HIROSHIMA MATH. J.
14 (1984), 115–124

Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics

Minoru SIOTANI and Yasunori FUJIKOSHI

(Received April 21, 1983)

§ 1. Introduction

Let $\boldsymbol{Y}=(Y_1,\dots, Y_k)'$ be a random vector with the multinomial distribution $\mathrm{M}_k(n, \boldsymbol{\pi})$, i.e.,

$$\Pr(Y_1=n_1,\dots, Y_k=n_k) = \begin{cases} n!\prod_{j=1}^k (\pi_j^{n_j}/n_j!), & n_j = 0, 1,\dots, n \ (j=1,\dots, k) \\ & \text{and} \ \ \sum_{j=1}^k n_j = n, \\ 0 & , \ \text{otherwise}, \end{cases}$$

where $\boldsymbol{\pi}=(\pi_1,\dots, \pi_k)'$, $\pi_j>0$, $\sum_{j=1}^k \pi_j=1$. For testing the simple hypothesis H: $\boldsymbol{\pi}=\boldsymbol{p}$ ($\boldsymbol{p}$ a fixed vector) against K: $\boldsymbol{\pi}\neq\boldsymbol{p}$, the following three statistics are commonly used:

(1) Pearson's chi-square statistic

$$T_1 = \sum_{j=1}^k (Y_j-np_j)^2/(np_j),$$

(2) Log-likelihood ratio statistic

$$T_2 = 2\sum_{j=1}^k Y_j \log\{Y_j/(np_j)\},$$

(3) Freeman-Tukey statistic

$$T_3 = 4\sum_{j=1}^k (\sqrt{Y_j}-\sqrt{np_j})^2,$$

where $\boldsymbol{p}=(p_1,\dots, p_k)'$, $p_j>0$ $(j=1,\dots, k)$ and $\sum_{j=1}^k p_j=1$.

It is well known (e.g., see Bishop, Fienberg and Holland [2, p. 313]) that under the null hypothesis these three statistics have the same chisquare distribution with $k-1$ degrees of freedom in the limit. We use the chi-square approximation when expected numbers np_j are not small. But this brings about the question of how small the numbers np_j can be without invalidating the chi-square approximation. There are many papers attempting to answer this question, in particular, for the case of T_1, but there is wide difference of the proposed numbers for np_j. Another question arises in some practical applications when these statistics show significantly different values for a finite sample, in particular, between T_1 and T_2 or T_3.

Yarnold [6] obtained an asymptotic expansion for the null distribution of T_1

and studied the accuracy of the chi-square and other approximations to it. In this paper we give asymptotic expansions for the null distributions of T_2 and T_3 similar to that of T_1. Based on the asymptotic expansions, we shall propose new approximations for T_2 and T_3.

§2. A preliminary lemma

In this paper we assume $\boldsymbol{\pi}=\boldsymbol{p}$, since we treat the null distributions of T_i. Define

$$X_j = (Y_j - np_j)/\sqrt{n}, \quad j = 1,\dots, k,$$

and let $\boldsymbol{X}=(X_1,\dots, X_r)'$, where $r=k-1$. Then the random variable $\boldsymbol{X}$ is a lattice random vector which takes values in

$$L= \{\boldsymbol{x} = (x_1,\dots, x_r)';\ \boldsymbol{x} = (1/\sqrt{n})(\boldsymbol{m}-n\boldsymbol{q}) \quad \text{and} \quad \boldsymbol{m}\in M\}$$

where $\boldsymbol{q}=(p_1,\dots, p_r)'$ and M is a set of integer vectors $\boldsymbol{m}=(n_1,\dots, n_r)'$ such that $n_j\geq 0$ and $\sum_{j=1}^r n_j\leq n$. We can express $\boldsymbol{X}$ as

$$\boldsymbol{X} = \frac{1}{\sqrt{n}} \sum_{\alpha=1}^n (\boldsymbol{Z}_\alpha - \mathrm{E}(\boldsymbol{Z}_\alpha))$$

where $\boldsymbol{Z}_1,\dots, \boldsymbol{Z}_n$ are independently identically distributed lattice random vectors having the distribution of a random vector obtained by deleting the kth component of the random vector with the multinomial distribution $\mathrm{M}_k(1, \boldsymbol{p})$.

