DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 141

February, 1968

ESTIMATION OF A CERTAIN FUNCTIONAL OF A PROBABILITY DENSITY FUNCTION

By

G.K. Bhattacharyya and G.G. Roussas

This research was supported in part by the Wisconsin Alumni Research Foundation.

Estimation of a Certain Functional Of a Probability Density Function

By

G. K. Bhattacharyya
and
G. G. Roussas

1. Introduction. In nonparametric inference, the importance of the functional $\Delta(F) = \int_{-\infty}^{\infty} f^2(x)dx$, where F is the population cdf with density f, could hardly be overemphasized. It is a fundamental quantity involved in the expressions for the asymptotic efficiency of rank tests for many problems like location shift, regression, dependence, analysis of variance, etc. Also in some cases, point as well as interval estimates derived from rank tests have asymptotic efficiency involving the above functional. By a variational argument Hodges and Lehmann [1] derived the lower bound of $\Delta(F)$ over the class of all absolutely continuous F with finite variance. Thus this bound provides the guaranteed asymptotic performance of many nonparametric procedures. To assess the suitability of such a procedure for a specific body of data in the absence of any knowledge of F, an estimate of $\Delta(F)$ would be much desirable.

In this note we consider first the problem of estimation of $\Delta(F)$ from a single sample. Let $X_1, X_2, \ldots, X_n$ be

(1.3) $$\hat{\Delta}(\underset{\sim}{X}) = \hat{\Delta}_{h,K}(\underset{\sim}{X}) = \int f_n^2(x)\,dx.$$

Example 1. Taking $K(x)=1/2(0)$ according as $|x|\leq(>)1$, we have $f_n(x)=(2nh)^{-1}\Sigma_{j=1}^{n}I_j(x)$, where $I_j(x)=1$ if $X_j-h\leq x\leq X_j+h$ and equals 0 otherwise. Noting that $\int I_j(x)dx=2h$ and that, for $i\neq j$, $I_i(x)I_j(x)=1$ if $|X_j-X_i|\leq 2h$ and $\max(X_i,X_j)-h\leq x\leq \min(X_i,X_j)+h$, and equals 0 otherwise, we obtain

(1.4) $$\hat{\Delta}(\underset{\sim}{X}) = (2nh)^{-2}[2nh+2\Sigma^*(2h-|X_j-X_i|)],$$

where Σ^* represents the sum over all $1\leq i<j\leq n$ such that $|X_j-X_i|\leq 2h$. Let $W^{(1)}$ $W^{(2)}$ ... $W^{(a)}$ denote the ordered values of the $a=\binom{n}{2}$ differences $|X_j-X_i|$, and set $W^{(0)}=0$. $W^{(a+1)}=\infty$. Denote by n_o the integer satisfying $W^{(n_o)}\leq 2h$ and $W^{(n_o+1)}>2h$. Then (1.4) readily simplifies to

$$(1.5) \qquad \hat{\Delta}(\underset{\sim}{X}) = (2nh)^{-2}[2nh+4hn_o-2\Sigma_{i=0}^{n_o}W^{(i)}].$$

To calculate the last term one need not compute all the differences $|X_j-X_i|$ and order them. It can be calculated easily from the order statistics $X_{(1)}<X_{(2)}<\ldots<X_{(n)}$ of $\underset{\sim}{X}$. To see this, for $i=1,2,\ldots,n-1$, let a_i be the integer $0\le a_i\le n-1$ such that $0\le X_{(i+a_i)}-X_{(i)}\le 2h$ and $X_{(i+a_i+1)}-X_{(i)}$ $<2h$, where we set $X_{(n+1)}=\infty$. Then we have $\Sigma_0^{n_o}W^{(i)}=\Sigma_{i=1}^{n-1}$ $\Sigma_{j=0}^{a_i}[X_{(i+j)}-X_{(i)}]$ and $n_o=\Sigma_{i=1}^{n-1}a_i$, and hence $\hat{\Delta}(\underset{\sim}{X})$ has apparently the form of a linear combination of order statistics. The coefficients involved, however, depend on h and $\underset{\sim}{X}$.

Example 2. Take $K(x)=\phi(x)$, the standard normal density function. From (1.2) we have $\int f_n^2(x)dx=(nh)^{-2}\Sigma_{i=1}^n\Sigma_{j=1}^n$ $\int\phi[(x-X_i)h^{-1}]\phi[(x-X_j)h^{-1}]dx$ which by straightforward integration yields

$$(1.6) \qquad \hat{\Delta}(\underset{\sim}{X}) = 2^{-1/2}h^{-1}n^{-2}\Sigma_{i=1}^n\Sigma_{j=1}^n\,\phi[(X_i-X_j)2^{-1/2}h^{-1}].$$

In this case $\hat{\Delta}(\underset{\sim}{X})$ is a constant times the average of n^2 standard normal densities evaluated at the points $(X_i-X_j)/(2^{1/2}h)$, and hence it can be computed easily with the help of a table of normal density.

