The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 1999–2020

ESTIMATION AND TESTING HYPOTHESES FOR ONE, TWO, OR SEVERAL SAMPLES FROM GENERAL MULTIVARIATE DISTRIBUTIONS

BY MASAO TANAKA

Nara Medical University, Japan

0. Introduction and summary. Multivariate statistical analysis has been centered around normal theory thus far. Most papers dealing with this topic are based either directly on the assumption of normality of underlying distributions or indirectly on asymptotic normality of statistics induced by some limiting theorem such as the central limit theorem when sample sizes are sufficiently large.

In many if not all fields of application, however, the assumption of normality is not guaranteed even approximately and also the sample sizes are not so large compared with the dimension of the variates in question as to provide a good approximation by normal distributions. Thus, it is required to develop a theory applicable to general multivariate distributions and there are some papers in literature approaching this problem from the standpoint of a distance: [1], [4], [10] and [11] among others.

The purpose of this paper is to formulate statistical inference, point and interval estimation as well as testing hypotheses, in terms of a distance or a pseudo distance defined in the family of all probability distributions over a multidimensional Euclidean space. Three specific cases, one of which does not actually give a pseudo distance but which may be useful in some situations, are discussed.

The paper consists of four sections; in Section 1 we treat inference with a random sample from one distribution; in Section 2 and Section 3 we treat inference with independent random samples from two and several distributions, respectively, and in the Appendix we present some mathematical results which are used in the preceding sections as tools more practical than [3], [5] and [9] and which may be interesting in themselves.

1. One sample case.

DEFINITION 1.1. A function ρ of two elements of a space S is called a pseudo distance if

(1.1) $$\rho(P, R) \leqq \rho(P, Q) + \rho(Q, R) \qquad \text{for any} \quad P, Q, R \in S.$$

Let Ω be the s-dimensional Euclidean space and S the set of all possible distribution functions over Ω with a pseudo distance ρ. Let $F^*(\mathbf{x})$ be the empirical distribution function of a random sample of size n from a distribution function $F_t(\mathbf{x})$ defined over Ω and $F_0(\mathbf{x})$ be a given distribution function.

LEMMA 1.2.

(1.2) $$-\rho(F_t, F^*) \leqq \rho(F_0, F_t) - \rho(F_0, F^*) \leqq \rho(F^*, F_t),$$
$$-\rho(F^*, F_t) \leqq \rho(F_t, F_0) - \rho(F^*, F_0) \leqq \rho(F_t, F^*).$$

Received December 16, 1968; revised April 19, 1970.

ASSUMPTION 1.3. Let $f_i(y) = f_{in}(y)$ $(i = 1,2)$ be real continuous decreasing functions satisfying the following inequalities

$$(1.3)\qquad \begin{aligned} &P[\rho(F^*, F_t) \geqq y] \leqq f_1(y) \qquad \text{for any } F_t \in S,\\ &P[\rho(F_t, F^*) \geqq y] \leqq f_2(y) \qquad \text{for any } F_t \in S, \end{aligned}$$

for any $y > 0$.

Such functions f_1 and f_2 will be called (a pair of) majorants with respect to the pseudo distance ρ.

THEOREM 1.4.

(i) *If* $\lim_{n\to\infty} f_{in}(\varepsilon) = 0$ *for any* $\varepsilon > 0$ $(i = 1,2)$, *then* $\rho(F_0, F^*)$ *and* $\rho(F^*, F_0)$ *are consistent estimates of* $\rho(F_0, F_t)$ *and* $\rho(F_t, F_0)$ *respectively.*

(ii) *A joint confidence interval of* $\rho(F_0, F_t)$ *and* $\rho(F_t, F_0)$ *with at least* $100(1-\alpha)\%$ *confidence coefficient, where* $\alpha = \alpha_1 + \alpha_2$, *is given by*

$$(1.4)\qquad \begin{aligned} &[\rho(F_0, F^*) - f_2^{-1}(\alpha_2), \rho(F_0, F^*) + f_1^{-1}(\alpha_1)],\\ &[\rho(F^*, F_0) - f_1^{-1}(\alpha_1), \rho(F^*, F_0) + f_2^{-1}(\alpha_2)]. \end{aligned}$$

PROOF. By Lemma 1.2 and Assumption 1.3, we have

$$(1.5)\qquad \begin{aligned} &P[|\rho(F_0, F_t) - \rho(F_0, F^*)| < \varepsilon \text{ and } |\rho(F_t, F_0) - \rho(F^*, F_0)| < \varepsilon]\\ &\qquad \geqq P[\rho(F^*, F_t) < \varepsilon \text{ and } \rho(F_t, F^*) < \varepsilon] \geqq 1 - f_1(\varepsilon) - f_2(\varepsilon), \end{aligned}$$

which implies part (i). Next, we obtain

$$(1.6)\qquad \begin{aligned} &P[-f_2^{-1}(\alpha_2) < \rho(F_0, F_t) - \rho(F_0, F^*) < f_1^{-1}(\alpha_1)\\ &\qquad \text{and } -f_1^{-1}(\alpha_1) < \rho(F_t, F_0) - \rho(F^*, F_0) < f_2^{-1}(\alpha_2)]\\ &\qquad\qquad \geqq P[\rho(F^*, F_t) < f_1^{-1}(\alpha_1) \text{ and } \rho(F_t, F^*) < f_2^{-1}(\alpha_2)]\\ &\qquad\qquad \geqq 1 - \alpha_1 - \alpha_2 = 1 - \alpha, \end{aligned}$$

which implies (ii).

THEOREM 1.5. *For the testing problem*

$$(1.7)\qquad H: \rho(F_0, F_t) \leqq \rho_0, \qquad A: \rho(F_0, F_t) > \rho_0$$

with a fixed constant ρ_0, *the test procedure given by*

(1.8) *When* $\rho(F_0, F^*) < \rho_0 + f_2^{-1}(\alpha)$, *we make no decision,*[1]

When $\rho(F_0, F^*) \geqq \rho_0 + f_2^{-1}(\alpha)$, *we reject* H,

has level not greater than α *and power not less than* $1 - f_1(y)$ *for* F_0 *such that* $\rho(F_0, F_t) = \rho_0 + f_2^{-1}(\alpha) + y$ *with* $y > 0$.

---

[1] Our test procedure is different from the standard notion in that it never accepts the null hypothesis, but this formulation seems more realistic to the present author.

Similarly, for the testing problem

$$H: \rho(F_t, F_0) \leqq \rho_0, \qquad A: \rho(F_t, F_0) > \rho_0, \tag{1.9}$$

the procedure:

$$\begin{aligned} &\textit{When} \quad \rho(F^*, F_0) < \rho_0 + f_1^{-1}(\alpha), \quad \textit{we make no decision},\\ &\textit{When} \quad \rho(F^*, F_0) \geqq \rho_0 + f_1^{-1}(\alpha), \quad \textit{we reject } H, \end{aligned} \tag{1.10}$$

has level not greater than α *and power not less than* $1-f_2(y)$ *when* $\rho(F_t, F_0) = \rho_0 + f_1^{-1}(\alpha) + y$.

Proof. The level of the test is given by

$$\begin{aligned} &P[\rho(F_0, F^*) \geqq \rho_0 + f_2^{-1}(\alpha) \mid H]\\ &\quad \leqq P[\rho(F_0, F^*) - \rho(F_0, F_t) \geqq f_2^{-1}(\alpha)]\\ &\quad \leqq P[\rho(F_t, F^*) \geqq f_2^{-1}(\alpha)] \leqq \alpha \end{aligned} \tag{1.11}$$

in view of (1.2) and (1.3).

When $\rho(F_0, F_t) = \rho_0 + f_2^{-1}(\alpha) + y$, the power of the test is

$$\begin{aligned} &P[\rho(F_0, F^*) \geqq \rho_0 + f_2^{-1}(\alpha)]\\ &\quad = P[\rho(F_0, F_t) - \rho(F_0, F^*) \leqq y]\\ &\quad \geqq P[\rho(F^*, F_t) \leqq y] \geqq 1 - f_1(y), \end{aligned} \tag{1.12}$$

which proves the former half of the theorem. The latter half can be proved similarly.

Definition 1.6. If a pseudo distance ρ is symmetric i.e. $\rho(F_1, F_2) = \rho(F_2, F_1)$, then it is called a distance and denoted by d.

Lemma 1.7.

$$|d(F_0, F_t) - d(F_0, F^*)| \leqq d(F_t, F^*). \tag{1.13}$$

Assumption 1.8. Let $f(y) = f_n(y)$ be a real continuous decreasing function satisfying

$$P[d(F^*, F_t) \geqq y] \leqq f(y) \qquad \text{for any } F_t \in S \text{ and any } y > 0. \tag{1.14}$$

Such a function f will be called a majorant with respect to d.

The following two theorems follow from this Assumption similarly as Theorem 1.4 and Theorem 1.5 followed from Assumption 1.3.

Theorem 1.9.

(i) *If* $\lim_{n\to\infty} f_n(\varepsilon) = 0$ *for any* $\varepsilon > 0$, *then* $d(F_0, F^*)$ *is a consistent estimate of* $d(F_0, F_t)$.

