A BAYESIAN ESTIMATION OF A MEASURE OF THE DIFFERENCE BETWEEN TWO CONTINUOUS DISTRIBUTIONS

著者	YAMATO Hajime
journal or publication title	鹿児島大学理学部紀要. 数学・物理学・化学
volume	8
page range	29-38
別言語のタイトル	2つの連続分布の差の尺度のベイズ推定
URL	http://hdl.handle.net/10232/00003960

Rep. Fac. Sci Kagoshima Univ., (Math. Rhys. Chem.)
No.8 pp.29-38, 1975

A BAYESIAN ESTIMATION OF A MEASURE OF THE DIFFERENCE BETWEEN TWO CONTINUOUS DISTRIBUTIONS

By

Hajime YAMATO

(Received September 30, 1975)

0. Summary.

A measure of the difference between two continuous distributions is estimated by a Bayesian method. The proposed estimators are consistent. The absolute value of the difference between one of our estimators and the U.M.V. unbiased estimator is smaller than $2(m/(m^2-1)+n/(n^2-1))$, where m and n are sample sizes.

1. Introduction

A measure of the difference between two distribution functions F and G is given by

$$d(F,G)=\int_{-\infty}^{+\infty}[F(t)-G(t)]^2 d\frac{F(t)+G(t)}{2}. \tag{1.1}$$

It is well-known that $F(t)\equiv G(t)$ if and only if $d(F, G)=0$, for continuous F and G. If the distribution functions F and G are continuous, then the measure $d(F, G)$ can be written as

$$d(F,G)=\frac{4}{3}-\left\{\int_{-\infty}^{+\infty}G(t)dF^2(t)+\int_{-\infty}^{+\infty}F(t)dG^2(t)\right\}, \tag{1.2}$$

(See, for example, Fraser [4], p. 164–165).

Let $X_1,\dots, X_m$ be a sample of size m from an unknown continuous distribution F and $Y_1,\dots, Y_n$ be a sample of size n from an unknown continuous distribution G. We want to estimate a measure of the difference $d(F, G)$. We shall derive an estimator by a Bayesian method. In our problem the parameter space Θ is the set of all continuous distribution functions on the real line, R. Let the action space be the interval $[0, 1]$. Let the loss function be $L(F, G, \hat{d})=(d(F, G)-\hat{d})^2$ for an action $\hat{d}$. Before we say a prior distribution on Θ, following Doksum [2] we define a linearized Dirichlet process. Let a, b be constants with $a<b$ and we choose a set of points $t_1,\dots, t_k$ with $a=t_1<t_2<\cdots<t_k=b$. Then the set of points $\Delta=(t_1, \dots, t_k)$ is called the division into subintervals and we denote $\max_{1\leq i\leq k-1}|t_{i+1}-t_i|$ by $|\Delta|$. Let α be a positive

and σ-additive finite measure on $(R, \mathscr{B})$ with support (a, b), where $\mathscr{B}$ is the σ-field of Borel sets. We recall α determines a Dirichlet process, which has as its realization a discrete distribution function H_0 such that $H_0(a)=0$ and $H_0(b)=1$ with probability one. Given a division Δ of the interval (a, b), the joint distribution of the corresponding increments of the distribution function being a Dirichlet process is denoted by a Dirichlet distribution, and we define a linearized Dirichlet process on Definition 1. For the Dirichlet process see Ferguson [3].

DEFINITION 1. *We say H is a linearized Dirichlet process with the parameters α, Δ when H is linear between the points $(t_1, H_0(t_1))$,...., $(t_k, H_0(t_k))$ and $H_0(t_i)$, $i=1$, ..., are the realization of the Dirichlet process with parameter α having support (a, b), where $a=t_1$, $b=t_k$* and $\Delta(t_1,\dots, t_k)$.

