The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 89–95

THE ASYMPTOTIC INADMISSIBILITY OF THE SAMPLE DISTRIBUTION FUNCTION

BY R. R. READ

Naval Postgraduate School

Given a sample of size n, a continuous estimator for a distribution F (based on Pyke's modified sample distribution) is shown to have the property that its expected squared error, for almost all x in the positive sample space of F, is no larger than that of the sample distribution function given F and n sufficiently large. Letting risk be given by the expected squared error integrated with respect to F, it is shown that this estimator dominates both the sample distribution and the other best invariant estimator found by Aggarwal, given F and n sufficiently large. Other common estimators cannot serve in this dominating role. Explicit calculation of risk is made when F is the uniform distribution. In this case the estimator strictly dominates the sample distribution for all $n \geqq 1$.

1. Notation and background. Let $X_1, X_2, \cdots, X_n$ be the order statistics of a random sample from an absolutely continuous distribution F having density f. Following Aggarwal [1], let us use the risk function

$$R(F, \hat{F}) = E \int |F - \hat{F}|^2 k(F)\, dF \tag{1.1}$$

where k is a positive weight function. Somewhat stronger results are obtained by considering the pointwise risk

$$R_x(F, \hat{F}) = E\{|F(x) - \hat{F}(x)|^2\} . \tag{1.2}$$

In order to contrast the two, the function (1.1) will be called the integrated risk.

Let N_x be the number of observations in $(-\infty, x]$. It is known (see [1] or [4]) that the sample distribution function, defined by

$$\hat{D}(x) = N_x/n \tag{1.3}$$

is the best invariant estimate if the weight function $k(t) = [t(1-t)]^{-1}$. Also it is known that the estimator

$$\hat{H}(x) = (N_x + 1)/(n + 2) \tag{1.4}$$

is best invariant if $k(t) \equiv 1$. Neither $\hat{D}$ nor $\hat{H}$ is continuous and $\hat{H}$ does not achieve the values 0 and 1.

Let us assume that the population sampled is bounded and contained in a finite interval. There is no loss in using the interval [0, 1]. It will be convenient to carry this assumption throughout the paper. It will be shown that the asymptotic results are not affected by it. Thus we can define $X_0 = 0$,

Received October 21, 1970.

$x_{n+1} = 1$ and let $U_x(V_x)$ be the distances from x to the nearest observation on the left (right). Letting the relative distance be

$$W_x = U_x/(U_x + V_x)\,, \tag{1.5}$$

define

$$\hat{C}(x) = (N_x + W_x)/(n+1)\,. \tag{1.6}$$

This function is continuous and consonant with Pyke's suggestion [6]. That is, the statistic C_n can be calculated from

$$C_n = \max_{0 \leqq x \leqq 1} |\hat{C}(x) - F(x)|\,. \tag{1.7}$$

We mention in passing that the small sample distribution of C_n is included in the works of Brunk [2], Durbin [3], and Steck [7], and has been tabled in [5]. The referee has pointed out that since $\hat{C}(x)$ lies between $\hat{D}(x-)$ and $\hat{D}(x)$ at all observations, C_n is stochastically smaller than Kolmogorov's statistic. The numerical effect of this is indicated in [5] and [7].

Since $\hat{C}$ is not a step function it is not invariant under the full group of strictly increasing continuous transformations as are $\hat{D}$ and $\hat{H}$. A weakening of invariance should provide better estimators. It can be shown that $\hat{C}$ is invariant under the subgroup of linear transformations having positive slope.

2. Asymptotic behavior of the pointwise risk. We work with the form

$$\begin{aligned}(n+1)^2 R_x(F, \hat{C}) &= E\{N_x + W_x - (n+1)F(x)\}^2 \\ &= \mathrm{Var}\,(N_x) + E\{W_x - F(x)^2 + 2\,\mathrm{Cov}\,\{N_x, W_x\}\,.\end{aligned} \tag{2.1}$$

Clearly N_x is a binomial variable $(n, F(x))$. The joint distribution of U_x, $V_x N_x$ is given by

$$\begin{aligned}P\{U_x > u, V_x > v, N_x = r\} &= \tbinom{n}{r} F^r(x-u)[1 - F(x+v)]^{n-r} \\ &\qquad 0 \leqq u < x\,, \quad 0 \leqq v < 1 - x\,,\end{aligned} \tag{2.2}$$

with singularities on the boundary given by

$$\begin{aligned}P\{U_x = x, V_x > v, N_x = 0\} &= [1 - F(x+v)]^n\,, \qquad 0 \leqq v < 1 - x \\ P\{U_x > u, V_x = 1 - x, N_x = n\} &= [F(x-u)]^n\,, \qquad 0 \leqq u < x\,.\end{aligned} \tag{2.3}$$

