DECISION RULES, BASED ON THE DISTANCE, FOR PROBLEMS OF FIT, TWO SAMPLES, AND ESTIMATION[1]

By Kameo Matusita

The Institute of Statistical Mathematics, Tokyo

1. Summary. Concrete decision rules are given for the problem of goodness of fit and the problem of two samples with a risk smaller than any preassigned value. The problem of estimation is also treated.

2. Introduction. In the theory of statistical decision functions it is very desirable to give the problem considered a concrete solution which evaluates the risk. We have previously given a concrete form of the decision function when the distribution functions involved are specified, so that the risk can be made smaller than an arbitrarily preassigned positive number by suitable choice of sample-size ([1], [2]). In that case, the following notions of affinity and distance played an important role. Let F_1, F_2 be simultaneously discrete or continuous distribution functions, so that by the aid of a suitable measure m, the probabilities of a set E under F_1 and F_2 can be written as:

$$F_1(E) = \int_E p_1(x)\, dm, \qquad F_2(E) = \int_E p_2(x)\, dm$$

respectively. Here E denotes an arbitrary measurable set for which the probability under F_1 or F_2 is defined. The *distance* between F_1 and F_2 is then given by

$$||F_1 - F_2|| = \left(\int_R (\sqrt{p_1(x)} - \sqrt{p_2(x)})^2\, dm\right)^{\frac{1}{2}}$$

and the affinity between F_1 and F_2 by

$$\rho = \int_R \sqrt{p_1(x)}\, \sqrt{p_2(x)}\, dm$$

where R denotes the whole sample space (of one dimension).

In the present paper, using the distance $||\quad||$, we shall give a concrete solution to the problem of goodness of fit and the two-sample problem and mention finally that the estimation problem can be treated similarly. Our treatment of the problem of fit is based on the following considerations.[2] It is not necessary to decide whether the random variable on which the observation is made has *exactly* the given distribution or not but to decide whether the variable has a distribution *near* the given one. On the other hand, from a finite number of ob-

Received July 20, 1953, revised April 15, 1955.

[1] The results of this paper were announced in March, 1953 at a meeting of the Institute of Statistical Mathematics.

[2] See [2]. These considerations and the formulation of the problem were also given in [3].

servations it is impossible, or at least very difficult, to discern efficiently whether the variable has exactly the given distribution when in the alternative class of distributions there exists one which lies within any distance from the given distribution. From this point of view, we formulate the problem as follows.[2]

When a distribution F_0 is given, decide whether the random variable has the distribution F_0 or a distribution outside some neighborhood of F_0.

The neighborhood of F_0 mentioned here is defined as the set of distributions $\{F: \|F - F_0\| < \epsilon\}$, where a positive number ϵ is determined according to the nature of the problem concerned. Throughout this paper we shall consider only discrete distributions with a finite number of possible outcomes. For practical purposes this does not involve an essential loss of generality since in most statistical applications the quantities observed can be grouped in a finite number of classes.

The actual method of our treatment of the problem is based on the following properties of distance which the distance $\| \quad \|$ possesses:

(1) Axioms of distance:

(i) $\delta(F, G) \geqq 0,$
$\delta(F, F) = 0,$
(ii) $\delta(F, G) = \delta(G, F),$
(iii) $\delta(F, G) + \delta(G, H) \geqq \delta(F, H),$

for any distributions F, G and H, where $\delta(F, G)$ denotes the distance between F and G.