Let F and G be independent and be linearized Dirichlet process with the parameters α, Δ and β, Δ respectively, with α, β dominated by the Lebesgue measure on R. Since we shall estimate a measure $d(F, G)$ with squared error loss, the Bayes estimate is given by

$$E[d(F, G)\,|\,X_1,\dots, X_m, Y_1,\dots, Y_n].$$

We shall derive the above estimate and its limit in section 2.

2. Estimators.

Since $\alpha((-\infty, t_1]\cup[t_k, \infty))=0$ and $\beta((-\infty, t_1]\cup[t_k, \infty))=0$, we have with probability one

$$F(t) = G(t) = 0 \qquad \text{for } t\leqq t_1$$
$$F(t) = G(t) = 1 \qquad \text{for } t\geqq t_k.$$

By the definition we have for $t_i<t<t_{i+1}$, $i=1,\dots, k-1$

$$F(t)=\frac{F(t_{i+1})-F(t_i)}{t_{i+1}-t_i}(t-t_i)+F(t_i),$$

$$G(t)=\frac{G(t_{i+1})-G(t_i)}{t_{i+1}-t_i}(t-t_i)+G(t_i).$$

Therefore we have easily with probability one

$$d(F,G)=\frac{4}{3}-\sum_{i=1}^{k-1}\left\{\int_{t_i}^{t_{i+1}}G(t)dF^2(t)+\int_{t_i}^{t_{i+1}}F(t)dG^2(t)\right\}, \tag{2.1}$$

where

$$\begin{aligned}\int_{t_i}^{t_{i+1}}G(t)dF^2(t)=&\frac{2}{3}[G(t_{i+1})-G(t_i)]\,[F(t_{i+1})-F(t_i)]^2\\&+[G(t_{i+1})+G(t_i)]F(t_i)\,[F(t_{i+1})-F(t_i)]\\&+G(t_i)\,[F(t_{i+1})-F(t_i)]^2\end{aligned} \tag{2.2}$$

and by replacing F with G and G with F we have the equality for $\int_{t_i}^{t_{i+1}} F(t)dG^2(t)$.

Let us put

$$\hat{F}_m(t) = p_{1,m}F_0(t)+(1-p_{1,m})F_m(t) \tag{2.3}$$

$$\hat{G}_n(t) = p_{2,n}G_0(t)+(1-p_{2,n})G_n(t), \tag{2.4}$$

where (i) $p_{1,m}=\alpha(R)/(\alpha(R)+m)$, $p_{2,n}=\beta(R)/(\beta(R)+n)$, (ii) $F_0(t)=\alpha((-\infty,t])/\alpha(R)$, $G_0(t)=\beta((-\infty, t])/\beta(R)$ and (iii) $F_m(t)$ and $G_n(t)$ are the empirical distribution functions of the samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, respectively. For our prior distributions we have

$$\hat{F}_m(t) = \hat{G}_n(t) = 0 \qquad \text{for } t \leqq t_1,$$

$$\hat{F}_m(t) = \hat{G}_n(t) = 1 \qquad \text{for } t \geqq t_k$$

with probability one. Corresponding to the division Δ, we define a distribution function $\hat{F}_{m,\Delta}$ which is linear between the points $(t_1, \hat{F}_m(t_1)), \dots (t_k, \hat{F}_m(t_k))$ and similarly we define a distribution function $\hat{G}_{n,\Delta}$ which is linear between the points $(t_1, \hat{G}_n(t_1)), \dots, (t_k, \hat{G}_n(t_k))$. Then we obviously have

$$\lim_{|\Delta|\to 0} \hat{F}_{m,\Delta}(t) = \hat{F}_m(t) \qquad \text{for } t \in C(\hat{F}_m), \tag{2.5}$$

$$\lim_{|\Delta|\to 0} \hat{G}_{n,\Delta}(t) = \hat{G}_n(t) \qquad \text{for } t \in C(\hat{G}_n), \tag{2.6}$$

where $C(H)$ denotes the all continuity points of H.