(ii) *A confidence interval of* $d(F_0, F_t)$ *with at least* $100(1-\alpha)\%$ *confidence coefficient is given by*

$$[d(F_0, F^*) - f^{-1}(\alpha), d(F_0, F^*) + f^{-1}(\alpha)]. \tag{1.15}$$

THEOREM 1.10. *For the testing problem*

$$H: d(F_0, F_t) \leqq d_0, \qquad A: d(F_0, F_t) > d_0, \tag{1.16}$$

with a fixed constant d_0, *the test procedure*:

$$\text{When} \quad d(F_0, F^*) < d_0 + f^{-1}(\alpha), \quad \text{we make no decision.} \tag{1.17}$$

$$\text{When} \quad d(F_0, F^*) \geqq d_0 + f^{-1}(\alpha), \quad \text{we reject } H,$$

has level not greater than α *and power not less than* $1 - f(y)$ *for* F_0 *such that* $d(F_t, F_0) = d_0 + f^{-1}(\alpha) + y$.

Now we shall introduce two examples of pseudo distance and another specific pseudo-distance-like quantity, for which more extensive results than in general formulation can be obtained.

CASE 1.

$$\rho_K(F_1, F_2) = \int_\Omega (F_1(\mathbf{x}) - F_2(\mathbf{x}))\, dK(\mathbf{x}) \tag{1.18}$$

and

$$d_K(F_1, F_2) = \int_\Omega |F_1(\mathbf{x}) - F_2(\mathbf{x})|\, dK(\mathbf{x}), \tag{1.19}$$

where K is a probability distribution over Ω, i.e. $\int_\Omega dK(\mathbf{x}) = 1$. Then, we have the following theorem which can be proved by Theorem A.4.

THEOREM 1.11. *Define*

$$f_1(y) = f_2(y) = e^{-2ny^2} \qquad (y > 0), \tag{1.20}$$

$$f(y) = 2\, e^{-2ny^2} \qquad (y > 0). \tag{1.21}$$

Then, for any probability distribution K *over* Ω, f_1 *and* f_2 *are majorants with respect to* ρ_K *defined by* (1.18), *while* f *is a majorant with respect to* d_K *defined by* (1.19).

CASE 2. For the pseudo distance and the distance defined by

$$\rho(F_1, F_2) = \sup_K \rho_K(F_1, F_2) = \sup_{\mathbf{x}} (F_1(\mathbf{x}) - F_2(\mathbf{x})), \tag{1.22}$$

$$d(F_1, F_2) = \sup_K d_K(F_1, F_2) = \sup_{\mathbf{x}} |F_1(\mathbf{x}) - F_2(\mathbf{x})|, \tag{1.23}$$

we obtain the following two theorems, of which the former is obtained from Corollary A.7.

THEOREM 1.12. *If* F_t *is a discrete probability distribution over* Ω, *and if*

$$f_1(y) = f_2(y) = e^{-2nz^2}, \tag{1.24}$$

where y *is the function of* z *satisfying* $y = z + \log(t-1)/4nz$,

$$f(y) = e^{-2nz^2}, \tag{1.25}$$

where y is the function of z satisfying $y = z+\log 2(t-1)/4nz$, $t = \prod_{i=1}^{s} t_i$, t_i is the number of mass points of the component x_i with respect to F_t, then f_1 and f_2 are majorants with respect to ρ defined by (1.22) *while f is a majorant with respect to the distance d defined by* (1.23).

THEOREM 1.13. *For the testing problem*

$$H: \rho(F_0, F_t) \leqq \rho_0, \qquad A: \rho(F_0, F_t) > \rho_0 \tag{1.26}$$

with ρ defined in (1.22) *and a fixed constant ρ_0, the test procedure*:

$$\begin{aligned} &\text{When} \quad \rho(F_0, F^*) < \rho_0 + f_2^{-1}(\alpha), \quad \text{we make no decision,} \\ &\text{When} \quad \rho(F_0, F^*) \geqq \rho_0 + f_2^{-1}(\alpha), \quad \text{we reject } H, \end{aligned} \tag{1.27}$$

has level not greater than α and power not less than $1-\exp(-2ny^2)$ when $\rho(F_0, F_t) = \rho_0 + f_2^{-1}(\alpha) + y$. For the dual test problem with F_0 and F_t interchanged, a similar result obtains. For the problem

$$H: d(F_0, F_t) \leqq d_0, \qquad A: d(F_0, F_t) > d_0 \tag{1.28}$$

with d defined in (1.23), *the procedure:*

$$\begin{aligned} &\text{When} \quad d(F^*, F_0) < d_0 + f^{-1}(\alpha), \quad \text{we make no decision,} \\ &\text{When} \quad d(F^*, F_0) \geqq d_0 + f^{-1}(\alpha), \quad \text{we reject } H, \end{aligned} \tag{1.29}$$

has level not greater than α and power not less than $1-\exp(-2ny^2)$, when $d(F_t, F_0) = d_0 + f^{-1}(\alpha) + y$.

PROOF. The level of the test is given by the same method as (1.11). From (1.22) it follows that there exists a sequence of points $\{\mathbf{x}_m\}$ such that $\rho(F_0, F_t) = \lim_{m\to\infty} [F_0(\mathbf{x}_m) - F_t(\mathbf{x}_m)]$. Let K_m denote the distribution concentrated on the single point $\mathbf{x}_m$. Then $\rho(F_0, F_t) = \lim_{m\to\infty} \rho_{K_m}(F_0, F_t)$, which implies by the triangular inequality that

$$\rho(F_0, F_t) - \rho(F_0, F^*) \leqq \liminf_{m\to\infty} \rho_{K_m}(F^*, F_t). \tag{1.30}$$

When $\rho(F_0, F_t) = \rho_0 + f_2^{-1}(\alpha) + y$, we have

$$\begin{aligned} &P[\rho(F_0, F^*) \geqq \rho_0 + f_2^{-1}(\alpha)] \\ &\quad = P[\rho(F_0, F_t) - \rho(F_0, F^*) \leqq y] \\ &\quad \geqq P[\liminf_{m\to\infty} \rho_{K_m}(F^*, F_t) \leqq y] \geqq 1 - e^{-2ny^2} \end{aligned} \tag{1.31}$$

because of Theorem A.5.

Next, from (1.22) and (1.23) it follows that $d(F_0, F_t) = \rho(F_0, F_t)$ or $d(F_0, F_t) = \rho(F_t, F_0)$. These two cases can be treated similarly. If the former relation holds and if $d(F_0, F_t) = d_0 + f^{-1}(\alpha) + y$, then

$$\begin{aligned} &P[d(F^*, F_0) \geqq d_0 + f^{-1}(\alpha)] \\ &\quad = P[\rho(F_0, F^*) \geqq d_0 + f^{-1}(\alpha) \quad \text{or} \quad \rho(F^*, F_0) \geqq d_0 + f^{-1}(\alpha)] \\ &\quad \geqq P[\rho(F_0, F^*) \geqq d_0 + f^{-1}(\alpha)] \end{aligned} \tag{1.32}$$

which is bounded by $1-\exp(-2ny^2)$ from below similarly as in (1.31).

CASE 3. We introduce

$$\rho_r(F_1, F_2) = \log\left[\int_A \int_\Omega \exp\{ra(F_1(\mathbf{x})-F_2(\mathbf{x}))\}\, dK(\mathbf{x})\, dL(a)\right]^{r^{-1}} \tag{1.33}$$

where $r > 0$ and K and L are probability distributions over Ω and $A = \{a \mid a = \pm 1\}$ respectively.

This ρ_r is not actually a pseudo distance but has similar properties as a pseudo distance, (1.37) below. If Lebesgue measure over Ω is absolutely continuous with respect to K, then ρ_r is a generalization of ρ and d defined in (1.22), (1.23). In fact,

$$\rho_\infty(F_1, F_2) = \sup_{\mathbf{x}}(F_1(\mathbf{x})-F_2(\mathbf{x})), \qquad \text{if} \quad \int_{a=1} dL(a) = 1, \tag{1.34}$$

$$\rho_\infty(F_1, F_2) = \sup_{\mathbf{x}}(F_2(\mathbf{x})-F_1(\mathbf{x})), \qquad \text{if} \quad \int_{a=-1} dL(a) = 1, \tag{1.35}$$

$$\rho_\infty(F_1, F_2) = \sup_{\mathbf{x}}|F_1(\mathbf{x})-F_2(\mathbf{x})|, \tag{1.36}$$

$$\text{if} \quad \int_{a=1} dL(a) > 0 \quad \text{and} \quad \int_{a=-1} dL(a) > 0.$$

Now, from Hölder's inequality, we have

$$\begin{aligned} \rho_r(F_0, F^*) - \rho_\infty(F_0, F_t) &\leqq \rho_r(F_t, F^*), \\ \rho_{r/2}(F_0, F_t) - \rho_r(F_0, F^*) &\leqq \rho_r(F^*, F_t). \end{aligned} \tag{1.37}$$

For any positive ε, there exists a real number $R = R(F_0, F_t, K, L, \varepsilon)$ such that

$$-\rho_r(F^*, F_t) - \varepsilon \leqq \rho_r(F_0, F^*) - \rho_\infty(F_0, F_t) \qquad \text{for any} \quad r > R. \tag{1.38}$$

In the sequel throughout Section 1, put

$$f(y) = \exp(-2ny^2). \tag{1.39}$$

For the function ρ_r we have the following three theorems, of which the first is the same as Theorem A.6.

