Statistics & Probability Letters 7 (1989) 179–185
North-Holland

December 1988

TESTING FOR DISPERSIVE ORDERING

Ibrahim A. AHMAD and Subhash C. KOCHAR *

Division of Statistics, Nothern Illinois University, DeKalb, IL 60115, USA

Received December 1987
Revised April 1988

Abstract: An asymptotically distribution-free test is proposed and studied for testing the null hypothesis H_0: $F \overset{\text{disp}}{=} G$ versus H_1: $F \overset{\text{disp}}{<} G$, that is $G^{-1}(\beta) - G^{-1}(\alpha) \geqslant F^{-1}(\beta) - F^{-1}(\alpha)$ for $0 \leqslant \alpha < \beta \leqslant 1$. The proposed test is based on an estimator of $\int f^2(x)\,\mathrm{d}z - \int g^2(x)\,\mathrm{d}x$, where $f(g)$ is the probability density function corresponding to the distribution function $F(G)$. The cases of two independent samples as well as that of paired samples are discussed in details.

AMS 1980 Subject Classification: Primary 62G05, 62E20; Secondary: 62G10.

Keywords: nonparametric, asymptotically distribution-free, density estimate, kernel, window size.

1. Introduction

Let X and Y be two random variables having absolutely continuous distribution functions (d.f.'s) F and G, respectively, with F^{-1} and G^{-1} as their left continuous inverses.

Definition 1.1. G is said to be more dispersed than F, written $F \overset{\text{disp}}{<} G$, if

$$G^{-1}(\beta) - G^{-1}(\alpha) \geqslant F^{-1}(\beta) - F^{-1}(\alpha) \quad \text{for } 0 \leqslant \alpha < \beta \leqslant 1, \tag{1.1}$$

i.e., if any two quantiles of G are at least as far apart as the corresponding quantiles of F.

The above definition is equivalent to saying that $G^{-1}F(x) - x$ is nondecreasing in x. Doksum (1969) calls this ordering as tail ordering.

Let f and g denote the probability density functions of F and G, respectively. The failure (hazard) rate of F is defined as

$$r_F(x) = f(x)/[1 - F(x)], \quad F(x) < 1.$$

* On leave from Panjab University, Chandigarh, INDIA.

Differentiating $G^{-1}F(x) - x$, we find that (1.1) is equivalent to

$$g\left[G^{-1}(u)\right] \leqslant f\left[F^{-1}(u)\right] \quad \text{for } 0 \leqslant u \leqslant 1. \tag{1.2}$$

Equivalently,

$$r_G\left[G^{-1}(u)\right] \leqslant r_F\left[F^{-1}(u)\right] \quad \text{for } 0 \leqslant u \leqslant 1. \tag{1.3}$$

The above several equivalent versions of dispersive ordering have been discussed by many authors including Doksum (1969), Bickel and Lehmann (1979), Oja (1981), Lewis and Thompson (1981), and Shaked (1982). For positive random variables, which are mainly applicable in reliability theory, Deshpande and Kochar (1983), Bartoszewicz (1986), and Ahmad, Alzaid, Bartoszewicz, and Kochar (1986), have discussed relations between dispersive ordering, star ordering, superadditive ordering, and failure rate ordering.

Note that if $F(x) = H((x - \theta_1)/\eta_1)$ and $G(x) = H((x - \theta_2)/\eta_2)$, then $F \overset{\text{disp}}{<} G$ if and only if $\eta_1 < \eta_2$.

It is possible sometimes to compare two distributions F and G not necessarily belonging to the same location-scale family. Doksum (1969) has shown that for distributions symmetric about 0, if F is ordered with respect to G in the sense of

Lawrence (1975), i.e., $G^{-1}F(x)/x$ is nondecreasing (nonincreasing) for x positive (negative) on the support of F, and if $g(0) < f(0)$, then $F \overset{\text{disp}}{<} G$. From this and the work of Rivest (1985), it follows that if t_n denotes the Student's t distribution with n degrees of freedom, then $t_n \overset{\text{disp}}{<} t_m$ for all $n > m$. Bickel and Lehmann (1979) and Shaked (1982) contain some more examples of pairs of distributions which do not belong to the same location-scale family but they are compatible according to dispersive ordering.