Lemma 1. *We have with probability one*

$$E\left[\int_{t_i}^{t_{i+1}} G(t)dF^2(t) \mid X_1, \cdots, X_m, Y_1, \cdots, Y_n\right]$$

$$= \frac{\alpha(R)+m}{\alpha(R)+m+1}\int_{t_i}^{t_{i+1}} \hat{G}_{n,\Delta}(t)d\hat{F}_{,\Delta}(t)$$

$$+\frac{1}{\alpha(R)+m+1}\left\{\frac{2}{3}\int_{t_i}^{t_{i+1}} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)\right.$$

$$\left.+\frac{1}{3}\hat{G}_n(t_{i+1})[\hat{F}_m(t_{i+1})-\hat{F}_m(t_i)]\right\}. \tag{2.7}$$

Proof. The posterior distribution of $(F(t_i, F(t_{i+1})\text{-}F(t_i), 1\text{-}F(t_{i+1}))$ given $X_1, \dots, X_m$ is a Dirichlet distribution with parameter $\left((\alpha+\sum_{j=1}^m \delta_{X_j})(-\infty,t_i]), (\alpha+\sum_{j=1}^m \delta_{X_j})((t_i, t_{i+1}]), (\alpha+\sum_{j=1}^m \delta_{X_j})((t_{i+1}, \infty))\right)$, where δ_z denotes the unite measure concentrating at the point z. The posterior distribution of $(G(t_i), G(t_{i+1})\text{-}G(t_i), 1\text{-}G(t_{i+1}))$ given Y_1,

$\ldots, Y_n$ is a Dirichelt distribution with parameter $\left((\beta+\sum_{j=1}^{n}\delta_{Y_j})((-\infty, t_i)), (\beta+\sum_{j=1}^{n}\delta_{Y_j})\right.$ $\left.((t_i, t_{i+1}]), (\beta+\sum_{j=1}^{n}\delta_{Y_j})((t_{i+1}, \infty))\right)$. Hence by taking the conditional expectation of (2.2) given $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ we have (2.7).

From the lemma 1 we have with probability one

$$E\left[\int_a^b G(t)dF^2(t)\,|\,X_1, \cdots, X_m, Y_1, \cdots, Y_n\right]$$

$$=\frac{\alpha(R)+m}{\alpha(R)+m+1}\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}^2_{m,\Delta}(t)$$

$$+\frac{1}{\alpha(R)+m+1}\left\{\frac{2}{3}\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)\right.$$

$$\left.+\frac{1}{3}\sum_{i=1}^{k-1}\hat{G}_n(t_{i+1})[\hat{F}_m(t_{i+1})-\hat{F}_m(t_i)]\right\}. \tag{2.8}$$

Thus the Bayes estimate is

$$E[d(F, G)\,|\,X_1, \cdots, X_m, Y_1, \cdots, Y_n]$$

$$=\frac{4}{3}-\frac{\alpha(R)+m}{\alpha(R)+m+1}\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}^2_{m,\Delta}(t)$$

$$-\frac{1}{\alpha(R)+m+1}\left\{\frac{2}{3}\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}_{m;\Delta}(t)\right.$$

$$\left.+\frac{1}{3}\sum_{i=1}^{k-1}\hat{G}(t_{i+1})[\hat{F}_m(t_{i+1})-\hat{F}_m(t_i)]\right\}$$

$$-\frac{\beta(R)+n}{\beta(R)+n+1}\int_a^b \hat{F}_{m,\Delta}(t)d\hat{G}^2_{n,\Delta}(t)$$

$$-\frac{1}{\beta(R)+n+1}\left\{\frac{2}{3}\int_a^b \hat{F}_{m,\Delta}(t)d\hat{G}_{n,\Delta}(t)\right.$$

$$\left.+\frac{1}{3}\sum_{i=1}^{k=1}\hat{F}_m(t_{i+1})[\hat{G}_n(t_{i+1})-\hat{G}_n(t_i)]\right\}.$$

Next we shall evaulate this estimate in the limit.