THEOREM 1.14. *For any $y > 0$, any $F_t \in S$ and any probability distributions K and L, it holds that*

$$\begin{aligned} P[\rho_{4ny}(F^*, F_t) \geqq y] &\leqq f(y), \\ P[\rho_{4ny}(F_t, F^*) \geqq y] &\leqq f(y). \end{aligned} \tag{1.40}$$

THEOREM 1.15.

(i) *If y is a function of n such that $ny \to \infty$, and $y \to 0$ as $n \to \infty$, then $\rho_{4ny}(F_0, F^*)$ is a consistent estimate of $\rho_\infty(F_0, F_t)$.*

(ii) *For $y = f^{-1}(\alpha)$, the interval* (1.41)

$$(\rho_{4ny}(F_0, F^*) - f^{-1}(\alpha), \infty) \tag{1.41}$$

is a confidence interval of $\rho_\infty(F_0, F_t)$ with at least $100(1-\alpha)\%$ confidence coefficient.

PROOF. From the first equation of (1.37) together with (1.38) it follows that for $4ny > R$,

$$\begin{aligned} & P[|\rho_{4ny}(F_0, F^*) - \rho_\infty(F_0, F_t)| < 2\varepsilon] \\ (1.42) \qquad & \geqq P[\rho_{4ny}(F_t, F^*) < \varepsilon \text{ and } \rho_{4ny}(F^*, F_t) < \varepsilon] \\ & \geqq 1 - 2\exp\{-4ny(\varepsilon - y/2)\} \end{aligned}$$

in view of Lemma A.3. This proves part (i), since $ny \to \infty$ and $y \to 0$ as $n \to \infty$.

Part (ii) holds true, since the first equation of (1.37) implies

$$\begin{aligned} (1.43) \qquad & P[\rho_\infty(F_0, F_t) > \rho_{4ny}(F_0, F^*) - f^{-1}(\alpha)] \\ & \geqq P[\rho_{4ny}(F_t, F^*) < f^{-1}(\alpha)] \geqq 1 - \alpha. \end{aligned}$$

THEOREM 1.16. *For the problem*

$$(1.44) \qquad H: \rho_\infty(F_0, F_t) \leqq \rho_0, \qquad A: \rho_\infty(F_0, F_t) > \rho_0,$$

the procedure:

$$\begin{aligned} (1.45) \qquad & \textit{When} \quad \rho_{4ny_0}(F_0, F^*) < \rho_0 + f^{-1}(\alpha), \quad \textit{we make no decision.} \\ & \textit{When} \quad \rho_{4ny_0}(F_0, F^*) \geqq \rho_0 + f^{-1}(\alpha), \quad \textit{we reject } H, \end{aligned}$$

where $y_0 = f^{-1}(\alpha)$, *has level not greater than* α *and power not less than* $1 - \exp(-2ny^2)$, *when* $\rho_{2ny_0}(F_0, F_t) = \rho_0 + f^{-1}(\alpha) + y$ *with* $y \geqq y_0$.

PROOF. The level of the test is given by

$$\begin{aligned} & P[\rho_{4ny_0}(F_0, F^*) \geqq \rho_0 + f^{-1}(\alpha)] \\ (1.46) \qquad & \leqq P[\rho_{4ny_0}(F_0, F^*) - \rho_\infty(F_0, F_t) \geqq f^{-1}(\alpha)] \\ & \leqq P[\rho_{4ny_0}(F_t, F^*) \geqq f^{-1}(\alpha)] \leqq \alpha. \end{aligned}$$

When $\rho_{2ny_0}(F_0, F_t) = \rho_0 + f^{-1}(\alpha) + y$, $(y \geqq y_0)$ the power of the test is

$$\begin{aligned} & P[\rho_{4ny_0}(F_0, F^*) \geqq \rho_0 + f^{-1}(\alpha)] \\ (1.47) \qquad & = P[\rho_{2ny_0}(F_0, F_t) - \rho_{4ny_0}(F_0, F^*) \leqq y] \\ & \geqq P[\rho_{4ny_0}(F^*, F_t) \leqq y] \geqq P[\rho_{4ny}(F^*, F_t) \leqq y] \geqq 1 - e^{-2ny^2}. \end{aligned}$$

REMARK 1.17. The fact that

$$(1.48) \qquad \lim_{r\to\infty} \rho_r(F^*, F_t) = \sup_x |F^*(x) - F_t(x)|$$

together with the following table which was prepared from Massey's table [6] suggests that the inequality

$$\begin{aligned} (1.49) \qquad & P[\log[\textstyle\int_A \int_\Omega \exp\{4\eta^{\frac{1}{2}}\lambda a(F^*(x) - F_t(x))\}\, dK(x)\, dL(a)]^{1/4\eta^{1/2}\lambda} \leqq \lambda/\eta^{\frac{1}{2}}] \\ & \geqq 1 - e^{-2\lambda^2} \end{aligned}$$

which is equivalent to the first equation of (1.40) provides actually a fairly good approximation.

TABLE 1

λ	.9	1.0	1.2	1.4
$P[\sup_x \lvert F^*(x)-F_t(x)\rvert \leqq \lambda/n^{\frac{1}{2}}],\ n=30$	.65	.76	.90	.96
$\lim_{n\to\infty} P[\sup_x \lvert F^*(x)-F_t(x)\rvert \leqq \lambda/n^{\frac{1}{2}}]$	.607	.730	.888	.960
$1-\exp(-2\lambda^2)$	.80210	.86466	.94387	.98016

2. Two sample case.

DEFINITION 2.1. Let S be the set of all possible $\mathbf{F}$,

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} F^{(1)}(\mathbf{x}) \\ F^{(2)}(\mathbf{x}) \end{bmatrix}, \tag{2.1}$$

where $F^{(j)}(\mathbf{x})$ $(j=1,2)$ denote probability distributions over Ω. We assume that a pseudo distance ρ is defined in the space S. We define

$$S_0 = \left\{ \mathbf{F}(\mathbf{x}) = \begin{bmatrix} F^{(1)}(\mathbf{x}) \\ F^{(2)}(\mathbf{x}) \end{bmatrix} \middle|\, F^{(1)} = F^{(2)} \right\} \tag{2.2}$$

and

$$\rho(\mathbf{F}, S_0) = \inf_{\mathbf{F}_1 \in S_0} \rho(\mathbf{F}, \mathbf{F}_1), \tag{2.3}$$

$$\rho(S_0, \mathbf{F}) = \inf_{\mathbf{F}_1 \in S_0} \rho(\mathbf{F}_1, \mathbf{F}). \tag{2.4}$$

Let

$$\mathbf{F}_t(\mathbf{x}) = \begin{bmatrix} F_t^{(1)}(\mathbf{x}) \\ F_t^{(2)}(\mathbf{x}) \end{bmatrix}, \qquad \mathbf{F}^*(\mathbf{x}) = \begin{bmatrix} F^{(1)*}(\mathbf{x}) \\ F^{(2)*}(\mathbf{x}) \end{bmatrix}, \tag{2.5}$$

where $F^{(j)*}(\mathbf{x})$ $(j=1,2)$ be the empirical distribution function of independent samples of size n_j from $F_t^{(j)}(\mathbf{x})$ and $n_1+n_2=n$. Then we have

LEMMA 2.2.

$$\begin{aligned} -\rho(\mathbf{F}_t, \mathbf{F}^*) &\leqq \rho(S_0, \mathbf{F}_t) - \rho(S_0, \mathbf{F}^*) \leqq \rho(\mathbf{F}^*, \mathbf{F}_t), \\ -\rho(\mathbf{F}^*, \mathbf{F}_t) &\leqq \rho(\mathbf{F}_t, S_0) - \rho(\mathbf{F}^*, S_0) \leqq \rho(\mathbf{F}_t, \mathbf{F}^*). \end{aligned} \tag{2.6}$$

ASSUMPTION 2.3. There exist majorants f_1 and f_2 with respect to the pseudo distance ρ which satisfy

$$\begin{aligned} P[\rho(\mathbf{F}^*, \mathbf{F}_t) \geqq y] &\leqq f_1(y) \qquad \text{for any } \mathbf{F}_t \in S, \\ P[\rho(\mathbf{F}_t, \mathbf{F}^*) \geqq y] &\leqq f_2(y) \qquad \text{for any } \mathbf{F}_t \in S. \end{aligned} \tag{2.7}$$

THEOREM 2.4.

(i) *If* $\lim_{n\to\infty} f_{in}(\varepsilon) = 0$ *for any* $\varepsilon > 0$ $(i=1,2)$, *then* $\rho(S_0, \mathbf{F}^*)$ *and* $\rho(\mathbf{F}^*, S_0)$ *are consistent estimates of* $\rho(S_0, \mathbf{F}_t)$ *and* $\rho(\mathbf{F}_t, S_0)$ *respectively.*

(ii) *A joint confidence interval of* $\rho(S_0, \mathbf{F}_t)$ *and* $\rho(\mathbf{F}_t, S_0)$ *with at least* $100\,(1-\alpha)\,\%$ *confidence coefficient, where* $\alpha = \alpha_1 + \alpha_2$, *is given by*

$$[\rho(S_0, \mathbf{F}^*) - f_2^{-1}(\alpha_2), \rho(S_0, \mathbf{F}^*) + f_1^{-1}(\alpha_1)], \qquad (2.8)$$
$$[\rho(\mathbf{F}^*, S_0) - f_1^{-1}(\alpha_1), \rho(\mathbf{F}^*, S_0) + f_2^{-1}(\alpha_2)].$$

THEOREM 2.5. *For the problem*

$$H: \rho(\mathbf{F}_t, S_0) \leqq \rho_0, \qquad A: \rho(\mathbf{F}_t, S_0) > \rho_0, \qquad (2.9)$$

the procedure:

(2.10) *When* $\rho(\mathbf{F}^*, S_0) < \rho_0 + f_1^{-1}(\alpha)$, *we make no decision.*
When $\rho(\mathbf{F}^*, S_0) \geqq \rho_0 + f_1^{-1}(\alpha)$, *we reject* H,

has level not greater than α *and power not less than* $1 - f_2(y)$ *when* $\rho(\mathbf{F}_t, S_0) = \rho_0 + f_1^{-1}(\alpha) + y$.