(2) For any integer n and any positive number η there can be found a sequence of numbers $\{B_n\}$ such that

$$\Pr\{\delta(F, S_n) > \eta\} \leqq B_n$$

for any F in the class of distribution functions under consideration, where S_n denotes the empirical distribution function based on n observations of a random variable with distribution F, and such that $B_n \to 0$ as $n \to \infty$. There are, of course, other distances having properties (1) and (2). For instance,

$$\delta_e(F, F') = \left(\sum_{i=1}^{k} (p_i - p_i')^2\right)^{\frac{1}{2}}$$

where F and F' are discrete distributions defined on the same events with probabilities $p_1, \cdots, p_k$, and $p_1', \cdots, p_k'$, respectively. Actually, our method can be applied with any definition of distance which has properties (1) and (2). Among the above properties, (1)–(ii) need not necessarily hold for the problem of fit, but must hold for the two-sample problem. (1)–(iii), that is, the triangle inequality, must always hold. A so-called directed distance, like

$$\left(\int_{-\infty}^{\infty} (F_1(x) - F_2(x))^2 \, dF_2(x)\right)^{\frac{1}{2}},$$

does not always satisfy these conditions, so that we cannot use it, at least for the two-sample problem. Further, χ^2, itself, does not have property (2).

Since the inference will be based on a finite number of observations, it is desirable that the distance we use has the property:

(3) The distance represents well the discrepancy between distributions at every point, or, for the problem of fit, it represents the discrepancy at the point with large probability (or density) better than that at the point with small probability (density).

Taking into account (3), the Fréchet distance $\delta_F(F, G) = \sup_x |F(x) - G(x)|$ is seen to be unsuitable, as is any distance similar to it, although it is useful in cases where the convergence or a property in the limit is concerned. The same can be said concerning χ^2. When we consider χ^2 as a quantity which expresses the discrepancy between the theoretical and the empirical distribution, the discrepancy at the point with small probability is liable to make an excessive contribution. This shows that, in general, χ^2 is not always suitable for the problem of fit.

On the other hand our distance $\|\quad\|$ seems to satisfy property (3) adequately and is also simple to compute.

With this distance we shall give a simple upper bound of the risk and at the same time show how to make the risk smaller than any preassigned positive number. This is not the case with the tests thus far presented, like the χ^2-test.

3. Properties of distance $\|\quad\|$ and affinity ρ. In this paper, we shall use only the distance $\|\quad\|$ explicitly, but for its calculation the affinity ρ will prove useful. Therefore, in the following we shall state the properties of ρ as well as those of the distance $\|\quad\|$.

First we have

$$0 \leqq \rho \leqq 1,$$

$$\|F_1 - F_2\|^2 = 2(1 - \rho(F_1, F_2)) \leqq 2,$$

$$\|F_1 - F_2\|^2 \leqq \int_R |p_1(x) - p_2(x)| \, dm \leqq 2\|F_1 - F_2\|.$$

From these relations it follows that

$$|F_1(E) - F_2(E)| \leqq 2\|F_1 - F_2\|$$

for any set E, and that for a sequence of distributions $\{F_n\}$ and a distribution F_0 it holds

$$\|F_n - F_0\| \to 0 \qquad (n \to \infty)$$

when and only when

$$|F_n(E) - F_0(E)| \to 0 \qquad (n \to \infty) \qquad \text{uniformly in } E. \tag{3.1}$$

Further, $\|F_1 - F_2\| = 0$ when and only when $F_1(E) = F_2(E)$ for every set E. We also note that $\|F_n - F_0\| \to 0$ $(n \to \infty)$ implies $\rho(F_n, F_0) \to 1$ $(n \to \infty)$ and vice versa.

Relation (3.1) is Wald's definition of regular convergence in finite-dimensional sample space. When the set of distributions Ω, therefore, is separable in the sense of regular convergence, this topology is equivalent to that given by our distance $\|\quad\|$. It follows from this fact that if Ω is compact with respect to the topology induced by regular convergence, it is also compact with respect to the topology induced by our distance.

4. Fundamental theorems. Let F be a discrete distribution with probabilities $p_1, p_2, \cdots, p_k$ for the events (1), (2), $\cdots$, (k), respectively. Let n_i be the number of occurrences of event (i) in n observations. We denote the empirical distribution $(n_1 / n, \cdots, n_k / n)$ by S_n. Then, by definition,

$$\|F - S_n\|^2 = \sum_{i=1}^{k} \left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right)^2 = 2\left(1 - \sum_{i=1}^{k} \sqrt{\frac{n_i}{n}} \sqrt{p_i}\right).$$

The last expression in terms of the affinity $\rho = \sum_{i=1}^{k} \sqrt{n_i/n}\sqrt{p_i}$ can be used for the calculation of the distance $\|F - S_n\|$.

