Table II gives the confidence γ that 100 P percent of the population lies between the largest and smallest of a random sample of n.

In the case where we are dealing with a multivariate population, we take m to be the number of blocks (See Tukey [8]) excluded from the tolerance region.

REFERENCES

[1] C. E. Eisenhart, M. W. Hastay, and W. A. Wallis, *Techniques of Statistical Analysis*, McGraw-Hill Book Company Inc., New York, 1947, Chapter 2.

[2] D. A. S. Fraser, "Sequentially determined statistically equivalent blocks," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 372–381.

[3] D. A. S. Fraser, "Nonparametric tolerance regions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 44–55.

[4] R. B. Murphy, "Nonparametric tolerance limits," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 581–589.

[5] Karl Pearson, *Tables of the Incomplete Beta-function*, Cambridge University Press (1934).

[6] Harry G. Romig, "50–100 Binomial Tables," John Wiley and Sons, New York (1947).

[7] Catherine M. Thompson, "Tables of the Percentage Points of the Incomplete Beta-function," *Biometrika*, Vol. 32 (1941–42), pp. 151–181.

[8] J. W. Tukey, "Non-parametric estimation, II; Statistically equivalent blocks and tolerance regions, the continuous case," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 529–539.

[9] A. Wald, "An extension of Wilks' method for setting tolerance limits," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 45–55.

[10] S. S. Wilks, "Determination of sample sizes for setting tolerance limits," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 91–96.

[11] Tables of the Cumulative Binomial Probability Distribution, *Annals of the Computation Laboratory of Harvard University*, Vol. XXXV (1955).

NONPARAMETRIC ESTIMATION OF SAMPLE PERCENTAGE POINT STANDARD DEVIATION

By John E. Walsh

Military Operations Research Division, Lockheed Aircraft Corporation, Burbank, California

1. Summary. The available data consists of a random sample $x(1) < \cdots < x(n)$ from a reasonably well-behaved continuous statistical population. The problem is to estimate the standard deviation of a specified $x(r)$ that is not in the tails of the sample. The estimates examined are of the form

$$a[x(r + i) - x(r - i)]$$

and the explicit problem consists of determining suitable values for a and i. The solution

$$a = (\tfrac{1}{2})(n + 1)^{-3/10}\{[r/(n + 1)][1 - r/(n + 1)]\}^{1/2}, \; i \doteq (n + 1)^{4/5}$$

Received November 15, 1957.

appears to be satisfactory. Then the expected value of the estimate equals the standard deviation of $x(r)$ plus $O(n^{-9/10})$; also the standard deviation of this estimate is $O(n^{-9/10})$. That is, the fixed and random errors for this point estimate are of the same order of magnitude with respect to n. Solutions can be obtained which decrease the order of one of these types of error. However, these solutions increase the order of the other type of error, so that the overall error magnitude exceeds $O(n^{-9/10})$.

2. Introduction and results. A sample percentage point $x(r)$ furnishes a point estimate of the corresponding population percentage point $\theta[r/(n+1)]$, where $\theta(p) = \theta[p]$ represents the $100p$ percent point of the population sampled. The appropriateness of $x(r)$ as an estimate depends on its variability. Thus an estimate of the standard deviation of $x(r)$ can be of value. This paper presents an easily computed nonparametric estimate of the standard deviation of $x(r)$ that is valid for most continuous populations of practical interest and has favorable properties compared to other estimates of the same type.

The estimate derived is based on the results of [1]. The expected value and variance-covariance expansions of [1] are assumed to be valid for the continuous statistical population sampled. In particular, this population is assumed to have a probability density function that is analytic and nonzero at all points of interest. These requirements appear to be satisfied for most practical situations that involve continuous populations.

The derived estimate properties are approximate in the sense that terms of specified orders of n are neglected. The order results stated are not applicable for the case of extreme observations. That is, $p_r = r/(n+1)$ and $q_r = 1 - p_r$ are assumed to be bounded away from 0 and 1. Also a standard deviation estimate is not necessarily reasonably accurate even for situations where the order relations are valid. In some cases the neglected terms may be important even though they are of the stated order with respect to n. For many commonly encountered types of populations (unimodal, etc.), the importance of the neglected terms tends to increase as p_r deviates from $\frac{1}{2}$. Examination of the expansions used in the derivations suggests that the standard deviation estimates presented are usually satisfactory if

$$p_r q_r (n+1)^{9/10} \geqq 3.$$

This relation implies that the magnitude of the increments with respect to which the final expansions are made never exceeds $(\frac{1}{3})p_r q_r$.

Let $\sigma\{w\}$ denote the population standard deviation of w. The statistic advocated for estimating $\sigma\{x(r)\}$ is

$$s[x(r)] = \tfrac{1}{2}(n+1)^{-3/10}\sqrt{p_r q_r}\{x[r + (n+1)^{4/5}] - x[r - (n+1)^{4/5}]\},$$

where $x[z] = x$ (largest integer contained in z). This statistic has the properties

$$\begin{aligned} E\{s[x(r)]\} &= \sigma\{x(r)\} + O(n^{-9/10}) \\ &= \sigma\{x(r)\}[1 + O(n^{-2/5})], \\ \sigma\{s[x(r)]\} &= O(n^{-9/10}). \end{aligned}$$

The statistic $s[x(r)]$ has the smallest order error of all expected value estimates of $\sigma\{x(r)\}$ that are of the form $a[x(r+i) - x(r-i)]$, where i is $o(n)$. Here the order of the error of an estimate is considered to be the larger of the order of $E(\text{estimate}) - \sigma\{x(r)\}$ and the order of $\sigma\{\text{estimate}\}$.