ASSUMPTION 2.6. When we speak of a distance d in the space S, we assume the existence of a majorant which satisfies

$$P[d(\mathbf{F}^*, \mathbf{F}_t) \geqq y] \leqq f(y) \qquad \text{for any} \quad \mathbf{F}_t \in S. \qquad (2.11)$$

LEMMA 2.7.

$$|d(S_0, \mathbf{F}_t) - d(S_0, \mathbf{F}^*)| \leqq d(\mathbf{F}^*, \mathbf{F}_t). \qquad (2.12)$$

THEOREM 2.8.
(i) *If* $\lim_{n\to\infty} f(\varepsilon) = 0$ *for any* $\varepsilon > 0$, *then* $d(S_0, \mathbf{F}^*)$ *is a consistent estimate of* $d(S_0, \mathbf{F}_t)$.
(ii) *A confidence interval of* $d(S_0, \mathbf{F}_t)$ *with at least* $100\,(1-\alpha)\,\%$ *confidence coefficient is given by*

$$[d(S_0, \mathbf{F}^*) - f^{-1}(\alpha), d(S_0, \mathbf{F}^*) + f^{-1}(\alpha)]. \qquad (2.13)$$

THEOREM 2.9. *For the problem*

$$H: d(S_0, \mathbf{F}_t) \leqq d_0, \qquad A: d(S_0, \mathbf{F}_t) > d_0, \qquad (2.14)$$

the procedure:

(2.15) *When* $d(S_0, \mathbf{F}^*) < d_0 + f^{-1}(\alpha)$, *we make no decision.*
When $d(S_0, \mathbf{F}^*) \geqq d_0 + f^{-1}(\alpha)$, *we reject* H,

has level not greater than α *and power not less than* $1 - f(y)$ *when* $d(\mathbf{F}_t, S_0) = d_0 + f^{-1}(\alpha) + y$.

Now we shall introduce three specific cases.

CASE 1. For a probability distribution K over Ω,

$$\begin{aligned}\rho_K(\mathbf{F}_1, \mathbf{F}_2) &= (n_1 n_2/n^2)^{\frac{1}{2}} \int_\Omega (1, -1)(\mathbf{F}_1 - \mathbf{F}_2)\, dK(\mathbf{x}) \\ &= (n_1 n_2/n^2)^{\frac{1}{2}} \int_\Omega \{(F_1^{(1)} - F_2^{(1)}) - (F_1^{(2)} - F_2^{(2)})\}\, dK(\mathbf{x})\end{aligned} \qquad (2.16)$$

and

$$d_K(\mathbf{F}_1, \mathbf{F}_2) = (n_1 n_2/n^2)^{\frac{1}{2}} \int_\Omega |(1, -1)(\mathbf{F}_1 - \mathbf{F}_2)|\, dK(\mathbf{x}). \qquad (2.17)$$

THEOREM 2.10. *Define*

$$f_1(y) = f_2(y) = e^{-2ny^2} \quad (y > 0), \tag{2.18}$$

$$f(y) = 2\,e^{-2ny^2} \quad (y > 0). \tag{2.19}$$

Then f_1 and f_2 are majorants with respect to the pseudo distance ρ_K defined by (2.16) *and f is a majorant with respect to d_K defined by* (2.17) (cf. Theorem A.8).

CASE 2. Define a pseudo distance and a distance

$$\rho(\mathbf{F}_1, \mathbf{F}_2) = \sup_K \rho_K(\mathbf{F}_1, \mathbf{F}_2) = (n_1 n_2/n^2)^{\frac{1}{2}} \sup_{\mathbf{x}} \{(F_1^{(1)} - F_2^{(1)}) - (F_1^{(2)} - F_2^{(2)})\} \tag{2.20}$$

$$d(\mathbf{F}_1, \mathbf{F}_2) = \sup_K d_K(\mathbf{F}_1, \mathbf{F}_2) = (n_1 n_2/n^2)^{\frac{1}{2}} \sup_{\mathbf{x}} \left|(F_1^{(1)} - F_2^{(1)}) - (F_1^{(2)} - F_2^{(2)})\right|.$$

THEOREM 2.11. *Define*

$$f_1(y) = f_2(y) = e^{-2nz^2}, \quad \textit{where } y = z + \frac{\log(t-1)}{4nz} \tag{2.21}$$

$$f(y) = e^{-2nz^2}, \quad \textit{where } y = z + \frac{\log 2(t-1)}{4nz} \tag{2.22}$$

$t = t_1 t_2 \cdots t_s$ and t_i is the number of mass points of x_i with respect to $F_t^{(1)}$ and $F_t^{(2)}$, then f_1 and f_2 are majorants with respect to ρ defined by (2.20) *while f is a majorant with respect to the distance d defined by* (2.20) (cf. Corollary A.11).

THEOREM 2.12. *For the problem*

$$H: \rho(\mathbf{F}_t, S_0) \leqq \rho_0, \qquad A: \rho(\mathbf{F}_t, S_0) > \rho_0, \tag{2.23}$$

the procedure:

$$\begin{aligned} &\textit{When} \quad \rho(\mathbf{F}^*, S_0) < \rho_0 + f_1^{-1}(\alpha), \quad \textit{we make no decision.} \\ &\textit{When} \quad \rho(\mathbf{F}^*, S_0) \geqq \rho_0 + f_1^{-1}(\alpha), \quad \textit{we reject } H, \end{aligned} \tag{2.24}$$

has level not greater than α and power not less than $1 - \exp(-2ny^2)$ *when* $\rho(\mathbf{F}_t, S_0) = \rho_0 + f_1^{-1}(\alpha) + y$. *For the problem*

$$H: d(S_0, \mathbf{F}_t) \leqq d_0, \qquad A: d(S_0, \mathbf{F}_t) > d_0, \tag{2.25}$$

the procedure:

$$\begin{aligned} &\textit{When} \quad d(\mathbf{F}^*, S_0) < d_0 + f^{-1}(\alpha), \quad \textit{we make no decision.} \\ &\textit{When} \quad d(\mathbf{F}^*, S_0) \geqq d_0 + f_1^{-1}(\alpha), \quad \textit{we reject } H, \end{aligned} \tag{2.26}$$

has level not greater than α and power not less than $1 - \exp(-2ny^2)$, *when* $d(\mathbf{F}_t, S_0) = d_0 + f^{-1}(\alpha) + y$.

CASE 3. We introduce

$$(2.27)\qquad \rho_r(\mathbf{F}_1,\mathbf{F}_2)=\log\left[\int_A\int_\Omega \exp\{ra(1,-1)(\mathbf{F}_1-\mathbf{F}_2)\}\,dK(\mathbf{x})\,dL(a)\right]^{1/r}$$

where $r>0$ and K and L are probability distributions over Ω and $A=\{a\,|\,a=\pm(n_1n_2/n^2)^{\frac{1}{2}}\}$ respectively. If Lebesgue measure is absolutely continuous with respect to K,

$$(2.28)\qquad \rho_\infty(\mathbf{F}_1,\mathbf{F}_2)=\sup_{\mathbf{x}}(n_1n_2/n^2)^{\frac{1}{2}}(1,-1)(\mathbf{F}_1-\mathbf{F}_2),\qquad \text{if } \int_{a=(n_1n_2/n^2)^{1/2}} dL(a)=1,$$

$$(2.29)\qquad \rho_\infty(\mathbf{F}_1,\mathbf{F}_2)=\sup_{\mathbf{x}}(n_1n_2/n^2)^{\frac{1}{2}}\left|(1,-1)(\mathbf{F}_1-\mathbf{F}_2)\right|,$$

if $\int_{a=(n_1n_2/n^2)^{1/2}} dL(a)>0$ and $\int_{a=-(n_1n_2/n^2)^{1/2}} dL(a)>0$.

Now, from Hölder's inequality

$$(2.30)\qquad \begin{aligned}\rho_r(S_0,\mathbf{F}^*)-\rho_\infty(S_0,\mathbf{F}_t)&\leqq\rho_r(\mathbf{F}_t,\mathbf{F}^*),\\ \rho_{r/2}(S_0,\mathbf{F}_t)-\rho_r(S_0,\mathbf{F}^*)&\leqq\rho_r(\mathbf{F}^*,\mathbf{F}_t).\end{aligned}$$

For any ε, there exists a real number $R=R(F_t,K,L,\varepsilon)$ such that

$$(2.31)\qquad -\rho_r(\mathbf{F}^*,\mathbf{F}_t)-\varepsilon\leqq\rho_r(S_0,\mathbf{F}^*)-\rho_\infty(S_0,\mathbf{F}_t).$$

In the sequel in Section 2, put

$$(2.32)\qquad f(y)=\exp(-2ny^2).$$

THEOREM 2.13. *For any $y>0$, and $\mathbf{F}_t\in S$ and any probability distributions K and L, it holds that*

$$(2.33)\qquad \begin{aligned}P[\rho_{4ny}(\mathbf{F}^*,\mathbf{F}_t)\geqq y]&\leqq f(y),\\ P[\rho_{4ny}(\mathbf{F}_t,\mathbf{F}^*)\geqq y]&\leqq f(y).\end{aligned}$$

Proof follows from Theorem A.10.

