[5] A WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[6] G. HUNT AND C. STEIN, "Most stringent tests of statistical hypotheses," unpublished.
[7] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[8] M. PEISAKOFF, "Transformation parameters," unpublished thesis, Princeton University, 1950.
[9] P. R. HALMOS, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 34–43.
[10] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), pp. 62–69.

# ONE-SIDED CONFIDENCE CONTOURS FOR PROBABILITY DISTRIBUTION FUNCTIONS[1]

By Z. W. BIRNBAUM AND FRED H. TINGEY[2]

*University of Washington*

**Summary.** Let $F(x)$ be the continuous distribution function of a random variable $X$, and $F_n(x)$ the empirical distribution function determined by a sample $X_1, X_2, \cdots, X_n$. It is well known that the probability $P_n(\epsilon)$ of $F(x)$ being everywhere majorized by $F_n(x) + \epsilon$ is independent of $F(x)$. The present paper contains the derivation of an explicit expression for $P_n(\epsilon)$, and a tabulation of the 10%, 5%, 1%, and 0.1% points of $P_n(\epsilon)$ for $n = 5, 8, 10, 20, 40, 50$. For $n = 50$ these values agree closely with those obtained from an asymptotic expression due to N. Smirnov.

**1. Introduction.** Let $X$ be a random variable with the continuous probability distribution function $F(x) = \text{Prob.}\ \{X \le x\}$. An ordered sample $X_1 \le X_2 \le \cdots \le X_n$ of $X$ determines the empirical distribution function

$$F_n(x) = \begin{cases} 0 & \text{for } x < X_1, \\ \dfrac{k}{n} & \text{for } X_k \le x < X_{k+1}, \qquad k = 1, 2, \cdots, n-1, \\ 1 & \text{for } X_n \le x. \end{cases}$$

The function

$$F^{+}_{n,\epsilon}(x) = \min\ [F_n(x) + \epsilon, 1],$$

also determined by the sample, will be called an *upper confidence contour*. It is well known [2] that the probability

$$P_n(\epsilon) = \text{Prob.}\ \{F(x) \le F^{+}_{n,\epsilon}(x) \text{ for all } x\}$$

of $F(x)$ being everywhere majorized by $F^{+}_{n,\epsilon}(x)$ is independent of the distribution $F(x)$. An expression for $P_n(\epsilon)$ in determinant form was given by A. Wald and

[1] Presented to the American Mathematical Society on April 28, 1951.
[2] Research under the sponsorship of the Office of Naval Research.

J. Wolfowitz [2]. N. Smirnov [1] obtained the asymptotic expression

$$\lim_{n\to\infty} P_n\left(\frac{z}{\sqrt{n}}\right) = 1 - e^{-2z^2}. \tag{1.1}$$

The present paper contains the derivation of an explicit expression for $P_n(\epsilon)$, and a tabulation of values $\epsilon_{n,\alpha}$ such that

$$P_n(\epsilon_{n,\alpha}) = 1 - \alpha \tag{1.2}$$

for $\alpha = .10, .05, .01, .001$, and $n = 5, 8, 10, 20, 40, 50$. For $n = 50$ these values agree very closely with those obtained from Smirnov's asymptotic expression (1.1).

**2. Two integral formulae.** For any integer $k$, $1 \leq k \leq n$, we have

$$f_{k-1}(X_{k-1}) = \int_{X_{k-1}}^1 \int_{X_k}^1 \cdots \int_{X_{n-1}}^1 dX_n \cdots dX_{k+1}\, dX_k = \frac{(1 - X_{k-1})^{n-k+1}}{(n-k+1)!}. \tag{2.1}$$

This formula is well known and may be obtained by an easy induction.

For any integer $k \geq 0$ we have

$$\int_0^\epsilon \int_{X_1}^{(1/n)+\epsilon} \cdots \int_{X_k}^{(k/n)+\epsilon} dX_{k+1} \cdots dX_2\, dX_1 = \frac{\epsilon}{(k+1)!}\left(\epsilon + \frac{k+1}{n}\right)^k. \tag{2.2}$$

To prove (2.2) one shows by induction that the left-hand expression is equal to

$$\frac{\epsilon}{(m+2)!}\sum_{j=1}^{m+2}\binom{m+2}{j}\left(\epsilon + \frac{m+2-j}{n}\right)^{m+1}(-1)^{j-1},$$

which is equal to the right-hand term in view of the identity

$$\sum_{j=0}^{m+2}\binom{m+2}{j}\left(\epsilon + \frac{m+2-j}{n}\right)^{m+1}(-1)^{j-1} = 0.$$

**3. An expression for $P_n(\epsilon)$.**

THEOREM. *For $0 < \epsilon \leq 1$ we have*

$$P_n(\epsilon) = 1 - \epsilon \sum_{j=0}^{[n(1-\epsilon)]}\binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}, \tag{3.0}$$