Since the measure β is dominated by the Lebesgue measure on R, $G_0(t)$ is an absolutely continuous distribution function on R and has its derivative $G'_0(t)$.

LEMMA 2. *If* $\max_{a\leq t\leq b} G'_0(t)$ *is finite, then we have with probability one*

$$\lim_{|\Delta|\to 0}\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)=\int_a^b \hat{G}_n(t)d\hat{F}_m(t)\,, \tag{2.9}$$

$$\lim_{|\Delta|\to 0}\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}^2_{m,\Delta}(t)=\int_a^b \hat{G}_n(t)d\hat{F}^2_m(t)\,. \tag{2.10}$$

PROOF. Let $y_1,\dots, y_n$ be the observations of $Y_1,\dots,Y_n$ and $y_0=a$, $y_{n+1}=b$. The points $y_1,\dots, y_n$ are discontinuity points of $\hat{G}_n$ and are with probability one continuity points of $\hat{F}_m$. Now we evaluate the integral

$$\int_a^b \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)=\sum_{j=0}^{n}\int_{y_j}^{y_{j+1}} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)\,. \tag{2.11}$$

Since y_j and y_{j+1} are continuity points of $\hat{F}_m$ with probability one $\hat{F}_m$ is continuous on $[y_j, y_j+\delta]$ and $[y_{j+1}-\delta, y_{j+1}]$ for a sutiable $\delta>0$ with probability one. For this δ, we choose an arbitrary ε with $0<\varepsilon<\delta$ and choose a division into subintervals Δ of (a, b) with $|\Delta|<\varepsilon$. The points in the division Δ which lie on the open interval $(y_j+\delta, y_{j+1}-\delta)$ are denoted by $t_1,\dots, t_{r-1}$, the point in the division Δ which is the largest one below $y_j+\delta$ is denoted by t_0 and the point in the division δ which is the smallest one above $y_{j+1}-\delta$ is denoted by t_r. Then $y_j<t_0\leqq y_j+\delta$ and $y_{j+1}-\delta\leqq t_r<y_{j+1}$. Since $\hat{G}_n$ has the derivative $p_{2,m}G_0'(t)$ on $[t_0, t_r]$, we have for $t\in[y_j+\delta, y_{j+1}-\delta]$

$$\begin{aligned}|\hat{G}_{n,\Delta}(t)-\hat{G}_n(t)|&\leqq \max_{i=0,\cdots,r-1}|\hat{G}_n(t_{i+1})-\hat{G}_n(t_i)|\\ &=\max_{i=0,\cdots,r-1}\hat{G}_n'(\xi_i)|t_{i+1}-t_i|\,,\end{aligned}$$

where $t_i\leqq\xi_i\leqq t_{i+1}$. We shall denote $\max_{a\leqq t\leqq b}\hat{G}_0'(t)$ by M. Then we have $|\hat{G}_{n,\Delta}(t)-\hat{G}_n(t)|<M\varepsilon$ for $t\in[y_j+\delta,y_{j+1}-\delta]$ and therefore

$$\left|\int_{y_j+\delta}^{y_{j+1}-\delta} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)-\int_{y_j+\delta}^{y_{j+1}-\delta} \hat{G}_n(t)d\hat{F}_{m,\Delta}(t)\right|<M\varepsilon. \tag{2.12}$$

On the other hand $\hat{G}_n$ is continuous on $[y_j+\delta, y_{j+1}-\delta]$ and $y_j+\delta, y_{j+1}-\delta\in C(\hat{F}_m)$. Therefore by (2.5) we have

$$\lim_{|\Delta|\to 0}\int_{y_j+\delta}^{y_{j+1}-\delta} \hat{G}_n(t)d\hat{F}_{m,\Delta}(t)=\int_{y_j+\delta}^{y_{j+1}-\delta} \hat{G}_n(t)d\hat{F}_m(t). \tag{2.13}$$

