The Annals of Statistics
1980, Vol. 8, No. 4, 851–864

AN ASYMPTOTIC EXPANSION FOR PERMUTATION TESTS WITH SEVERAL SAMPLES

BY J. ROBINSON

University of Sydney

Let V_n be the standardized sum of squares of the means of $r + 1$ random samples of sizes $s_0, s_1, \cdots, s_r$, where $n = s_0 + s_1 + \cdots + s_r$, taken without replacement from n numbers. Then using an approximation to the characteristic function of the means, an asymptotic expansion is obtained for the distribution of V_n with first term being the distribution function of χ_r^2 and with error of approximation generally of smaller order than $1/n$. When the numbers are the first n integers, V is the Kruskal-Wallis statistic and the approximation is compared with the exact distribution in some examples of this special case.

1. Introduction. Permutation tests for the case of several samples or treatments arise in testing the null hypothesis that the distribution of a set of observations is invariant under all permutations of the observations, against the alternative hypothesis that there are several different samples or treatments. This arises either from independent sampling of several populations and the null hypothesis is that these populations are the same or from a physical randomization process in a completely randomized design with several treatments applied at random and the null hypothesis is that all treatments have the same effect. The most common tests of significance are based on sums of squares of sample means of scores. In particular, the permutation test in the analysis of variance uses observations as scores and the Kruskal-Wallis test uses ranks as scores. We are interested in obtaining an asymptotic approximation to the distribution of these test statistics as the minimum number in the samples becomes large. The results are an extension of the asymptotic expansions under the null hypothesis for the two-sample case, obtained by Bickel and van Zwet (1978) and Robinson (1978).

Let $\{a_{nk} : k = 1, \cdots, n, n = 2, 3, \cdots\}$ be a triangular array of real numbers and suppose $\Sigma_k a_{nk} = 0$ and $\Sigma_k a_{nk}^2 = 1$, where here and in the sequel Σ_k denotes summation over k from 1 to n. Let

$$X_{nj} = [(n-1)/n]^{\frac{1}{2}} \Sigma_{k=S_{j-1}+1}^{S_j} a_{nR_{nk}}, \qquad j = 0, 1, \cdots, r,$$

where $s_0, s_1, \cdots, s_r$ are integers (the sample sizes) such that $s_0 + \cdots + s_r = n$ and $S_j = s_0 + \cdots + s_j$ with $S_{-1} = 0$ and $(R_{n1}, \cdots, R_{nn})$ is a random vector taking each permutation of $(1, \cdots, n)$ with probability $1/n!$. Let $p_j = s_j/n$ and $q_j = 1 - p_j, j = 0, 1, \cdots, r$.

$$EX_{nj} = 0, \operatorname{Cov}(X_{nj}, X_{nj'}) = p_j q_j, \qquad j = j',$$
$$= -p_j p_{j'}, \qquad j \neq j'.$$

Received February 1978; revised August 1978.

AMS 1970 *subject classifications.* Primary 60F05; secondary 62G10.

Key words and phrases. Asymptotic expansions, the k sample problem, permutation tests, completely randomized design, sampling without replacement.

Let

$$V_n = \Sigma_j X_{nj}^2 / p_j,$$

where here and in the sequel Σ_j denotes summation over j from 0 to r.

V_n is the form of test statistic used in the case of $r + 1$ samples. It can be shown that V_n converges in distribution to a chi-squared variate with r degrees of freedom, if $b_n = \max_k |a_{nk}|$ tends to zero and $np = n \min_j p_j$ tends to infinity. This follows from the results of Rosén (1965) and Hájek and Šidák (1967, Chapter V) where a functional limit theorem for partial sums of the a_{nk} is given.

Here we will obtain an approximation for $F_n(x) = P(V_n \leqslant x)$ by the expansion (1)

$$G_n(x) = P(\chi_r^2 \leqslant x) - \Big\{ \tfrac{1}{24}(\Sigma_k a_{nk}^4 - 3/n)(\Sigma_j p_j^{-1} - r^2 - 4r - 1)\Big(\frac{3x}{r+2} - 3\Big) - \frac{r(r+2)}{4n}\Big(\frac{x}{r+2} - 1\Big) + \tfrac{1}{72}(\Sigma_k a_{nk}^3)^2 (15\Sigma_j p_j^{-1} - 9r^2 - 36r - 15) \times \Big(\frac{x^2}{(r+4)(r+2)} - \frac{2x}{r+2} + 1\Big)\Big\} \frac{x^{\frac{1}{2}r} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}r}\Gamma\big(\frac{r+2}{2}\big)}.$$

Let $A_{mn} = \Sigma_k |a_{nk}|^m$. In Section 2 it will be shown that the expansion above is a valid approximation with an error of smaller order than A_{4n} subject to a certain condition discussed at the beginning of Section 2. An approximation to the characteristic function of $X_{n1}, \cdots, X_{nr}$ is obtained in Section 2. The formal inversion of the approximating Fourier transform is given in Section 3, while the remainder of Section 2 is devoted to estimating the error of the approximation. Section 4 gives tables comparing the approximation to exact probabilities for the Kruskal-Wallis test statistic in the case $r = 2$.

2. The approximation theorem. If λ denotes Lebesgue measure, then write

$$\gamma(\zeta) = \lambda\{x : \exists k \text{ with } |x - a_{nk}| < \zeta\}.$$

Albers, Bickel and van Zwet assume that the following condition holds for some $\zeta \geqslant n^{-2} \log n$.

CONDITION (B). $\exists \delta > 0$ such that $\gamma(\zeta) \geqslant 4\delta n \zeta$.

They show, under the conditions $A_{2n} \geqslant c$, $A_{4n} \leqslant Cn^{-1}$, for some positive c, C, that Condition (B) implies the condition below, with A_{5n} replaced by $n^{-\frac{3}{2}}$, but their proof can be extended easily to this case. We will assume that these conditions hold for $\zeta \geqslant A_{5n}^{\frac{4}{3}} \log A_{5n}^{-1}$.

CONDITION (C). Given c', $\exists C$, δ depending only on c', such that for any fixed x the number of indices k, for which $|a_{nk}t - x - 2\pi l| > C\zeta A_{5n}^{-1}$, for all $t \in [c' A_{5n}^{-\frac{1}{3}}, CA_{5n}^{-1}]$ and all $l = 0, \pm 1, \pm 2, \cdots$, is greater than δn. Condition (B) is verified for a number of important special cases in Albers, Bickel and van Zwet

(1976, page 120 and pages 133 – 134). In most cases A_{5n} is of order $n^{-\frac{3}{2}}$ and then the conditions and the bound in the following theorem can be given in terms of powers of n. If $b_n \nrightarrow 0$, and so $A_{5n} \nrightarrow 0$, the theorem is true but uninteresting.

