ON DISTRIBUTION FREE CONFIDENCE BOUNDS FOR $\Pr(Y < X)$

BY

HANS K. URY

TECHNICAL REPORT NO. 172

FEBRUARY 23, 1971

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
N00014-67-A-0112-0053 (NR-042-267)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON DISTRIBUTION-FREE CONFIDENCE BOUNDS FOR $\Pr\{Y < X\}$

by

HANS K. URY

TECHNICAL REPORT NO. 172
February 23, 1971

PREPARED UNDER CONTRACT N00014-67-A-0112-0053
(NR-042-267)
OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON DISTRIBUTION-FREE CONFIDENCE BOUNDS FOR $\Pr(Y < X)$

by

Hans K. Ury

1. INTRODUCTION AND SUMMARY

Let $X$ and $Y$ be independent random variables with cumulative distribution functions $F$ and $G$, respectively. Examples of practical situations where one is interested in estimating $p = \Pr(Y < X)$ on the basis of random samples $X_1, X_2, \ldots, X_m$ from $F(x)$ and $Y_1, Y_2, \ldots, Y_n$ from $G(y)$ are given in [1], [3], and [6], for instance, in some cases with distributional assumptions on $F$ or $G$.

We shall deal entirely with distribution-free confidence bounds for $p$ in the case in which both $F$ and $G$ are unknown. For this situation, Z.W. Birnbaum [1] showed that the Mann-Whitney statistic $U$ could be used to estimate $p$ from the samples for continuous $F$ and $G$, giving the minimum-variance unbiased estimate $\hat{p} = U/mn$, and he obtained one-sided confidence intervals for $p$. A numerical procedure for computing the sample sizes needed for these confidence bounds, for specified width and confidence level, was given by Z.W. Birnbaum and R.C. McCarty in [2].

D.B. Owen, K.J. Craswell and D.L. Hanson [6] showed that the continuity assumption on $F$ and $G$ could be dropped. They also gave a table for use in computing confidence intervals by means of the Birnbaum-McCarty procedure and, for the case $m = n$, a table of sample sizes corresponding to various confidence levels and one-sided interval widths.
Z. Govindarajulu [5] noted that the sample sizes associated with the Birnbaum-McCarty bounds are quite large. He obtained a distribution-free upper bound for the asymptotic variance of $\hat{p}$, equivalent to van Dantzig's upper bound [4],

$$\sigma^2 \hat{p} \leq \frac{1}{4v},$$  

(1)

where $v = \min(m,n)$. Using this and the asymptotic normality of $U$, he was able to obtain substantial asymptotic improvements over the Birnbaum-McCarty bounds by means of asymptotic confidence intervals of the form

$$\Pr(\left| \hat{p} - p \right| \leq \frac{\phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\frac{1}{2}\left(4v\right)^{1/2}}) \geq \gamma, \ 0 < \gamma < 1,$$

(2)

for all $F$ and $G$ and non-random and large $m$ or $n$, where $\phi$ denotes the standard normal distribution and $\gamma$ is the confidence coefficient.

J.D. Church and B. Harris [3] pointed out that for $p$ near unity, the rate of convergence to normality should be quite slow and the use of (2) should tend to overestimate the confidence coefficient; as a result, Govindarajulu's method should only be adequate for very large sample sizes. They verified this by means of Monte Carlo methods for $p = .9, .95, \text{ and } .999$ with sample sizes of 25, 75, and 225.

For smaller samples there is thus a need for reliable procedures giving distribution-free confidence intervals that are shorter than the Birnbaum-McCarty bounds. For confidence coefficients not exceeding .925 and sample sizes that are not too unequal (this is discussed below), the Chebyshev inequality, used in conjunction with the upper
TABLE 1
Sample Sizes for the Chebyshev Intervals

