The Annals of Statistics
1978, Vol. 6, No. 4, 923–925

A NOTE ON THE EDGEWORTH EXPANSION FOR THE KENDALL RANK CORRELATION COEFFICIENT

BY W. ALBERS

Technological University Twente

In this note it is shown how to some extent the Edgeworth expansion for the distribution function of Kendall's τ can be established by using a well-known general result on such expansions.

Let $X_1, Y_1, \cdots, X_N, Y_N$ be independent random variables (rv's), the X_i with a continuous distribution function (df) F, the Y_i with a continuous df G. Let R_i and S_i be the ranks of X_i and Y_i, respectively, then Kendall's rank correlation coefficient is defined as

$$\tau_N = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \operatorname{sign}(R_i - R_j) \operatorname{sign}(S_i - S_j) . \tag{1}$$

Using a direct approach, Prašková-Vizková (1976) establishes the Edgeworth expansion (for a definition see, e.g., Feller (1971), page 542) for the df of τ_N. In the present note we shall point out how to some extent this result can be obtained more easily by applying the following standard theorem due to Feller (1971) (see page 548).

THEOREM. *Let $Z_1, \cdots, Z_N$ be independent* rv's *with zero mean and let $T_N = \sum_{j=1}^N Z_j$. Suppose that for some integer $r \geqq 3$ there exist positive constants c and C such that for $v = 1, \cdots, r+1$ and $j = 1, \cdots, N$,*

$$c < E|Z_j|^v < C . \tag{2}$$

Moreover, assume that the characteristic functions (ch.f.'s) *ρ_j of Z_j, $j = 1, \cdots, N$, satisfy*

$$|\textstyle\prod_{j=1}^N \rho_j(t)| = o(N^{-r-1}) , \tag{3}$$

uniformly in $|t| > \delta$ for all $\delta > 0$. Then $\sup_x |F_N(x) - G_{rN}(x)| = o(N^{-r/2+1})$, where F_N is the df *of $T_N/\sigma(T_N)$ and G_{rN} is its Edgeworth expansion to $O(N^{-r/2+1})$.*

REMARK. As in Feller (1971), the theorem is formulated here for a single sequence of rv's Z_j, $j = 1, \cdots, N$. However, from Feller's proof it is clear that the theorem also holds for a triangular array Z_{jN}, $j = 1, \cdots, N$, $N = 1, 2, \cdots$, provided that conditions (2) and (3) hold uniformly for such Z_{jN}.

To apply the theorem to τ_N in (1), we note that τ_N has the same df as $4T_N/(N-1)$, where $T_N = \sum_{j=1}^{N-1} Z_{jN}$, in which the Z_{jN} are independent rv's with

Received June 1977; revised October 1977.

AMS 1970 *subject classifications.* 62G10, 62G20.

Key words and phrases. Edgeworth expansion, Kendall rank correlation coefficient, characteristic function.

$P(Z_{jN} = k/N) = 1/(j+1)$, $k = -j/2, -j/2+1, \cdots, j/2-1, j/2$ and $j = 1, \cdots, N-1$ (this follows immediately from Hájek & Šidák (1967), page 115 and Hájek (1955)). The problem is, however, that neither (2) nor (3) is satisfied for these Z_{jN}. We shall demonstrate how these obstacles can be removed. As concerns (2), this is quite simple: just write $T_N = \sum_{j=1}^{[N/2]} V_{jN}$, where $V_{jN} = Z_{jN} + Z_{(N-j)N}$, $j = 1, \cdots, [(N-1)/2]$, and for N even, $V_{[N/2]N} = Z_{[N/2]N}$. For these V_{jN} condition (2) holds for all r.

The real problem lies in condition (3). Let ρ_N and ρ_{jN} be the ch.f. of T_N and Z_{jN}, respectively, then $\rho_N(t) = \prod_{j=1}^{N-1} \rho_{jN}(t) = \prod_{j=1}^{N} (\sin\{jt/(2N)\}/(j \sin\{t/(2N)\}))$ (see, e.g., Prašková-Vizková (1976), page 599). Clearly, $|\rho_N(2k\pi N)| = 1$ for $k = \pm 1, \pm 2, \cdots$ and hence (3) does not hold. For this reason we introduce $\tilde{T}_N = T_N + U_N$, with $U_N = \sum_{i=1}^{[\log^2 N]} U_{iN}$. Here the U_{iN} are independent rv's, also independent of the Z_{jN} and all uniformly distributed on $(-1/(2N), 1/(2N))$. The difference $U_N = \tilde{T}_N - T_N$ is small with respect to T_N for two reasons: in the first place U_N has $[\log^2 N]$ rather than N terms and furthermore the support of the U_{iN} is of a smaller order of magnitude than the supports of (most of) the Z_{jN}. Nevertheless, adding U_N to T_N suffices to overcome (3): $\tilde{T}_N$ has ch.f.

$$\tilde{\rho}_N(t) = \prod_{j=1}^{N-[\log^2 N]} \frac{\sin\{jt/(2N)\}}{(j \sin\{t/(2N)\})} \prod_{j=N-[\log^2 N]+1}^{N} \frac{\sin\{jt/(2N)\}}{\{jt/(2N)\}}$$

and it follows that for $\tilde{\rho}_N$ condition (3) holds for all r. Hence the replacement of some of the lattice rv's Z_{jN} in T_N by smooth rv's $Z_{jN} + U_{iN}$—which in fact are uniformly distributed on $(-(j+1)/(2N), (j+1)/(2N))$—enables us to apply the theorem for arbitrary r to the resulting rv $\tilde{T}_N$.