THEOREM. *If Condition* (B) (*and so Condition* (C)) *holds for some* $\zeta \geqslant A_{5n}^{\frac{4}{3}} \log A_{5n}^{-1}$ *and if for some* $\frac{1}{2} > \varepsilon > 0$, $\varepsilon < p$, *then for any* $\eta > 0$,

$$|F_n(u) - G_n(u)| < BA_{5n}^{1-\eta},$$

where B *depends only on* η *and* ε.

PROOF. The joint characteristic function of $X_{n1}, \cdots, X_{nr}$ is

$$f_n^*(u) = (n!)^{-1}\Sigma^*\exp\{[(n-1)/n]^{\frac{1}{2}}[iu_1(a(i_{11}) + \cdots + a(i_{1s_1})) + \cdots + \\ + iu_r(a(i_{r1}) + \cdots + a(i_{rs_r}))]\},$$

where Σ^* denotes summation over all permutations $(i_{01}, \cdots, i_{0s_0}, i_{11}, \cdots, i_{rs_r})$ of $(1, \cdots, n)$ and $a_{ni} = a(i)$. Using an extension of the method of Erdös and Rényi (1959) (also given in Rényi (1970 page 460)), this can be written as

$$f_n^*(u) = [(2\pi)^r B_n(p)]^{-1}\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi}\Pi_{k=1}^n[\Sigma_j p_j e^{i(\eta_j + v_j a_{nk})}]e^{-i\Sigma_j\eta_j s_j}d\eta_1 \cdots d\eta_r,$$

where $\eta_0 = u_0 = 0$, $v_j = [(n-1)/n]^{\frac{1}{2}}u_j$ and

$$B_n(p) = n!\Pi_{j=0}^r(p_j^{s_j}/s_j!).$$

This can be put into a more convenient form by setting $n^{-\frac{1}{2}}\psi_0 = -\Sigma_j p_j\eta_j$, $n^{-\frac{1}{2}}\psi_j = \eta_j + n^{-\frac{1}{2}}\psi_0$, $t_0 = -\Sigma_j p_j u_j$, $t_j = u_j + t_0$ $j = 1, \cdots, r$. Then $\Sigma_j p_j\psi_j = \Sigma_j p_j t_j = 0$. Also set $\tau_j = [(n-1)/n]^{\frac{1}{2}}t_j$, $j = 0, \cdots, r$. Then, if $f_n(t) = f_n^*(u)$

$$f_n(t) = p_0^{-1}[(2\pi n^{\frac{1}{2}})^r B_n(p)]^{-1}\int\Pi_{k=1}^n[\Sigma_j p_j e^{i\xi_{jk}}]\,d\psi_1 \cdots d\psi_r \tag{2}$$

where $\xi_{jk} = n^{-\frac{1}{2}}\psi_j + \tau_j a_{nk}$ and the integral is over the region $-\pi n^{\frac{1}{2}} < \psi_j - \psi_0 < \pi n^{\frac{1}{2}}$, $j = 1, \cdots, r$, $\Sigma_j p_j\psi_j = 0$. The theorem will be proved by means of a number of lemmas.

LEMMA 2.1. *For some* $c > 0$ *and* $(\Sigma_j p_j t_j^2)^{\frac{1}{2}} = \|t\| < pcb_n^{-1}$,

$$|f_n(t) - g_n(t)| < A_{5n}P(\|t\|)e^{-\frac{1}{4}\|t\|^2},$$

where $P(\|t\|)$ *is a polynomial of degree* 15,

$$g_n(t) = e^{-\frac{1}{2}\Sigma_j p_j t_j^2}(1 + Q_1(t) + Q_2(t)), \tag{3}$$

$$Q_1(t) = \frac{i^3}{6}\Sigma_j p_j t_j^3\Sigma_k a_{nk}^3,$$

$$Q_2(t) = \frac{1}{24}[\Sigma_j p_j t_j^4 - 3(\Sigma_j p_j t_j^2)^2][\Sigma_k a_{nk}^4 - 3/n] - \frac{1}{4n}(\Sigma_j p_j t_j^2)^2 \\ - \frac{1}{72}(\Sigma_j p_j t_j^3)^2(\Sigma_k a_{nk}^3)^2.$$

PROOF. The integral in (2) can be considered in two parts. Let $f_n(t) = I_1 + I_2$, where I_1 and I_2 contain the integrals in (2) over the regions $\|\psi\| = (\Sigma_j p_j \psi_j^2)^{\frac{1}{2}} \leqslant 2cn^{\frac{1}{2}}$ and $\|\psi\| > 2c(n)^{\frac{1}{2}}$ but $|\psi_j - \psi_0| < \pi n^{\frac{1}{2}}$ for all $j = 1, \cdots, r$, respectively, where we will choose c later. In the sequel θ will denote a quantity with $|\theta| \leqslant 1$ and B will denote a bounded positive quantity depending only on ε and r. Consider I_1 first and notice that

$$(4) \qquad \Pi_{k=1}^n(\Sigma_j p_j e^{i\xi_{jk}}) = \exp[\Sigma_k \log(\Sigma_j p_j e^{i\xi_{jk}})]$$

$$= \exp\left[\Sigma_k \log\left(1 - \tfrac{1}{2}\Sigma_j p_j \xi_{jk}^2 + \frac{i^3}{3!}\Sigma_j p_j \xi_{jk}^3 + \frac{1}{4!}\Sigma_j p_j \xi_{jk}^4 + \frac{\theta}{5!}\Sigma_j p_j |\xi_{jk}|^5\right)\right],$$

since for all real x

$$|e^{ix} - \Sigma_{l=0}^h (ix)^l / l!| \leqslant |x|^{h+1}/(h+1)!.$$

Since $\|t\| < pcb_n^{-1}$, $|t_j| < p^{\frac{1}{2}}cb_n^{-1}$ for all $j = 0, 1, \cdots, r$, and when $\|\psi\| < 2cn^{\frac{1}{2}}$, $|\psi_j| < 2p^{-\frac{1}{2}}cn^{\frac{1}{2}}$ for all $j = 0, 1, \cdots, r$. Thus we can choose c so that $|\xi_{jk}| < \frac{1}{2}$, in which case the sum of the terms in ξ_{jk} in (4) has modulus less than $\frac{1}{4}$, so

$$(5) \qquad \Pi_{k=1}^n(\Sigma_j p_j e^{i\xi_{jk}}) = \exp\left[-\tfrac{1}{2}\Sigma_j p_j \Sigma_k \xi_{jk}^2 + \frac{i^3}{3!}\Sigma_j p_j \Sigma_k \xi_{jk}^3 + \frac{1}{4!}\Sigma_j p_j \Sigma_k \xi_{jk}^4 - \tfrac{1}{8}\Sigma_k(\Sigma_j p_j \xi_{jk}^2)^2 + \tfrac{1}{4}\theta \Sigma_j p_j \Sigma_k |\xi_{jk}|^5\right],$$

