Note on the Decision Problem

By Kameo MATUSITA and Hirotugu AKAIKE

(Received July 5, 1952)

1 One of the authors has previously obtained the following result:

*Theorem**) Let $\Omega = \{F_1, \ldots, F_s\}$ be a finite set of admissible distribution functions, and let ρ be the maximum among the affinities $\rho(F_i, F_j)\ (i \neq j)$. Further, for an arbitrarily chosen positive number ε let k be an integer such that $(s-1)\rho^k < \varepsilon$, and in the k-dimensional sample space $R^{(k)}$ put

$$E_{ij} = \{(x_1, x_2, \ldots, x_k)\,;\, p_i(x_1).\ldots.p_i(x_k) > p_j(x_1).\ldots.p_j(x_k) \quad (i \neq j, i, j = 1, \ldots, s)\}$$

and

$$E_i = \bigcap_{j \neq i} E_{ij}$$

where $p_i(x)$ denote the density or probability of F_i for each i, i, e., $F_i(E) = \int_E p_i(x)dm$. Then $E_1, \ldots, E_s$ are mutually disjoint and we have

$$F_i^{(k)}(E_i) > 1 - \varepsilon \qquad (i = 1, \ldots, s)$$

where $F_i^{(k)}$ represent the extended distribution on $R^{(k)}$ of F_i.

In virtue of the theorem, we can take E_1 as a criterion region in the case where we are concerned with the problem, whether the true distribution function is F_1 or is contained in $\Omega_2 = \{F_2, F_3, \ldots, F_s\}$. That is, if we decide that the true distribution function is F_1 when the obtained sample point $(x_1, \ldots, x_k)$ lies in E_1, and the true distribution function is contained in Ω_2 in all other cases, then the risk is smaller than ε, provided that the weight function does not take any value larger than 1. Now, we have clearly

$$E_1 = \{(x_1.\ldots.x_k)\,;\, p_1(x_1).\ldots.p_1(x_k) > \sup_{2 \leq j \leq k} p_j(x_1).\ldots.p_j(x_k)\}$$

This shows that in this case our decision rule is apparently similar to that by the λ-principle, but it is essentially different. Though the above theorem is restricted only to the case where Ω_2 is finite, we shall remark in the following section (2) that our method is also valid in some cases where Ω_2 consists of an infinite number of distributions. Further, in section 3

*) Cf. K. MATUSITA: On the Theory of Statistical Decision Functions. This Annals Vol. II

we shall illustrate this method by some numerical examples.

2 Given a distribution F_0 and a set of distributions $\Omega = \{F_\nu\}$, we have the following

Theorem If $\sup_\nu \rho\ (F_0, F_\nu) < 1$ and if there exist a finite number of distributions $\{F_1, \ldots, F_n\}$ in Ω such that for any F_ν in Ω

$$E_{0\nu} \supset E_0 = E_{01} \frown E_{02} \frown \cdots \frown E_{0n}$$

holds, where $E_{0\nu}$ denotes the above-mentioned set in k-dimensional sample space, then E_0 can be used as a criterion region for F_0 and Ω in the sense of decision rule. *The risk can be made arbitrarily small by taking k large enough. (The true distribution function may or may not be F_0.)*

The proof runs just as in the case of finite Ω and is omitted here. The theorem can also be extended to the case where we have any number of distributions instead of F_0.

In the following some examples will be given.

(1) Let F_0 be a Gaussian distribution $N(m_0, \sigma^2)$ and Ω a family $\{N(m, \sigma^2)\ ;\ |m - m_0| \geqq \delta > 0\}$. As

$$\rho(N(m_0, \sigma^2),\ N(m, \sigma^2)) = e^{-\frac{|m-m_0|^2}{8\sigma^2}}$$

we have

$$\rho(N(m_0, \sigma^2),\ N(m, \sigma^2)) \leqq e^{-\frac{\delta^2}{8\sigma^2}} < 1$$

$$F_1 \equiv N(m_0 + \delta, \sigma^2)$$

$$F_2 \equiv N(m_0 - \delta, \sigma^2)$$

Then

$$F_0^{(k)}(E_{01} \frown E_{02}) > 1 - \varepsilon$$

$$F_i^{(k)}(E_{01} \frown E_{02}) < \varepsilon \qquad (i = 1, 2)$$

when $k > -\dfrac{8\sigma^2}{\delta^2} \log \varepsilon$. Now, for any $F_\nu = N(m_\nu, \sigma^2)$ in Ω, we have

$$E_{0\nu} \supset E_{01} \qquad \text{if} \quad m_\nu > m_0,$$

and

$$E_{0\nu} \supset E_{02} \qquad \text{if} \quad m_\nu < m_0$$

Thus, we have

$$E_{0\nu} \supset E_0 = E_{01} \frown E_{02}$$

which shows that our theorem is applicable.