We say that F is equivalent to G in the dispersive ordering sense ($F \overset{\text{disp}}{=} G$) if and only if $F \overset{\text{disp}}{<} G$ and $G \overset{\text{disp}}{<} F$. It is easy to see that $F \overset{\text{disp}}{=} G$ is equivalent to $F(x) = G(x+\theta)$ for some θ and all real x.

In this paper, we discuss the problem of testing the null hypothesis H_0: $F \overset{\text{disp}}{=} G$ against the alternative H_1: $F \overset{\text{disp}}{<} G$ and $F(x) \neq G(x+\theta)$. We present test statistics for the two-sample case and also for the paired-sample case. The tests presented are based on consistent estimators of $\int f^2(x)\,\mathrm{d}x$ and $\int g^2(x)\,\mathrm{d}x$, where the integration is taken over the whole real line whenever the limits are not given.

2. Asymptotically distribution-free test for dispersive ordering

Observe that in order to test H_0: $F \overset{\text{disp}}{=} G$ versus H_1: $F \overset{\text{disp}}{<} G$ and $F(x) \neq G(x+\theta)$, one needs a measure of deviation from the null hypothesis. This will follow from the following lemma.

Lemma 2.1. *If* $F \overset{\text{disp}}{<} G$ *then* $\int f^2(x)\,\mathrm{d}x \geqslant \int g^2(x)\,\mathrm{d}x$, *whenever the densities exist.*

Proof. Note that since $F \overset{\text{disp}}{<} G$ is equivalent to $f(F^{-1}(u)) \geqslant g(G^{-1}(u))$ for all $0 \leqslant u \leqslant 1$, we get $f(x) \geqslant g(G^{-1}F(x))$, for all x and hence

$$\int f(x)\,\mathrm{d}F(x) \geqslant \int g\left(G^{-1}(F(x))\right)\mathrm{d}F(x)$$
$$= \int g(w)\,\mathrm{d}G(w). \qquad \square$$

Thus from the above lemma we use the following measure of deviation from H_0:

$$\Delta(F, G) = \int f^2(x)\,\mathrm{d}x - \int g^2(x)\,\mathrm{d}x \overset{\text{def}}{=} \delta_F - \delta_G. \tag{2.1}$$

Note that under H_0, $\Delta(F, G) = 0$ and under H_1, $\Delta(F, G) > 0$. Now, if we estimate $\Delta(F, G)$, then this estimate can be used as a test statistc that rejects H_0 for large values. Thus we proceed to do in two cases, two independent samples case and the case of paired (dependent) observations.

(A) The two-sample problem: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ denote independent random samples from F and G respectively. In order to estimate δ_F and δ_G (thus estimate $\Delta(F, G)$), we use the estimate proposed in Ahmad (1976),

$$\hat{\delta}_F = \left[m^2 a_m\right]^{-1} \sum_{i=1}^{m} \sum_{j=1}^{m} k\left(\frac{X_i - X_j}{a_m}\right), \tag{2.2}$$

where k is a known pdf which is symmetric and bounded such that $\lim_{|u| \to \infty} |u| k(u) = 0$, and $\{a_m\}$ is a sequence of reals (called window size) such that $a_m \to 0$ as $m \to \infty$. In the same way we estimate δ_G by

$$\hat{\delta}_G = \left[n^2 b_n\right]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} k\left(\frac{Y_i - Y_j}{b_n}\right), \tag{2.3}$$

where $\{b_n\}$ is a sequence of reals such that $b_n \to 0$ as $n \to \infty$. Hence we estimate $\Delta(F, G)$ by:

$$\hat{\Delta}(F, G) = \hat{\delta}_F - \hat{\delta}_G. \tag{2.4}$$

Bhattacharyya and Roussas (1969) proposed to estimate δ_F by $\tilde{\delta}_F = \int \hat{f}^2(x)\,\mathrm{d}x$ where

$$\hat{f}(x) = \left[m a_m\right]^{-1} \sum_{i=1}^{m} k\left(\frac{x - X_i}{a_m}\right).$$

It is not difficult to see that the estimate $\tilde{\delta}_F$ is a special case of $\hat{\delta}_F$, since we can write

$$\tilde{\delta} = \int \hat{f}^2(x)\,\mathrm{d}x$$
$$= \left[m^2 a_m\right]^{-1} \sum_{i=1}^{m} \sum_{i=1}^{m} k^{(2)}\left(\frac{X_i - X_j}{a_m}\right), \tag{2.5}$$

where $k^{(2)}(u)$ is the convolution of $k(u)$ with itself.