since for $|x| < \frac{1}{4}$, we have from Taylor's theorem

$$|\log(1 + x) - x + x^2/2| \leqslant |x|^3.$$

For any z,

$$(6) \qquad |e^z - \Sigma_{l=0}^h z^j/j!| \leqslant |z|^{h+1}e^{|z|}/(h+1)!,$$

so if z is the sum of the terms in the exponent of (5) excluding the first, we have from (6), taking $h = 2$,

$$(7) \quad \Pi_{k=1}^n(\Sigma_j p_j e^{i\xi_{jk}}) = \exp(-\tfrac{1}{2}\Sigma_j p_j \Sigma_k \xi_{jk}^2)\left[1 + \frac{i^3}{3!}\Sigma_j p_j \Sigma_k \xi_{jk}^3 + \frac{1}{4!}\Sigma_j p_j \Sigma_k \xi_{jk}^4 - \tfrac{1}{8}\Sigma_k(\Sigma_j p_j \xi_{jk}^2)^2 - \tfrac{1}{72}(\Sigma_j p_j \Sigma_k \xi_{jk}^3)^2 + R_1 e^{R_2}\right],$$

where for $|\xi_{jk}| < \frac{1}{2}$,

$$(8) \qquad |R_2| = |z| \leqslant \frac{1}{3!}\Sigma_j p_j \Sigma_k |\xi_{jk}|^3 + \frac{1}{4!}\Sigma_j p_j \Sigma_k |\xi_{jk}|^4 + \tfrac{1}{8}\Sigma_k(\Sigma_j p_j \xi_{jk}^2)^2 + \tfrac{1}{4}\Sigma_j p_j \Sigma_k |\xi_{jk}|^5$$

$$\leqslant \tfrac{1}{6}\Sigma_j p_j \Sigma_k \xi_{jk}^2$$

and an estimate of R_1 is obtained by collecting terms of z^2 and $|z|^3$ not written out in (7); noting that, since $\Sigma_k|a_{nk}|^h > n^{-\frac{1}{2}(h-2)}$,

$$\Sigma_k|\xi_{jk}|^h \leqslant (|\psi_j| + |\tau_j|)^h \Sigma_k|a_{nk}|^h$$

and from the Holder inequality, for $2 < h \leqslant 5$,

$$A_{hn} = \Sigma_k|a_{nk}|^h \leqslant (\Sigma_k a_{nk}^2)^{(5-h)/3}(\Sigma_k|a_{nk}|^5)^{(h-2)/3} = A_{5n}^{(h-2)/3},$$

then considering each term, we have

$$|R_1| \leqslant A_{5n}P_1(|\psi_j| + |\tau_j|),$$

where P_1 denotes a polynomial in $r + 1$ variables of degree 15, with coefficients depending only on ε or r.

To obtain an estimate for I_1 notice first that, since $\Sigma_j p_j \Sigma_k \xi_{jk}^2 = \Sigma_j p_j \psi_j^2 + \Sigma_j p_j \tau_j^2$,

$$(9)\quad |\int \exp(-\tfrac{1}{2}\Sigma_j p_j \Sigma_k \xi_{jk}^2) R_1 e^{R_2} d\psi_1 \cdots d\psi_r| \leqslant A_{5n}|\int \exp(-\tfrac{1}{3}\Sigma_j p_j \psi_j^2 - \tfrac{1}{3}\Sigma_j p_j \tau_j^2) \times P_1(|\psi_j| + |\tau_j|)\, d\psi_1 \cdots d\psi_r| \leqslant A_{5n}P(\|t\|)e^{-\frac{1}{3}\Sigma_j p_j \tau_j^2},$$

where here and in the sequel, $P(x)$ denotes a polynomial of degree at most 15 with coefficients depending only on ε and r, which may be different at each occurrence and where the integral is over the range $\|\psi\| < 2cn^{\frac{1}{2}}$. If the other terms in (7) are integrated over the range, $-\infty < |\psi_j| < \infty$, $j = 1, \cdots, r$, the integral will be of the form

$$(10)\qquad p_0(2\pi)^{\frac{1}{2}r}[p_0 p_1 \cdots p_r]^{-\frac{1}{2}} \exp(-\tfrac{1}{2}\Sigma_j p_j \tau_j^2)[1 + Q_1^*(\tau) + Q_2^*(\tau)]$$

where $Q_1^*(\tau)$, $Q_2^*(\tau)$ are polynomials to be obtained explicitly below. The difference between this integral and the integral of the same terms over the range $\|\psi\| < 2cn^{\frac{1}{2}}$, is less than

$$B \exp(-c^2 n - \tfrac{1}{2}\Sigma_j p_j \tau_j^2) \leqslant BA_{5n} \exp(-\tfrac{1}{2}\Sigma_j p_j \tau_j^2).$$

As a particular case of (9) and (10), we have, taking $t = 0$,

$$(11)\quad (2\pi)^r B_n(p) = (2\pi)^{\frac{1}{2}r}[p_0 \cdots p_r]^{-\frac{1}{2}}[1 + Q_1^*(0) + Q_2^*(0)] + 0(n^{-\frac{3}{2}}),$$

where the last term is obtained by considering R_1 and R_2 when $\tau_j = 0$, $j = 1, \cdots, r$. Further

$$\exp(-\tfrac{1}{2}\Sigma_j p_j \tau_j^2) = \exp(-\tfrac{1}{2}\Sigma_j p_j t_j^2)\left(1 + \frac{1}{2n}\Sigma_j p_j t_j^2 + \frac{\theta}{n^2}(\Sigma_j p_j t_j^2)^2 \exp\left(-\frac{1}{2n}\Sigma_j p_j t_j^2\right)\right),$$

$$Q_1^*(\tau) = Q_1^*(t) + 0(A_{5n})P(\|t\|),$$

and

$$Q_2^*(\tau) = Q_2^*(t) + 0(A_{5n})P(\|t\|).$$

So using these results together with (9), (10) and (11), we have

$$(12)\quad |I_1 - \exp(-\tfrac{1}{2}\Sigma_j p_j t_j^2)[1 + Q_1(t) + Q_2(t)]| < A_{5n}P(\|t\|)\exp -\left(((2n-3)/6n)\Sigma_j p_j t_j^2\right) < A_{5n}P(\|t\|)e(-\tfrac{1}{4}\Sigma_j p_j t_j^2)$$

for $n \geqslant 6$, where $Q_1(t) = Q_1^*(t) - Q_1^*(0)$, $Q_2(t) = Q_2^*(t) - Q_2^*(0) + (1/2n)\Sigma_j p_j t_j^2$. For $2 \leqslant n < 6$, the constant may be adjusted to make this result hold.