\( p = \text{Pr}(Y < X) \) for Equal Sample Sizes

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>.50</th>
<th>.45</th>
<th>.40</th>
<th>.35</th>
<th>.30</th>
<th>.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>.500</td>
<td>2 (6)</td>
<td>3 (8)</td>
<td>4 (10)</td>
<td>5 (13)</td>
<td>6 (17)</td>
<td>8 (24)</td>
</tr>
<tr>
<td>.700</td>
<td>4 (9)</td>
<td>5 (11)</td>
<td>6 (14)</td>
<td>7 (18)</td>
<td>10 (25)</td>
<td>14 (35)</td>
</tr>
<tr>
<td>.750</td>
<td>4 (10)</td>
<td>5 (12)</td>
<td>7 (16)</td>
<td>9 (20)</td>
<td>12 (27)</td>
<td>16 (39)</td>
</tr>
<tr>
<td>.800</td>
<td>5 (11)</td>
<td>7 (14)</td>
<td>8 (17)</td>
<td>11 (22)</td>
<td>14 (30)</td>
<td>20 (43)</td>
</tr>
<tr>
<td>.850</td>
<td>7 (13)</td>
<td>9 (15)</td>
<td>11 (19)</td>
<td>14 (25)</td>
<td>19 (34)</td>
<td>27 (49)</td>
</tr>
<tr>
<td>.900</td>
<td>10 (15)</td>
<td>13 (18)</td>
<td>16 (22)</td>
<td>21 (29)</td>
<td>28 (39)</td>
<td>40 (57)</td>
</tr>
<tr>
<td>.925</td>
<td>14 (16)</td>
<td>17 (19)</td>
<td>21 (24)</td>
<td>28 (32)</td>
<td>38 (43)</td>
<td>54 (62)</td>
</tr>
<tr>
<td>.950</td>
<td>20 (18)</td>
<td>25 (22)</td>
<td>32 (27)</td>
<td>41 (36)</td>
<td>56 (48)</td>
<td>80 (69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>.20</th>
<th>.15</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>.500</td>
<td>13 (38)</td>
<td>23 (66)</td>
<td>50 (149)</td>
<td>200 (593)</td>
</tr>
<tr>
<td>.700</td>
<td>21 (55)</td>
<td>38 (97)</td>
<td>84 (218)</td>
<td>334 (869)</td>
</tr>
<tr>
<td>.750</td>
<td>25 (61)</td>
<td>45 (107)</td>
<td>100 (241)</td>
<td>400 (961)</td>
</tr>
<tr>
<td>.800</td>
<td>32 (67)</td>
<td>56 (119)</td>
<td>125 (268)</td>
<td>500 (1071)</td>
</tr>
<tr>
<td>.850</td>
<td>42 (76)</td>
<td>75 (135)</td>
<td>167 (303)</td>
<td>667 (1210)</td>
</tr>
<tr>
<td>.900</td>
<td>63 (88)</td>
<td>112 (156)</td>
<td>250 (351)</td>
<td>1000 (1402)</td>
</tr>
<tr>
<td>.925</td>
<td>84 (96)</td>
<td>149 (171)</td>
<td>334 (384)</td>
<td>1334 (1535)</td>
</tr>
<tr>
<td>.950</td>
<td>125 (108)</td>
<td>223 (191)</td>
<td>500 (430)</td>
<td>2000 (1719)</td>
</tr>
</tbody>
</table>
bound on the variance of \( \hat{p} \) given by (1), yields shorter intervals. Since (1), which is simply based on the worst possible case of \( p = \frac{1}{2} \), does not depend on the sample sizes or the estimate of \( p \), these intervals have the desired reliability.

The comparison with the Birnbaum-McCarty intervals is carried out in terms of the required sample size for several values of \( \gamma \) and of the interval width for the case \( m = n \). The sample size is compared with the corresponding entry for the Birnbaum-McCarty procedure given in Table 2 of Owen, Craswell and Hanson [6]. For the general case, limits are given on the ratio of the smaller sample size to the combined sample size which ensure that the intervals will remain shorter than the Birnbaum-McCarty bounds when \( \gamma \leq .925 \).