In order to provide an asymptotically distribution-free test statistic for H_0, we will discuss the asymptotic distribution of $\hat{\Delta}(F, G)$ and present a consistent estimate of its variance. This we do in the next two theorems whose proofs are deferred to the Appendix.

Theorem 2.1. *Let* $v = \min(m, n)$. *If* $ma_m^2 \to \infty$ *and* $ma_m^4 \to 0$ ($nb_n^2 \to \infty$ *and* $nb_n^4 \to 0$) *as* $v \to \infty$, *if* $f(g) \in L^4(-\infty, \infty)$), *and if* $f(g)$ *has bounded second derivative, then as* $v \to \infty$,

$$\sqrt{v}\left[\hat{\Delta}(F, G) - \Delta(F, G)\right]$$

is asymptotically normally distributed with mean 0 *and variance* σ^2 *given by*

$$\sigma^2 = 4\left\{\int f^3(x)\,\mathrm{d}x - \left(\int f^2(x)\,\mathrm{d}x\right)^2 + \int g^3(x)\,\mathrm{d}x - \left(\int g^2(x)\,\mathrm{d}x\right)^2\right\}, \qquad (2.6)$$

From the above theorem, in order to perform the test, one needs a consistent estimate of σ^2. This we propose as follows:

$$\begin{aligned}\hat{\sigma}^2 &= 4\left\{\int \hat{f}^2(x)\,\mathrm{d}F_m(x) - \left(\hat{\delta}_F\right)^2 + \int \hat{g}^2(x)\,\mathrm{d}G_n(x) - \left(\hat{\delta}_G\right)^2\right\} \\ &= \left[m^3 a_m^2\right]^{-1} \times \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{l=1}^{m} k\left(\frac{X_l - X_j}{a_m}\right) k\left(\frac{X_j - X_i}{a_m}\right) - \hat{\delta}_F^2 \\ &\quad + \left[n^3 b_n^2\right]^{-1} \times \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{l=1}^{n} k\left(\frac{Y_l - Y_j}{b_n}\right) k\left(\frac{Y_j - Y_i}{b_n}\right) \\ &\quad - \hat{\delta}_G^2. \end{aligned} \qquad (2.7)$$

Theorem 2.2. *If* $ma_m^2 \to \infty$ *and* $nb_n^2 \to \infty$ *as* $v \to \infty$, *and if* $f \in L^4(-\infty, \infty)$ *and* $g \in L^4(-\infty, \infty)$, *then* $\hat{\sigma}^2 \to \sigma^2$ *in probability as* $v \to \infty$.

Thus under the conditions of the above two theorems, an asymptotically distribution free test for testing H_0 against H_1 is to reject H_0 if $\sqrt{v}\,\hat{\Delta}(F, G)/\hat{\sigma} > z_\alpha$, where z_α is the $(1-\alpha)$-th quantile of the standard normal distribution. This test is obviously consistent for testing H_0 against H_1. Under some additional conditions $\hat{\sigma}^2$ converges to σ^2 with probability one thus establishing the strong consistency of $\hat{\sigma}^2$. Precisely we have:

Proposition 2.1. *If* k *is a function of bounded variation and if for any* $\varepsilon > 0$, $\sum_{m=1}^{\infty} \exp(-\varepsilon m a_m^2) < \infty$ *and* $\sum_{n=1}^{\infty} \exp(-\varepsilon n b_n^2) < \infty$, *then* $\hat{\sigma}^2$ *converges to* σ^2 *with probability one as* $v \to \infty$.