Now we will obtain $Q_1(t)$ and $Q_2(t)$ explicitly. The terms in (7) can be expanded to give

$$\Sigma_j p_j \Sigma_k \xi_{jk}^3 = n^{-\frac{1}{2}}\Sigma_j p_j \psi_j^3 + 3n^{-\frac{1}{2}}\Sigma_j p_j \psi_j \tau_j^2 + \Sigma_j p_j \tau_j^3 \Sigma_k a_{nk}^3,$$

$$\Sigma_j p_j \Sigma_k \xi_{jk}^4 = n^{-1}\Sigma_j p_j \psi_j^4 + 6n^{-1}\Sigma_j p_j \psi_j^2 \tau_j^2 + 4n^{-\frac{1}{2}}\Sigma_j p_j \psi_j \tau_j^3 \Sigma_k a_{nk}^3 + \Sigma_j p_j \tau_j^4 \Sigma_k a_{nk}^4,$$

$$\Sigma_k(\Sigma_j p_j \xi_{jk}^2)^2 = n^{-1}(\Sigma_j p_j \psi_j^2)^2 + 4n^{-1}(\Sigma_j p_j \psi_j \tau_j)^2 + (\Sigma_j p_j \tau_j^2)^2 \Sigma_k a_{nk}^4 + 2n^{-1}(\Sigma_j p_j \psi_j^2)(\Sigma_j p_j \tau_j^2) + 4n^{-\frac{1}{2}}(\Sigma_j p_j \psi_j \tau_j)(\Sigma_j p_j \tau_j^2)\Sigma_k a_{nk}^3,$$

$$(\Sigma_j p_j \Sigma_k \xi_{jk}^3)^2 = n^{-1}(\Sigma_j p_j \psi_j^3)^2 + 9n^{-1}(\Sigma_j p_j \psi_j \tau_j^2)^2 + (\Sigma_j p_j \tau_j^3)^2(\Sigma_k a_{nk}^3)^2 + 6n^{-1}(\Sigma_j p_j \psi_j^3)(\Sigma_j p_j \psi_j \tau_j^2) + 2n^{-\frac{1}{2}}(\Sigma_j p_j \psi_j^3)(\Sigma_j p_j \tau_j^3)\Sigma_k a_{nk}^3 + 6n^{-\frac{1}{2}}(\Sigma_j p_j \psi_j \tau_j^2)(\Sigma_j p_j \tau_j^3)\Sigma_k a_{nk}^3.$$

The integrals involving these terms may now be written down using Lemma 3.2. After some reduction we obtain

$$Q_1(t) = \frac{i^3}{6}\Sigma_j p_j t_j^3 \Sigma_k a_{nk}^3$$

$$\begin{aligned}Q_2(t) = {} & \tfrac{1}{24}\left[6n^{-1}\Sigma_j(1-p_j)t_j^2 + \Sigma_j p_j t_j^4 \Sigma_k a_{nk}^4\right] \\ & -\tfrac{1}{8}\left[4n^{-1}\Sigma_j p_j(1-p_j)t_j^2 - 4n^{-1}\Sigma_{j\neq j'} p_j p_{j'} t_j t_{j'} + (\Sigma_j p_j t_j^2)^2 \Sigma_k a_{nk}^4 + 2n^{-1} r \Sigma_j p_j t_j^2\right] \\ & -\tfrac{1}{72}\left[9n^{-1}\Sigma_j p_j(1-p_j)t_j^4 - 9n^{-1}\Sigma_{j\neq j'} p_j p_{j'} t_j^2 t_{j'}^2 + (\Sigma_j p_j t_j^3)(\Sigma_k a_{nk}^3)^2 \right. \\ & \left. + 18n^{-1}\Sigma_j(1-p_j)^2 t_j^2 - 18n^{-1}\Sigma_{j\neq j'} p_j(1-p_{j'})t_j^2\right] + \frac{1}{2n}\Sigma_j p_j t_j^2 \\ = {} & \tfrac{1}{24}\left[\Sigma_j p_j t_j^4 - 3(\Sigma_j p_j t_j^2)^2\right]\left[\Sigma_k a_{nk}^4 - 3/n\right] - \frac{1}{4n}(\Sigma_j p_j t_j^2)^2 \\ & -\tfrac{1}{72}(\Sigma_j p_j t_j^3)^2(\Sigma_k a_{nk}^3)^2.\end{aligned}$$

It remains only to show that for $\|t\| < pcb_n^{-1}$,

$$(13) \qquad |I_2| < BA_{5n} \exp\left(-\tfrac{1}{4}\Sigma_j p_j t_j^2\right).$$

If $\|\psi\| > 2cn^{\frac{1}{2}}$, then for some $j = 0, 1, \cdots, r$, $|\psi_j| > 2cn^{\frac{1}{2}}$, and if $\|t\| < pcb_n^{-1}$, then $|t_j| < p^{\frac{1}{2}}cb_n^{-1}$, for all $j = 0, 1, \cdots, r$. Thus $\max_{0 \leqslant j \leqslant r}|\xi_{jk}| > c$, for $k = 1, \cdots, n$. Now

$$|\Sigma_j p_j e^{i\xi_{jk}}|^2 = |\Sigma_j \Sigma_{j'} p_j p_{j'} \cos(\xi_{jk} - \xi_{j'k})|.$$

Since $\Sigma_j p_j \xi_{jk} = 0$ and $\max_{0 \leqslant j \leqslant r}|\xi_{jk}| > c$, $|\xi_{jk} - \xi_{j'k}| > c$ for some pair $j, j' = 0, 1, \cdots, r$. Also $|\psi_j - \psi_0| < \pi n^{\frac{1}{2}}$, for all $j = 1, \cdots, r$, so $|\xi_{jk} - \xi_{j'k}| < 2\pi - c$, unless either $c < |\xi_{jk} - \xi_{0k}| < 2\pi - c$ or $c < |\xi_{j'k} - \xi_{0k}| < 2\pi - c$. So for some pair $j, j' = 0, 1, \cdots, r$, $c < |\xi_{jk} - \xi_{j'k}| < 2\pi - c$. Thus

$$|\Sigma_j p_j e^{i\xi_{jk}}|^2 < 1 - 2p^2(1 - \cos(c)).$$

Now

$$1 - \cos(c) \geqslant \frac{c^2}{2} - \frac{c^4}{24} \geqslant \frac{c^2}{3}$$

for $c < 2$. Thus for $\|t\| < pcb_n^{-1} < pcn^{\frac{1}{2}}$,

$$(14) \qquad |\Pi_{k=1}^n(\Sigma_j p_j e^{i\xi_{jk}})| \leqslant e^{-p^2c^2n/3} \leqslant \exp\left(-\tfrac{1}{4}\Sigma_j p_j t_j^2 - \tfrac{1}{12}p^2c^2n\right).$$

Now (13) follows immediately and the lemma follows from (12) and (13).