2. Properties.

In this section we study some properties of the class of estimates given by (1.3). Denoting by F and G the cdf's of the random variables X and Y, respectively, one can easily verify the following properties of the functional $\Delta(F)$: (i) for any constand d, $Y=X+d$ implies $\Delta(G)=\Delta(F)$, (ii) for any constant $d>0$, $Y=dX$ implies $\Delta(G)=d^{-1}\Delta(F)$ and (iii) $Y=-X$ implies $\Delta(G)=\Delta(F)$. The following theorem states the similar invariance properties of the estimates (1.3).

Theorem 2.1. Let $\hat{\Delta}_{h,K}(\underset{\sim}{X})$ be the estimate (1.3) based upon h,K and the random variables $\underset{\sim}{X}$, and let $\underset{\sim}{\nu}=(1,1,\ldots,1)$ be the unit n-vector. Then the following properties hold:

(i) for any constant d, $\hat{\Delta}_{h,K}(\underset{\sim}{X}+d\underset{\sim}{\nu})=\hat{\Delta}_{h,K}(\underset{\sim}{X})$

(ii) for any constant $d>0$, $\hat{\Delta}_{h,K}(d\underset{\sim}{X})=d^{-1}\hat{\Delta}_{(h/d),K}(\underset{\sim}{X})$

(iii) if K is symmetric about 0, $\hat{\Delta}_{h,K}(-\underset{\sim}{X})=\hat{\Delta}_{h,K}(\underset{\sim}{X})$.

Proof. Denoting the expression for $f_n(x)$ given in (1.3) by $q_h(x,\underset{\sim}{X})$ we have $q_h(x,\underset{\sim}{X}+d\underset{\sim}{\nu})=q_h(x-d,\underset{\sim}{X})$, $q_h(x,d\underset{\sim}{X})=d^{-1}q_{h/d}(x/d,\underset{\sim}{X})$. Also $K(v)=K(-v)$ implies $q_h(x,-\underset{\sim}{X})=q_h(-x,\underset{\sim}{X})$. Using these results, the proof follows by integration and simple transformation of variables.

The rest of this section is devoted to the study of asymptotic properties of $\hat{\Delta}(\underset{\sim}{X})$. To denote explicitly the sample size we shall henceforth write $\hat{\Delta}(\underset{\sim}{X}_n)$. The following regularity conditions will be needed in the sequel:

(A) $\sup\{K(y);-\infty<y<\infty\}<\infty$ and $yK(y)\to 0$, as $|y|\to\infty$

(B_1) $\lim h(n)=0$ and $\lim nh(n)=\infty$, as $n\to\infty$.

(B_2) $\lim nh^2(n)=\infty$, as $n\to\infty$.

__Theorem 2.2.__ _If the function_ $K(\cdot)$ _satisfies the condition_ (A) _and_ $h=h(n)$ _satisfies_ (B_1), _the estimate_ $\hat{\Delta}(\underset{\sim}{X}_n)$ _given in_ (1.3) _is consistent for_ $\Delta(F)$ _in the mean, that is,_ $E|\hat{\Delta}(\underset{\sim}{X}_n)-\Delta(F)|\to 0$, _as_ $n\to\infty$.

__Proof.__ By Cauchy-Schwarz inequality,

$$E|\int[f_n^2(x)-f^2(x)dx|\le E^{1/2}\int[f_n(x)+f(x)]^2dx \cdot E^{1/2}\int[f_n(x)-f(x)]^2dx \tag{2.1}$$

Using Fubini's theorem and (1.2), we have

$$\begin{aligned} E\int f_n(x)f(x)dx &= h^{-1}\int\int f(x)K[(x-v)h^{-1}]f(v)dxdv \\ &= \int\int f(x)K(z)f(x-hz)dxdz \\ &\le[\int f^2(x)dx]^{1/2}[\int\int f^2(x-hz)K(z)dxdz]^{1/2} \\ &= \int f^2(x)dx. \end{aligned} \tag{2.2}$$

The last equality follows from $\int K(z)dz=1$ and $\int f^2(x-hz)dx=\int f^2(x)dx$. (2.2) implies $\limsup E\int f_n(x)f(x)dx\le\int f^2(x)dx$.