(B) The paired sample case. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote a random sample from a bivariate distribution with pdf h. Assume that $F(f)$ is the marginal distributions (pdf) of X and that $G(g)$ in the marginal distribution (pdf) of Y. On the basis of the above sample we want to test H_0 versus H_1 as specified above.

Again, we base our test on the statistic $\hat{\Delta}(F, G)$ with $m = n$ and $a_m = b_n$. The following results establish the asymptotic behavior of $\hat{\Delta}$ in this case.

Theorem 2.3. *Under the conditions of Theorem* 2.1, *as* $n \to \infty$, $\sqrt{n}(\hat{\Delta}(F, G) - \Delta(F, G))$ *is asymptotically normal with mean* 0 *and variance* Σ^2 *given by*

$$\begin{aligned}\Sigma^2 &= 4\left[\int f^3(x)\,\mathrm{d}x - \left(\int f^2(x)\,\mathrm{d}x\right)^2 + \int g^3(x)\,\mathrm{d}x - \left(\int g^2(x)\,\mathrm{d}x\right)^2 \right. \\ &\quad - 2\int\int f(x) g(y) h(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad \left. - \left(\int f^2(x)\,\mathrm{d}x\right)\left(\int g^2(x)\,\mathrm{d}x\right)\right] \\ &= \sigma^2 - 8\left\{\int\int f(x) g(y) h(x, y)\,\mathrm{d}x\,\mathrm{d}y - \delta_F \delta_G\right\}. \end{aligned} \qquad (2.8)$$

Thus to proceed with our procedure we need to consistently estimate Σ^2. This is done as follows: Consider

$$\hat{\Sigma}^2 = \hat{\sigma}^2 - 8\left\{\int\int \hat{f}(x)\hat{g}(x)\,\mathrm{d}H_n(x, y) - \hat{\delta}_F \hat{\delta}_G\right\}. \qquad (2.9)$$

Note that

$$\int\int \hat{f}(x)\hat{g}(y)\,\mathrm{d}H_n(x, y)$$

$$= \left(n^3 a_n^2\right)^{-1} \sum_i \sum_j \sum_l k\left(\frac{X_i - X_l}{a_n}\right) k\left(\frac{Y_j - Y_l}{a_n}\right). \tag{2.10}$$

Theorem 2.4. *If $na_n^2 \to \infty$ as $n \to \infty$ and if f and g are in $L^4(-\infty, \infty)$, then $\hat{\Sigma}^2 \to \Sigma^2$ in probability as $n \to \infty$.*

Thus again we can reject H_0 if $\sqrt{n}\,\hat{\Delta}(F, G)/\hat{\Sigma} > z_\alpha$. Also, as in Proposition 2.1, it is possible to establish the strong consistency of $\hat{\Sigma}^2$ under slightly more general conditions than those of Theorem 2.4. Precisely, under conditions of Proposition 2.1, $\hat{\Sigma}^2 \to \Sigma^2$ with probability one as $n \to \infty$.

3. Some remarks

(i) In the statements of the above theorems we employed the customary conditions used to obtain the consistency and asymptotic normality of the so-called kernel method density estimators. For details and literature review, see the books by Prakasa Rao (1983) and Silverman (1986). As pointed out by many authors, the choice of the kernel k is not very crucial but the choice of the window is a serious problem that has been addressed in the literature extensively. The optimal (in the sense of minimizing mean square error or integrated mean square error) choice of the window depends on the unknown pdf f or g. Thus other methods have been suggested. Some are ad hoc (see Scott and Factor (1981) for details and references) while others are based on resampling techniques, e.g. cross validation (see Marron (1987) and references therein).