The next lemma simply gives the formal inversion of $g_n(t)$. It follows as a direct consequence of Lemma 3.3.

LEMMA 2.2. *If $G_n(u)$ and $g_n(t)$ are defined in* (1) *and* (3), *respectively, then*

$$G_n(u) = p_0^{-1}(2\pi)^{-r}\int\left[\int e^{-i\Sigma_j t_j x_j} g_n(t)\, dt_1 \cdots dt_r\right] dx_1 \cdots dx_r,$$

where the first integral is over the region $\Sigma_j x_j^2/p_j \leqslant u$, $\Sigma_j x_j = 0$ and the second integral is over the region $-\infty < t_j < \infty, j = 1, \cdots, r$, $\Sigma_j p_j t_j = 0$.

The next lemma gives an estimate of the error of the approximation for large values of u.

LEMMA 2.3. *If $u > 3(r + 1)(\log A_{5n}^{-1})$, then*

$$|F_n(u) - G_n(u)| < BA_{5n}.$$

PROOF. From (1) it is clear that for $u > 3(r + 1)(\log A_{5n}^{-1})$

$$|1 - G_n(u)| \leqslant B(\log A_{5n}^{-1})^{\frac{1}{2}r+2} \exp\left(-\tfrac{3}{2}(r + 1)(\log A_{5n}^{-1})\right) \leqslant BA_{5n}.$$

Also

$$(15) \qquad 1 - F_n(u) = P\left(\Sigma_j X_{nj}^2/p_j > u\right) \leqslant \Sigma_j P\left(|X_{nj}| > \left[p_j u/(r + 1)\right]^{\frac{1}{2}}\right).$$

From Robinson (1978) we have

$$(16) \qquad |P\left(X_{nj} \leqslant v(p_j q_j)^{\frac{1}{2}}\right) - G_{ns}(v)| \leqslant BA_{5n}$$

where

$$G_{ns}(v) = \Phi(v) - a_2H_2(v)\phi(v) - a_3H_3(v)\phi(v) - a_5H_5(v)\phi(v)$$

for bounded coefficients a_2, a_3, a_5 depending only on p_j, where

$$\phi(v) = \Phi'(v) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}v^2}, \qquad H_i(v)\phi(v) = (-1)^i \frac{d^i}{dv^i}\phi(v).$$

If $v = [u/(r+1)]^{\frac{1}{2}}$ and $u > 3(r+1)(\log A_{5n}^{-1})$, then $\frac{1}{2}v^2 > \frac{3}{2}(\log A_{5n}^{-1})$ and so

$$|1 - G_{ns}(v)| < BA_{5n}. \tag{17}$$

Now (15), (16) and (17) imply that

$$|1 - F_n(u)| < BA_{5n},$$

and the lemma follows immediately.

The following smoothing lemma is of the same type as that used in the one dimensional case. Its proof is omitted since it is almost identical to the proof of Von Bahr (1966) for rectangles. It may also be obtained as a consequence of Lemma 8 of Bhattacharya (1970).

LEMMA 2.4. *Let x and y denote vectors $(x_0, x_1, \cdots, x_r)$ and $(y_0, y_1, \cdots, y_r)$ such that $\Sigma_j x_j = \Sigma_j y_j = 0$. Let $Q(x)$ be a density function, let $H(x) = F(x) - G(x)$, where $F(x)$ is a distribution function and $G(x)$ is of bounded variation and let*

$$H_T(x) = \int_{R^r} Q(y)H(x + y/T)\, dy_1 \cdots dy_r.$$

Let $S(y, u) = \{x : \Sigma_j(x_j - y_j)^2/p_j \leqslant u\}$, $0 < p_j < 1$, $j = 0, 1, \cdots, r$, and if L is any function of bounded variation, write

$$L(S(y, u)) = \int_{S(y,u)} dL.$$

If $\delta = \sup_y H(S(y, u))$ and $\delta_T = \sup_y H_T(S(y, u))$, then

$$\delta \leqslant 3\delta_T + 3\alpha(G, a/T),$$

where

$$\alpha(G, a/T) = \sup_y \int d|G|$$

where the integral is over the set $\{x : u - a/T \leqslant \Sigma_j(x_j - y_j)^2/p_j \leqslant u + a/T\}$ and a is chosen so that

$$\int_{S(0,a)} Q(y)\, dy_1 \cdots dy_r = \tfrac{2}{3}.$$

We are now able to complete the proof of the theorem. Let F_n^r and G_n^r be the functions in R^r whose characteristic functions are f_n and g_n. Let

$$H_T(x) = \int Q(y)[F_n^r(x + y/T) - G_n^r(x + y/T)]\, dy_1 \cdots dy_r.$$

Then $F_n(u) = F_n^r(S(0, u))$, $G_n(u) = G_n^r(S(0, u))$. If $q(t)$ is the characteristic function corresponding to the density $Q(x)$, then since $\Sigma_j p_j t_j = 0$,

$$H_T(x) = p_0^{-1}(2\pi)^{-r}\int_{R^r} e^{-i\Sigma_j t_j x_j}(f_n(t) - g_n(t))q(-t/T)\, dt_1 \cdots dt_r.$$

The integral of this over the region $S(y, u)$ is

$$P_0^{-1}(2\pi)^{-\frac{1}{2}r}\int\left[\frac{u}{\|t\|}\right]^{\frac{1}{2}r} J_{r/2}(u\|t\|)e^{-i\Sigma_j t_j y_j}(f_n(t) - g_n(t))q(-t/T)\, dt_1 \cdots dt_r,$$

where $J_{r/2}(z)$ is the Bessel function of order $r/2$. Write this integral as the sum of the integrals I_1^*, I_2^*, I_3^*, I_4^* and I_5^* which are integrals over the regions $\|t\| \leqslant (\log A_{5n}^{-1})^{-1}$, $(\log A_{5n}^{-1})^{-1} \leqslant \|t\| \leqslant pcb_n^{-1}$, $pcb_n^{-1} \leqslant \|t\| \leqslant pcA_{5n}^{-\frac{1}{3}}$, $pcA_{5n}^{-\frac{1}{3}} \leqslant \|t\| \leqslant \frac{1}{2}pCA_{5n}^{-1}$ and $\|t\| \geqslant \frac{1}{2}pCA_{5n}^{-1}$, respectively, where C is chosen later.