The notation $f[x]$ is used to represent the probability density function of the population sampled. For the situation considered,

$$\sigma\{x(r)\} = \sqrt{p_r q_r}/\sqrt{n+1}f[\theta(p_r)] + O(n^{-3/2}).$$

This relation shows that a modification of $s[x(r)]$ can be used in estimating the value of the density function at the point $x = \theta(p_r)$. Explicitly,

$$\sqrt{n+1}s[x(r)]/\sqrt{p_r q_r}$$

furnishes an expected value estimate of $1/f[\theta(p_r)]$ that is accurate to terms of order $n^{-2/5}$.

3. Derivations. This section contains a verification of the properties stated for $s[x(r)]$ in the preceding sections.

Consider any integer t such that $1 \leqq t \leqq n$. From the results of [1],

$$(1)\qquad \begin{aligned} E[x(t)] &= \theta(p_t) - \frac{p_t q_t f'[\theta(p_t)]}{2(n+2)f[\theta(p_t)]^2} + O(n^{-2}), \\ \sigma\{x(t)\} &= \sqrt{p_t q_t}/\sqrt{n+1}f(\theta(p_t)] + O(n^{-3/2}). \end{aligned}$$

Here $p_t = t/(n+1)$ and $q_t = 1 - p_t$. These expansions, combined with appropriate use of Taylor series expansions, form the basis for the derivations.

Let $i = \epsilon(n+1)^\alpha$ = integer, where $0 \leqq \alpha < 1$ and both ϵ and α are $0(1)$. Using (1) and expanding around r in Taylor series,

$$E[x(r+i)] = \theta(p_r) + \frac{\epsilon}{(n+1)^{1-\alpha}f[\theta(p_r)]} - \left[\frac{\epsilon^2}{(n+1)^{2-2\alpha}} + \frac{p_r q_r}{2(n+2)}\right]\frac{f'[\theta(p_r)]}{f[\theta(p_r)]^3} + O(n^{-3+3\alpha}) + O(n^{-2+\alpha})$$

$$E[x(r-i)] = \theta(p_r) - \frac{\epsilon}{(n+1)^{1-\alpha}f[\theta(p_r)]} - \left[\frac{\epsilon^2}{(n+1)^{2-2\alpha}} + \frac{p_r q_r}{2(n+2)}\right]\frac{f'[\theta(p_r)]}{f[\theta(p_r)]^3} + O(n^{-3+3\alpha}) + O(n^{-2+\alpha})$$

$$E\{a[x(r+i) - x(r-i)]\} = \frac{2a\epsilon}{(n+1)^{1-\alpha}f[\theta(p_r)]} + O(an^{-3+3\alpha}) + O(an^{-2+\alpha})$$

$$\sigma\{a[x(r+i) - x(r-i)]\} = a\sqrt{2\epsilon}/(n+1)^{1-\alpha/2}f[\theta(p_r)] + O(an^{-3/2+\alpha}).$$

The problem is to use these relations to determine suitable values for ϵ, α, and a.

Since $a[x(r+i) - x(r-i)]$ is an expected value estimate of $\sigma\{x(r)\}$,

$$2a\epsilon/(n+1)^{1-\alpha} = \sqrt{p_r q_r}/(n+1)^{1/2}, \quad \text{or} \quad a = (1/2\epsilon)\sqrt{p_r q_r}(n+1)^{1/2-\alpha}.$$

Using this expression for a,

$$E\{a[x(r + i) - x(r - i)]\} = \sigma\{x(r)\} + O(n^{-5/2+2\alpha}) + O(n^{-3/2})$$
$$\sigma\{a[x(r + i) - x(r - i)]\} = O(n^{-1/2-\alpha/2}).$$

Thus increasing α decreases the order of magnitude of

$$\sigma\{a[x(r + i) - x(r - i)]\},$$

but increases the order of

$$E\{a[x(r + i) - x(r - i)]\} - \sigma\{x(r)\}.$$

Hence the order of the error is minimized when

$$-1/2 - \alpha/2 = -5/2 + 2\alpha.$$

Thus $\alpha = 4/5$ appears to be the most desirable choice for α.

In $\sigma\{a[x(r + i) - x(r - i)]\}$, the parameter ϵ appears predominantly as the factor $1/\sqrt{\epsilon}$. In $E\{a[x(r + i) - x(r - i)]\} - \sigma\{x(r)\}$ the predominant factor is ϵ^2. Solution of the equation

$$\epsilon^2 = 1/\sqrt{\epsilon}$$

suggests that $\epsilon = 1$ is an appropriate compromise choice for ϵ.

Use of $\alpha = 4/5$, $\epsilon = 1$, and the expression for a yields the results

$$i = (n + 1)^{4/5}, \qquad a = \tfrac{1}{2}(n + 1)^{-3/10}\sqrt{p_r q_r},$$

and verifies the properties stated for $s[x(r)]$.

REFERENCE

[1] F. N. DAVID AND N. L. JOHNSON, "Statistical treatment of censored data. Part I. Fundamental formulae," *Biometrika*, Vol. 41 (1954), p. 228–240.

A UNIQUENESS PROPERTY NOT ENJOYED BY THE NORMAL DISTRIBUTION

BY GEORGE P. STECK

Sandia Corporation

1. Summary. It is well known that if X and Y (or $1/X$ and $1/Y$) are independently normally distributed with mean zero and variance σ^2, then X/Y has a Cauchy distribution. It is the purpose of this note to show that the converse statement is not true. That is, the fact that the ratio of two independent, identically distributed, random variables X and Y follows a Cauchy distribution is not sufficient to imply that X and Y (or $1/X$ and $1/Y$) are normally distributed. This will be shown by exhibiting several counterexamples.

Received October 21, 1957; revised December 26, 1957.