(ii) A different method for estimating δ_F (and then $\Delta(F, G)$) is possible based on orthogonal series expansion of $f(x)$. Since f is in $L^2(-\infty, \infty)$ then $f(x) = \sum_{j=0}^{\infty} \theta_j \phi_i(x)$, where $\theta_j = \int f(x)\phi_j(x)\,\mathrm{d}x$, and $\{\phi_j\}$ is an orthonormal basis, $j = 0, 1, 2, \ldots$. Thus θ_j can be unbiasedly estimated by

$$\hat{\theta}_j = m^{-1} \sum_{i=1}^{m} \phi_j(X_i)$$

and $f(x)$ by

$$\tilde{f}(x) = \sum_{j=0}^{q(m)} \hat{\theta}_i \phi_i(x),$$

where $q(m)$ is an integervalued function such that $q(m) \to \infty$ as $n \to \infty$. Hence, an estimate of δ_F is

$$\tilde{\delta}_F = m^{-1} \sum_{j=0}^{q(m)} \sum_{i=1}^{m} \hat{\theta}_j \phi_j(X_i)$$

$$= m^{-2} \sum_{i=1}^{m} \sum_{i^*=1}^{m} \sum_{j=0}^{q(m)} \phi_j(X_i)\phi_j(X_{i^*}). \tag{3.1}$$

Define $\tilde{\delta}_G$ similarly, thereby giving an estimate $\tilde{\Delta}(F, G) = \tilde{\delta}_F - \tilde{\delta}_G$ for $\Delta(F, G)$.

Note that $\tilde{\delta}_F = \int \tilde{f}(x)\,\mathrm{d}F_m(x)$. Thus $\tilde{\delta}_F$ resembles $\hat{\delta}_F$ after we replace $\hat{f}(x)$ by $\tilde{f}(x)$. Hence it is possible to discuss the large sample properties of $\tilde{\delta}_F$ and $\tilde{\Delta}(F, G)$ but we shall not present the details here. One observation is in order here; the estimate $\hat{f}(x)$ is a density function while $\tilde{f}(x)$ is not. In fact $\tilde{f}(x)$ may assume negative values, thus making the estimate of $\int f^2(x)\,\mathrm{d}x$ awkward. This is why we opted to work with $\hat{\delta}_F$.

(iii) Bartoszewicz (1986) initiated a test for testing H_0 against H_1 in the one sample case (assuming that G is completely known). However, he has not studied the properties of his test in detail. A competitor to his test would be

$$\hat{\Delta}_0(F, F_0) = \left(\hat{\delta}_F - \delta_{F_0}\right)\sqrt{m}\,/\hat{\sigma}_0,$$

where $\hat{\sigma}_0^2 = 4\{\int \hat{f}^2(x)\,\mathrm{d}F_m(x) - (\hat{\delta}_F)^2\}$. Reject H_0: $F \overset{\text{disp}}{=} F_0$ (completely known) in favor of H_1: $F \overset{\text{disp}}{<} F_0$, if $\sqrt{m}\,\hat{\Delta}_0/\hat{\sigma}_0 > z_\alpha$. Thus our procedure offers a complete solution to this problem.

(iv) Bagai and Kochar (1986) have shown that if $F \overset{\text{disp}}{<} G$ and F or G is IFR, then $r_G(x) \leqslant r_F(x)$. Kochar (1979, 1981) and Bagai and Kochar (1986) have proposed tests for testing the null, hypothesis H_0': $r_F(x) = r_G(x)$ against the alternative H_1: $r_G(x) \leqslant r_F(x)$ for every x.

Appendix

Proof of Theorem 2.1. It follows from the proof of Theorem 2.2 of Ahmad (1976) that

$$\sqrt{m}\left(\hat{\delta}_F - E\hat{\delta}_F\right) = (\sqrt{m})^{-1} \sum_{ii=1}^{m} V_{mi} + o_p(1),$$

where $V_{mi} = q_m(X_i) - Eq_m(X_i)$ with

$$q_m(x) = E\hat{f}(x) = a_m^{-1} \int_{-\infty}^{\infty} k\left(\frac{x-u}{a_m}\right) f(u) \, du.$$

Similarly,

$$\sqrt{n}\left(\hat{\delta}_G - E\hat{\delta}_G\right) = (\sqrt{n})^{-1} \sum_{j=1}^{n} W_{nj} + o_p(1)$$

where $W_{nj} = p_n(Y_j) - Ep_n(Y_j)$ with

$$p_n(y) = E\hat{g}(y) = b_n^{-1} \int_{-\infty}^{\infty} k\left(\frac{y-w}{b_n}\right) g(w) \, dw.$$

But again as in Ahmad (1976, Theorem 2.2),

$$E\,|V_{mi} + W_{nj}|^3 / \sqrt{v}\left[V\left(V_{mi} + W_{nj}\right)\right]^{3/2} \to 0$$
as $v \to \infty$.