Using the inequalities (see, e.g., Esseen (1945))

$$J_{r/2}(z)z^{r/2} < B \text{ and } |q(t)| \leqslant 1,$$

and Lemma 2.1, we have, taking $\|t\| = \rho$,

$$(18) \qquad |I_1^*| \leqslant u^r A_{5n}\int_{\rho<(\log A_{5n}^{-1})^{-1}}\rho^{r-1}P(\rho)e^{-\frac{1}{4}\rho^2}\, d\rho$$
$$< Bu^4(\log A_{5n}^{-1})^{-r}A_{5n} < BA_{5n},$$

for $u \leqslant 3(r + 1)(\log A_{5n}^{-1})$.

From the inequalities

$$J_{r/2}(z) < B \text{ for } z \geqslant 0 \text{ and } |q(t)| \leqslant 1,$$

and Lemma 2.1 again, we have for $u \leqslant 3(r + 1)(\log A_{5n}^{-1})$,

$$(19) \qquad |I_2^*| \leqslant u^{\frac{1}{2}r}A_{5n}\int_0^\infty \rho^{\frac{1}{2}r-1}e^{-\frac{1}{4}\rho^2}P(\rho)\, d\rho$$
$$\leqslant B(\log A_{5n}^{-1})^{\frac{1}{2}r}A_{5n}$$
$$< BA_{5n}^{1-\eta}.$$

For $pcb_n^{-1} \leqslant \|t\| \leqslant pcA_{5n}^{-\frac{1}{3}}$, it is clear that

$$(20) \qquad |g_n(t)| < BA_{5n}e^{-\frac{1}{4}t^2}$$

since $b_n^{-1} \geqslant A_{5n}^{-\frac{1}{3}}$. Further, if $\|\psi\| < 2cn^{\frac{1}{2}}$ then $|\psi_j| < 2p^{-\frac{1}{2}}cn^{\frac{1}{2}}$ for all $j = 0, 1, \cdots, r$, so for $k \in K_d = \{k : |a_{nk}| < dA_{5n}^{\frac{1}{3}}\}$, $|\xi_{jk}| < 2p^{-\frac{1}{2}}c + p^{\frac{1}{2}}cd < \frac{1}{2}$, by choice of c, since $|t_j| < p^{\frac{1}{2}}cA_{5n}^{-\frac{1}{3}}$ when $\|t\| < pcA_{5n}^{-\frac{1}{3}}$. Then, using the trivial inequality $|\Sigma_j p_j \exp i\xi_{jk}| < 1$ for $k \notin K_d$ and the estimate from (5) for $k \in K_d$, we have, writing K for K_d,

$$|\Pi_{k=1}^n(\Sigma_j p_j e^{i\xi_{jk}})| \leqslant \exp\left(-\tfrac{1}{2}\Sigma_j p_j \Sigma_{k\in K}\xi_{jk}^2 + |R_2|\right)$$

where R_2 is given in (8). Thus

$$(21) \qquad |\Pi_{k=1}^n(\Sigma_j p_j e^{i\xi_{jk}})| \leqslant \exp\Big[-\tfrac{1}{2}mn^{-1}\Sigma_j p_j\psi_j^2 - n^{-\frac{1}{2}}\Sigma_j p_j\psi_j t_j\Sigma_{k\in K}a_{nk}$$
$$-\tfrac{1}{2}\Sigma_j p_j t_j^2\Sigma_{k\in K}a_{nk}^2 + \tfrac{1}{6}\Sigma_j p_j\psi_j^2 + \tfrac{1}{6}\Sigma_j p_j t_j^2\Big],$$

where m is the number of elements in K_d. Now

$$1 = \Sigma_k a_{nk}^2 \geqslant \Sigma_{k\notin K}a_{nk}^2 \geqslant (n - m)d^2A_{5n}^{\frac{2}{3}} \geqslant (n - m)d^2n^{-1}.$$

So $m \geqslant n(d^2 - 1)/d^2$. Also

$$A_{5n} \geqslant \Sigma_{k \notin K} |a_{nk}|^5 \geqslant d^3 A_{5n} \Sigma_{k \notin K} a_{nk}^2.$$

Thus

$$\Sigma_{k \notin K} a_{nk}^2 \leqslant d^{-3} \quad \text{and} \quad \Sigma_{k \in K} a_{nk}^2 \geqslant 1 - d^{-3}.$$

Further

$$|\Sigma_{k \in K} a_{nk}| = |\Sigma_{k \notin K} a_{nk}| \leqslant d^{-1} A_{5n}^{-\frac{1}{3}} \Sigma_{k \notin K} a_{nk}^2 \leqslant d^{-4} A_{5n}^{-\frac{1}{3}} \leqslant d^{-4} n^{\frac{1}{2}}.$$

Using these results and the inequality $|\psi_j t_j| \leqslant \frac{1}{2}\psi_j^2 + \frac{1}{2}t_j^2$ in (21), and taking $d = 3$, we have

$$|\Pi_{k=1}^n (\Sigma_j p_j e^{i\xi_{jk}})| \leqslant \exp\left(-\tfrac{1}{4}\Sigma_j p_j \psi_j^2 - \tfrac{1}{4}\Sigma_j p_j t_j^2\right) < BA_{5n} \exp\left(-\tfrac{1}{4}\Sigma_j p_j \psi_j^2 - \tfrac{1}{8}\Sigma_j p_j t_j^2\right),$$

since $b_n^{-1} > A_{5n}^{-\frac{1}{5}}$ and $\|t\| > pcb_n^{-1}$. Also if $\|\psi\| > 2cn^{\frac{1}{2}}$ then for some $j = 0, 1, \cdots, r$, $|\psi_j| > 2cn^{\frac{1}{2}}$. Further, if $pcb_n^{-1} < \|t\| < pcA_{5n}^{-\frac{1}{3}}$ then $|t_j| < p^{\frac{1}{2}}cA_{5n}^{-\frac{1}{3}}$, for all $j = 0, 1, \cdots, r$. So, if $k \in K_{\frac{3}{2}}$, $\max_{0 \leqslant j \leqslant r} |\xi_{jk}| > \frac{1}{2}c$ and so as in (14)

$$|\Sigma_j p_j e^{i\xi_{jk}}|^2 \leqslant 1 - 2p^2\left(1 - \cos\left(\tfrac{1}{2}c\right)\right).$$

Using this estimate for $k \in K_{\frac{3}{2}}$ and the estimate 1 for $k \notin K_{\frac{3}{2}}$, we have

$$|\Pi_{k=1}^n (\Sigma_j p_j e^{i\xi_{jk}})| < e^{-p^2c^2n/24}$$

since the number of terms in $K_{\frac{3}{2}}$ is greater than $\frac{1}{2}n$. So since $|\psi_j - \psi_0| < \pi n^{\frac{1}{2}}$ and $pcb_n^{-1} < \|t\| < pcA_{5n}^{-\frac{1}{3}}$,