2. THE CHEBYSHEV INTERVALS FOR \( \Pr[X < X] \)

Using the upper bound for the variance of \( \hat{p} \) given by (1), we obtain intervals of the following type

\[
\Pr\left( \left| \hat{p} - p \right| \leq \frac{1}{\frac{1}{2} \left( \frac{1}{2} \right) \left( \frac{1}{2} \right)} \right) \geq \gamma. \quad (3)
\]

For a given interval width, \( 2\varepsilon \), we then calculate the sample size from

\[
\varepsilon = \frac{1}{\left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \left( \frac{1}{2} \right)} \quad (4)
\]

and thus the required size of the smaller sample is

\[
v = \frac{1}{4\varepsilon^2 (1-\gamma)}. \quad (5)
\]
TABLE 2

Minimum Values of $\lambda$ for which the Chebyshev Intervals are Shorter Than the Birnbaum-McCarty Upper Bounds

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\lambda$</th>
<th>$\gamma$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>.033</td>
<td>.85</td>
<td>.18</td>
</tr>
<tr>
<td>.70</td>
<td>.041</td>
<td>.90</td>
<td>.32</td>
</tr>
<tr>
<td>.75</td>
<td>.06</td>
<td>.925</td>
<td>.43</td>
</tr>
<tr>
<td>.80</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In Table 1, for \( m = n = v \), these sample sizes are given for \( \gamma = .50 \) and \( .70 (.05) .90 (.025) .95 \) and for \( \varepsilon = .05 (.05) .50 \). The entries in parentheses are the corresponding values for the Birnbaum-McCarty upper bound, obtained from Table 2 of Owen, Craswell and Hanson [6]. At \( \gamma = .50 \), these intervals require only about \( 1/3 \) the sample size of the Birnbaum-McCarty bounds. This advantage decreases with increasing \( \gamma \) and vanishes between \( \gamma = .925 \) and \( \gamma = .95 \). For \( \varepsilon < .05 \), the sample size increases rapidly; for instance, when \( \varepsilon = .025 \) and \( \gamma = .50 \), \( v = 800 \) is required. Hence, Govindarajulu's method can probably be safely used in this case.

The Chebyshev intervals are, of course, two-sided bounds. As a result, using them for an upper bound only is a quite conservative procedure. Under appropriate symmetry conditions, the confidence coefficient in Table 1 can be increased from \( \gamma \) to \((1+\gamma)/2\); this will be the case, for example, when \( F \) and \( G \) are symmetric and \( \hat{p} = 0.5 \). But we have not obtained sharper bounds for completely general \( F \), \( G \), and \( \hat{p} \).

When the sample sizes are unequal, the Chebyshev intervals, which are based on the size of the smaller sample, eventually lose their advantage over the Birnbaum-McCarty bounds, which take into consideration the size of the combined sample. Let \( \lambda = v/(m+n) \). Then in order to determine, for a given value of \( \gamma \leq .925 \), the \( \lambda \)-value at which the Birnbaum-McCarty bound becomes smaller than the Chebyshev interval width \( \varepsilon(\gamma,v) \), it is necessary to find both \( \lambda \) and \( \delta \), where \( \delta \) is computed from the equation
\[
\gamma = 1 - \lambda e^{-2(1-\lambda)\delta^2} - (1-\lambda)e^{-2\lambda\delta^2} - 2\sqrt{2\pi} \lambda(1-\lambda)\delta e^{-2\lambda(1-\lambda)\delta^2} \frac{1}{\sqrt{2\pi}} \left[ \int_{-2(1-\lambda)\delta}^{2\delta} e^{-t^2/2} dt \right],
\]

such that

\[
\frac{\delta}{\sqrt{\nu+\nu/\lambda}} = \epsilon.
\]

This follows from expression (4.2) of [2]. The resulting lower bounds on \( \lambda \) for the seven \( \gamma \)-values of interest are given in Table 2 to three decimals for \( \gamma = .5 \) and .7 and to two decimals otherwise.

It is evident that the Chebyshev intervals remain superior over a wide range of \( \lambda \) when \( \gamma \) does not exceed .80. For \( \gamma = .50 \), the change is rather slow with respect to \( \lambda \). For example, when \( \lambda = .030 \), the Chebyshev intervals are less than 1% longer than the Birnbaum-McCarty bounds, and even for \( \lambda = .01 \) the increase is less than 7%. Therefore, if a large number of upper bounds on \( p \) have to be computed at this confidence level, the Chebyshev intervals might well also be considered for cases with \( .01 \leq \lambda \leq .033 \) in view of the extreme ease with which they can be computed.