THEOREM I. *When the random variable concerned has the discrete distribution F, then we have*

$$\Pr\left\{\|F - S_n\|^2 < \frac{k-1}{n} t\right\} \geqq 1 - \frac{1}{t}$$

for any positive number t.

PROOF. Let $p_1, \cdots, p_{k'} > 0$, and $p_{k'+1} = \cdots = p_k = 0$ $(k' \leqq k)$. Then, clearly

$$\Delta^2 = \|F - S_n\|^2 = \sum_{i=1}^{k} \left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right)^2 \leqq \sum_{i=1}^{k'} \frac{1}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 + \sum_{i=k'+1}^{k} \frac{n_i}{n}.$$

Accordingly,

$$E(\Delta^2) \leqq \sum_{i=1}^{k'} \frac{1}{p_i} E\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{k'} \frac{1 - p_i}{n} = \frac{k' - 1}{n} \leqq \frac{k-1}{n}$$

where E denotes the expectation with respect to F. Now, an inequality of Markov shows

$$\Pr\{\Delta^2 < E(\Delta^2)t\} \geqq 1 - \frac{1}{t} \qquad \text{for any positive } t,$$

and we have

$$\Pr\left\{\Delta^2 < \frac{k-1}{n} t\right\} \geqq 1 - \frac{1}{t}$$

which we wanted to prove.

When $p_i > 0$, $i = 1, 2, \cdots, k$, and n is large, $\chi^2 = \sum_{i=1}^{k}(n_i - np_i)^2 / np_i$ is asymptotically distributed according to the chi-square distribution with $k - 1$ degrees of freedom. Therefore, from the relation

$$\|F - S_n\|^2 = \frac{1}{n}\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}\left(1 + \sqrt{\frac{n_i}{np_i}}\right)^{-2}$$

we have:

THEOREM II. *When the random variable concerned has the discrete distribution F with $p_i > 0$ $(i = 1, 2, \cdots, k)$ and n is large, we have an asymptotic equality*

$$\Pr\{\|F - S_n\|^2 < \delta^2\} \doteq \Pr\{\chi^2_{(k-1)} < 4n\delta^2\}$$

for any number δ, where $\chi^2_{(k-1)}$ is a random variable having the chi-square distribution with $k - 1$ degrees of freedom.

When applying such an inequality or equality to the problem of fit, for instance, it must hold for any distribution under consideration (see below). Therefore, in case the alternative class of distributions against the specified distribution includes one with any small probability p_i, we cannot use an inequality or equality containing $D^2(\chi^2)$, like

$$\Pr\{\|F - S_n\|^2 < \eta\} \geqq 1 - (1 / n^2\eta^2)\{(k - 1)^2 + D^2(\chi^2)\},$$

which is derived from the relation $\|F - S_n\|^2 < \chi^2 / n$ when

$$p_i > 0 \ (i = 1, 2, \cdots, k),$$

although it may seem more precise than the inequality in Theorem I at a glance. For we have then $\sup_F D^2(\chi^2) = \infty$ and cannot obtain an adequate inequality or equality for our purpose. On the other hand, the inequality in Theorem I holds in any case and is applicable to the problem. When there is a positive lower bound for all p_i, we can obtain more precise inequalities. For example, let p_0 be such a lower bound. Then we have

$$\Pr\{\|F - S_n\|^2 < \eta\} \geqq 1 - \frac{1}{n^2\eta^2}\left\{k^2 - 1 + \frac{1}{n}\left(\frac{k}{p_0} - k^2 - 2k + 2\right)\right\}$$

or, when

$$n\eta > k + 1 + \frac{1}{(k-1)\cdot n}\left(\frac{k}{p_0} - k^2 - 2k + 2\right),$$

$$\Pr\{\|F - S_n\|^2 < \eta\} \geqq 1 - \left(1 + \frac{(n\eta - k + 1)^2}{2(k-1) + \frac{1}{n}\left(\frac{k}{p_0} - k^2 - 2k + 2\right)}\right)^{-1}$$