PROPOSITION 1. *If x is a point of continuity of f and $f(x) > 0$, then*

$$(n+1)^2 R_x(F, \hat{C}) = (n-1)F(x)[1 - F(x)] + o(1)\,. \tag{2.4}$$

PROOF. Use (2.1). Clearly $\mathrm{Var}\,(N_x) = nF(x)[1 - F(x)]$. The variables nU_x and nV_x are asymptotically independent exponential variables with mean $1/f(x)$ and it follows that W_x is asymptotically a uniform random variable. Thus

$$E\{W_x - F(x)\}^2 = \tfrac{1}{3} - F(x)[1 - F(x)] + o(1)\,. \tag{2.5}$$

Similarly, letting D_{uv} denote the second partial derivative operation with respect to u and v, and $n^{(r)} = n!/(n-r)!$

$$\begin{aligned}
&\operatorname{Cov}\{N_x, W_x\} \\
&= \textstyle\sum_{r=1}^{n-1} [r - nF(x)]\binom{n}{r} \int_0^x\int_0^{1-x} \frac{u}{u+v} \\
&\quad \times D_{uv}\{F^r(x-u)[1-F(x+v)]^{n-r}\}\, du\, dv + o(1) \\
&= n^{(2)} \iint \frac{u}{u+v} f(x-u)f(x+v)\{1 - F(x+v) + F(x-u)\}^{n-3} \qquad (2.6)\\
&\quad \times \{(n-2)F(x-u) + (1-nF(x))[1 - F(x+v) \\
&\quad + F(x-u)]\}\, du\, dv + o(1) \\
&= f^2(x) \iint we^{-yf(x)}[1 - wyf(x)]y\, dy\, dw + o(1) = -\tfrac{1}{6} + o(1)
\end{aligned}$$

upon summing the binomial, and making the change $y = u + v$, $w = u/y$. Inserting these three quantities in (2.1) proves the proposition.

It is noted that if one modifies functions $\hat{D}$ and $\hat{H}$ by connecting the steps with straight lines, the resulting pointwise risk functions have asymptotic forms which can be obtained from

$$E\{n\hat{D}(x) + W_x - nF(x)\}^2 = nF(x)[1 - F(x)] + o(1)\,, \tag{2.7}$$

$$\begin{aligned}
&E\{(n+2)\hat{H}(x) + W_x - (n+2)F(x)\}^2 \\
&\qquad = (n-2)F(x)[1-F(x)] + 2[1-F(x)]^2 + o(1)
\end{aligned} \tag{2.8}$$

where (2.7) and (2.8) are obtained analogously to (2.4).

3. Calculation of risk when F is the uniform distribution. Define

$$h(x) = \int_0^x \left[\frac{x-w}{1-w}\right]^n w\, dw \tag{3.1}$$

and note that

$$h(1-x) = \int_x^1 \left[1 - \frac{x}{w}\right]^n (1-w)\, dw\,.$$

PROPOSITION 2. *If F_0 is the uniform distribution, then*

$$\begin{aligned}
(n+1)^2 R_x(F_0, \hat{C}) &= (n-1)x(1-x) + 2(n+1)\{(1-x)h(x) \\
&\quad + xh(1-x)\}\,.
\end{aligned} \tag{3.2}$$

PROOF. Clearly $\operatorname{Var}(N_x) = nx(1-x)$. Using (2.2) summed over r and (2.3)

$$\begin{aligned}
E\{W - x\}^2 &= n^{(2)} \iint \left(\frac{u}{u+v} - x\right)[1 - (v+u)]^{n-2}\, du\, dv \\
&\quad - \int_0^{1-x} \left(\frac{x}{x+v} - x\right)^2 d[1 - (x+v)]^n \\
&\quad - \int_0^x \left(\frac{u}{u+1-x} - x\right)^2 d(x-u)^n\,.
\end{aligned}$$

Making the change $y = v + u$, $w = u/y$ in the first term, defining $L(w) = \min\{(1-x)/(1-w), x/w\}$ for each x, and obvious changes in the other two terms leads to the representation

$$E\{W - x\}^2 = n^{(2)} \int_0^1 \int_0^{L(w)} (w-x)^2(1-y)^{n-2} y\, dy\, dw - \int_x^1 \left(\frac{x}{y} - x\right)^2 d(1-y)^n + \int_0^x \left(\frac{x-y}{1-y} - x\right)^2 dy^n .$$