*where $[n(1 - \epsilon)]$ = greatest integer contained in $n(1 - \epsilon)$.*

PROOF. Since $P_n(\epsilon)$ does not depend on $F(x)$, we will assume that $X$ has the probability distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ x & \text{for } 0 \leq x < 1, \\ 1 & \text{for } 1 \leq x. \end{cases}$$

For this random variable, $P_n(\epsilon)$ is the probability that the ordered sample

$$0 \leq X_1 \leq X_2 \leq \cdots \leq X_n \leq 1 \tag{3.1}$$

falls into the region

$$\begin{aligned} &X_{j-1} \leq X_j \leq \frac{j-1}{n} + \epsilon \qquad \text{for } j = 1, \cdots, K+1, \\ &X_{j-1} \leq X_j \leq 1 \qquad \text{for } j = K+2, \cdots, n, \end{aligned} \tag{3.2}$$

where $X_0 = 0$ and $K = [n(1 - \epsilon)]$. Since the probability density of an ordered sample $(X_1, X_2, \cdots, X_n)$ is equal to $n!$ in the region (3.1) and to zero elsewhere, the probability of (3.2) is equal to

$$P_n(\epsilon) = n!J(\epsilon, n, K), \tag{3.3}$$

where

$$J(\epsilon, n, K) = \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \int_{X_2}^{(2/n)+\epsilon} \cdots \int_{X_K}^{(K/n)+\epsilon} \int_{X_{K+1}}^{1} \int_{X_{K+2}}^{1} \cdots \int_{X_{n-1}}^{1} dX_n \cdots dX_{K+3}\, dX_{K+2}\, dX_{K+1} \cdots dX_3\, dX_2\, dX_1. \tag{3.4}$$

By (2.1) we see that

$$J(\epsilon, n, k) = \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \int_{X_2}^{(2/n)+\epsilon} \cdots \int_{X_k}^{(k/n)+\epsilon} \frac{(1 - X_{k+1})^{n-k-1}}{(n - k - 1)!} dX_{k+1} \cdots dX_3\, dX_2\, dX_1 . \tag{3.5}$$

We will prove by induction

$$J(\epsilon, n, k + 1) = J(\epsilon, n, k) - \frac{\epsilon}{n!}\binom{n}{k + 1}\left(1 - \epsilon - \frac{k + 1}{n}\right)^{n-k-1} \cdot \left(\epsilon + \frac{k + 1}{n}\right)^{k}, \tag{3.6}$$

for any integer $0 \leq k \leq n - 1$. For $k = 0$, (3.6) can be verified directly. Assuming (3.6) for $k \leq m$, we obtain

$$\begin{aligned} J(\epsilon, n, m + 1) &= \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \int_{X_m}^{(m/n)+\epsilon} \int_{X_{m+1}}^{((m+1)/n)+\epsilon} \frac{(1 - X_{m+2})^{n-m-2}}{(n - m - 2)!} dX_{m+2}\, dX_{m+1} \cdots dX_2\, dX_1 \\ &= \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \cdots \int_{X_m}^{(m/n)+\epsilon} \frac{(1 - X_{m+1})^{n-m-1}}{(n - m - 1)!} dX_{m+1} \cdots dX_2\, dX_1 \\ &\quad - \frac{\left(1 - \epsilon - \frac{m + 1}{n}\right)^{n-m-1}}{(n - m - 1)!} \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \cdots \int_{X_m}^{(m/n)+\epsilon} dX_{m+1} \cdots dX_2\, dX_1, \end{aligned}$$

and, by the assumption of induction and (2.2), this is

$$J(\epsilon, n, m) - \frac{\epsilon}{n!}\binom{n}{m + 1}\left(1 - \epsilon - \frac{m + 1}{n}\right)^{n-m-1}\left(\epsilon + \frac{m + 1}{n}\right)^{m},$$

which proves (3.6).

Noting that $J(\epsilon, n, 0) = \frac{1}{n!}[1 - (1 - \epsilon)^n]$, one obtains from (3.6)

$$J(\epsilon, n, k) = \frac{1}{n!}[1 - (1 - \epsilon)^n] - \frac{\epsilon}{n!}\sum_{j=1}^{k}\binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}.$$

This, together with (3.3) completes the proof of (3.0).

*Remark.* Setting $F_{n,\epsilon}^{-}(x) = \max [F_n(x) - \epsilon, 0]$, one easily verifies that Prob. $\{F(x) \geq F_{n,\epsilon}^{-}(x)$ for all $x\}$ is equal to $P_n(\epsilon)$, and hence also is given by (3.0).