Then Layaponouff's central limit theorem applies with

$$V\left(V_{mi} + W_{nj}\right) \to \sigma^2 \quad \text{as } v \to \infty.$$

Finally it follows from Theorem 2.3 of Ahmad (1976) that $\sqrt{v}\,(E\hat{\delta}_F - \delta_F) \to 0$ and $\sqrt{v}\,(E\hat{\delta}_G - \delta_G) \to 0$ as $v \to \infty$. The proof is complete.

Proof of Theorem 2.2. Clearly, cf. Ahmad (1976) and Bhattacharyya and Roussas (1969), $\hat{\delta}_F$ converges in probability to δ_F and so does $\hat{\delta}_G$. Thus it suffices to show that $\int \hat{f}^2(x) \, dF_m(x)$ converges in probability to $\int f^3(x) \, dx$ as $m \to \infty$. But

$$\begin{aligned}
&\left|\int \hat{f}^2(x) \, dF_m(X) - \int f^2(x) \, dF(x)\right| \\
&\leqslant \left|\int \left\{\hat{f}^2(x) - \left[E\hat{f}(x)\right]^2\right\} dF_m(x)\right| \\
&\quad + \left|\int \left(E\hat{f}(x)\right)^2 \left[dF_m(x) - dF(x)\right]\right| \\
&\quad + \left|\int \left[E\hat{f}(x))^2 - f^2(x)\right] dF(x)\right| \\
&= I_{1m} + I_{2m} + I_{3m}, \quad \text{say.}
\end{aligned}$$

Now,

$$I_{1m} \leqslant \sup_x |\hat{f}(x) - E\hat{f}(x)| \times \left\{\int \hat{f}(x) \, dF_m(x) + \int \left[E\hat{f}(x)\right] dF_m(x)\right\}.$$

By Theorem 4.1 of Parzen (1962), $\sup_x |\hat{f}(X) - E\hat{f}(x)|$ converges to 0 in probability as $m \to \infty$ provided that $ma_m^2 \to \infty$ as $m \to \infty$. Also

$$\hat{\delta}_F = \int \hat{f}(x) \, dF_m(x) \to \int f^2(x) \, dx = \delta_F$$

in probability as $m \to \infty$. Finally by the weak law of large numbers

$$\int \left[E\hat{f}(x)\right] dF_m(x) \to \delta_F$$

in probability as $m \to \infty$. Hence $I_{1m} \to 0$ in probability as $m \to \infty$. Writing $\phi_m(x) = E\hat{f}(x)$ and applying the weak law of large numbers we get

$$I_{2m} = \left|\int \phi_m^2(x) \, dF_m(x) - \int \phi_m^2(x) \, dF(x)\right| \to 0$$

in probability as $m \to \infty$. Finally, since $E\hat{f}(x) \to f(x)$ (cf. Parzen (1962)) as $m \to \infty$ for each continuity point x of f, then by the Lebesegue Dominated Convergence Theorem, $I_{3m} \to 0$ as $m \to \infty$.

Proof of Proposition 2.1. Recall the definition of $\hat{\sigma}^2$. Note that by Theorem 2.3 of Ahmad (1976), $\hat{\delta}_F \to \delta_F$ with probability one as $m \to \infty$, under the stated conditions. We need only show that $\int \hat{f}^2(x) \, dF_m(x) \to \int f^3(x) \, dx$ with probability one as $m \to \infty$. But from the proof of Theorem 2.2 we need only show that $I_{1m} \to 0$ and $I_{2m} \to 0$ with probability one. But $I_{2m} \to 0$ with probability one by the strong law of large numbers. Also since $\sup_x |\hat{f}(x) - E\hat{f}(x)| \to 0$ with probability one as $m \to \infty$ by Nadaraya (1965) and $\hat{\delta}_F \to \delta_F$ with probability one as $n \to \infty$ by Ahmad (1976), $I_{1m} \to 0$ with probability one as $m \to \infty$.