Using integration by parts twice on the first term, once in the other terms and reducing, yields

$$\begin{aligned} E\{W_x - x\}^2 &= \int_0^1 (w-x)^2\, dw - 2\int_0^x (1-L(w))^n (x-w) w\, dw \\ &\quad - 2\int_x^1 (1-L(w))^n (w-x)(1-w)\, dw \\ &= \tfrac{1}{3} - x(1-x) + 2(1-x)h(x) + 2xh(1-x) \\ &\quad - 2\int_0^1 (1-L(w))^n w(1-w)\, dw . \end{aligned} \tag{3.3}$$

The determination of the contribution of the covariance term is similar but more lengthy. Using (2.2) and ignoring the terms involving the singular part we have

$$\begin{aligned} \operatorname{Cov}(N_x, W_x) &\doteq n^{(3)} \iint \frac{u}{u+v} (x-u)[1-v-u]^{n-3}\, du\, dv \\ &\quad + n^{(2)}(1-nx) \iint \frac{u}{u+v} (1-v-u)^{n-2}\, du\, dv \\ &= n^{(3)} \int_0^1 w \int_0^{L(w)} (x - wy) y (1-y)^{n-3}\, dy \\ &\quad + (1-nx) n^{(2)} \int_0^1 w \int_0^{L(w)} y(1-y)^{n-2}\, dy \end{aligned}$$

using the same change as before. The inner integrals can be treated by repeated integrations by parts. This yields, after reducing,

$$\begin{aligned} &- n^{(2)} \int_0^1 w(x - wL(w))(1-L(w))^{n-2} L(w)\, dw \\ &- (1-nx) n \int_0^1 w(1-L(w))^{n-1} L(w)\, dw \\ &- n \int_1^0 w(x - 2wL(w))(1-L(w))^{n-1} \\ &- (1-nx) \int_0^1 w(1-L(w))^n + 2\int_0^1 w^2(1-L(w))^n\, dw - \tfrac{1}{6} . \end{aligned} \tag{3.4}$$

Using the facts that

$$\begin{aligned} &x - wL(w) = 1 - L(w) \quad \text{and} \quad L(w)\, dw = (1-w)\, dL(w) \quad &&\text{for } 0 \leqq w < x \\ &x - wL(w) = 0 \quad \text{and} \quad L(w)\, dw = -w\, dL(w) \quad &&\text{for } x \leqq w \leqq 1 \end{aligned} \tag{3.5}$$

we proceed to integrate the terms in (3.4) by parts until $(1 - L(w))$ appears to the power n in all terms. Treating the first two terms first this process yields

$$- n(1-x)\int_0^x (1-L(w))^n(1-2w)\,dw - (1-nx)(1-x)^n + 2(1-nx)\int_x^1 (1-L(w))^n w\,dw\,.$$

Similarly the third and fourth terms of (3.4) can be represented as

$$\int_0^x (1-L(w))^n(2w-3w^2)\,dw + (1-x)^n - 3\int_x^1 (1-L(w))^n w^2\,dw - n(1-x)\int_0^x (1-L(w))^n w\,dw + nx\int_x^1 (1-L(w))^n w\,dw - \int_0^1 (1-L(w))^n w\,dw\,.$$

The contribution of the singular part of the distribution to the covariance is

$$- nx(1-x)^n + nx\int_x^1 (1-L(w))^n\,dw + n(1-x)\int_0^x (1-L(w))^n\,dw\,.$$

Collecting all the parts and reducing yields

$$\text{Cov}\,(N_x, W_x) = -\tfrac{1}{6} + \int_0^1 (1-L(w))^n w(1-w)\,dw + n(1-x)h(x) + nxh(1-x) \tag{3.6}$$

and upon applying the basic formula (2.1'), Proposition 2 is proved.

It seems desirable to record the pointwise mean

$$(n+1)E\{\hat{C}(x)\} = nx + \tfrac{1}{2} + h(x) - h(1-x)\,. \tag{3.7}$$

4. Results. Asymptotic inadmissibility using pointwise risks is shown in Proposition 3.

PROPOSITION 3. *Let F be a distribution on* $[0, 1]$ *and let x be a point of increase of F. For n sufficiently large for this F,*

$$(n+1)^2\{R_x(F, \hat{C}) - R_x(F, \hat{D})\} < -\,3F(x)[1-F(x)]\,. \tag{4.1}$$

PROOF. It follows easily from (2.4). The relaxation of the condition that f be continuous at x is permitted because the continuous functions are dense in L^1.