LEMMA 2.1. *Let* $\boldsymbol{x}=(1/\sqrt{n})(\boldsymbol{m}-n\boldsymbol{q})$. *Then for any* $\boldsymbol{m}\in M$,

$$\Pr(\boldsymbol{X}=\boldsymbol{x}) = n^{-r/2}\phi(\boldsymbol{x}) \left\{1+\frac{1}{\sqrt{n}}h_1(\boldsymbol{x}) + \frac{1}{n}h_2(\boldsymbol{x}) + O(n^{-3/2})\right\} \tag{2.1}$$

where $\phi(\boldsymbol{x}) = (2\pi)^{-r/2}|\Omega|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}'\Omega^{-1}\boldsymbol{x}\right)$, $\Omega=\mathrm{diag}\,(p_1,\dots, p_r)-\boldsymbol{q}\boldsymbol{q}'$,

$$h_1(\boldsymbol{x}) = -\frac{1}{2}\sum_{j=1}^k \frac{x_j}{p_j} + \frac{1}{6}\sum_{j=1}^k x_j\left(\frac{x_j}{p_j}\right)^2,$$

$$h_2(\boldsymbol{x}) = \frac{1}{2}h_1(\boldsymbol{x})^2 + \frac{1}{12}\left(1 - \sum_{j=1}^k \frac{1}{p_j}\right) + \frac{1}{4}\sum_{j=1}^k\left(\frac{x_j}{p_j}\right)^2 - \frac{1}{12}\sum_{j=1}^k x_j\left(\frac{x_j}{p_j}\right)^3$$

and $x_k= -\sum_{j=1}^r x_j$.

PROOF. Let $Q(\boldsymbol{t})$ be the characteristic function of $\boldsymbol{Y}^*=(Y_1,\dots, Y_r)'$, which is given by

$$Q(\boldsymbol{t}) = \sum_{\boldsymbol{m}\in M} \exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{m}) \Pr(\boldsymbol{Y}^* = \boldsymbol{m})$$
$$= \{\sum_{j=1}^{r} p_j \exp(\mathrm{i}t_j) + p_k\}^n$$

where $\boldsymbol{t}=(t_1,\ldots, t_r)'$. Then, for any $\boldsymbol{x}=(1/\sqrt{n})(\boldsymbol{m}-n\boldsymbol{q})\in L$, we have

$$\Pr(\boldsymbol{X}=\boldsymbol{x}) = \Pr(\boldsymbol{Y}^* = \boldsymbol{m})$$
$$= (2\pi)^{-r}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi} Q(\boldsymbol{t}) \exp(-\mathrm{i}\boldsymbol{t}'\boldsymbol{m})\, d\boldsymbol{t}$$
$$= (2\pi\sqrt{n})^{-r}\int_{-\sqrt{n}\pi}^{\sqrt{n}\pi}\cdots\int_{-\sqrt{n}\pi}^{\sqrt{n}\pi} \gamma(\boldsymbol{t}) \exp(-\mathrm{i}\boldsymbol{t}'\boldsymbol{x})\, d\boldsymbol{t}$$

where $\gamma(\boldsymbol{t})=Q(\boldsymbol{t}/\sqrt{n})\exp(-\mathrm{i}\sqrt{n}\boldsymbol{t}'\boldsymbol{q})$. For large n and fixed $\boldsymbol{t}$, $\gamma(\boldsymbol{t})$ can be expanded as

$$\gamma(\boldsymbol{t}) = \exp\left(-\frac{1}{2}\boldsymbol{t}'\Omega\boldsymbol{t}\right)\{1 + \sum_{j=1}^{2} n^{-j/2}b_j(\boldsymbol{t}) + O(n^{-3/2})\}$$

where

$$b_1(\boldsymbol{t}) = \frac{\mathrm{i}^3}{6}\{\sum_{j=1}^{r} p_j t_j^3 - 3(\boldsymbol{t}'\boldsymbol{q})\,\boldsymbol{t}'\Omega\boldsymbol{t} - (\boldsymbol{t}'\boldsymbol{q})^3\},$$

$$b_2(\boldsymbol{t}) = \frac{1}{2}b_1(\boldsymbol{t})^2 + \frac{\mathrm{i}^4}{24}\{\sum_{j=1}^{r} p_j t_j^4 - 4(\boldsymbol{t}'\boldsymbol{q})\sum_{j=1}^{r} p_j t_j^3$$
$$- 3(\boldsymbol{t}'\Omega\boldsymbol{t})^2 + 6(\boldsymbol{t}'\boldsymbol{q})^2\boldsymbol{t}'\Omega\boldsymbol{t} + 3(\boldsymbol{t}'\boldsymbol{q})^4\}.$$