By (2.12) and (2.13) we have

$$\lim_{|\Delta|\to 0}\int_{y_j+\delta}^{y_{j+1}-\delta} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)=\int_{y_j+\delta}^{y_{j+1}-\delta} \hat{G}_n(t)d\hat{F}_m(t). \tag{2.14}$$

In the inequality, $0\leqq\int_{y_j}^{y_j+\delta} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)\leqq\int_{y_j}^{y_j+\delta} d\hat{F}_{m,\Delta}(t)$, the right hand term converges to $\int_{y_j}^{y_j+\delta} d\hat{F}_m(t)$ because $\hat{F}_m$ is continuous at y_j, $y_j+\delta$. Since $\int_{y_j}^{y_j+\delta} d\hat{F}_m(t)$ becomes arbitrary small when we choose an arbitraily small δ, we have

$$\lim_{\delta,|\Delta|\to 0}\int_{y_j}^{y_j+\delta} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t)=0, \tag{2.15}$$

$$\lim_{\delta\to 0}\int_{y_j}^{y_j+\delta} \hat{G}_n(t)d\hat{F}_m(t)=0. \tag{2.16}$$

Similarly we have

$$\lim_{\delta,|\Delta|\to 0}\int_{y_{j+1}-\delta}^{y_{j+1}} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t) = 0, \tag{2.17}$$

$$\lim_{\delta\to 0}\int_{y_{j+1}-\delta}^{y_{j+1}} \hat{G}_n(t)d\hat{F}_m(t) = 0. \tag{2.18}$$

By (2.14), (2.15), (2.16), (2.17) and (2.18), we have

$$\lim_{|\Delta|\to 0}\int_{y_j}^{y_{j+1}} \hat{G}_{n,\Delta}(t)d\hat{F}_{m,\Delta}(t) = \int_{y_j}^{y_{j+1}} \hat{G}_n(t)d\hat{F}_m(t)$$

with probability one and consequently by (2.11) we obtain (2.9) with probability one. Similarly we can show (2.10) with probability one. Thus the lemma is proved.

By applying Lemma 2 to (2.8), under the condition $\max_{a\leq t\leq b} G_0'(t)<\infty$ we have with probability one

$$\lim_{|\Delta|\to 0} E\left[\int_a^b G(t)dF^2(t)\,|X_1,\cdots,X_m, Y_1,\cdots,Y_n\right.$$

$$=\frac{1}{\alpha(R)+m+1}\left\{\int_a^b \hat{G}_n(t)d\hat{F}_m(t)+(\alpha(R)+m)\int_a^b \hat{G}_n(t)d\hat{F}_m^2(t)\right\}.$$

Under the condition $\max_{a\leq t\leq b} F_0'(t)<\infty$ we also have the similar result about the conditional expectation of $\int_a^b F(t)\,dG^2(t)$ given the samples. By using these results to the conditional expectation of (1.2) given the samples, we have

PROPOSITION 1. *If $\max_{a\leq t\leq b} F_0'(t)$ and $\max_{a\leq t\leq b} G_0'(t)$ are finite then we have with probability one*

$$\lim_{|\Delta|\to 0} E[d(F,G)\,|X_1,\cdots,X_m, Y_1,\cdots,Y_n]$$

$$=\frac{4}{3}-\frac{1}{\alpha(R)+m+1}\left\{\int_a^b \hat{G}_n(t)d\hat{F}_m(t)+(\alpha(R)+m\int_a^b \hat{G}_n(t)d\hat{F}_m^2(t)\right\}$$

$$-\frac{1}{\beta(R)+n+1}\left\{\int_a^b \hat{F}_m(t)d\hat{G}_n(t)+(\beta(R)+n)\int_a^b \hat{F}_m(t)d\hat{G}_n^2(t)\right\}.$$