Let F_N and $\tilde{F}_N$ be the df of $T_N/\sigma(T_N)$ and $\tilde{T}_N/\sigma(\tilde{T}_N)$, respectively, and let G_{rN} and $\tilde{G}_{rN}$ be their Edgeworth expansions. According to the above, $\sup_x |\tilde{F}_N(x) - \tilde{G}_{rN}(x)| = o(N^{-r/2+1})$ for all r. It remains to find out what this means for $\sup_x |F_N(x) - G_{rN}(x)|$. We note in the first place that $E|U_N|^k = O(N^{-k} \log^k N)$, $k = 1, 2, \cdots, \sigma^2(T_N) = (N-1)(2N+5)/(72N)$ and $\sigma^2(\tilde{T}_N) = \sigma^2(T_N) + O(N^{-2} \log^2 N)$. Then we observe that $P(T_N/\sigma(T_N) \leqq x) \leqq P(\tilde{T}_N/\sigma(T_N) \leqq x+\varepsilon) + P(|U_N|/\sigma(T_N) \geqq \varepsilon)$ for all $\varepsilon > 0$. Combining this with a similar inequality in the opposite direction, we obtain that

$$\begin{aligned} &|F_N(x) - \tilde{F}_N(x\sigma(T_N)/\sigma(\tilde{T}_N))| \\ &\qquad \leqq P(x - \varepsilon \leqq \tilde{T}_N/\sigma(T_N) \leqq x + \varepsilon) + P(|U_N| \geqq \varepsilon\sigma(T_N)) . \end{aligned}$$

Using the results above and Chebyshev's inequality, we find for $r \geqq 3$ and k sufficiently large that

$$\begin{aligned} \sup_x &|F_N(x) - \tilde{F}_N(x\sigma(T_N)/\sigma(\tilde{T}_N))| \\ &\leqq \sup_x |\tilde{G}_{rN}((x+\varepsilon)\sigma(T_N)/\sigma(\tilde{T}_N)) \\ &\qquad - \tilde{G}_{rN}((x-\varepsilon)\sigma(T_N)/\sigma(\tilde{T}_N))| + o(N^{-r/2+1}) + \{\varepsilon\sigma(T_N)\}^{-k} E|U_N|^k \\ &= O(\varepsilon + \varepsilon^{-k} N^{-3k/2} \log^k N) + o(N^{-r/2+1}) = O(N^{-\frac{3}{2}+\eta}) + o(N^{-r/2+1}) \end{aligned}$$

for all $\eta > 0$, where the last step follows by choosing $\varepsilon = N^{-\frac{3}{2}+\eta}$.

Next we note that $\sup_x |G_{rN}(x) - \tilde{G}_{rN}(x)| = O(N^{-\frac{3}{2}})$ for all r and that $\sup_x |\tilde{G}_{rN}(x\sigma(T_N)/\sigma(\tilde{T}_N)) - \tilde{G}_{rN}(x)| = O(|\sigma(T_N)/\sigma(\tilde{T}_N) - 1|) = O(N^{-\frac{3}{2}})$ for all r. Hence, for all $r \geqq 3$, we have

$$\begin{aligned}\sup_x &|\tilde{F}_N(x\sigma(T_N)/\sigma(\tilde{T}_N)) - G_{rN}(x)| \\ &\leqq \sup_x |\tilde{F}_N(x\sigma(T_N)/\sigma(\tilde{T}_N)) - \tilde{G}_{rN}(x\sigma(T_N)/\sigma(\tilde{T}_N))| \\ &\qquad + \sup_x |\tilde{G}_{rN}(x) - \tilde{G}_{rN}(x\sigma(T_N)/\sigma(\tilde{T}_N))| + \sup_x |\tilde{G}_{rN}(x) - G_{rN}(x)| \\ &= o(N^{-r/2+1}) + O(N^{-\frac{3}{2}}) .\end{aligned}$$

Combining the results above we finally arrive at the conclusion that $\sup_x |F_N(x) - G_{rN}(x)| = O(N^{-\frac{3}{2}+\eta}) + o(N^{-r/2+1})$ for all $r \geqq 3$ and $\eta > 0$. As G_{rN} is continuous while F_N has jumps of order $N^{-\frac{3}{2}}$, this is, apart from η, the best result possible when ordinary Edgeworth expansions are used. However, it should be clear that the present result is weaker than the one obtained by Prašková-Vizková (1976), using methods of Esseen (1945). She adds terms to the G_{rN} to account for the lattice character. With the generalized Edgeworth expansions thus obtained one can approximate F_N to every order desired by using sufficiently many terms of the expansion, whereas with the G_{rN} the rate of convergence will never be better than $O(N^{-\frac{3}{2}})$.

REFERENCES

[1] Esseen, C. G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77** 1-125.

[2] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, **2**. Wiley, New York.

[3] Hájek, J. (1955). Some rank distributions and their use. *Časopis Pěst. Mat.* **80** 17-31. (In Czechoslovakian.)

[4] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* Academia, Prague.

[5] Prašková-Vizková, Z. (1976). Asymptotic expansion and a local limit theorem for a function of the Kendall rank correlation coefficient. *Ann. Statist.* **4** 597-606.

Department of Mathematics
Twente University of Technology
P. O. Box 217
Enschede, The Netherlands