Form a discussion on the asymptotic expansions of the density functions of sums of independent identically distributed random vectors (e.g., see Bhattacharya and Ranga Rao [1, p. 231]) it follows that

$$\Pr(\boldsymbol{X}=\boldsymbol{x}) = (2\pi\sqrt{n})^{-r}\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{x})\exp\left(-\frac{1}{2}\boldsymbol{t}'\Omega\boldsymbol{t}\right)\right.$$
$$\left.\times\{1 + \sum_{j=1}^{2} n^{-j/2}b_j(\boldsymbol{t})\}\,d\boldsymbol{t} + O(n^{-3/2})\right].$$

Now then the formula (2.1) is obtained by substituting the above expressions of $b_1(\boldsymbol{t})$ and $b_2(\boldsymbol{t})$ and carrying out the integration with the aid of formulae of the inverse Fourier transforms for the normal density and its derivatives.

Let $D=\mathrm{diag}(p_1,\ldots,p_k)$, $\sqrt{\boldsymbol{p}}=(\sqrt{p_1},\ldots,\sqrt{p_k})'$, and $A=(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k)'$ be a $k\times r$ matrix such that $(A, \sqrt{\boldsymbol{p}})$ is an orthogonal matrix. Define

$$\boldsymbol{z} = (z_1,\ldots, z_k)' = H\boldsymbol{x}$$

(2.2)
$$= A'D^{-1/2}\begin{bmatrix} I_r \\ -1,\ldots,-1\end{bmatrix}\boldsymbol{x}.$$

Then, noting that $H\Omega H' = I_r$ and $\sqrt{p_j}(\boldsymbol{a}_j'\boldsymbol{z}) = x_j$, we can express (2.1) as

$$\Pr(X=x) = n^{-r/2}\,|\Omega|^{-1/2}\{f(z) + O(n^{-3/2})\} \tag{2.3}$$

where

$$f(z) = (2\pi)^{-r/2}\exp\left(-\frac{1}{2}z'z\right)\left\{1 + \frac{1}{\sqrt{n}}g_1(z) + \frac{1}{n}g_2(z)\right\} \tag{2.4}$$

and

$$g_1(z) = -\frac{1}{2}\sum_{j=1}^k \frac{1}{\sqrt{p_j}}(\boldsymbol{a}_j'z) + \frac{1}{6}\sum_{j=1}^k \frac{1}{\sqrt{p_j}}(\boldsymbol{a}_j'z)^3,$$

$$g_2(z) = \frac{1}{2}g_1(z)^2 + \frac{1}{12}\left(1 - \sum_{j=1}^k \frac{1}{p_j}\right)$$

$$+ \frac{1}{4}\sum_{j=1}^k \frac{1}{p_j}(\boldsymbol{a}_j'z)^2 - \frac{1}{12}\sum_{j=1}^k \frac{1}{p_j}(\boldsymbol{a}_j'z)^4.$$

§3. Asymptotic expansions for $\Pr(X \in B)$

In order to get an asymptotic expansion for $\Pr(X \in B)$, it is necessary to sum the local expansion (2.1) over all the points in $B \cap L$. It is known (Esséen [3], Ranga Rao [5]) that such a lattice sum can be expressed as a Stieltjes integral when B is a Borel set. Yarnold [6] gave a reduction for the Stieltjes integral when B is an "extended convex set". It is convenient to summarize here Yarnold's result, since we will use it in the subsequent sections. A set B is called an extended convex set if B has the following representation for every $l \in \{1,\dots, r\}$:

$$B = \{\boldsymbol{x} = (x_1,\dots, x_r)' : \lambda_l(\boldsymbol{x}^*) < x_l < \theta_l(\boldsymbol{x}^*) \quad \text{and}$$
$$\boldsymbol{x}^* = (x_1,\dots, x_{l-1}, x_{l+1},\dots, x_r)' \in B_l\}$$

where $B_l \subset R^{r-1}$ and λ_l, θ_l are continuous functions on R^{r-1}. If B is an extended convex set, then