(See [4].) The result of Vora ([6]) could also be used. In this paper, however, we shall not assume that there exists such a lower bound. In the following we shall denote generally by $(C_{n,k-1})$ or for short (C_n) a class of distributions such that χ^2 based on n observations of the random variable has asymptotically the chi-square distribution with $k - 1$ degrees of freedom for any F in $(C_{n,k-1})$. A set of distributions $\{(p_1, p_2, \cdots, p_k)\}$ in the same finite discrete space, for which there exists a positive number p_0 such that $p_i > p_0$, defines such a class (C_n) for n sufficiently large.

5. The problem of goodness of fit. As stated in the introduction, our formulation of the problem is to find a rule according to which, for any given finite discrete distribution F_0 and $\delta_0 > 0$, one can decide whether a random variable has distribution F_0 or a distribution F with $\|F - F_0\| \geqq \delta_0$.

Now, let X be the random variable of interest. When X has the distribution F_0 we have by Theorem I

$$\Pr\left\{ \|F_0 - S_n\|^2 < \frac{k-1}{n} t \right\} \geqq 1 - \frac{1}{t},$$

and when X has a distribution F with $\|F - F_0\| \geqq \delta_0$,

$$\|F_0 - S_n\| \geqq \|F_0 - F\| - \|F - S_n\| \geqq \delta_0 - \|F - S_n\|$$

and

$$\Pr\left\{ \|F - S_n\|^2 < \frac{k-1}{n} t \right\} \geqq 1 - \frac{1}{t}.$$

Consequently

$$\Pr\left\{ \|F_0 - S_n\| > \delta_0 - \sqrt{\frac{k-1}{n} t} \right\} \geqq 1 - \frac{1}{t}.$$

On the other hand, when

$$n \geqq 4 \frac{k-1}{\delta_0^2} t,$$

we have

$$\delta_0 - \sqrt{\frac{k-1}{n} t} \geqq \sqrt{\frac{k-1}{n} t}.$$

Therefore, we have:

THEOREM III. *For any positive number t, let $n \geqq 4(k-1)t / \delta_0^2$. Then, when X has the distribution F_0, we have*

$$\Pr\left\{ \|F_0 - S_n\|^2 < \frac{k-1}{n} t \right\} \geqq 1 - \frac{1}{t},$$

and when X has a distribution F with $\|F - F_0\| \geqq \delta_0$,

$$\Pr\left\{ \|F_0 - S_n\|^2 \geqq \frac{k-1}{n} t \right\} \geqq 1 - \frac{1}{t}.$$

According to this theorem we can answer the problem formulated above. The risk then can be made smaller than any preassigned value ϵ by taking $t > 1/\epsilon$. Here we assume that the weight function is zero when the decision is correct and less than or equal to 1 when the decision is wrong.

If it is known that F_0 and any alternative distribution $(p_1, \cdots, p_k)$ are contained in a class (C_n), Theorem II is applicable instead of Theorem I, and we have the following theorem.

THEOREM IV. *Let F_0 and a distribution F with $\|F - F_0\| \geqq \delta_0$ be in a class (C_n), and let η be any positive number smaller than δ_0. Then, when X has the distribution F_0, we have*

$$\Pr\{\|F_0 - S_n\|^2 < \eta^2\} \doteq \Pr\{\chi^2_{(k-1)} < 4n\eta^2\}$$

and when X has a distribution F

$$\Pr\{\|F_0 - S_n\|^2 \geqq \eta^2\} \dot{\geqq} \Pr\{\chi^2_{(k-1)} < 4n(\delta_0 - \eta)^2\}$$

where $\chi^2_{(k-1)}$ *is a random variable having the chi-square distribution with* $k - 1$ *degrees of freedom.*