Proof of Theorem 2.3. From Theorem 2.1, it is enough to show that $\sqrt{n}\,(\hat{\Delta}(F, G) - E\hat{\Delta}(F, G))$ is asymptotically normal with mean 0 and variance Σ^2. Again,

$$\sqrt{n}\,(\hat{\Delta} - E\hat{\Delta})$$
$$= (\sqrt{n})^{-1}\left\{ \sum_{i=1}^{n} V_{ni} - \sum_{j=1}^{n} W_{nj} \right\} + o_p(1),$$

where V_{in} and W_{jn} are as in Theorem 2.1. Now,

$$n^{-1}V\left(\sum_i V_{ni} - \sum_j W_{nj}\right) = V(V_{n1} - W_{n1}) = V(V_{n1}) + V(W_{n1}) - 2\operatorname{cov}(V_{n1}, W_{n1}).$$

But we have already seen that

$$\lim_n V(V_{n1}) = \int f^3(x)\,\mathrm{d}x - \left(\int f^2(x)\,\mathrm{d}x\right)^2$$

and

$$\lim_n V(W_{n1}) = \int g^3(x)\,\mathrm{d}x - \left(\int g^2(x)\,\mathrm{d}x\right)^2$$

Also it is not difficult to see that

$$\lim_n \operatorname{cov}(V_{n1}, W_{n1}) = \int\int f(x)g(y)h(x, y)\,\mathrm{d}x\,\mathrm{d}y - \int f^2(x)\,\mathrm{d}x \int g^2(y)\,\mathrm{d}y.$$

For, note that

$$\operatorname{cor}(V_{n1}, W_m) = Eq_n(X_i)p_n(Y_1) - Eq_n(X)Ep_n(Y_1),$$

and

$$Eq_n(X_1)p_n(Y_1) = \int\int q_n(x)p_n(y)h(x, y)\,\mathrm{d}x\,\mathrm{d}y \to \int\int f(x)g(y)h(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

as $n \to \infty$. Now using Layaponouff's central limit theorem we have

$$E\,|V_{m1} + W_{n1}|^3/\sqrt{n}\,[V(V_{n1} + W_{ni})]^{3/2} \to 0 \quad \text{as } n \to \infty$$

This completes the proof.

Proof of Theorem 2.4. From Theorem 2.2 and the representation of Σ^2 it is enough to show that $\int\int \hat{f}(x)\hat{g}(y)\,\mathrm{d}H_n(x, y)$ converges in probability to $\int\int f(x)g(y)h(x, y)\,\mathrm{d}x\,\mathrm{d}y$ as $n \to \infty$ but, as $n \to \infty$ we easily see that

$$\{\int\int E\hat{f}(x)E\hat{f}(y)\,\mathrm{d}H(x, y) \to \int\int f(x)g(y)\,\mathrm{d}H(x, y)$$

as $n \to \infty$, under the stated conditions. Next, look at

$$\left|\int\int \hat{f}(x)\hat{g}(y)\,\mathrm{d}H_n(x, y) - \int\int [E\hat{f}(x)][E\hat{g}(y)]\,\mathrm{d}H(x, y)\right|$$
$$\leqslant \left|\int\int \{\hat{f}(x)\hat{g}(y) - [E\hat{f}(x)]\,\hat{g}(y)\}\,\mathrm{d}H_n(x, y)\right|$$
$$+ \left|\int\int \{[E\hat{f}(x)]\,\hat{g}(y) - [E\hat{f}(x)][E\hat{g}(y)]\}\,\mathrm{d}H_n(x, y)\right|$$
$$+ \left|\int\int [E\hat{f}(x)][E\hat{g}(y)] \times [\mathrm{d}H_n(x, y) - \mathrm{d}H(x, y)]\right|$$
$$= J_{1n} + J_{2n} + J_{3n}, \quad \text{say.}$$