(2) Let F_0 be $N(m, \sigma_0^2)$ and Ω $\{N(m, \sigma^2)\ ;\ |\sigma - \sigma_0| \geqq \delta\}$ In this case, taking

$$F_1 \equiv N(m, (\sigma_0 + \delta)^2)$$

$$F_2 \equiv N(m, (\sigma_0 - \delta)^2)$$

we proceed as in (1).

(3) Let F_0 be a Poisson distribution $P(\alpha_0)$ with mean α_0, and Ω $\{P(\alpha); |\sqrt{\alpha} - \sqrt{\alpha_0}| \geqq \delta\}$. Then, taking

$$F_1 \equiv P((\sqrt{\alpha_0} + \delta)^2)$$
$$F_2 \equiv P((\sqrt{\alpha_0} - \delta)^2)$$

we proceed as in the above examples.

3 Numerical Examples

(1) Let F_0 be $N(m_0; \sigma^2)$ and Ω $\{N(m, \sigma^2); |m - m_0| \geqq 3\sigma\}$. Then

$$\rho = \max_{F \in \Omega} \rho(F_0, F) = e^{-\frac{9}{8}}$$

Consequently

$$\rho^3 = e^{-\frac{27}{8}} \doteqdot 0.034$$

$$\rho^4 = e^{-\frac{36}{8}} \doteqdot 0.011$$

which shows that a sample of size 3 or 4 is enough for our decision with the risk less than 4% or 2%.

(2) Let F_0 be a χ^2-distribution with degree of freedom m and Ω a single distribution $F_1 \equiv N(m, 2m)$.

Then we have

$$\rho \equiv \rho(F_0, F_1) = \frac{e^{-\frac{m}{8}}(2m)^{\frac{m}{8}}\Gamma\left(\frac{m}{8} + \frac{1}{4}\right)}{\left\{(2\pi)^{\frac{1}{2}} \cdot 2 \cdot \Gamma\left(\frac{m}{2}\right)\right\}^{\frac{1}{2}}}$$

When $m = 10$, ρ is approximately equal to 0.3644, and we have

$$\rho^3 < 0.05 \qquad \rho^5 < 0.01$$

which has as a result that a sample of size 3 or 5 is enough for deciding whether the sample is drawn from a population with χ^2- or Gaussian distribution, with the risk less than 5%, 1%, respectively.

(3) Let F_0 be $P(1)$ (Poisson distribution with mean 1) and Ω a family $\{P(\alpha'); \alpha' \geqq 2\}$. Then

$$\rho \equiv \max_{F \in \Omega} (F_0, F) = e^{-\frac{1}{4}} \doteqdot 0.7788$$

and

$$\rho^{13} < 0.05 \qquad \rho^{20} < 0.01$$

What these inequalities mean are obvious.

(4) Let F_0 be $N(m, \sigma_0^2)$ and Ω $\{N(m, \sigma^2); \sigma \geqq 1.5\sigma_0\}$. In this case

we have the following results.

$$\rho \equiv \max_{F\in\Omega} \rho(F_0, F) = \sqrt{\frac{3\sigma_0^2}{3.25\sigma_0^2}} = \left(\frac{3}{3.25}\right)^{\frac{1}{2}} \doteqdot 0.9608$$

$$\rho^{75} < 0.05 < \rho^{74} \qquad \rho^{116} < 0.01 < \rho^{115}$$

These results shows intuitively that the discrimination of two Gaussian distributions with the same mean and different variances is quite difficult compared with the discrimination of two Gaussian distributions with the same variance and different means from the standpoint of our two-way decision. In fact, comparison of $\rho's$ ascertains it generally.

(5) Given a sample of size n from a population with Gaussian distribution, we want to decide whether this distribution has variance σ^2 or τ^2. Now, when the sample

$$x_1, x_2, \ldots, x_n$$

is drawn from $N(m, \sigma^2)$, then

$$\chi^2 \equiv \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} \qquad \text{where} \quad \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

obeys χ^2-distribution with degree of freedom $n-1$. From this fact follows that $z \equiv \sigma^2\chi^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$ has the distribution with probability density

$$f_\sigma(z) = \frac{\left(\frac{z}{2\sigma^2}\right)^{\frac{n-1}{2}-1}}{2\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{e^{-\frac{z}{2\sigma^2}}}{\sigma^2}$$

Denoting this distribution by F_σ, we have the following two-way decision problem. Let a sample $(x_1 \ldots, x_n)$ and two numbers σ, τ be given. Is $\sum_{i=1}^{n}(x_i - \bar{x})^2$ distributed according to F_σ or F_τ? Following the procedure carried out in the above examples, we have in this case

$$\rho \equiv \rho(F_\sigma, F_\tau) = \left(\frac{2\sigma\tau}{\sigma^2 + \tau^2}\right)^{\frac{n-1}{2}}$$

and the risk can be calculated. Further, when the means are known, the affinity ρ is obtained by replacing $n-1$ by n in the above formula.

Institute of Statistical Mathematics