We have derived the proposition 1 based on a particular prior distribution. The author propose a following estimator $\hat{d}(F, G)$ of a measure $d(F, G)$ with continuous F and G,

$$\hat{d}(F,G)=\frac{4}{3}-\frac{1}{\alpha(R)+m+1}\left\{\int_{-\infty}^{\infty} \hat{G}_n(t)d\hat{F}_m(t)+(\alpha(R)+m)\int_{-\infty}^{\infty} \hat{G}_n(t)d\hat{F}_m^2(t)\right\}$$

$$-\frac{1}{\beta(R)+n+1}\left\{\int_{-\infty}^{\infty} \hat{F}_m(t)d\hat{G}_n(t)+(\beta(R)+n)\int_{-\infty}^{\infty} \hat{F}_m(t)d\hat{G}_n^2(t)\right\},$$

where (i), α, β are non-negative, finitely additive and finite measures on $(R, \mathscr{B})$ and (ii) $\hat{F}_m$ (t), $\hat{G}_n$ (t) are given by (2.3), (2.4) with $p_{1,m}=\alpha(R)/(\alpha(R)+(m)$, $p_{2,n}=\beta(R)/(\beta(R)+n)$, F_0 $(t)=\alpha$ $((-\infty, t])/\alpha(R)$, G_0 $(t)=\beta((-\infty, t])/\beta(R)$ and the empirical distribution functions $F_m(t)$, $G_n(t)$. In the estimator $\hat{d}(F, G)$, we may put the prior information about F and G into α and β, respectively. By letting $\alpha(R)$ and $\beta(R)$ tend to zero in $\hat{d}(F, G)$, we have an estimator

$$d^*(F,G)=\frac{4}{3}-\frac{1}{m+1}\left\{\int_{-\infty}^{\infty} G_n(t)dF_m(t)+m\int_{-\infty}^{\infty} G_n(t)dF_m^2(t)\right\}$$

$$-\frac{1}{n+1}\left\{\int_{-\infty}^{\infty} F_m(t)dG_n(t)+n\int_{-\infty}^{\infty} F_m(t)dG_n^2(t)\right\}.$$

This estimator is written as follows,

$$d^*(F,G)=\frac{4}{3}-\frac{1}{mn(m+1)}\left[U_1+\sum_{i=1}^{m} i\max\{j: Y_{(j)}\leqq X_{(i)}\}\right]$$

$$-\frac{1}{mn(n+1)}\left[U_2+\sum_{j=1}^{n} j\max\{i: X_{(i)}\leqq Y_{(j)}\}\right],$$

where

$$U_1=no.\ of\{(i,j):X_i<Y_j\}, \qquad U_2=no.\ of\{(i,j):Y_j<X_i\},$$

$X_{(i)}$ is the i-th smallest order statistic of $X_1,\ldots, X_m$ and $Y_{(j)}$ is the j-th smallest order statistic of $Y_1,\ldots, Y_n$. In the next section, we shall investigate the properties of the estimators $\hat{d}(F, G)$ and d^* (F, G).

The $d(F, G)$ given by (1.2) implies a measure of the difference between two distribution functions F and G if and only if F and G are continuous. Therefore the author did not choose Dirichlet processes as prior distributions to estimate a measure

$d(F, G)$ given by (1.2). By using linearized Dirichlet processes as prior distributions, the author derived the Bayes estimate of a measure $d(F, G)$. The limit of the Bayes estimate suggested the estimators $\hat{d}(F, G)$ and $d^*(F, G)$.

If we regard the $d(F, G)$ given by (1.2) as a quantity made by two distributions F and G which are not necessarily continuous and if we choose Dirichlet processes as prior distributions, then we can directly compute the Bayes estimate of a quantity $d(F, G)$ given by (1.2). This estimate is equal to $\hat{d}(F, G)$.