On the other hand, by Theorem 1A of [3], $Ef_n(x) \to f(x)$ at every continuity point of f, and hence an application of Fatou-Lebesgue theorem yields $\int f^2(x)dx \le \lim\inf \int f(x)Ef_n(x)dx$. Combining the last two inequalities, we have

$$(2.3) \qquad \lim E\int f_n(x)f(x)dx = \int f^2(x)dx.$$

Again from (1.2) we have

$$(2.4) \qquad E\int f_n^2(x)dx = (nh^2)^{-1}E\int K^2[(x-X_1)h^{-1}]dx + (n-1)(nh^2)^{-1}\int E^2K[(x-X_1)h^{-1}]dx$$

$$= (nh)^{-1}\int K^2(v)dv + (n-1)n^{-1} \int[\int K(v)f(x-hv)dv]^2dx.$$

From (A) and (B_1), we have $\int K^2(y)dy<\infty$ and $nh\to\infty$. So the first term on the right hand side of (2.4) tends to zero. Writing the integral in the second term as

$$\int\int\int K(v)K(w)f(x-hv)f(x-hw)dvdwdx$$

and applying Cauchy-Schwarz inequality, we obtain

$$\lim\sup E\int f_n^2(x)dx \le \int f^2(x)dx.$$

On the other hand, $h^{-2}E^2K[(x-X_1)h^{-1}] \to f^2(x)$ at every continuity point of f. These results and an application of Fatou-Lebesgue theorem yield

$$(2.5) \qquad \lim E\int f_n^2(x)\,dx = \int f^2(x)\,dx \ .$$

Use of (2.3) and (2.5) in (2.1) completes the proof of the theorem. As a consequence of the above theorem we have immediately

Corollary 2.1. *Under the conditions* (A) *and* (B_1), $\hat{\Delta}(\underset{\sim}{X}_n)$ *is an asymptotically unbiased and consistent estimate of* $\Delta(F)$ *in the probability sense*.

Theorem 2.3. *If* $K(\cdot)$ *satisfies* (A) *and* $h=h(n)$ *satisfies the condition* (B_2) *in addition to* (B_1), *then* $\mathrm{Var}[\hat{\Delta}(\underset{\sim}{X}_n)]\to 0$, *as* $n\to\infty$.

Proof. From (1.2)

$$(2.6) \qquad E\left(\int f_n^2(x)\,dx\right)^2 = (nh)^{-4}\Sigma_1 E\{\int K[(x-X_i)h^{-1}]\, K[(x-X_j)h^{-1}]\,dx \cdot \int K[(x-X_k)h^{-1}]K[(x-X_r)h^{-1}]\,dx\}$$

where Σ_1 is the sum over all $1\le i,j,k,r\le n$. In this sum the total contribution from the terms with i,j,k,r all different is

$$(n)_4 (nh)^{-4} \{E\int K[(x-X_1)h^{-1}]K[(x-X_2)h^{-1}]dx\}^2,$$

where $(n)_k = \binom{n}{k} k!$. As $n\to\infty$, the above quantity converges to $(\int f^2(x)dx)^2$. Since $0 \le h\int K(x-u)K(x-v)dx \le h\int K^2(v)dv < \infty$, the contribution from all the remaining terms is at most $(\int K^2(y)dy)^2 n^{-4} h^{-2} [n^4 - (n)_4]$ which tends to 0, as $n\to\infty$. Hence we have

(2.7) $$\lim E[\int f_n^2(x)dx]^2 = (\int f^2(x)dx)^2 .$$

(2.5) and (2.7) together complete the proof of the theorem.

As an immediate consequence of the above two theorems we have

Corollary 2.2. *Under the conditions* (A), (B_1) *and* (B_2) *the estimate* $\hat{\Delta}(\underset{\sim}{X}_n)$ *is consistent for* $\Delta(F)$ *in quadratic mean*.

3. An Extension and Remarks. A simple extension of our estimation procedure is considered here for the case of several samples from populations differing only in location. For $i=1,2,\ldots,c$, consider random samples $\underset{\sim}{Y}_i = (Y_{i1},Y_{i2},\ldots,Y_{in_i})$ of sizes n_i from cdf's $F_i \varepsilon \mathfrak{F}$. Denote the complete set of $n=\Sigma_{i=1}^{c} n_i$ observations by $\underset{\sim}{Y} = (\underset{\sim}{Y}_1,\underset{\sim}{Y}_2,\ldots,\underset{\sim}{Y}_c)$. For the case $c=2$ and $F_i(x)=F(x-\theta_i)$, $i=1,2$ Lehmann [2] derived a confidence interval for $(\theta_2-\theta_1)$ using the Wilcoxon test. If $D^{(1)}<D^{(2)}<\ldots<D^{(n_1n_2)}$ denote the ordered differences $(Y_{2j}-Y_{1i})$, the $100(1-\alpha)\%$ confidence interval is given by $[\eta_L,\eta_U]$ where $\eta_L=D^{(b)}$, $\eta_U=D^{(n_1n_2+1-b)}$, and for large n_1 and n_2, $b \approx n_1n_2/2 - \tau_{\alpha/2}[nn_1n_2/12]^{1/2}+o(nn_1n_2)^{1/2}$. It is further proved in [2] that