THEOREM 2.14.

(i) *If y is a function of n such that $ny\to\infty$ and $y\to 0$ as $n\to\infty$, then $\rho_{4ny}(S_0,\mathbf{F}^*)$ is a consistent estimate of $\rho_\infty(S_0,\mathbf{F}_t)$.*

(ii) *For $y=f^{-1}(\alpha)$, the interval* (2.42) *below is a confidence interval of $\rho_\infty(S_0,\mathbf{F}_t)$ with at least* 100 $(1-\alpha)\%$ *confidence coefficient.*

$$(2.34)\qquad (\rho_{4ny}(S_0,\mathbf{F}^*)-f^{-1}(\alpha),\infty).$$

THEOREM 2.15. *For the problem*

$$(2.35)\qquad H:\rho_\infty(S_0,\mathbf{F}_t)\leqq\rho_0,\qquad A:\rho_\infty(S_0,\mathbf{F}_t)>\rho_0$$

the procedure:

(2.36) *When* $\rho_{4ny_0}(S_0, \mathbf{F}^*) < \rho_0 + f^{-1}(\alpha)$, *we make no decision.*

When $\rho_{4ny_0}(S_0, \mathbf{F}^*) \geqq \rho_0 + f^{-1}(\alpha)$, *we reject* H,

where $y_0 = f^{-1}(\alpha)$, *has level not greater than* α *and power not less than* $1-\exp(-2ny^2)$, *when* $\rho_{2ny_0}(S_0, \mathbf{F}_t) = \rho_0 + f^{-1}(\alpha) + y$ $(y \geqq y_0)$.

REMARK 2.16. The fact that $\lim_{r\to\infty} \rho_r(\mathbf{F}^*, \mathbf{F}_t) = \sup(n_1 n_2/n^2)^{\frac{1}{2}} |(1, -1)(\mathbf{F}^* - \mathbf{F}_t)|$ together with the following table which was prepared from Massey's table [7] suggest that the inequality

$$P[\log[\textstyle\int_A \int_\Omega \exp\{4ny\, a(F^{(1)*} - F^{(2)*})\}\, dK(x)\, dL(a)]^{1/4ny} \tag{2.37}$$

$$\leqq (n_1 n_2/n^2)^{\frac{1}{2}} y] \geqq 1 - e^{-2my^2}, \qquad \text{where } n = 2m$$

which is equivalent to (2.29) provides actually a fairly good approximation.

TABLE 2

m	y	$P[\sup_x \lvert F_1{}^*(x) - F_2{}^*(x)\rvert \leqq y]$, $n_1 = n_2 = m$	$1 - \exp\{-2my^2\}$
5	4/5	.992063	.959238
10	6/10	.987659	.972676
15	7/15	.973752	.961841
20	8/20	.966458	.959238
30	9/30	.929113	.932794
40	10/40	.90293	.917915

3. Several sample case.

DEFINITION 3.1. Put

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} F^{(1)}(\mathbf{x}) \\ F^{(2)}(\mathbf{x}) \\ \vdots \\ F^{(k)}(\mathbf{x}) \end{bmatrix} \tag{3.1}$$

where $F^{(j)}(\mathbf{x})$ $(j = 1, 2 \cdots k)$ are probability distributions over Ω and denote by S the set of all possible $\mathbf{F}$. Assume that a pseudo distance ρ is defined in the space S.

Define

$$(3.2)\qquad S_0 = \left\{ \mathbf{F}(\mathbf{x}) = \begin{bmatrix} F^{(1)}(\mathbf{x}) \\ F^{(2)}(\mathbf{x}) \\ \vdots \\ F^{(k)}(\mathbf{x}) \end{bmatrix} \middle| F^{(1)} = \cdots = F^{(k)} \right\},$$

$$(3.3)\qquad \rho(\mathbf{F}, S_0) = \inf_{\mathbf{F}_1 \in S_0} \rho(\mathbf{F}, \mathbf{F}_1),$$

$$(3.4)\qquad \rho(S_0, \mathbf{F}) = \inf_{\mathbf{F}_1 \in S_0} \rho(\mathbf{F}_1, \mathbf{F})$$

and

$$\mathbf{F}_t = \begin{bmatrix} F_t^{(1)}(\mathbf{x}) \\ \vdots \\ F_t^{(k)}(\mathbf{x}) \end{bmatrix}, \qquad \mathbf{F}^* = \begin{bmatrix} F^{(1)*}(\mathbf{x}) \\ \vdots \\ F^{(k)*}(\mathbf{x}) \end{bmatrix},$$

where $F^{(j)*}$ $(j = 1, 2, \cdots, k)$ are empirical distribution functions of independent random samples of size n_j from $F_t^{(j)}$.

Then we have

LEMMA 3.2.

$$(3.5)\qquad \begin{aligned} -\rho(\mathbf{F}_t, \mathbf{F}^*) &\leqq \rho(S_0, \mathbf{F}_t) - \rho(S_0, \mathbf{F}^*) \leqq \rho(\mathbf{F}^*, \mathbf{F}_t), \\ -\rho(\mathbf{F}^*, \mathbf{F}_t) &\leqq \rho(\mathbf{F}_t, S_0) - \rho(\mathbf{F}^*, S_0) \leqq \rho(\mathbf{F}_t, \mathbf{F}^*). \end{aligned}$$

All results in Section 2 from Assumption 2.3 to Theorem 2.9 remain true in the present situation, which are omitted here but will be referred to as Assumption 3.3 to Theorem 3.9; while three specific cases can be described as follows.

CASE 1.

$$(3.6)\qquad \begin{aligned} \rho_{LK}(\mathbf{F}_1, \mathbf{F}_2) &= \textstyle\int_A \int_\Omega \mathbf{a}'(\mathbf{F}_1 - \mathbf{F}_2)\, dK(\mathbf{x})\, dL(\mathbf{a}), \\ d_{LK}(\mathbf{F}_1, \mathbf{F}_2) &= \textstyle\int_A \int_\Omega |\mathbf{a}'(\mathbf{F}_1 - \mathbf{F}_2)|\, dK(\mathbf{x})\, dL(\mathbf{a}), \end{aligned}$$

where K and L are probability distributions over Ω and A respectively and A is the set of all points $(a_1, \cdots, a_k)$ which satisfy

$$\textstyle\sum_{j=1}^k a_j = 0, \qquad \sum_{j=1}^k b_j^2 = 1, \qquad a_j = (n_j/n)^{\frac{1}{2}} b_j, \qquad j = 1, \cdots, k.$$

THEOREM 3.10. *Define*

$$(3.7)\qquad f_1(y) = f_2(y) = e^{-2ny^2}, \qquad f(y) = 2e^{-2ny^2}.$$

Then f_1 and f_2 are majorants with respect to ρ_{LK} and f is a majorant with respect to d_{LK} (cf. Theorem A.12).

CASE 2. Define

$$(3.8)\qquad d(\mathbf{F}_1, \mathbf{F}_2) = \sup_{\mathbf{x}} \sup_{\mathbf{a} \in A} \mathbf{a}'(\mathbf{F}_1 - \mathbf{F}_2),$$

which implies by Theorem A.16

$$d(\mathbf{F}_t, S_0) = \sup_{\mathbf{x}} \{\textstyle\sum_{j=1}^k (n_j/n)(F_t^{(j)} - F_t^{(\cdot)})^2\}, \tag{3.9}$$

$$d(\mathbf{F}^*, S_0) = \sup_{\mathbf{x}} \{\textstyle\sum_{j=1}^k (n_j/n)(F^{(j)*} - F^{(\cdot)*})^2\},$$

where $F_t^{(\cdot)} = \sum_{j=1}^k (n_j/n) F_t^{(j)}$ and $F^{(\cdot)*} = \sum_{j=1}^k (n_j/n) F^{(j)*}$.

THEOREM 3.11. *If $F_t^{(j)}$ $(j = 1, 2, \cdots, k)$ are discrete probability distributions, $t = t_1 \cdots t_s$ and t_i is the number of mass points of x_i with respect to $F_t^{(1)}, \cdots, F_t^{(k)}$, then*

$$P\left[\sup_{\mathbf{x}}\left\{\sum_{j=1}^k (n_j/n)[(F^{(j)*} - F_t^{(j)}) - (F^{(\cdot)*} - F_t^{(\cdot)})]^2\right\}^{\frac{1}{2}} \geqq (k-1)^{\frac{1}{2}}\left\{z + \frac{\log 2(t-1)(k-1)}{4nz}\right\}\right] \leqq e^{-2nz^2}. \tag{3.10}$$

(cf. Corollary A.15.)