$$\Pr(X \in B) = J_1 + J_2 + J_3 + O(n^{-3/2}) \tag{3.1}$$

where

$$J_1 = \int\cdots\int_B \phi(\boldsymbol{x})\left\{1 + \frac{1}{\sqrt{n}}h_1(\boldsymbol{x}) + \frac{1}{n}h_2(\boldsymbol{x})\right\}d\boldsymbol{x},$$

$$J_2 = -\frac{1}{\sqrt{n}}\sum_{l=1}^r n^{-(r-l)/2}\sum_{x_{l+1}\in L_{l+1}}\cdots\sum_{x_r \in L_r}$$
$$\cdot [S_1(\sqrt{n}x_l + np_l)\phi(\boldsymbol{x})]_{\lambda_l(\boldsymbol{x}^*)}^{\theta_l(\boldsymbol{x}^*)}\,dx_1\cdots dx_{l-1},$$

$$J_3 = \frac{1}{n}\sum_{l=1}^r n^{-(r-l)/2}\sum_{x_{l+1}\in L_{l+1}}\cdots\sum_{x_r\in L_r}\int\cdots\int_{B_l}$$

$$\cdot\Big[-S_1(\sqrt{n}\,x_l+np_l)h_1(\boldsymbol{x})\phi(\boldsymbol{x})$$

$$+ S_2(\sqrt{n}\,x_l+np_l)\frac{\partial}{\partial x_l}\phi(\boldsymbol{x})\Big]_{\lambda_l(\boldsymbol{x}^*)}^{\theta_l(\boldsymbol{x}^*)} dx_1\cdots dx_{l-1},$$

$L_j = \{x_j : x_j = (1/\sqrt{n})\,(n_j - np_j)$ and n_j is integer$\}$,

$S_1(x) = x - [x] - \frac{1}{2}$,

$S_2(x)$ is the real-valued periodic function of period one such that $S_2(x) = \frac{1}{2}\left(x^2 - x + \frac{1}{6}\right)$ on $0 \le x < 1$,

$$[h(\boldsymbol{x})]_{\lambda_l(\boldsymbol{x}^*)}^{\theta_l(\boldsymbol{x}^*)} = h(x_1,\dots,x_{l-1},\theta_l(\boldsymbol{x}^*),x_{l+1},\dots,x_r)$$

$$- h(x_1,\dots,x_{l-1},\lambda_l(\boldsymbol{x}^*),x_{l+1},\dots,x_r).$$

The J_1 term can be regarded as the Edgeworth expansion for a continuous distribution, while the J_2 term is a term to account for the discontinuity in $\boldsymbol{X}$. It is known that $J_2 = O(n^{-1/2})$ and $J_3 = O(n^{-1})$. Since Pearson's chi-square statistic T_1 is expressed as $T_1 = \boldsymbol{X}'\Omega^{-1}\boldsymbol{X}$, it holds that

$$\Pr(T_1 < c) = \Pr(\boldsymbol{X} \in B_1)$$

where $B_1 = \{\boldsymbol{x} = (x_1,\dots,x_r)' : \boldsymbol{x}'\Omega^{-1}\boldsymbol{x} < c\}$. Hoel [4] evaluated the J_1 term for the case of $B = B_1$. Yarnold [6] evaluated the J_2 term for the case of $B = B_1$ and showed that $J_1 + J_2$ provides very accurate approximation to $\Pr(T_1 < c)$.

§4. Log-likelihood ratio statistic

We can express the null distribution of T_2 as

$$\Pr(T_2 < c) = \Pr(\boldsymbol{X} \in B_2) \tag{4.1}$$

where $B_2 = \{\boldsymbol{x} = (x_1,\dots,x_r)' : T_2(\boldsymbol{x}) < c\}$ and

$$\mathrm{T}_2(\boldsymbol{x}) = 2\sum_{j=1}^{k}(np_j + \sqrt{n}x_j)\log\{1 + x_j/(\sqrt{n}p_j)\}. \tag{4.2}$$

Observing that the set B_2 is an extended convex set, we can write (4.1) as the formula (3.1) with $B = B_2$. In the following we shall evaluate the J_1 and J_2 terms.