**4. Tabulation of $\epsilon_{n,\alpha}$ and comparison with asymptotic values.** Table 1 contains numerical solutions $\epsilon_{n,\alpha}$ of equation (1.2), computed to a number of digits sufficient to assure that $|P_n(\epsilon_{n,\alpha}) - (1 - \alpha)| < 5 \cdot 10^{-5}$.

TABLE 1.[3]

*Solutions $\epsilon_{n,\alpha}$ of equation (1.2)*

| $n$ \ $\alpha$ | .100 | .050 | .010 | .001 |
|---|---|---|---|---|
| 5 | .4470 | .5094 | .6271 | .7480 |
| 8 | .3583 | .4096 | .5065 | .6130 |
| 10 | .3226 | .3687 | .4566 | .5550 |
| 20 | .23155 | .26473 | .3285 | .4018 |
| 40 | .16547 | .18913 | .2350 | .2877 |
| 50 | .14840 | .16959 | .2107 | .2581 |

Setting $z/\sqrt{n} = \tilde{\epsilon}_{n,\alpha}$ in (1.1), one obtains for large $n$ the asymptotic values

$$\tilde{\epsilon}_{n,\alpha} = \sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}. \tag{4.1}$$

These values are presented in Table 2.

TABLE 2

*Values of $\tilde{\epsilon}_{n,\alpha} = \sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}$*

| $n$ \ $\alpha$ | .100 | .050 | .010 | .001 |
|---|---|---|---|---|
| 5 | .4799 | .5473 | .6786 | .8311 |
| 8 | .3794 | .4327 | .5365 | .6571 |
| 10 | .3393 | .3870 | .4799 | .5877 |
| 20 | .2399 | .2737 | .3393 | .4156 |
| 40 | .1697 | .1935 | .2399 | .2938 |
| 50 | .1517 | .1731 | .2146 | .2628 |

A comparison of the two tables indicates that, for the probability levels $.001 \leq \alpha \leq .1$, the asymptotic values $\tilde{\epsilon}_{n\alpha}$, are greater than the "exact" values

[3] The authors wish to express their appreciation to the National Bureau of Standards, Institute for Numerical Analysis, for performing the computations which are summarized in this table.

$\epsilon_{n,\alpha}$ so that the error committed by using $\tilde{\epsilon}_{n,\alpha}$ instead of $\epsilon_{n,\alpha}$ would be in the safe direction, and that this error becomes already very small for $n = 50$.

REFERENCES

[1] N. SMIRNOV, "Sur les écarts de la courbe de distribution empirique," *Rec. Math.* (*Mat. Sbornik*), N. S. Vol. 6 (48) (1939), pp. 3–26.

[2] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.

# ON THE ESTIMATION OF CENTRAL INTERVALS WHICH CONTAIN ASSIGNED PROPORTIONS OF A NORMAL UNIVARIATE POPULATION

BY G. E. ALBERT AND RALPH B. JOHNSON

*University of Tennessee and Clemson Agricultural College*

**Summary.** For samples of any given size $N \geq 2$ from a normal population, Wilks [1] has shown how to choose the parameter $\lambda_p$ so that the expected coverage of the interval $\bar{x} \pm \lambda_p s$ will be $1 - p$. The present paper treats the choice of the minimal sample size $N$ necessary to effect a certain type of statistical control on the fluctuation of that coverage about its expected value; a brief table of such minimal sample sizes is given.

**1. Introduction.** Let $F(y)$ denote the normal cumulative distribution function

$$F(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(u-m)^2/(2\sigma^2)}\, du. \tag{1}$$

If $p$ is any number in the range $0 < p < 1$, factors $\lambda(p)$ are well known such that the proportion

$$A = F(m + \lambda\sigma) - F(m - \lambda\sigma) \tag{2}$$

of the probability between $m \pm \lambda\sigma$ will equal $1 - p$.

If $m$ and $\sigma$ are unknown, it is natural to consider the random variable

$$A(\bar{y}, s; \lambda) = F(\bar{y} + \lambda s) - F(\bar{y} - \lambda s), \tag{3}$$

where $\bar{y} = \sum_{n=1}^{N} y_n/N$ and $s = \left\{\sum_{i=1}^{N} (y_i - \bar{y})^2/(N - 1)\right\}^{\frac{1}{2}}$.

Obviously $\lambda$ cannot be chosen to guarantee $A(\bar{y}, s; \lambda) = 1 - p$. S. S. Wilks [1] has shown that, for a random sample of size $N$, the expectation of (3) is $1 - p$,

$$EA(\bar{y}, s; \lambda) = 1 - p, \tag{4}$$

if the parameter $\lambda$ is chosen as

$$\lambda = t_p \sqrt{\frac{N+1}{N}}. \tag{5}$$