(22) $$|I_3^*| < BA_{5n}.$$

If $pcA_{5n}^{-\frac{1}{3}} < \|t\| < \frac{1}{2}pCA_{5n}^{-1}$, then $|t_j| > pcA_{5n}^{-\frac{1}{3}}$ for some $j = 0, 1, \cdots, r$ and so for some pair j, j', $|t_j - t_{j'}| > pcA_{5n}^{-\frac{1}{3}}$ since in addition $\Sigma_j p_j t_j = 0$. Also $|t_j - t_{j'}| < (p_j^{-1} + p_{j'}^{-1})^{\frac{1}{2}} \|t\| < CA_{5n}^{-1}$ for all pairs j, j'. Thus if Condition (B), and so Condition (C), holds with $\zeta \geqslant A_{5n}^{\frac{4}{3}} \log A_{5n}^{-1}$,

$$|\Sigma_j p_j e^{i\xi_{jk}}|^2 \leqslant 1 - 2p^2(1 - \cos(\zeta CA_{5n}^{-1}))$$

for at least δn indices k, where δ and C are chosen to satisfy Condition (C) for $c' = pc$. Bounding the other terms by 1, we have, as in the proof of (14),

$$|\Pi_{k=1}^n (\Sigma_j p_j e^{i\xi_{jk}})| < \exp\left(-\tfrac{1}{3}p^2 \delta n \zeta^2 C^2 A_{5n}^{-2}\right)$$
$$< \exp\left(-\tfrac{1}{3}p^2 C^2 \delta (\log A_{5n}^{-1})^2\right).$$

So

(23) $$|I_4^*| < BA_{5n}.$$

We can choose Q so that $q(t) = 0$ for $\|t\| > 1$ (see for example, Von Bahr (1966)). Then

(24) $$I_5^* = 0,$$

if we take $T = \frac{1}{2}pCA_{5n}^{-1}$.

Combining (18), (19), (22), (23) and (24), we have for $u < 3(r+1)\log A_{5n}^{-1}$,

$$\delta_T = \sup H_T(S(y, u)) < BA_{5n}^{1-\eta}.$$

Further, since $G_n^r(x)$ is the Fourier-Stieltjes transform of $g_n(t)$, it is readily seen that for $T = \frac{1}{2}pCA_{5n}^{-1}$,

$$\alpha(G_n^r, a/T) < BA_{5n}.$$

For $u < 3(r+1)(\log A_{5n}^{-1})$, the theorem follows from Lemmas 2.2 and 2.4 while it follows for $u \geqslant 3(r+1)(\log A_{5n}^{-1})$ from Lemma 2.3.

REMARK. Condition (C) is weaker than the corresponding condition of Robinson (1978). The method of proof used to obtain a bound for I_3^* could be used there to obtain a weakening of the condition to that given here.

3. Some results for multivariate normal distributions. The purpose of this section is to obtain some formal results on integrals of densities of multivariate normal type which occur in Section 2.

LEMMA 3.1. *If* $Y_1, \cdots, Y_r$ *are* NID(0, 1) *and* I_A *is the indicator function of the set* A, *then*

$$EY_1^\nu I_{\{\Sigma_i Y_i^2 < a\}} = \frac{2^{\frac{1}{2}\nu}\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{1}{2}\right)} P\left(\chi^2_{r+\nu} < a\right), \qquad \nu \text{ even},$$

$$= 0, \qquad \nu \text{ odd},$$

and

$$EY_1^{\nu_1}Y_2^{\nu_2} I_{\{\Sigma_i Y_i^2 < a\}} = \frac{2^{\frac{1}{2}(\nu_1+\nu_2)}\Gamma\left(\frac{1}{2}(\nu_1+1)\right)\Gamma\left(\frac{1}{2}(\nu_2+1)\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)} P\left(\chi^2_{r+\nu_1+\nu_2} < a\right),$$

ν_1 *and* ν_2 *even*,

$$= 0, \qquad \nu_1 \text{ or } \nu_2 \text{ odd}.$$

PROOF.

$$EY_1^\nu I_{\{\Sigma_i Y_i^2 < a\}} = EU^{\nu/2} I_{\{U+V<a\}}$$

where U and V are chi-squared variates with 1 and $r-1$ degrees of freedom, respectively. Upon integrating this we have

$$EY_1^\nu I_{\{\Sigma_i Y_i^2 < a\}} = \frac{2^{\frac{1}{2}\nu}\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{1}{2}\right)} P\left(\chi^2_{r+\nu} < a\right)$$

for ν even and 0 for ν odd. The second result follows in the same way.

LEMMA 3.2. *Let* X_j *be* NID$(0, p_j)$, $j = 0, 1, \cdots, r$. *Given* $\Sigma_j X_j = 0$, *the conditional frequency function of* $X_1, \cdots, X_r$ *is*

$$\left[(2\pi)^r p_0 p_1 \cdots p_r\right]^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\Sigma_j x_j^2/p_j\right),$$

where $\Sigma_j x_j = 0$. *If* E_0 *denotes conditional expectation given* $\Sigma_j X_j = 0$, *then if* I *is a random variable taking the value* 1 *when* $\Sigma_j X_j^2/p_j < a$ *and* 0 *otherwise and* $q_j = 1 - p_j, j = 0, 1, \cdots, r$,

$$\begin{aligned}
E_0 X_j^2 I &= p_j q_j P(\chi_{r+2}^2 < a),\\
E_0 X_j^4 I &= 3p_j^2 q_j^2 P(\chi_{r+4}^2 < a),\\
E_0 X_j X_{j'} I &= -p_j p_{j'} P(\chi_{r+2}^2 < a),\\
E_0 X_j^3 X_{j'} I &= -3p_j^2 q_j p_{j'} P(\chi_{r+4}^2 < a),\\
E_0 X_j^2 X_{j'}^2 I &= p_j p_{j'}(1 - p_j - p_{j'} + 3p_j p_{j'}) P(\chi_{r+4}^2 < a),\\
E_0 X_j^3 X_{j'}^3 I &= -3p_j^2 p_{j'}^2 (3 - 3p_j - 3p_{j'} + 5p_j p_{j'}) P(\chi_{r+6}^2 < a),\\
E_0 X_j^h X_{j'}^k I &= 0, \text{ if } h + k \text{ is odd},
\end{aligned}$$

$$\text{for } j, j' = 0, 1, \cdots, r, j \neq j'.$$

PROOF. Consider a transformation to Y_0, Y_1, Y_2, where $Y_0 = \Sigma_j X_j$, $Y_1 = X_1 - p_1 \Sigma_j X_j$, $Y_2 = X_1 - p_1 \Sigma_j X_j + (X_2 - p_2 \Sigma_j X_j) q_1/p_2$. Then Y_0, Y_1, Y_2 are independent with variances $1, p_1 q_1$ and $q_1(1 - p_1 - p_2)/p_2$, respectively. Given $\Sigma_j X_j = 0$, $Y_0 = 0$ and $X_1 = Y_1$, $X_2 = (Y_2 - Y_1)p_2/q_1$. The results follow by direct application of Lemma 3.2.