Making the transformation (2.2), we can write J_1 as

$$J_1 = \int\cdots\int_{\tilde{B}_2} f(\boldsymbol{z})d\boldsymbol{z} \tag{4.3}$$

where $f(\boldsymbol{z})$ is given by (2.4) and $\tilde{B}_2 = \{\boldsymbol{z} = (z_1,\dots,z_r)' : T_2(H^{-1}\boldsymbol{z}) < c\}$. We may regard J_1 as the distribution function of $T_2(H^{-1}\boldsymbol{Z})$ when $\boldsymbol{Z}$ has a continuous density function $f(z)$. Then the characteristic of $T_2(H^{-1}Z)$ is defined by

$$C(t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\{itT_2(H^{-1}z)\}f(z)dz. \tag{4.4}$$

We can expand $T_2(H^{-1}z)$ as

$$\begin{aligned}T_2(H^{-1}z) &= z'z-\frac{1}{3\sqrt{n}}\textstyle\sum_{j=1}^{k}\frac{1}{\sqrt{p_j}}(a_j'z)^3 \\ &\quad+\frac{1}{6n}\textstyle\sum_{j=1}^{k}\frac{1}{p_j}(a_j'z)^4+O(n^{-3/2})\end{aligned} \tag{4.5}$$

in the set Θ_n of z for which $|a_j'z/\sqrt{n}p_j|<1$, $j=1,\dots,r$. Substituting (4.5) into (4.4), we obtain

$$\begin{aligned}C(t) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp(it\,z'z)\Big[1-\frac{it}{3\sqrt{n}}\textstyle\sum_{j=1}^{k}\frac{1}{\sqrt{p_h}}(a_j'z)^3 \\ &\quad+\frac{it}{6n}\textstyle\sum_{j=1}^{k}\frac{1}{p_j}(a_j'z)^4+\frac{(it)^2}{18n}\Big\{\sum_{j=1}^{k}\frac{1}{\sqrt{p_j}}(a_j'z)^3\Big\}^2\Big]f(z)dz+O(n^{-3/2}).\end{aligned} \tag{4.6}$$

The validity of this reduction is obtained by controling the errors of approximations. For example, if we define the set Λ_n of z by $|z_j|<2\sqrt{2\log n}$, $j=1,\dots,r$, then it can be checked that for sufficiently large n,

(i) $\Lambda_n\subset\Theta_n$,

(ii) $\int_{\Lambda_n^c}|f(z)|dz=o(n^{-2})$.

The formula (4.6) follows by dividing the region of the integral in (4.4) into Λ_n and Λ_n^c and using the properties (i) and (ii). Carrying out the integral (4.6) with the aid of the moment formulae for a multivariate normal variate, we obtain

$$\begin{aligned}C(t) &= (1-2it)^{-r/2}\Big[1+\frac{1}{12n}\Big(1-\textstyle\sum_{j=1}^{k}\frac{1}{p_j}\Big)\{1-(1-2it)^{-1}\}\Big] \\ &\quad+O(n^{-3/2}).\end{aligned} \tag{4.7}$$

Inverting (4.7) we obtain

$$\begin{aligned}J_1 &= \Pr(\chi_r^2<c)+\frac{1}{12n}\Big(1-\textstyle\sum_{j=1}^{k}\frac{1}{p_j}\Big) \\ &\quad\times\{\Pr(\chi_r^2<c)-\Pr(\chi_{r+2}^2<c)\}+O(n^{-3/2}).\end{aligned} \tag{4.8}$$

Next we consider the J_2 term in (3.1) with $B=B_2$. Approximating $[S_1(\sqrt{n}x_l+np_l)\phi(x)]_{\lambda_l(x_*)}^{\theta_l(x_*)}$ by its asymptotic approximation

$$(2\pi)^{-r/2}|\Omega|^{-1/2}\exp\Big(-\frac{1}{2}c\Big)[S_1(\sqrt{n}x_l+np_l)]_{\lambda_l(x_*)}^{\theta_l(x_*)},$$

and using the same argument as in Yarnold [6], we obtain an asymptotic approximation to J_2,