$$(3.1) \qquad n^{1/2}(\eta_U-\eta_L) \to \tau_{\alpha/2}[\{3\lambda(1-\lambda)\}^{1/2}\int f^2(x)dx]^{-1}$$

in probability, if $n_1/n \to\lambda$, $0<\lambda<1$. Hence

$$\tau_{\alpha/2}[n/(3n_1n_2)]^{1/2}(\eta_U-\eta_L)^{-1}$$

provides a consistent estimate of $\int f^2(x)dx$. An extension of this method to the case $c>2$ was considered by Sen[4] cf. p. 1768. No other property of the estimate besides consistency in probability is known. For $c>2$, the method is computationally cumbrous due to the fact that the solution of the system of equations (5.11) of [4] requires repeated ranking of $(Y_{11},\dots,Y_{1n_1},Y_{21}+a_2,\dots,Y_{2n_2}+a_2,\dots,Y_{c1}+a_c,\dots,Y_{cn_c}+a_c)$ with different trial combinations of $a_2,a_3,\dots,a_c$. This is extremely tedious when c and the n_i's are even moderately large.

The extension of our estimate (1.3) to the above situation is straightforward. From each sample $\underset{\sim}{Y}_i$ one can construct an estimate $\hat{\Delta}_{n_i}=\int f_{n_i}^2(x)dx$, where

$$(3.2) \qquad f_{n_i}(x) = (n_i h)^{-1}\Sigma_{j=1}^{n_i}K[(x-Y_{ij})h^{-1}] .$$

Due to translation invariance, $\hat{\Delta}_{n_i}$, $i=1,2,\dots,c$ all estimate $\Delta(F)=\int f^2(x)dx$, if the model $F_i(x)=F(x-\theta_i)$, $i=1,2,\dots,c$ holds. Moreover, these are independently distributed. As the natural combined estimate, we propose

$$(3.3) \qquad \hat{\Delta}_n(\underset{\sim}{Y}) = n^{-1}\Sigma_{i=1}^{c}n_i\hat{\Delta}_{n_i} .$$

Using Theorems 2.1 and 2.2, we have at once the following:

<u>Corollary 3.1</u>. <u>If, for every</u> $i=1,2,\dots,c$, $n_i\to\infty$, <u>as</u> $n\to\infty$, <u>and</u> K <u>and</u> h <u>satisfy the conditions</u> (A) <u>and</u> (B_1), <u>respectively, then</u> $\hat{\Delta}_n(\underset{\sim}{Y})$ <u>given in</u> (3.3) <u>converges in</u>

the mean to $\Delta(F)$. If, in addition, h satisfies (B_2), the convergence holds also in quadratic mean.

Finally, we remark that if the individual cdf F_i are not, in fact, translates of one another, it is hard to interpret the estimates in [2] and [4], in the sense that one has no idea of what these are estimating. The properties of $\hat{\Delta}_n(\underset{\sim}{Y})$ given by (3.3), however, remain clear even when the translation model does not hold. If, as $n\to\infty$, $n_i/n\to\lambda_i$, $i=1,2,\ldots,c$, then $\hat{\Delta}_n(\underset{\sim}{Y})$ estimates $\Sigma_{i=1}^{c}\lambda_i\int f_{io}^2(x)\,dx$, where f_{io} is the density of F_i. Consistency in the mean and mean square still hold under regularity conditions stated in Corollary 3.1.

Estimates of more complicated functionals of f which also often occur in the same situations as those described in Section 1, will be discussed in a forthcoming paper.

REFERENCES

[1] Hodges, J. L., Jr. and Lehmann, E. L. (1956). Efficiency of some nonparametric competitors of the t-test. Ann. Math. Statist. **27**, pp. 324-335.

[2] Lehmann, E. L. (1963). Nonparametric confidence interval for a shift parameter. Ann. Math. Statist. **34**, pp. 1507-1512.

[3] Parzen, E. (1962). On estimation of probability density function and mode. Ann. Math. Statist. **33**, pp. 1065-1076.

[4] Sen, P. K. (1966). On a distribution-free method of estimating asymptotic efficiency of a class of nonparametric tests. Ann. Math. Statist. **37**, pp. 1759-1770.