REMARK. Neither the estimator $\hat{H}$ nor the polygonal versions of $\hat{H}$ and $\hat{D}$ whose asymptotic risks are given in (2.8) and (2.7) can replace $\hat{C}$ in the pointwise dominating role exhibited in (4.1). This is due to the terms $F(x)^2$ and $[1-F(x)]^2$ in the pointwise risk of $\hat{H}$ which makes the inequality reverse near 0 and 1; to the term $2[1-F(x)]^2$ in (2.8) which makes the inequality reverse near 0; and to the fact that the risk of the polygonal version of $\hat{D}$ behaves the same as $R_x(F, \hat{D})$.

The estimator $\hat{C}$ does not dominate $\hat{H}$ in the pointwise sense, but it does in the integrated sense. This is stated below without proof.

PROPOSITION 4. *If $k(t) \equiv 1$, then*

$$(4.2)\qquad (n+1)^2\{R(F,\hat{C}) - R(F,\hat{H})\} = -\tfrac{1}{6} + o(1)\,.$$

It is also noted that $\hat{C}$ is itself dominated in this asymptotic sense. For example, consider the estimator $\hat{S}(x) = (N_x + W_x)/(n + 5/4)$ whose integrated risk is $(n + 5/4)^{-2}(8n - 5)/48 + o(1)$. It is easily shown that this is smaller than $R(F, \hat{C})$ for sufficiently large n.

Exact comparisons when F is the uniform distribution are of interest.

PROPOSITION 5. *If $k(t) = [t(1-t)]^{-1}$ then for all $n \geqq 1$,*

$$(4.3)\qquad (n+1)^2\{R(F_0,\hat{C}) - R(F_0,\hat{D})\} < -2.$$

PROOF. Consider

$$(4.4)\qquad (n+1)\int_0^1 \frac{1}{x}h(x)\,dx = \int_0^1 \frac{u\,du}{(1-u)^n}\int_u^1 \frac{1}{x}\,d(x-u)^{n+1}$$

$$< \int_0^1 u(1-u)\,du + \int\int \frac{u(x-u)}{x^2}\,du\,dx = \tfrac{1}{4}\,.$$

Obviously $(n+1)\int_0^1 (1-x)^{-1}h(1-x)\,dx < \frac{1}{4}$. Then the appropriate integral of (3.2) yields $(n+1)^2R(F_0, \hat{C}) < n$ and (4.3) follows.

Let us now show that the asymptotic results are unaffected by the assumption that the population sampled is bounded. Without such prior knowledge we will have either $X_0 = -\infty$ or $X_{n+1} = +\infty$ or both. In the former case $W_x = 1$ with probability $[1 - F(x)]^n$, in the latter case $W_x = 0$ with probability $[F(x)]^n$ and (1.6) is no longer a distribution. The definition of $\hat{C}$ (or any other estimator) can be modified rather arbitrarily in these "tails" because the corresponding change in the risk will tend to zero exponentially fast. Thus Propositions 1, 3 and 4 remain valid. Propositions 2 and 5, however, depend upon the use of finite endpoints for defining $\hat{C}$ and appropriate modifications would be in order.

Acknowledgment. I am grateful to the referee whose suggestions led to much-improved and succinct presentations in Sections 2 and 3.

REFERENCES

[1] AGGARWAL, OM. P. (1955). Some minimax invariant procedures for estimating a cumulative distribution function. *Ann. Math. Statist.* **26** 450–463.

[2] BRUNK, H. D. (1962). On the range of the difference between hypothetical distribution function and Pyke's modified empirical distribution function. *Aun. Math. Statist.* **33** 525–532.

[3] DURBIN, J. (1968). The probability that the sample distribution function lies between two parallel straight lines. *Ann. Math. Statist.* **39** 398–411.

[4] FERGUSON, THOMAS, S. (1967). *Mathematical Statistics.* Academic Press, New York.

[5] HENDREN, J. P. (1969). A comparison of some statistics of the Kolmogorov type. Master's thesis, Department of Operations Research, Naval Postgraduate School. AD 704780.

[6] Pyke, R. (1959). The supremum and infimum of the Poisson process. *Ann. Math. Statist.* **30** 568–576.

[7] Steck, G. P. (1971). Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. *Ann. Math. Statist.* **42** 1–11.