3. Properties of the estimators.

The estimators $\hat{d}(F, G)$ and $d^*(F, G)$ can be used to estimate $d(F, G)$ with continuous distribution functions F and G. A reason for this is the following

PROPOSIOION 2. *Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be samples of size m and n from continuous distribution functions F and G, respectively. Then the estimator $\hat{d}(F, G)$ converges to $d(F, G)$ as m and n tend to infinity with probability one.*

PROOF. $\hat{G}_n(t)$ converges to $G(t)$ uniformly as n tends to infinity with probability one (See, Ferguson [2], p. 223). Hence in the inequality

$$\left|\int_{-\infty}^{\infty} \{\hat{G}_n(t)-G(t)\}d\hat{F}_m(t)\right| < \sup_t |G_n(t)-G(t)|,$$

the right hand converges to zero as n tends to infinity with probability one. Since G is bounded and continuous, we have

$$\lim_{m\to\infty}\int_{-\infty}^{\infty} G(t)d\hat{F}_m(t) = \int_{-\infty}^{\infty} G(t)dF(t)$$

with probabilityone. Therefore the integral

$$\int_{-\infty}^{\infty} \hat{G}_n(t)d\hat{F}_m(t) = \int_{-\infty}^{\infty} \{\hat{G}_n(t)-G(t)\}d\hat{F}_m(t) + \int_{-\infty}^{\infty} G(t)d\hat{F}_m(t)$$

converges to $\int_{-\infty}^{\infty} G(t)dF(t)$ as m and n tend to infinity with probability one. Similarly we have with probability one

$$\lim_{m,n\to\infty}\int_{-\infty}^{\infty} \hat{G}_n(t)d\hat{F}_m^2(t) = \int_{-\infty}^{\infty} G(t)dF^2(t),$$

$$\lim_{m.n\to\infty}\int_{-\infty}^{\infty} \hat{F}_m(t)d\hat{G}_n(t) = \int_{-\infty}^{\infty} F(t)dG(t),$$

$$\mathrm{lin}_{m,n\to\infty}\int_{-\infty}^{\infty} \hat{F}_m(t)d\hat{G}_n^2(t) = \int_{-\infty}^{\infty} F(t)dG^2(t).$$

From the above four convergences, we have with probability one

$$\lim_{m,n\to\infty} \hat{d}(F,G) = d(F,G). \tag{3.1}$$

Thus the proposition is proved.

It follows that $\hat{d}(F, G)$ and $d^*(F, G)$ are consistent estimators of $d(F, G)$ with continuous F and G. Next we shall compare the estimator $d^*(F, G)$ with the U.M.V. unbiased estimator of $d(F, G)$ in absolutely continuous case. If F and G are absolutely continuous, then the U.M.V. unbiased estimator of $d(F, G)$ is given by

$$\tilde{d}(F,G) = \frac{1}{2}\left\{\frac{1}{\binom{m}{2}\binom{n}{2}}\sum_{i_1<i_2}\sum_{j_1<j_2}\phi(X_{i_1}, X_{i_2}, Y_{j_1}, Y_{j_2}) - \frac{1}{3}\right\}, \tag{2.3}$$

where ϕ

$$\phi(X_1, X_2, Y_1, Y_2) = \begin{cases} 1 & \text{if } \max(X_1, X_2)<\min(Y_1, Y_2) \text{ or} \\ & \quad \max(Y_1, Y_2)<\min(X_1, X_2) \\ 0 & \text{otherwise} \end{cases}$$

(See, Zacks [5], p. 155). Before a comparison we prepare

LEMMA 3. *If F and G are absolutely continuous, then we have with probability one*

$$\tilde{d}(F,G) = \frac{4}{3} - \frac{1}{m-1}\left\{m\int_{-\infty}^{\infty} G_n(t)dF_m^2(t) - \int_{-\infty}^{\infty} G_n(t)dF_m(t)\right\}$$
$$- \frac{1}{n-1}\left\{n\int_{-\infty}^{\infty} F_m(t)dG_n^2(t) - \int_{-\infty}^{\infty} F_m(t)dG_n(t)\right\}. \tag{3.3}$$

PROOF. Since the right hand of (3.3) is an unbiased estimator of $d(F, G)$, symmetric in $X_1, \ldots, X_m$ and symmetric in $Y_1, \ldots, Y_n$, the right hand of (3.3) is identical with $\tilde{d}(F, G)$ with probability one. Thus the lemma is proved.