$J_{1n} \leqslant \sup_x |\hat{f}(x) - E\hat{f}(x)| \int \hat{g}(y)\,\mathrm{d}G_n(y) \to 0$ in probability as $n \to \infty$, since $\sup_x |\hat{f}(x) - E\hat{f}(x)| \to 0$ in probability by Parzen (1962) and $\int \hat{g}(y)\,\mathrm{d}G_n(y) = \hat{\delta}_G \to \delta_G$ in probability as $n \to \infty$. Secondly, by the weak law of large numbers with $\phi_n(x, y) = E\hat{f}(x) \cdot E\hat{g}(y)$,

$$J_{3n} = \left|\int\int \phi_n(x, y)\,\mathrm{d}H_n(x, n) - \int\int \phi_n(x, y)\,\mathrm{d}H(x, y)\right| \to 0$$

as $n \to \infty$.

Finally,

$$J_{2n} \leqslant \sup_y |\hat{g}(y) - E\hat{g}(y)| \int (E\hat{f}(x))\,\mathrm{d}F_n(x) \to 0$$

in probability as $n \to \infty$.

References

Ahmad, I.A. (1976), On asymptotic properties of an estimate of a functional of a probability density, *Scand. Actuarial J.* **3**, 176–181.

Ahmed, A.N., A. Alzaid, J. Bartoszewicz and S.C. Kochar (1986), Dispersive and superadditive ordering. *Adv. Appl. Probability* **18** 1019–1022.

Bagai, I. and S.C. Kochar (1986), On tail-ordering and comparison of failure rates, *Commun. Statist.- Theor. Meth.* **15**, 1377–1398.

Bartoszewicz, J. (1986), Dispersive ordering and the total time on test transform, *Statist. Prob. Letters* **4**, 285–288.

Bhattacharryya, G.K. and G.G. Roussas (1969), Estimation of a certain functional of a probability density function, *Skand. Aktuar. Tidskr.* **52**, 201–206.

Bickel, P.J. and E.L. Lehmann (1979), Discriptive statistics for nonparametric models IV: Spread, in: J. Jureckova, ed., *Contribution to Statistics* (Academia, Prague) pp. 33–40.

Deshpande, J.V. and S.C. Kochar (1983), Dispersive ordering is the same as tail-ordering, *Adv. Appl. Prob.* **15**, 686–687.

Doksum, K. (1969), Starshaped transformations and the power of rank tests, *Ann. Math. Statist.* **40**, 1167–1176.

Kochar, S.C. (1979), Distribution-free comparisons of two probability distributions with reference to their hazard rates, *Biometrika* **66**, 437–441.

Kochar, S.C. (1981), A new distribution-free test for the equality of two failure rates, *Biometrika* **68**, 423–426.

Lawrence, M.J. (1975), Inequalities of *s*-ordered distributions, *Ann. Statist.* **3**, 413–428.

Lewis, T. and W.J. Thompson (1981), Dispersive distributions and the connection between dispersivity and strong unimodality, *J. Appl. Prob.* **18**, 76–90.

Marron, J.S. (1987), A comparison of cross-validation techniques in density estimation, *Ann. Statist.* **15**, 152–162.

Nadaraya, E.A. (1965), On nonparametric estimation of density functions and regression curves, *Theor. Probability Appl.* **10**, 186–190.

Oja, H. (1981), On location, scale, skewness, and Kurtosis of univariate distributions, *Scand. J. Statist.* **8**, 154–168.

Parzen, E. (1962), On the estimation of a probability density function and mode, *Ann. Math. Statist.* **33**, 1065–1076.

Prakasa Rao, B.L.S. (1983), *Nonparametric Functional Estimation* (Academic Press, New York)

Rivest, L.P. (1982), Product of random variables and star-shaped ordering, *Canad. J. Statist.* **10**, 219–224.

Scott, D.W. and L.E. Factor (1981), Monte Carlo study of three data-based nonparametric probability density estimators, *J. Amer. Statist. Assoc.* **76**, 9–15.

Shaked, M. (1982), Dispersive ordering of distributions, *J. Appl. Probability* **19**, 310–320.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London).