6. Two-sample problem. Let X and Y be (not necessarily independent) random variables which have discrete (marginal) distributions $F = (p_1^0, \cdots, p_k^0)$ and $G = (q_1^0, \cdots, q_k^0)$ on the same events (1), $\cdots$, (k), respectively, that is, $p_i^0 = \Pr\{X = (i)\}$, $q_i^0 = \Pr\{Y = (i)\}$; $i = 1, \cdots, k$. Further, let $(n_1, \cdots, n_k)$ $(\sum n_i = n)$ and $(m_1, \cdots, m_k)(\sum m_i = m)$ be observations on X and Y, respectively. Then we want to decide from the two sets of observations $(n_1, \cdots, n_k)$, $(m_1, \cdots, m_k)$ whether $F = G$ or $\|F - G\| \geqq \delta_0$, δ_0 being a preassigned positive number. In this problem we are also interested in whether F and G lie near each other. Denote the empirical distributions $(n_1/n, \cdots, n_k/n)$, $(m_1/m, \cdots, m_k/m)$ by S_n, S'_m, respectively. Then we have:

THEOREM V. *Let* η *be any positive number smaller than* δ_0. *When* $\|F - G\| = 0$,

$$\text{(a)} \qquad \Pr\{\|S_n - S'_m\| < \eta\} \geqq 1 - \frac{k-1}{\eta^2}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)^2,$$

and when $\|F - G\| \geqq \delta_0$,

$$\text{(b)} \qquad \Pr\{\|S_n - S'_m\| \geqq \eta\} \geqq 1 - \frac{k-1}{(\delta_0 - \eta)^2}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)^2.$$

Moreover, if X *and* Y *are independent of each other, we have*

$$\text{(c)} \qquad \Pr\{\|S_n - S'_m\| < \eta\} \geqq 1 - \frac{4(k-1)}{\eta^2}\left(\frac{1}{n} + \frac{1}{m}\right) + \frac{16(k-1)^2}{\eta^4 nm}$$

when $\|F - G\| = 0$, *and*

$$\text{(d)} \qquad \Pr\{\|S_n - S'_m\| \geqq \eta\} \geqq 1 - \frac{4(k-1)}{(\delta_0 - \eta)^2}\left(\frac{1}{n} + \frac{1}{m}\right) + \frac{16(k-1)^2}{(\delta_0 - \eta)^4 nm}$$

when $\|F - G\| \geqq \delta_0$. *In this case,* (a) *is more precise than* (c) *when and only when* $3(m + n) - 2\sqrt{mn} > 16(k - 1) / \eta^2$, *and* (b) *is more precise than* (d) *when and only when* $3(m + n) - 2\sqrt{mn} > 16(k - 1) / (\delta_0 - \eta)^2$.

PROOF. When $\|F - G\| = 0$, then

$$\|S_n - S'_m\| \leqq \|F - S_n\| + \|G - S'_m\|$$

Therefore, according to an inequality of Markov we have

$$\begin{aligned}
\Pr\{\|S_n - S'_m\| < \eta\} &\geqq \Pr\{\|F - S_n\| + \|G - S'_m\| < \eta\} \\
&\geqq 1 - \frac{1}{\eta^2} E(\|F - S_n\| + \|G - S'_m\|)^2 \\
&\geqq 1 - \frac{1}{\eta^2}\left(\sqrt{E(\|F - S_n\|^2)} + \sqrt{E(\|G - S'_m\|^2)}\right)^2 \\
&\geqq 1 - \frac{k-1}{\eta^2}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)^2
\end{aligned}$$

and moreover, if X and Y are independent,

$$\begin{aligned}\Pr\{\|S_n - S'_m\| < \eta\} &\geqq \Pr\{\|F - S_n\| < \tfrac{1}{2}\eta, \|G - S'_m\| < \tfrac{1}{2}\eta\}\\ &= \Pr\{\|F - S_n\| < \tfrac{1}{2}\eta\} \Pr\{\|G - S'_m\| < \tfrac{1}{2}\eta\}\\ &\geqq \left(1 - \frac{4(k-1)}{n\eta^2}\right)\left(1 - \frac{4(k-1)}{m\eta^2}\right)\\ &= 1 - \frac{4(k-1)}{\eta^2}\left(\frac{1}{n} + \frac{1}{m}\right) + \frac{16(k-1)^2}{\eta^4 nm}.\end{aligned}$$