PROPOSITION 3. *Let $X_1, \ldots, X_m$ be a sample of size m from a distribution F and $Y_1, \ldots, Y_n$ be a sample of size n from a distribution G. If F and G are absolutely continuous, then we have with probability one*

$$|d^*(F, G) - \tilde{d}(F, G)| \leqq 2\left(\frac{m}{m^2-1} + \frac{n}{n^2-1}\right). \tag{4.3}$$

PROOF. By the inequality

$$0 \leqq \int_{-\infty}^{\infty} G_n(t)dF_m^2(t) \leqq 2\int_{-\infty}^{\infty} G_n(t)dF_m(t),$$

we have

$$\left| \int_{-\infty}^{\infty} G_n(t)dF_m^2(t) - \int_{-\infty}^{\infty} G_n(t)dF_m(t) \right| \leqq \int_{-\infty}^{\infty} G_n(t)dF_m(t) \leqq 1. \tag{3.5}$$

Similarly we have

$$\left| \int_{-\infty}^{\infty} F_m(t)dG_n^2(t) - \int_{-\infty}^{\infty} F_m(t)dG_n(t) \right| \leqq 1. \tag{3.6}$$

By applying (3.5) and (3.6) to the equation

$$d^*(F,G) - \hat{d}(F,G) = \frac{2m}{m^2-1}\left\{ \int_{-\infty}^{\infty} G_n(t)dF_m^2(t) - \int_{-\infty}^{\infty} G_n(t)dF_m(t) \right\}$$

$$+ \frac{2n}{n^2-1}\left\{ \int_{-\infty}^{\infty} F_m(t)dG_n^2(t) - \int_{-\infty}^{\infty} F_m(t)dG_n(t) \right\},$$

we have (3.4) with probability one. Thus the proposition is proved.

If $m \leqq n$ and $\lim_{m\to\infty} \frac{m}{n}$ exists, then from Proposition 3

$$\lim_{m\to\infty} \left[\sqrt{m}\{d^*(F,G) - d(F,G)\} - \sqrt{m}\{\hat{d}(F,G) - d(F,G)\} \right] = 0$$

with probability one and then by Theorem 5.6 of Fraser [4], p. 229, $\sqrt{m}\,\{\hat{d}(F, G) - d(F, G)\}$ has a limiting normal distribution with mean zero. Therefore by Theorem 4.1 of Billingsley [1], p. 25, if $m \leqq n$ and $\lim_{m\to\infty} \frac{m}{n}$ exists, then $\sqrt{m}\{d^*(F, G) - d(F, G)\}$ has a limiting normal distribution with mean zero.

The author wishes to thank Prof. A. Kudo of Kyushu University for his encouragments and advices.

References

[1] P. BILLINGSLEY (1968), *Convergence of Probability Measures*, John Wiley Sons.

[2] K.A. DOKSUM (1972), *Decision Theory for Some Nonparametric Models*, Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I: Theory of Statistics, p. 331–343.

[3] T.S. FERGUSON (1973), *A Bayesian Analysis of Some Nonparametric Problems*, Annals of Statistics, Vol. 1, No. 2, p. 209–230.

[4] D.A.S. FRASER (1963), *Nonparametric Methods in Statistics*, John Wiley Sons.

[5] S. ZACKS (1971), *The Theory of Statistical Inference*, John Wiley Sons.