3. SOME NUMERICAL EXAMPLES

We shall give some examples of the type considered in [2] and [6]. Suppose that we have 40 values of \( Y \), the strength at failure of a steel rod. We are interested in the probability of failure of this steel rod when subjected to a force \( X \). Failure occurs whenever
$Y < X$, and both $X$ and $Y$ can be considered random variables with unknown distributions. Suppose that we also have a sample of 10 values of $X$, the forces which are applied to the rod.

We begin by computing $U$. This is done by comparing all $X$-observations with all $Y$-observations and counting the number of comparisons in which a $Y$-value is smaller than an $X$-value. If there are no ties, a quicker method for moderate or large $m$ and $n$ consists of computing $T$, the sum of the ranks of the $Y$'s in the combined sample of $m+n$ observations. Then

\[ U = mn + \frac{n(n+1)}{2} - T. \]

If there are some ties between $X$'s and $Y$'s (but the $X$'s are all different and the $Y$'s are all different), we can still compute $T$ by assigning the higher of the tied ranks to the $Y$-observation.

Suppose, then, that we obtain $U = 24$ and hence $\hat{p} = 0.015$.

We shall consider some cases in which the desired confidence coefficient is 0.90 or less. (For $\gamma \geq 0.95$, the Birnbaum-McCarty bounds should be calculated, as shown in [6].)

(a) For $\gamma = 0.90$ we find $\epsilon = 0.25$ directly from Table 1. Hence a 90\% upper confidence limit (UCL) on the probability of failure of the steel rod when subjected to forces measured by $X$ is $0.015+0.25 = 0.265$.

(b) If $\gamma = 0.75$, we use (3) and obtain

\[ \Pr \left[ \left| 0.015 - p \right| \leq \frac{1}{\sqrt{40}} \right] \geq 0.75, \]

which yields $\epsilon = 0.1581$ and a 75\% UCL of 0.1731.
(c) If \( U = 560 \) and hence \( \hat{p} = 0.35 \), one can also use Table 1 to obtain a 90\% lower confidence limit (LCL) of 0.10, while (3) yields a 75\% LCL of 0.1919. For (a) and (b) the LCL is obviously 0.

(d) Suppose we have \( m > 40 \) X-observations. The confidence limits obtained by means of the Chebyshev intervals remain unchanged, and it can be seen from Table 2 that for \( \gamma = 0.90 \) they will be shorter than the Birnbaum-McCarty upper bound provided \( m \leq 80 \). For instance, when \( m = 60 \) a computation using Table 1 of [6] yields a Birnbaum-McCarty \( \epsilon \) of 0.269; when \( m = 120 \), their \( \epsilon = 0.237 \).

(e) If \( \gamma = 0.75 \) the Chebyshev intervals will be shorter than the Birnbaum-McCarty bounds as long as \( m \leq 626 \).

(f) Finally, if a closer upper bound is needed, say, \( \epsilon = 0.15 \), Table 1 shows that the smaller sample size must be 112 if \( \gamma = 0.90 \) and 45 if \( \gamma = 0.75 \). The confidence coefficient corresponding to our smaller sample size of 40 can be found by using (4). From

\[
\frac{1}{(1-\gamma)^2} = \frac{1}{\frac{1}{(0.15)(160)}}
\]

we obtain \( \gamma = 0.72 \).
REFERENCES


ON DISTRIBUTION-FREE CONFIDENCE BOUNDS FOR $\Pr(Y < X)$

For the case in which $X$ and $Y$ are both unknown, distribution-free confidence intervals for $\Pr(Y < X)$ are obtained by means of the Chebyshew inequality and van Dantzig's [4] upper bound for the variance of the Mann-Whitney statistic $U$. The (two-sided) intervals are shorter than the upper bounds obtained by Birnbaum and McCarty [2], provided the confidence coefficient does not exceed $0.925$ and the sample sizes are not too unequal; and they are more reliable for small sample sizes than the shorter intervals given by Govindaraju [5], especially when $\Pr(Y < X)$ is close to 0 or 1. For equal sample sizes and for several confidence levels and interval widths, the required sample size is compared with the corresponding Birnbaum-McCarty entry. For unequal sizes, lower bounds are given for the ratio of smaller sample size to combined sample size which will ensure that the intervals remain shorter than the Birnbaum-McCarty bounds. Some numerical examples are given.
CONFIDENCE INTERVALS

RELIABILITY

PR(Y < X)