THEOREM 3.12. *For the problem*

$$H: d(\mathbf{F}_t, S_0) \leqq d_0, \qquad A: d(\mathbf{F}_t, S_0) > d_0, \tag{3.11}$$

the procedure:

(3.12) *When* $d(\mathbf{F}^*, S_0) < d_0 + f^{-1}(\alpha)$, *we make no decision.*

When $d(\mathbf{F}^*, S_0) \geqq d_0 + f^{-1}(\alpha)$, *we reject H,*

has level not greater than α and power not less than $1 - \exp(-2ny^2)$ when $d(S_0, \mathbf{F}_t) = d_0 + f^{-1}(\alpha) + y$.

CASE 3. We introduce

$$\rho_r(\mathbf{F}_1, \mathbf{F}_2) = \log [\textstyle\int_A \int_\Omega \exp\{r\mathbf{a}'(\mathbf{F}_1 - \mathbf{F}_2)\}\, dK\, dL]^{1/r}. \tag{3.13}$$

If Lebesgue measure is absolutely continuous with respect to K and $L(U) > 0$ for any neighborhood U of the point $\mathbf{a}$, then we have

$$\rho_\infty(\mathbf{F}_1, \mathbf{F}_2) = \sup_{\mathbf{x}} \sup_{\mathbf{a} \in A} \mathbf{a}'(\mathbf{F}_1 - \mathbf{F}_2), \tag{3.14}$$

and

$$\rho_\infty(S_0, \mathbf{F}) = \sup_{\mathbf{x}} \{\textstyle\sum_{j=1}^k (n_j/n)(F^{(j)} - F^{(\cdot)})^2\}^{\frac{1}{2}}. \tag{3.15}$$

Next, we obtain

$$\rho_r(S_0, \mathbf{F}^*) - \rho_\infty(S_0, \mathbf{F}_t) \leqq \rho_r(\mathbf{F}_t, \mathbf{F}^*), \tag{3.16}$$

$$\rho_{r/2}(S_0, \mathbf{F}_t) - \rho_r(S_0, \mathbf{F}^*) \leqq \rho_r(\mathbf{F}^*, \mathbf{F}_t).$$

Therefore, for any positive ε, there exists $R = R(\mathbf{F}_t, K, L, \varepsilon)$ such that

(3.17) $$-\rho_r(\mathbf{F}^*, \mathbf{F}_t) - \varepsilon \leqq \rho_r(S_0, \mathbf{F}^*) - \rho_\infty(S_0, \mathbf{F}_t) \qquad \text{for any} \quad r > R.$$

THEOREM 3.13. *For any $y > 0$, any $\mathbf{F}_t \in S$ and any probability distributions K and L, it holds that*

(3.18) $$P[\rho_{4ny}(\mathbf{F}^*, \mathbf{F}_t) \geqq y] \leqq f(y),$$
$$P[\rho_{4ny}(\mathbf{F}_t, \mathbf{F}^*) \geqq y] \leqq f(y).$$

(Cf. Theorem A.14.) Theorem 2.14 and Theorem 2.15 can be easily generalized to the present situation.

Acknowledgment. The author expresses his appreciation of indebtedness to Professor Masashi Okamoto, Osaka University, for the help rendered to him when this paper was in preparation. Thanks are also due to Mr. Shiro Yamazoe for valuable discussions.

APPENDIX

LEMMA A.1. *Define $\varphi(\alpha)$ by*

(A.1) $$\varphi(\alpha) = (1/\alpha)\log((e^\alpha - 1)/\alpha) \qquad \textit{for any} \quad \alpha > 0,$$

then $0 < \varphi(\alpha) < 1$ and $\varphi(\alpha)$ is an increasing function.

PROOF. First, it holds that

(A.2) $$0 < \varphi(\alpha) < 1 \Leftrightarrow 1 < (e^\alpha - 1)/\alpha < e^\alpha \Leftrightarrow \alpha < e^\alpha - 1 < \alpha e^\alpha.$$

We can easily obtain the last inequalities, so that we have $0 < \varphi(\alpha) < 1$. Next, the derivative of $\varphi(\alpha)$ is written as

(A.3) $$\varphi'(\alpha) = (1/\alpha^2)\, g(\alpha) \qquad \text{where}$$

(A.4) $$g(\alpha) = -\alpha\varphi(\alpha) + \exp(\alpha - \alpha\varphi(\alpha)) - 1.$$

Now, it holds that

(A.5) $$\lim_{\alpha\to 0} g(\alpha) = 0 \qquad \text{and}$$

(A.6) $$g'(\alpha) = (1/(e^\alpha - 1)^2\alpha)\, h(\alpha) \qquad \text{where}$$

(A.7) $$h(\alpha) = (e^\alpha - 1)^2 - \alpha^2 e^\alpha.$$

Since we have $\lim_{\alpha\to 0} h(\alpha) = 0$ and

(A.8) $$h'(\alpha) = 2\, e^\alpha\{e^\alpha - (1 + \alpha + \alpha^2/2)\} > 0,$$

it follows that $h(\alpha) > 0$ for any $\alpha > 0$, which implies $g'(\alpha) > 0$ in view of (A.6). This and (A.5) yields $g(\alpha) > 0$, which implies in turn $\varphi'(\alpha) > 0$ because of (A.3), thus completing the proof.

LEMMA A.2.

$$\alpha^2/8 \geqq -1+\alpha/(e^\alpha-1)+\log((e^\alpha-1)/\alpha) \qquad \textit{for any} \quad \alpha \neq 0. \tag{A.9}$$

PROOF. First, for $\alpha > 0$, put

$$f(\alpha) = \alpha^2/8+1-\alpha/(e^\alpha-1)-\log((e^\alpha-1)/\alpha). \tag{A.10}$$

Then, we have $\lim_{\alpha\to 0} f(\alpha) = 0$ and

$$\begin{aligned} f'(\alpha) &= \alpha/4+\{(e^\alpha-1-\alpha e^\alpha)/(e^\alpha-1)\}\{1/\alpha-1/(e^\alpha-1)\} \\ &= \alpha/4-\alpha\varphi(\alpha)[1-\varphi(\alpha)]\exp\{\alpha\varphi(\alpha)\varphi(\alpha\varphi(\alpha)) \\ &\qquad +\alpha[1-\varphi(\alpha)]\varphi(\alpha[1-\varphi(\alpha)])-\alpha\varphi(\alpha)\} \end{aligned} \tag{A.11}$$

using $\varphi(\alpha)$ defined by (A.1). Since, by Lemma A.1,

$$\varphi(\alpha[1-\varphi(\alpha)]) < \varphi(\alpha), \qquad \varphi(\alpha\varphi(\alpha)) < \varphi(\alpha), \tag{A.12}$$

we have

$$f'(\alpha) \geqq \alpha/4-\alpha\varphi(\alpha)[1-\varphi(\alpha)] \geqq 0, \tag{A.13}$$

which implies $f(\alpha) > 0$ for any $\alpha > 0$.

Next, for the case $\alpha = -\beta < 0$, put

$$g(\beta) = f(-\beta) = \beta^2/8+1-\exp\{\beta[1-\varphi(\beta)]\}+\beta[1-\varphi(\beta)]. \tag{A.14}$$

Since $\lim_{\beta\to 0} g(\beta) = 0$ and

$$\begin{aligned} g'(\beta) &= \beta/4+[1-\varphi(\beta)-\beta\varphi'(\beta)]\{1-\exp[\beta(1-\varphi(\beta))]\} \\ &= \beta\{1/4-[1-\varphi(\beta)-\beta\varphi'(\beta)][\varphi(\beta)+\beta\varphi'(\beta)]\} \geqq 0, \end{aligned} \tag{A.15}$$

we have $g(\beta) \geqq 0$ for $\beta > 0$, which completes the proof.

Let Ω be the s-dimensional Euclidean space and in the sequel until Theorem A.16 (inclusive) let $F^*(\mathbf{x})$ be the empirical distribution function of a random sample of size n from $F_t(\mathbf{x})$.

LEMMA A.3.

$$E[\exp\{n\alpha(F^*(\mathbf{x})-F_t(\mathbf{x}))\}] \leqq \exp(n\alpha^2/8) \qquad \textit{for any} \quad \mathbf{x}\in\Omega \tag{A.16}$$

and $\alpha > 0$.

PROOF. It is easily seen that the first member of the inequality (A.16) can be written as

$$E = \{F_t(\mathbf{x})e^\alpha+(1-F_t(\mathbf{x}))\}^n \exp\{-n\alpha F_t(\mathbf{x})\}. \tag{A.17}$$

This is a function of $F_t(\mathbf{x})$ and takes the maximum value at the point

$$F_t = (1/\alpha)-1/(e^\alpha-1). \tag{A.18}$$

Substituting this into the right-hand side of (A.17) yields

$$E \leqq \exp[n\{-1+\alpha^2/(e^\alpha-1)+\log((e^\alpha-1)/\alpha)\}] \leqq \exp(n\alpha^2/8). \tag{A.19}$$

by Lemma A.2.

The following theorem is a generalization of Okamoto's Theorem 1 [8] because the latter deals with a single binomial variate, while the former a linear combination of such variates.

THEOREM A.4. *We have*

$$P[\textstyle\int_\Omega (F^*(\mathbf{x})-F_t(\mathbf{x}))\,dK(\mathbf{x}) \geqq y] \leqq e^{-2ny^2}, \tag{A.20}$$

$$P[\textstyle\int_\Omega (F_t(\mathbf{x})-F^*(\mathbf{x}))\,dK(\mathbf{x}) \geqq y] \leqq e^{-2ny^2}, \tag{A.21}$$

$$P[\textstyle\int_\Omega |F^*(\mathbf{x})-F_t(\mathbf{x})|\,dK(\mathbf{x}) \geqq y] \leqq 2\,e^{-2ny^2}, \tag{A.22}$$

for any $y > 0$ and any probability distribution K over Ω.

