A DISTRIBUTION-FREE UPPER CONFIDENCE BOUND FOR $\Pr\{Y < X\}$, BASED ON INDEPENDENT SAMPLES OF X AND Y

By Z. W. Birnbaum and R. C. McCarty

University of Washington and *Boeing Airplane Company*

1. Summary. A solution for the problem of obtaining a distribution-free one-sided confidence interval for $p = \Pr\{Y < X\}$ has been proposed in [1]. At present a numerical procedure is given for computing the sample sizes needed for such a confidence interval with given width and confidence level.

2. Introduction and formulation of the problem. The problem discussed in this paper arises in practical situations such as the following: structural components of a mechanism are mass-produced and each component has a strength at failure Y which, in view of unavoidable variability of the product, is a random variable. A component is then installed in a mechanism and exposed to a stress which reaches its maximum value X, again a random variable. If, due to chance, the values of Y and X are so paired off that $Y < X$, then the component fails in use. It is therefore of considerable importance to have an upper bound for

$$p = \Pr\{Y < X\}. \tag{2.1}$$

Our problem is: Can p be estimated from samples of X and Y alone and, in particular, is there an upper confidence bound for p, i.e., a statistic ψ based on a sample of X and a sample of Y, such that for any $\epsilon > 0$, $\alpha > 0$ there exists a pair of numbers $M_{\epsilon,\alpha}$, $N_{\epsilon,\alpha}$, so that

$$\Pr\{p \leqq \psi + \epsilon\} \geqq 1 - \alpha \tag{2.2}$$

when the sample of X is of size $m \geqq M_{\epsilon,\alpha}$ and the sample of Y of size $n \geqq N_{\epsilon,\alpha}$?

The following answer to this question was proposed in [1].

We assume that X and Y are independent random variables with continuous cumulative distribution functions $F(s) = \Pr\{X < s\}$, $G(s) = \Pr\{Y < s\}$. Let $X_1 \leqq X_2 \leqq \cdots \leqq X_m$ be an ordered sample of X and $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_n$ an ordered sample of Y, and let $F_m(s)$, $G_n(s)$ be the empirical distribution functions corresponding to these samples.

Using the Wilcoxon-Mann-Whitney statistic,

$$U = \text{number of pairs } (X_i, Y_k) \text{ such that } Y_k < X_i,$$

we write

$$\hat{p} = U/mn. \tag{2.3}$$

Received June 7, 1957. Research sponsored by the Office of Naval Research and the Boeing Airplane Company.

It is easily verified that $\hat{p}$ is an unbiased estimate of p and

$$p = \int_{-\infty}^{+\infty} G(s)\, dF(s),$$

$$\hat{p} = \int_{-\infty}^{+\infty} G_n(s)\, dF_m(s),$$

hence

$$(2.4)\qquad \begin{aligned} p - \hat{p} &= \int_{-\infty}^{+\infty} G\, d(F - F_m) + \int_{-\infty}^{+\infty} (G - G_n)\, dF_m \\ &= \int_{-\infty}^{+\infty} (F_m - F)\, dG + \int_{-\infty}^{+\infty} (G - G_n)\, dF_m \end{aligned}$$

and

$$(2.5)\qquad p - \hat{p} \leqq D_m^- + D_n^+,$$

where

$$D_m^- = \sup_{-\infty < s < +\infty} \{F_m(s) - F(s)\},$$

$$D_n^+ = \sup_{-\infty < s < +\infty} \{G(s) - G_n(s)\}.$$

It is well known [2] that

$$\Pr\{D_m^- < v\} = \Pr\{D_m^+ < v\} = P_m(v)$$

and

$$\Pr\{D_n^+ < v\} = P_n(v)$$

are cumulative distribution functions which depend on the sample sizes m, n, but not on the c.d.f.'s F and G. It follows from (2.5) that

$$(2.6)\qquad \Pr\{p \leqq \hat{p} + \epsilon\} \geqq \Pr\{D_m^+ + D_n^+ \leqq \epsilon\} = P_{m,n}(\epsilon),$$

where $P_{m,n}(\epsilon)$ is the convolution of P_m and P_n, hence does not depend on F and G. The statistic $\hat{p}$ has, therefore, the property required of ψ in (2.2) provided one can, for given ϵ, α, determine numbers $M_{\epsilon,\alpha}$, $N_{\epsilon,\alpha}$ so that

$$(2.7)\qquad P_{m,n}(\epsilon) \geqq 1 - \alpha \qquad \text{for } m \geqq M_{\epsilon,\alpha},\ n \geqq N_{\epsilon,\alpha}.$$

Some further properties of $\hat{p}$ are discussed in [1].