When $\|F - G\| \geqq \delta_0$, then

$$\begin{aligned}\|S_n - S'_m\| &\geqq \|F - G\| - \|F - S_n\| - \|G - S'_m\|\\ &\geqq \delta_0 - \|F - S_n\| - \|G - S'_m\|\end{aligned}$$

from which we can obtain (b) and (d).

The remaining part of the theorem can easily be seen.

On the basis of this theorem we can make decisions about the two-sample problem.

Further, if it is known that X and Y are independent of each other and F and G belong to a class (C_n), then we can make use of the following theorem.

THEOREM VI. *Let X and Y be independent, and F and G belong to a class (C_n), and let η be any positive number smaller than δ_0. Then, when $\|F - G\| = 0$, we have the following asymptotic inequalities*

$$\begin{aligned}\Pr\{\|S_n - S'_m\| < \eta\} &\dot{\geqq} \Pr\left\{\frac{1}{2n}\chi^2_{(k-1)} + \frac{1}{2m}\chi'^2_{(k-1)} < \eta^2\right\}\\ &\geqq 1 - \frac{k-1}{2\eta^2}\left(\frac{1}{n} + \frac{1}{m}\right)\end{aligned}$$

and

$$\Pr\{\|S_n - S'_m\| < \eta\} \dot{\geqq} \Pr\{\chi^2_{(k-1)} < n\eta^2\}\Pr\{\chi'^2_{(k-1)} < m\eta^2\},$$

and when $\|F - G\| \geqq \delta_0$,

$$\begin{aligned}\Pr\{\|S_n - S'_m\| \geqq \eta\} &\dot{\geqq} \Pr\left\{\frac{1}{2n}\chi^2_{(k-1)} + \frac{1}{2m}\chi'^2_{(k-1)} \leqq (\delta_0 - \eta)^2\right\}\\ &\geqq 1 - \frac{k-1}{2(\delta_0 - \eta)^2}\left(\frac{1}{n} + \frac{1}{m}\right)\end{aligned}$$

and

$$\Pr\{\|S_n - S'_m\| \geqq \eta\} \dot{\geqq} \Pr\{\chi^2_{(k-1)} \leqq n(\delta_0 - \eta)^2\}\Pr\{\chi'^2_{(k-1)} \leqq m(\delta_0 - \eta)^2\}$$

where $\chi^2_{(k-1)}$ and $\chi'^2_{(k-1)}$ are independent random variables each having the chi-square distribution with $k - 1$ degrees of freedom, respectively.

7. Estimation problem. Now, let us turn to the problem of estimation. In this case too, as mentioned in the introduction, we confine ourselves to discrete distributions. Then, if $(n_1, \cdots, n_k)$, S_n, and F denote n observations, the empirical distribution, and the true distribution, respectively, we have for any positive t

$$\Pr\left\{\|F - S_n\|^2 < \frac{k-1}{n} t\right\} \geqq 1 - \frac{1}{t}.$$

This means that

$$\|F - S_n\|^2 = \sum_{i=1}^{k} \left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right)^2 < \frac{k-1}{n} t \tag{7.1}$$

is a confidence band for an unknown F with confidence coefficient at least $1 - 1/t$. If the form of F is known and only the parameters $\alpha_1, \cdots, \alpha_s$ involved in F are unknown, the confidence intervals for $\alpha_1, \cdots, \alpha_s$ are also obtained from (7.1).

When it is known that the unknown F is in a class (C_n), the relation

$$\Pr\{\|F - S_n\|^2 < \delta^2\} \doteq \Pr\{\chi^2_{(k-1)} < 4n\delta^2\}$$

can be used to obtain a confidence band.