PROOF. We shall use Chernoff's method [2]. For arbitrary positive θ, the first member of (A.20) can be written as

$$\begin{aligned} P &= P[\exp\{\theta \textstyle\int_\Omega (F^*-F_t)\,dK\} \geqq e^{\theta y}] \\ &\leqq P[\textstyle\int_\Omega \exp\{\theta(F^*-F_t)\}\,dK \geqq e^{\theta y}], \end{aligned} \tag{A.23}$$

since the exponential function is convex. Hence, for $\alpha > 0$, it holds that

$$\begin{aligned} P &\leqq \inf_{\alpha>0} e^{-n\alpha y} E[\textstyle\int_\Omega \exp\{n\alpha(F^*-F_t)\}\,dK] \\ &= \inf_{\alpha>0} e^{-n\alpha y} \textstyle\int_\Omega E[\exp\{n\alpha(F^*-F_t)\}]\,dK \end{aligned} \tag{A.24}$$

by Fubini's theorem. Thus

$$\begin{aligned} P &\leqq \inf_{\alpha>0} e^{-n\alpha y} \sup_{\mathbf{x}} E[\exp\{n\alpha(F^*-F_t)\}] \\ &\leqq \inf_{\alpha>0} \exp\{n(-\alpha y+\alpha^2/8)\} = e^{-2ny^2} \end{aligned} \tag{A.25}$$

by Lemma A.3, which has proved (A.20).

The inequality (A.21) can be proved similarly. As for (A.22), the first member is bounded from above by

$$\begin{aligned} &\inf_{\alpha>0} e^{-n\alpha y} \sup_{\mathbf{x}} E[\exp\{n\alpha\,|F^*-F_t|\,\}] \\ &\leqq \inf_{\alpha>0} e^{-n\alpha y}[\sup_{\mathbf{x}} E[\exp\{n\alpha(F^*-F_t)\}] \\ &\qquad + \sup_{\mathbf{x}} E[\exp\{n\alpha(F_t-F^*)\}]] \\ &\leqq \inf_{\alpha>0} e^{-n\alpha y}\, 2e^{n\alpha^2/8} = 2e^{-2ny^2}, \end{aligned} \tag{A.26}$$

which completes the proof.

THEOREM A.5. *For any sequence* $\{K_m\}$ *of probability distributions over* Ω *and for any* $y > 0$, *it holds that*

$$P[\liminf_{m\to\infty} \textstyle\int_\Omega (F^* - F_t)\, dK_m \geqq y] \leqq e^{-2ny^2}, \tag{A.27}$$

$$P[\liminf_{m\to\infty} \textstyle\int_\Omega (F_t - F^*)\, dK_m \geqq y] \leqq e^{-2ny^2}. \tag{A.28}$$

PROOF.[2] Put $X_m = \int_\Omega (F^* - F_t)\, dK_m$, then for any $\varepsilon > 0$ it holds that

$$\begin{aligned} P[\liminf_{m\to\infty} X_m \geqq y] &\leqq P[\liminf_{m\to\infty} X_m > y - \varepsilon] \\ &\leqq \liminf_{m\to\infty} P[X_m > y - \varepsilon] \end{aligned} \tag{A.29}$$

by Fatou's lemma. By Theorem A.4, the last expression is bounded from above by $\exp\{-2n(y-\varepsilon)^2\}$, which tends to $\exp\{-2ny^2\}$ as $\varepsilon \to 0$ and proves (A.27). A similar argument yields (A.28).

THEOREM A.6. *We have*

$$\begin{aligned} &P[\log [\textstyle\int_A \int_\Omega \exp\{4ny\, a(F^* - F_t)\}\, dK(\mathbf{x})\, dL(a)]^{1/4ny} \geqq y] \\ &\leqq e^{-2ny^2} \qquad \textit{for any} \quad y > 0, \end{aligned} \tag{A.30}$$

where $A = \{a \mid a = \pm 1\}$.

PROOF. The first member of the inequality (A.30) is written as

$$\begin{aligned} &P[\textstyle\int_A \int_\Omega \exp\{4ny\, a(F^* - F_t)\}\, dK\, dL \geqq e^{4ny^2}] \\ &\leqq e^{-4ny^2} \textstyle\int_A \sup_{\mathbf{x}} E[\exp\{4ny\, a(F^* - F_t)\}]\, dL \\ &\leqq e^{-4ny^2} \textstyle\int_A e^{n(4ya)^2/8}\, dL = e^{-2ny^2}, \end{aligned}$$

where the second inequality sign follows from Lemma A.3.

COROLLARY A.7. *If* F_t *is a discrete probability distribution over* Ω, *then for any* $z > 0$,

$$P\left[\sup_{\mathbf{x}} (F^*(\mathbf{x}) - F_t(\mathbf{x})) \geqq z + \frac{\log(t-1)}{4nz}\right] \leqq e^{-2nz^2}, \tag{A.31}$$

$$P\left[\sup_{\mathbf{x}} (F_t(\mathbf{x}) - F^*(\mathbf{x})) \geqq z + \frac{\log(t-1)}{4nz}\right] \leqq e^{-2nz^2}, \tag{A.32}$$

$$P\left[\sup_{\mathbf{x}} |F^*(\mathbf{x}) - F_t(\mathbf{x})| \geqq z + \frac{\log 2(t-1)}{4nz}\right] \leqq e^{-2nz^2}, \tag{A.33}$$

where $t = \prod_{i=1}^s t_i$ *and* t_i *is the number of mass points of the component* x_i *with respect to* F_t.

---

[2] The proof has been simplified to the present form by a suggestion of Professor Masashi Okamoto.

PROOF. Let K be the probability distribution which has the same probability $1/(t-1)$ at each point for which every coordinate is a mass point of the component variate, except for the maximum point which has the maximum value of the mass points for each component. Suppose now

$$\sup_{\mathbf{x}}(F^*(\mathbf{x})-F_t(\mathbf{x})) \geqq z+\log\{(t-1)/4nz\}.$$

Then it follows

$$\int_\Omega \exp[4nz(F^*(\mathbf{x})-F_t(\mathbf{x}))]\,dK(\mathbf{x}) = (t-1)^{-1}\sum_{\mathbf{x}}\exp[4nz(F^*(\mathbf{x})-F_t(\mathbf{x}))]$$

$$\geqq (t-1)^{-1}\exp[4nz\{z+\log\{(t-1)/4nz\}\}] = \exp(4nz^2).$$

This implies

$$P[\sup_{\mathbf{x}}(F^*-F_t) \geqq z+\log\{(t-1)/4nz\}]$$

$$\leqq P[\log(\int_\Omega \exp\{4nz(F^*-F_t)\}\,dK)^{1/4nz} \geqq z] \leqq e^{-2nz^2}$$

in view of Theorem A.6. Equation (A.32) can be proved similarly, while (A.33) follows from Theorem A.6 by using the same K and the probability distribution L which has probability $\frac{1}{2}$ at the two points $a = 1$ and $a = -1$.

All these theorems have their analogues for the two sample case which will be stated without proof, since they are all, except for Corollary A.11, special cases of the several sample problems considered later. Let $F^{(j)*}$ $(j = 1,2)$ be the empirical distribution functions of independent random samples of size n_j from $F_t^{(j)}(x)$. Put $n = n_1+n_2$.

THEOREM A.8. *For any* $y > 0$,

$$P\left[\int_\Omega (n_1 n_2/n^2)^{\frac{1}{2}}(1,-1)\begin{bmatrix}F^{(1)*}-F_t^{(1)}\\F^{(2)*}-F_t^{(2)}\end{bmatrix}dK(\mathbf{x}) \geqq y\right] \leqq e^{-2ny^2}, \tag{A.34}$$

$$P\left[\int_\Omega \left|(n_1 n_2/n^2)^{\frac{1}{2}}(1,-1)\begin{bmatrix}F^{(1)*}-F_t^{(1)}\\F^{(2)*}-F_t^{(2)}\end{bmatrix}dK(\mathbf{x})\right| \geqq y\right] \leqq 2e^{-2ny^2}. \tag{A.35}$$

THEOREM A.9. *For any sequence* $\{K_m\}$ *and any* $y > 0$,

$$P\left[\liminf_{m\to\infty}\int_\Omega (n_1 n_2/n^2)^{\frac{1}{2}}(1,-1)\begin{bmatrix}F^{(1)*}-F_t^{(1)}\\F^{(2)*}-F_t^{(2)}\end{bmatrix}dK_m(\mathbf{x}) \geqq y\right] \tag{A.36}$$

$$\leqq e^{-2ny^2},$$

$$P\left[\liminf_{m\to\infty}\int_\Omega (n_1 n_2/n^2)^{\frac{1}{2}}(1,-1)\begin{bmatrix}F_t^{(1)}-F^{(1)*}\\F_t^{(2)}-F^{(2)*}\end{bmatrix}dK_m(\mathbf{x}) \geqq y\right] \tag{A.37}$$

$$\leqq e^{-2ny^2}.$$

THEOREM A.10. *For any $y > 0$,*

$$\text{(A.38)}\quad P\left[\log\left[\int_A\int_\Omega \exp\left\{4nya(1,-1)\begin{bmatrix}F^{(1)*}-F_t^{(1)}\\ F^{(2)*}-F_t^{(2)}\end{bmatrix}\right\}dK(\mathbf{x})\,dL(a)\right]^{1/4ny} \geqq y\right] \leqq e^{-2ny^2}.$$

COROLLARY A.11. *For any $z > 0$, if $F_t^{(1)}$ and $F_t^{(2)}$ are discrete probability distributions over Ω, then*

$$\text{(A.39)}\quad P[\sup_{\mathbf{x}}(n_1 n_2/n^2)^{\frac{1}{2}}\{(F^{(1)*}-F_t^{(1)})-(F^{(2)*}-F_t^{(2)})\} \geqq z+(4nz)^{-1}\log(t-1)] \leqq e^{-2nz^2},$$

$$\text{(A.40)}\quad P[\sup_{\mathbf{x}}(n_1 n_2/n^2)^{\frac{1}{2}}|(F^{(1)*}-F_t^{(1)})-(F^{(2)*}-F_t^{(2)})| \geqq z+(4nz)^{-1}\log 2(t-1)] \leqq e^{-2nz^2},$$

where $t = \prod_{i=1}^{s} t_i$ and t_i is the number of mass points of x_i with respect to $F_t^{(1)}$ and $F_t^{(2)}$.

