vec{e}\cdot 2n\log(1/\alpha)}\right\}\ .$$

It remains to demonstrate that

$$(107)\qquad \bigwedge_{\substack{P\\ P\in\mathfrak{P}}} \quad P^{(n)}(W)\leqq P_0^{(n)}(W)$$

$$(108)\qquad \bigwedge_{\substack{Q\\ Q\in\mathfrak{Q}}} \quad Q^{(n)}(W)\geqq Q_0^{(n)}(W)\ .$$

If H is true we have $\vec{X}\sim N(\vec{\mu}, V)$ which gives

$$(109)\qquad \vec{X}_i^T\vec{e}\sim N(a_1, \vec{e}^T V\vec{e}) \qquad i=1,\cdots,n$$

and

$$(110)\qquad \sum_{i=1}^{n}\sum_{j=1}^{m}X_{ij}=\sum_{i=1}^{n}\vec{X}_i^T\vec{e}\sim N(na_1, n\vec{e}^T V\vec{e})\ .$$

Hence

$$(111)\qquad P^{(n)}(W)=\Pr\left(Z\geqq\sqrt{2\log(1/\alpha)}\right)=P_0^{(n)}(W)$$

where $Z\sim N(0,1)$. If G is true it is clear that $\vec{X}\sim N(\vec{\rho}, V)$ and

$$(112)\qquad \sum_{i=1}^{n}\sum_{j=1}^{m}X_{ij}=\sum_{i=1}^{n}\vec{X}_i\vec{e}\sim N(na_2, n\vec{e}^T V\vec{e})$$

from which we can conclude that

$$(113)\qquad Q^{(n)}(W)=\Pr\left(Z\geqq\sqrt{n}\,\frac{a_1-a_2}{\sqrt{\vec{e}^T V\vec{e}}}+\sqrt{2\log(1/\alpha)}\right)=Q_0^{(n)}(W)\,.$$

From (70) and (94) it follows directly that

$$(114)\qquad 1-Q^{(n)}(W)\leqq\exp\left[-\frac{1}{2}\left((a_2-a_1)\sqrt{\frac{n}{\vec{e}^T V\vec{e}}}-\sqrt{2\log(1/\alpha)}\right)^2\right].$$

The special problem

$$V=\begin{pmatrix}\sigma_1^2 & & 0\\ & \ddots & \\ 0 & & \sigma_m^2\end{pmatrix}$$

can be solved in the same way as in Example 1.

We should now discuss the quality of the preceding (of course not uniformly most powerful) tests. We shall do this only in the case of Example 2. We can easily see from (111) that

$$(115)\qquad \bigwedge_{\substack{P\\ P\in\mathfrak{P}}} P^{(n)}(W)<\alpha\,.$$

The following table shows the relationship between the given α and the real probability of comitting the error of the I. kind:

α	$\sqrt{2\log(1/\alpha)}$	$\Pr(Z\geqq\sqrt{2\log(1/\alpha)})$
0.001	3.7169	0.0001
0.01	3.0349	0.0012
0.03	2.6482	0.0041
0.05	2.4478	0.0072
0.0084	3.0902	0.001
0.0668	2.3263	0.01
0.1706	1.8808	0.03
0.2585	1.6449	0.05

But as the table indicates it is possible by choosing a bigger α^* than the desired α, to achieve that

$$(116)\qquad \bigwedge_{\substack{P\\ P\in\mathfrak{P}}} P^{(n)}(W)=\alpha\,.$$

One of the great advantages of these tests is the possibility of an easy computation of an upper bound for Pr (Error of the II. kind) for every n. Moreover there is no table of the χ^2-distribution or of any other distribution needed, a fact which makes the test rather handy.

Remark. With the principles of this paper one can easily test (with known σ) for example:

$$H:\ \mu \leqq -a \qquad G:\ \mu \geqq a$$

where $a>0$ is a given constant or

$$H:\ |\mu| \leqq a_1 \qquad G:\ |\mu| \geqq a_2$$

a_1, a_2 given constants $a_2-a_1>0$. These tests will also improve those proposed by Matusita, Suzuki and Hudimoto [1]. Details can be taken from [5].

INSTITUT FÜR QUANTITATIVE ÖKONOMIK UND STATISTIK, FACHRICHTUNG ANGEWANDTE STATISTIK

REFERENCES

[1] Matusita, K., Suzuki, Y. and Hudimoto, H. (1954/55). On testing statistical hypotheses, *Ann. Inst. Statist. Math.*, **6**, 133-141.

[2] Hogg, R. V. and Craig, A. T. (1965). *Introduction to Mathematical Statistics*, The Macmillan Company, New York.

[3] Köthe, G. (1966). *Topologische Lineare Räume I*, Springer, New York.

[4] Wetzel, W., Jöhnk, M. D. and Naeve, P. (1967). *Statistische Tabellen*, Walter de Gruyter, Berlin.

[5] Trenkler, G. (1973). Über eine abstandserzeugende Funktion von Wahrscheinlichkeitsmaßen, Dissertation, Berlin, to be published.