The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1753–1755

A CENTRAL LIMIT THEOREM WITH NONPARAMETRIC APPLICATIONS[1]

BY GOTTFRIED E. NOETHER

The University of Connecticut

Many nonparametric test and confidence interval procedures can be based on statistics of the type

$$S = \sum_{i \in I} \sum_{j \in J} v_{ij},$$

where $v_{ij} = 1$ or 0 depending on whether some other variable u_{ij} is $<$ or $> u_0$, u_0 being an appropriate constant, and where two variables u_{ij} and u_{gh} are independent if no subscript in the (i, j)-pair matches a subscript in the (g, h)-pair. If the number of elements in I and the number of elements in J depend linearly on some index N, we are interested in a simple central limit theorem for statistics S as N increases indefinitely.

Let $\mu = ES$ and $\sigma^2 = \operatorname{var} S$. We then have the following

THEOREM. *A sufficient condition for the asymptotic normality of* $(S-\mu)/\sigma$ *is that* Var S *is of order* N^3. (*We shall express this condition by writing* $\sigma^2 = \Omega(N^3)$.)

PROOF. Let $w_{ij} = v_{ij} - p_{ij}$, where $p_{ij} = P(v_{ij} = 1) = P(u_{ij} < u_0)$, and set

$$W = \sum_{i \in I} \sum_{j \in J} w_{ij}.$$

If μ_r denotes the rth moment of W, $r = 3, 4, \cdots$, we shall show that for $k = 2, 3, \cdots$, $\lim_{N\to\infty} \mu_{2k-1}/\sigma^{2k-1} = 0$ and $\lim_{N\to\infty} \mu_{2k}/\sigma^{2k} = (2k-1)(2k-3)\cdots 3$. These are the moments of the standard normal distribution, so that the theorem follows.

We have

$$\sigma^2 = \operatorname{var} W = EW^2 = \sum_{i,j} \sum_{g,h} Ew_{ij} w_{gh}. \tag{1}$$

Under *ties* among the subscripts we shall understand that one or both subscripts in the (i, j)-pair are tied with one or both subscripts in the (g, h)-pair. A *simple* tie is one in which exactly one subscript in one pair ties exactly one subscript in the other pair. Since for untied subscripts $Ew_{ij} w_{gh} = 0$, we have Var $W = O(N^3)$. The requirement Var $W = \Omega(N^3)$ implies that not all covariances between two simply tied variables w_{ij} and w_{gh} vanish.

Before we investigate the general moments of order $2k-1$ and $2k$, let us fix ideas by considering the case $k = 2$. We have

$$\mu_3 = EW^3 = \sum_{i_1,j_1} \sum_{i_2,j_2} \sum_{i_3,j_3} Ew_{i_1 j_1} w_{i_2 j_2} w_{i_3 j_3}.$$

$Ew_{i_1 j_1} w_{i_2 j_2} w_{i_3 j_3}$ can differ from 0 only if there are ties involving subscripts from each of the three subscript pairs. Thus $\mu_3 = O(N^4)$ and $\mu_3/\sigma^3 = O(N^4)/\Omega(N^{9/2}) = O(N^{-\frac{1}{2}})$. For μ_4 the contribution arising from ties involving three or four subscript

Received November 21, 1969.

[1] Research supported by NSF Grant GP-11470.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.*
www.jstor.org

groups can be ignored compared to the contribution arising from two simple ties among the four subscript pairs. Since there are three possible pairings, we have asymptotically

$$\mu_4 \sim 3\sum_{i_1,j_1}\sum_{i_2,j_2} Ew_{i_1j_1}w_{i_2j_2}\sum_{i_3,j_3}\sum_{i_4,j_4} Ew_{i_3j_3}w_{i_4j_4} = 3\sigma^2\sigma^2 = 3\sigma^4,$$

so that $\mu_4/\sigma^4 \sim 3$.

More generally,

$$\mu_{2k-1} = EW^{2k-1} = O(N^{3(k-2)+4})$$

resulting from $k-2$ simple and one triple tie among the $2k-1$ subscript groups. As a consequence, $\mu_{2k-1}/\sigma^{2k-1} = O(N^{-\frac{1}{2}})$. Finally the asymptotic value of $\mu_{2k} = EW^{2k}$ is determined by the contribution of k simple ties among k pairs of subscript groups. There are $(2k-1)(2k-3)\cdots 3$ such pairings, each contributing $(\sigma^2)^k$. Thus

$$\mu_{2k} \sim (2k-1)(2k-3)\cdots 3\sigma^{2k}.$$

This completes the proof.

We observe that this proof does not specifically make use of the fact that the v_{ij} are (0, 1)-variables, only that they are uniformly bounded. The theorem remains valid under this more general assumption. Another generalization consists in considering statistics of the type $S = \sum_{i_1}\cdots\sum_{i_e} v_{i_1\cdots i_e}$. The condition for asymptotic normality then becomes $\operatorname{Var} S = \Omega(N^{2e-1})$.

In conclusion we illustrate the statistic S with three well-known nonparametric procedures. In all three examples it follows immediately from (1) that the condition $\sigma^2 = \Omega(N^3)$ is satisfied not only under the null hypothesis but also under reasonable alternatives.

The Wilcoxon one-sample test. Let $x_1, \cdots, x_N$ be a random sample from a symmetric population with center of symmetry at θ. For $1 \leqq i \leqq j \leqq N$, set $u_{ij} = \frac{1}{2}(x_i + x_j)$. The Wilcoxon one-sample test of the hypothesis $\theta = \theta_0$ can be based on the statistic

$$S = \#(u_{ij} < \theta_0) = \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}$$

where v_{ij} equals 1 or 0 depending on whether u_{ij} is smaller or greater than θ_0. If c is the lower critical value of a two-sided test at significance level α, a confidence interval for θ having confidence coefficient $1-\alpha$ is bounded by the dth smallest and the dth largest of the u_{ij}, where $d = c+1$.

Nonparametric regression. Consider the regression model $y = \alpha + \beta x + e$, where e is a random variable with median 0. We have observations corresponding to N x-values $x_1 < \cdots < x_N$. For $1 \leqq i < j \leqq N$, set $u_{ij} = (y_j - y_i)/(x_j - x_i)$. The test of the hypothesis $\beta = \beta_0$ based on the statistic

$$S = \#(u_{ij} < \beta_0) = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} v_{ij}$$

corresponds to the Mann test of randomness derived from the Kendall rank correlation statistic. A confidence interval for β is found as in the first example.

The Wilcoxon two-sample test. Let $x_1, \cdots, x_m; y_1, \cdots, y_n$ be two random samples from populations with distributions $F(z)$ and $G(z)$, respectively. Set $u_{ij} = y_j - x_i$, $i = 1, \cdots, m$; $j = 1, \cdots, n$, where $m = \lambda N$, $n = (1-\lambda)N$, $0 < \lambda < 1$. The Mann–Whitney form of the Wilcoxon two-sample test of the hypothesis $F(z) = G(z)$ is based on the statistic

$$S = \#(u_{ij} < 0) = \sum_{i=1}^{m} \sum_{j=1}^{n} v_{ij}.$$

A confidence interval for the shift parameter Δ in $G(z) = F(z-\Delta)$ is found as in the first example.

While the asymptotic normality of the test statistics in all three examples is known from other limit theorems, the present theorem seems useful because of its great simplicity.