LEMMA 3.3. *For a function* $h(t)$ *of* $(t_0, \cdots, t_r)$, *let*

$$A(h(t)) = p_0^{-1}(2\pi)^{-r} \int \left[\int \exp\left(-i\Sigma_j t_j x_j - \tfrac{1}{2}\Sigma_j p_j t_j^2\right) h(t)\, dt_1 \cdots dt_r \right] dx_1 \cdots dx_r,$$

where the first integral is over the region $\Sigma_j x_j^2/p_j \leqslant u$, $\Sigma_j x_j = 0$ *and the second integral is over the region* $-\infty < t_j < \infty, j = 1, \cdots, r$, $\Sigma_j p_j t_j = 0$. *Then*

$$\begin{aligned}
A(1) &= P(\chi_r^2 \leqslant u),\\
A\left(\Sigma_j p_j t_j^2\right) &= r\gamma_r(u),\\
A\left(\Sigma_j p_j t_j^3\right) &= 0,\\
A\left(\Sigma_j p_j t_j^4\right) &= 3\left(\Sigma_j p_j^{-1} - 2r - 1\right)\gamma_r(u)\left[1 - u/(r+2)\right],\\
A\left[\left(\Sigma_j p_j t_j^2\right)^2\right] &= r(r+2)\gamma_r(u)\left[1 - u/(r+2)\right],\\
A\left[\left(\Sigma_j p_j t_j^3\right)^2\right] &= \left(15\Sigma_j p_j^{-1} - 9r^2 - 36r - 15\right)\gamma_r(u)\\
&\quad \times \left[u^2/(r+2)(r+4) - 2u/(r+2) + 1\right],
\end{aligned}$$

where $\gamma_r(u) = u^{\frac{1}{2}r} e^{-\frac{1}{2}u} / 2^{\frac{1}{2}r} \Gamma[\frac{1}{2}(r+2)]$.

PROOF. The results follow by a direct application of Lemma 3.2. For example,

$$\begin{aligned}
A\left(\Sigma_j p_j t_j^4\right) &= \left[(2\pi)^r p_0 p_1 \cdots p_r\right]^{-\frac{1}{2}} \int \exp\left(-\tfrac{1}{2}\Sigma_j x_j^2/p_j\right)\left[3\Sigma_j\left(q_j^2/p_j\right) - 6\Sigma_j\left(q_j x_j^2/p_j^2\right)\right.\\
&\quad \left. + \Sigma_j\left(x_j^4/p_j^3\right)\right] dx_1 \cdots dx_r,
\end{aligned}$$

from Lemma 3.2, applied to the random variables $Y_j = X_j/p_j, j = 0, 1, \cdots, r$, conditional on $\Sigma_j p_j Y_j = 0$, and with a infinite, where the integral is over the region $\Sigma_j x_j^2/p_j \leqslant u$ and $\Sigma_j x_j = 0$. So from Lemma 3.2 again

$$\begin{aligned} A(\Sigma_j p_j t_j^4) &= \Sigma_j(q_j^2/p_j)[3P(\chi_r^2 \leqslant u) - 6P(\chi_{r+2}^2 \leqslant u) + 3P(\chi_{r+4}^2 \leqslant u)] \\ &= 3(\Sigma_j p_j^{-1} - 2r - 1)\gamma_r(u)[1 - u/(r+2)]. \end{aligned}$$

The other results are obtained in a similar manner.

4. Numerical comparisons for Kruskal-Wallis test. It seems worthwhile to consider the accuracy of the approximation for small sample values in a particular case. Tables of the exact probabilities for the Kruskal-Wallis test for $r = 2$ are now available in a number of text books, so below we give a table of values of u throughout the range and give P_1, the exact probability that the statistic is greater than or equal to u, P_2, the chi-squared approximation, and P_3, the approximation given in this paper, for four choices of (s_0, s_1, s_2).

Inspection of the table reveals quite good approximation in the cases (3, 4, 5), (4, 5, 5) and (5, 5, 5), especially in the tails of the distribution for probabilities in the range .01 to .2. As might be expected the approximation is somewhat inadequate for (2, 3, 4).

TABLE 1

	(2, 3, 4)				(3, 4, 5)		
u	P_1	P_2	P_3	u	P_1	P_2	P_3
.111	.944	.946	.954	.118	.953	.943	.949
1.000	.660	.606	.644	1.004	.654	.605	.633
3.011	.256	.222	.236	3.010	.238	.222	.232
4.000	.149	.135	.135	4.015	.140	.134	.134
5.078	.057	.079	.070	5.041	.072	.080	.074
6.000	.024	.050	.038	6.026	.038	.049	.040
7.000	.005	.030	.017	7.004	.015	.031	.020
				8.030	.005	.018	.009

	(4, 5, 5)				(5, 5, 5)		
u	P_1	P_2	P_3	u	P_1	P_2	P_3
.111	.958	.946	.951	.140	.954	.932	.938
1.000	.650	.606	.629	1.040	.620	.594	.616
3.023	.229	.221	.228	3.020	.231	.221	.228
4.043	.133	.133	.132	4.020	.132	.134	.134
5.023	.075	.081	.076	5.040	.075	.081	.076
6.031	.040	.049	.041	6.000	.044	.049	.043
7.000	.019	.030	.022	7.020	.020	.030	.022
8.006	.009	.018	.011	8.000	.009	.018	.011

REFERENCES

[1] ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one sample problem. *Ann. Statist.* **4** 108–156.

[2] VON BAHR, B. (1966). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88.

[3] BHATTACHARYA, R. N. (1970). Rates of weak convergence for the multi-dimensional central limit theorem. *Teor. Verojatnost. i Primenen.* **15** 69–85.

[4] BICKEL, P. J. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution free tests in the two sample problem. *Ann. Statist.* **6** 937–1004.

[5] ERDÖS, P. and RÈNYI, A. (1959). On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **4** 49–61.

[6] ESSEEN, C. G. (1945). Fourier analysis of distribution functions. *Acta Math.* **77** 1–125.

[7] HAJÉK, J. and SIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[8] RÉNYI, A. (1970). *Probability Theory*. North Holland, Amsterdam.

[9] ROBINSON, J. (1978). An asymptotic expansion for samples from a finite population. *Ann. Statist.* **6** 1005–1011.

[10] ROSÉN, B. (1965). Limit theorems for sampling from a finite population. *Ark. Mat.* **5** 383–424.

DEPARTMENT OF MATHEMATICAL STATISTICS
UNIVERSITY OF SYDNEY
SYDNEY, NEW SOUTH WALES 2006
AUSTRALIA