Finally, for the several sample case, let $F^{(j)*}(\mathbf{x})$ $(j = 1, 2, \cdots, k)$ be the empirical distribution functions of independent random samples of size n_j from $F_t^{(j)}(\mathbf{x})$, $n = \sum_{j=1}^{k} n_j$ and A be the set of all points $(a_1, \cdots, a_k)$ which satisfy

$$\text{(A.41)}\quad \sum_{j=1}^{k} a_j = 0, \qquad \sum_{j=1}^{k} b_j^2 = 1, \qquad a_j = (n_j/n)^{\frac{1}{2}} b_j,\quad j = 1, 2, \cdots, k.$$

Let K and L denote probability distributions over Ω and A respectively.

THEOREM A.12. *For any K, L and any $y > 0$, we obtain*

$$\text{(A.42)}\quad P[\int_A\int_\Omega \sum_{j=1}^{k}(n_j/n)^{\frac{1}{2}} b_j(F^{(j)*}-F_t^{(j)})\,dK(\mathbf{x})\,dL(\mathbf{a}) \geqq y] \leqq e^{-2ny^2},$$

$$\text{(A.43)}\quad P[\int_A\int_\Omega |\sum_{j=1}^{k}(n_j/n)^{\frac{1}{2}} b_j(F^{(j)*}-F_t^{(j)})|\,dK(\mathbf{x})\,dL(\mathbf{a}) \geqq y] \leqq 2e^{-2ny^2}.$$

PROOF. From Jensen's inequality, the first member of (A.42) is not greater than $P[\int_A\int_\Omega \exp\{n\alpha\sum_{j=1}^{k}(n_j/n)^{\frac{1}{2}} b_j(F^{(j)*}-F_t^{(j)})\}\,dK\,dL \geqq e^{n\alpha y}]$ for any $\alpha > 0$, and hence not greater than

$$\text{(A.44)}\quad \inf_{\alpha>0} e^{-n\alpha y}\int_A\int_\Omega \prod_{j=1}^{k} E[\exp\{n\alpha(n_j/n)^{\frac{1}{2}} b_j(F^{(j)*}-F_t^{(j)})\}]\,dK\,dL.$$

Using Lemma A.3 as in the proof of Theorem A.4, we find that (A.44) is not greater than $\inf_{\alpha>0}\exp(-n\alpha y)\exp(n\alpha^2/8) = \exp(-2ny^2)$, which has proved (A.42).

Next, the first member of (A.43) is not greater than

$$\inf_{\alpha>0} e^{-n\alpha y}\int_A\{\sup_{\mathbf{x}} E[\exp\{n\alpha\sum_{j=1}^{k}(n_j/n)^{\frac{1}{2}} b_j(F^{(j)*}-F_t^{(j)})\}] + \sup_{\mathbf{x}} E[\exp\{n\alpha\sum_{j=1}^{k}(n_j/n)^{\frac{1}{2}} b_j(F_t^{(j)}-F^{(j)*})\}]\}\,dL$$
$$\leqq \inf_{\alpha>0} 2e^{-n\alpha y}e^{n\alpha^2/8} = 2e^{-2ny^2}.$$

The following two theorems can be proved similarly as Theorem A.5 and A.6.

THEOREM A.13. *For any sequence* $\{K_m\}$ *and any* $y > 0$,

$$\text{(A.45)} \quad P[\liminf_{m\to\infty} \int_\Omega \sum_{j=1}^k (n_j/n)^{\frac{1}{2}} b_j(F^{(j)*} - F_t^{(j)})\, dK_m(\mathbf{x}) \geqq y] \leqq e^{-2ny^2}.$$

THEOREM A.14. *For any* $y > 0$,

$$\text{(A.46)} \quad P[\log [\int_A \int_\Omega \exp \{4ny \sum_{j=1}^k (n_j/n)^{\frac{1}{2}} b_j(F^{(j)*} - F_t^{(j)})\}\, dK\, dL]^{1/4ny} \geqq y] \leqq e^{-2ny^2}.$$

COROLLARY A.15. *For any* $z > 0$,

$$\text{(A.47)} \quad P\left[\sup_{\mathbf{x}} \left(\sum_{j=1}^k (n_j/n)^{\frac{1}{2}} b_j\{(F^{(j)*} - F_t^{(j)}) - (F^{(\cdot)*} - F_t^{(\cdot)})\}^2\right)^{\frac{1}{2}} \geqq (k-1)^{\frac{1}{2}}\left(z + \frac{\log 2(t-1)(k-1)}{4nz}\right)\right] \leqq e^{-2nz^2},$$

where $t = t_1 \cdots t_s$ *and* t_i *is the number of mass points of* x_i *of* $F_t^{(1)}, \cdots, F_t^{(k)}$.

PROOF. Let K have equal probability $1/(t-1)$ on the mass points except the maximum point and $L(\mathbf{a})$ have equal probability $1/2(k-1)$, on the set of orthogonal vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{k-1}$ and $-\mathbf{a}_1, -\mathbf{a}_2, \cdots, -\mathbf{a}_{k-1}$, then

$$P\left[\sup_{\mathbf{x}} \left(\sum_{j=1}^k (n_j/n)\{(F^{(j)*} - F_t^{(j)}) - (F^{(\cdot)*} - F_t^{(\cdot)})\}^2\right)^{\frac{1}{2}} \geqq (k-1)^{\frac{1}{2}}\left(z + \frac{\log 2(t-1)(k-1)}{4nz}\right)\right]$$

$$= P\left[\sup_{\mathbf{x}} \sup_{\mathbf{a}\in A} \mathbf{a}(F^* - F_t) \geqq (k-1)^{\frac{1}{2}}\left(z + \frac{\log 2(t-1)(k-1)}{4nz}\right)\right]$$

$$\leqq P\left[\exists\, \mathbf{a}_i, \sup_{\mathbf{x}} \mathbf{a}_i(F^* - F_t) \geqq z + \frac{\log 2(t-1)(k-1)}{4nz}\right]$$

$$\leqq P[\log^{1/4nz} (\int_A \int_\Omega \exp\{4nza(F^* - F_t)\}\, dK\, dL)^{1/4nz} \geqq z] \leqq e^{-2nz^2}.$$

THEOREM A.16. *If* $\sum_{j=1}^k (n_j/n)^{\frac{1}{2}} b_j = 0$ *and* $\sum_{j=1}^k b_j^2 = 1$, *then, we have*

$$\text{(A.48)} \quad \max_{\mathbf{b}} \sum_{j=1}^k (n_j/n)^{\frac{1}{2}} b_j x_j = \{\sum_{j=1}^k (n_j/n)(x_j - \bar{x})^2\}^{\frac{1}{2}},$$

where $\bar{x} = \sum_{j=1}^k (n_j/n) x_j$.

REFERENCES

[1] BICKEL, P. J. (1969). A distribution-free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Statist.* **40** 1–23.

[2] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[3] DARLING, D. A. (1957). The Kolmogorov Smirnov, Cramér von Mises tests. *Ann. Math. Statist.* **28** 823–838.

[4] HOEFFDING, W. (1967). On probabilities of large deviation. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 203–220.

[5] KIEFER, K. (1959). K-sample analogues of the Kolmogorov–Smirnov and Cramér–von Mises tests. *Ann. Math. Statist.* **30** 420–449.

[6] MASSEY, F. J. (1950). A note on the estimation of a distribution function by confidence limits. *Ann. Math. Statist.* **21** 116–119.

[7] MASSEY, F. J. (1951). The distribution of the maximum deviation between two sample cumulative step functions. *Ann. Math. Statist.* **22** 125–128.

[8] OKAMOTO, M. (1958). Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.* **10** 29–35.

[9] SCHMIDT, P. (1958). On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions. *Ann. Math. Statist.* **29** 1011–1027.

[10] TANAKA, M. (1962). Distance between distributions. *Sem. Rep. Osaka Statist. Assoc.* **6** 53–84. In Japanese.

[11] TANAKA, M. (1963). Statistical inference for general distributions. *Sem. Rep. Osaka Statist. Assoc.* **7** 101–127. In Japanese.