A numerical procedure for computing $M_{\epsilon,\alpha}$, $N_{\epsilon,\alpha}$ is presented in the next sections.

3. An approximate expression for $P_{m,n}(\epsilon)$. It was shown by N. Smirnov [3] that

$$(3.1)\qquad \lim_{n\to\infty} \Pr\{D_n^+ \leqq z/\sqrt{n}\} = \lim_{n\to\infty} P_n(z/\sqrt{n}) = 1 - e^{-2z^2} = L(z).$$

Since, for fixed n, $P_n(z/\sqrt{n}) = H_n(z)$ is a cumulative distribution function, and $L(z) = 1 - e^{-2z^2}$ is a continuous c.d.f., it follows by a well-known argument (see, e.g., [4], p. 276) that $H_n(z) \to L(z)$ uniformly. We may, therefore, conclude that

$$\lim_{n\to\infty} [\Pr \{D_n^+ \leqq v\} - L(v\sqrt{n})] = \lim_{n\to\infty} [H_n(v\sqrt{n}) - L(v\sqrt{n})] = 0 \tag{3.2}$$

uniformly for $0 \leqq v \leqq 1$. Writing

$$\begin{aligned} P_{m,n}(\epsilon) &= \Pr \{D_m^+ + D_n^+ \leqq \epsilon\} = \int_0^\epsilon P_n(\epsilon - u)\, dP_m(u), \\ Q_{m,n}(\epsilon) &= \int_0^\epsilon L[(\epsilon - u)\sqrt{n}]\, dL(u\sqrt{m}), \end{aligned} \tag{3.3}$$

we have

$$\begin{aligned} | P_{m,n}(\epsilon) - Q_{m,n}(\epsilon) | &\leqq \left| \int_0^\epsilon \{P_n(\epsilon - u) - L[(\epsilon - u)\sqrt{n}]\}\, dP_m(u) \right| \\ &\quad + \left| \int_0^\epsilon \{P_m(\epsilon - v) - L[(\epsilon - v)\sqrt{m}]\}\, dL(v\sqrt{n}) \right| \\ &\leqq \operatorname*{Max}_{0\leqq u\leqq\epsilon} | P_n(\epsilon - u) - L[(\epsilon - u)\sqrt{n}] | \\ &\quad + \operatorname*{Max}_{0\leqq v\leqq\epsilon} | P_m(\epsilon - v) - L[(\epsilon - v)\sqrt{m}] |, \end{aligned} \tag{3.4}$$

which in view of (3.2) shows that

$$\lim_{\substack{m\to\infty \\ n\to\infty}} | P_{m,n}(\epsilon) - Q_{m,n}(\epsilon) | = 0 \tag{3.5}$$

uniformly for $0 \leqq \epsilon \leqq 1$. This justifies the use of $Q_{m,n}(\epsilon)$ as an approximation to $P_{m,n}(\epsilon)$ for m, n sufficiently large. Some observations on the goodness of this approximation are presented in Section 5.

By straightforward integration one obtains for $Q_{m,n}(\epsilon)$ the expression

$$\begin{aligned} Q_{m,n}(\epsilon) &= 1 - \frac{n}{m+n} e^{-2m\epsilon^2} - \frac{m}{m+n} e^{-2n\epsilon^2} \\ &\quad - \frac{2\sqrt{2\pi}\, mn\epsilon^2}{(m+n)^{3/2}} e^{-2mn\epsilon^2/(m+n)} \frac{1}{\sqrt{2\pi}} \int_{-2n\epsilon/\sqrt{m+n}}^{2m\epsilon/\sqrt{m+n}} e^{-t^2/2}\, dt. \end{aligned} \tag{3.6}$$

4. Sample sizes m, n which satisfy $Q_{m,n}(\epsilon) = 1 - \alpha$. With the notations

$$\begin{aligned} m + n &= N \\ m / (m + n) = \lambda, &\qquad n / (m + n) = 1 - \lambda \\ \epsilon\sqrt{m + n} &= \delta \end{aligned} \tag{4.1}$$

TABLE I

Values $\delta_{\lambda,\alpha}$ such that $Q(\delta_{\lambda,\alpha}; \lambda) = 1 - \alpha$

α	λ				
	.1	.2	.3	.4	.5
.10	4.1185	3.2027	2.8501	2.6928	2.6468
.05	4.6115	3.5667	3.1641	2.9844	2.9317
.01	5.5700	4.2745	3.7770	3.5524	3.4870
.005	5.9300	4.5405	4.0050	3.7665	3.6960
.001	6.6800	5.0980	4.4880	4.2150	4.1360

the expression (3.6) for $Q_{m,n}(\epsilon)$ may be written in the form

$$Q(\delta;\lambda) = 1 - \lambda e^{-2(1-\lambda)\delta^2} - (1-\lambda)e^{-2\lambda\delta^2} - 2\sqrt{2\pi}\,\lambda(1-\lambda)\,\delta e^{-2\lambda(1-\lambda)\delta^2} \frac{1}{\sqrt{2\pi}}\int_{-2(1-\lambda)\delta}^{2\lambda\delta} e^{-t^2/2}\,dt. \tag{4.2}$$