As to point estimation (see also [4], [5], [7], [8]), one can estimate parameters by minimizing $\|F - S_n\|^2$ under the restriction $\sum p_i = 1$. Let $F_{e,n}$ be the distribution with these estimates replacing the parameters. Then one can show that these estimates converge stochastically to the parameters $\alpha_1, \cdots, \alpha_s$ in F, respectively,[3] by means of the following relations:

$$\|F_{e,n} - F\| \leqq \|F_{e,n} - S_n\| + \|F - S_n\| \leqq 2\|F - S_n\|,$$

$$\Pr\left\{\|F_{e,n} - F\| < \sqrt{\frac{k-1}{n} t}\right\} \geqq \Pr\left\{\|F - S_n\| < \frac{1}{2}\sqrt{\frac{k-1}{n} t}\right\} \geqq 1 - \frac{4}{t}$$

Of course, we assume here that the parameters depend continuously on the distribution F. Further, the inequality

$$\Pr\left\{\|F_{e,n} - S_n\|^2 < \frac{k-1}{n} t\right\} \geqq 1 - \frac{1}{t} \qquad (t > 0)$$

or the asymptotic equality

$$\Pr\{\|F_{e,n} - S_n\|^2 < \delta^2\} \doteq \Pr\{\chi^2_{(k-s-1)} < 4n\delta^2\}$$

can be used for the problem of fit when the specified distribution contains certain (say, s) unknown parameters. This inequality and equality can easily be proved. For example, the last equality is obtained as follows. When $F = (p_1, \cdots, p_k)$ with $p_i > 0$ $(i = 1, \cdots, k)$ and n is sufficiently large, we have

$$\|F - S_n\|^2 \doteq \chi^2 / 4n$$

and

$$\|F_{e,n} - S_n\|^2 \doteq \chi^2_{e,n} / 4n$$

---

[3] We can, further, prove that the convergence here is almost sure. See [4], [5].

where $\chi^2_{e,n}$ is χ^2 with the minimum χ^2 estimates of the unknown parameters replacing the parameters. As $\chi^2_{e,n}$ has asymptotically the chi-square distribution with $k - s - 1$ degrees of freedom, we obtain the above asymptotic equality.

8. Acknowledgments. I should like to express my gratitude to Professor Wassily Hoeffding for his useful comments and suggestions, by which this paper was revised in many points, in particular the introduction and Theorems II, IV, V, VI were made more precise. Thanks are also due to Professor Erich L. Lehmann as well as to the referees whose suggestions helped to improve the paper. I also wish to express my gratitude to Professor J. Wolfowitz for his interest in this work.

REFERENCES

[1] KAMEO MATUSITA, "On the theory of statistical decision functions," *Ann. Inst. Stat. Math.*, Tokyo, Vol. III (1951), pp. 17–35.
[2] KAMEO MATUSITA AND HIROTUGU AKAIKE, "Note on the decision problem," *Ann. Inst. Stat. Math.*, Tokyo, Vol. IV (1952), pp. 11–14.
[3] Z. W. BIRNBAUM, "Distribution-free tests of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 1–8.
[4] KAMEO MATUSITA, "On the estimation by the minimum distance method," *Ann. Inst. Stat. Math.*, Tokyo, Vol. V (1954), pp. 59–65.
[5] KAMEO MATUSITA, "A remark to 'On the estimation by the minimum distance method,'" *Ann. Inst. Stat. Math.*, Tokyo, Vol. VI (1954), p. 124.
[6] SHANTI A. VORA, "Bounds on the distribution of chi square," *Sankhyā*, Vol. 11 (1951), pp. 365–378.
[7] J. WOLFOWITZ, "Estimation by the minimum distance method," *Ann. Inst. Stat. Math.*, Tokyo, Vol. V (1953), pp. 9–23.
[8] J. WOLFOWITZ, "Estimation by the minimum distance method in nonparametric stochastic difference equations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 203–217.