Table I contains solutions $\delta_{n,\alpha}$ of the equation.

$$Q(\delta;\lambda) = 1 - \alpha \tag{4.3}$$

for $\alpha = .001, .005, .01, .05, .10$, and $\lambda = .1\ (.1)\ .5$. These solutions were obtained on a desk calculator, using the National Bureau Bureau of Standards Tables of the Exponential Function [5], Descending Exponential [6], and the Normal Distribution Function [7].

The use of the quantities N, λ, δ instead of the original m, n, ϵ has not only the advantage of reducing the computations to a table with double entry, but also makes it possible to design an experiment with a given ratio $\lambda = m/N$. This ratio is often dictated by considerations of cost or time.

Example. We wish to use four times as many Y's as X's, i.e., $\lambda = .2$, and require $\epsilon = .10$, $\alpha = .05$. From Table I we have $\delta_{.2,.05} = 3.5667$; hence, by (4.1), $(.10)\sqrt{N} = 3.5667$, and $N = 1272.13$, $m = 254.43$, $n = 1017.70$. The rounded-up sample sizes are therefore 255 for X, 1018 for Y.

5. Concluding remarks. The sample sizes computed for given λ, ϵ, α by the use of Table I are conservative, i.e., too large, for two reasons. The first is that, instead of finding sample sizes m, n such that $P\{p \leqq \hat{p} + \epsilon\} = 1 - \alpha$, we used inequality (2.6) and looked for m, n satisfying $P_{m,n}(\epsilon) = 1 - \alpha$, a step which certainly yields larger values. The second reason is that in equation $P_{m,n}(\epsilon) = 1 - \alpha$ the exact expression $P_{m,n}(\epsilon)$ was replaced by the approximate expression $Q_{m,n}(\epsilon)$ and then only m, n were computed. This step was justified by (3.5) which, however, does not indicate which way the sample sizes are affected. The following arguments are offered in favor of the contention that the solutions m, n of $P_{m,n}(\epsilon) = 1 - \alpha$ differ little from those of $Q_{m,n}(\epsilon) = 1 - \epsilon$ and that the solutions of the second equation are more conservative (greater) than those of the first.

The exact form of $P\{D_n^+ \leqq v\}$ for finite n is known and numerical computations, some of which are reproduced in [8], show that already for $n \geqq 50$ the approximation of $P\{D_n^+ \leqq v\}$ by $L(v\sqrt{n})$ is uniformly very good. Since the sample sizes computed from Table I are in all practical situations much larger than 50, (3.4) assures very close agreement between $P_{m,n}(\epsilon)$ and $Q_{m,n}(\epsilon)$.

Furthermore, the following conjecture appears to be substantiated by considerable numerical computations and some analytical considerations, although no proof for it is available: for every integer $n \geqq 1$ and for $0 \leqq v \leqq 1$,

$$L(v\sqrt{n}) = 1 - e^{-2nv^2} \leqq P\{D_n^+ \leqq v\}. \tag{5.1}$$

From (5.1) would follow that $P_{m,n}(\epsilon) \geqq Q_{m,n}(\epsilon)$ for $0 \leqq \epsilon \leqq 1$, hence $Q_{m,n}(\epsilon) = 1 - \alpha$ would yield sample sizes larger than $P_{m,n}(\epsilon) = 1 - \alpha$.

REFERENCES

[1] Z. W. Birnbaum, "On a use of the Mann-Whitney statistic," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, University of California Press, 1956, pp. 13–17.

[2] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[3] N. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Recueil Math. (Mat. Sbornik)*, N.S., Vol. 6 (1939), pp. 3-26.

[4] M. Fréchet, "Generalités sur les probabilités. Variables aléatoires," *Traité de calcul des probabilités* (edited by E. Borel), Gauthier-Villars, Paris, 1937.

[5] National Bureau of Standards, "Tables of the exponential function," *N.B.S. Applied Mathematics Series No. 14*.

[6] National Bureau of Standards, "Tables of the descending exponential," *N.B.S. Applied Mathematics Series No. 46*.

[7] National Bureau of Standards, "Tables of Normal Probability Functions," *N.B.S. Applied Mathematics Series No. 23*.

[8] Z. W. Birnbaum and Fred H. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951) pp. 592–596.