(4.9) $$\hat{J}_2 = (N_2 - n^{r/2}V_2)\exp\left(-\frac{1}{2}c\right)\Big/\{(2\pi n)^r \textstyle\prod_{j=1}^{k} p_j\}^{1/2}$$

where N_2 is the number of lattice points in B_2, i.e.,

(4.1) $$N_2 = \#\{x;\ x \in L \quad \text{and} \quad T_2(x) < c\}$$

and V_2 is the volume of B_2. We shall give an expansion for

(4.11) $$V_2 = \int\cdots\int_{B_2} dx$$
$$= |\Omega|^{1/2}\int\cdots\int_{\tilde{B}_2} dz$$

where $z = Hx$ is defined by (2.2) and $\tilde{B}_2 = \{z : z = Hx \text{ and } x \in B_2\}$. Consider the transformation $z \to u$ such that $T_2(H^{-1}z) = u'u$. Using (4.5) we can express z in terms of u as

$$z = u + \frac{1}{6\sqrt{n}}\sum_{j=1}^{k}\frac{1}{\sqrt{p_j}}(a_j'u)^2 a_j$$
$$-\frac{1}{72n}\left\{5(u'u)u + \sum_{j=1}^{k}\frac{1}{p_j}(a_j'u)^3 a_j\right\} + O(n^{-3/2})$$

for sufficiently large n. It is seen that the Jacobian of the transformation is

$$\left|\frac{\partial z}{\partial u}\right| = \Bigg| I_r + \frac{1}{3\sqrt{n}}\sum_{j=1}^{k}\frac{1}{\sqrt{p_j}}(a_j' a_j)(a_j u')$$
$$-\frac{1}{72n}\left\{10uu' + 5(u'u)I_r + 3\sum_{j=1}^{k}\frac{1}{p_j}(a_j u')^2 a_j a_j'\right\} + O(n^{-3/2})\Bigg|$$
$$= 1 + \frac{1}{3\sqrt{n}}\sum_{j=1}^{k}\frac{1}{\sqrt{p_j}}a_j'u + \frac{1}{72n}\Bigg\{(1-5r)u'u$$
$$-7\sum_{j=1}^{k}\frac{1}{p_j}(a_j'u)^2 + 4\left(\sum_{j=1}^{k}\frac{1}{\sqrt{p_j}}a_j'u\right)^2\Bigg\} + O(n^{-3/2}).$$

From (4.11) we obtain

(4.12) $$V_2 = |\Omega|^{1/2}\int\cdots\int_{u'u<c}\left|\frac{\partial z}{\partial u}\right| du$$
$$= V_1\Bigg[1 + \frac{c}{n}\{72(k+1)\}^{-1}\Bigg\{-9k^2 + 15k - 6$$
$$-3\sum_{j=1}^{k}\frac{1}{p_j}\Bigg\} + O(n^{-3/2})\Bigg]$$

where V_1 is the volume of B_1, i.e.,

(4.13)
$$V_1 = |\Omega|^{1/2}\int\cdots\int_{u'u<c} du$$
$$= \{(\pi c)^r \textstyle\prod_{j=1}^k p_j\}^{1/2}/\Gamma\left(\frac{1}{2}r + 1\right).$$

The J_3 term is very complicated. However, from the general result in Section 3 it follows that $J_3 = O(n^{-1})$. Neglecting the J_3 term, it is suggested to use

(4.14)
$$J_1 + \hat{J}_2$$

as an approximation to $\Pr(T_2 < c)$, where J_1 and $\hat{J}_2$ are defined by (4.8) and (4.9), respectively.

§5. Freeman-Tukey statistic

The null distribution of the Freeman-Tukey statistic T_3 can be expressed as

(5.1)
$$\Pr(T_3 < c) = \Pr(X \in B_3)$$

where $B_3 = \{x = (x_1, \ldots, x_r)' : T_3(x) < c\}$ and

(5.2)
$$T_3(x) = 4\textstyle\sum_{j=1}^k \{(np_j + \sqrt{n}x_j)^{1/2} - \sqrt{np_j}\}^2.$$

It is easily seen that B_3 is an extended convex set. Therefore we can write (5.1) as the formula (3.1) with $B = B_3$. In the following we shall evaluate the J_1 and J_2 terms in (3.1) with $B = B_3$. When $|x_j/(\sqrt{n}p_j)| = |(a_j'z)/\sqrt{n}p_j| < 1$ and $|1/(np_j)| < 1$, we can expand $T_3(x)$ as

(5.3)
$$T_3(x) = T_3(H^{-1}z)$$
$$= z'z - \frac{1}{2\sqrt{n}}\sum_{j=1}^k \frac{1}{\sqrt{p_j}}(a_j'z)^3$$
$$+ \frac{5}{16n}\sum_{j=1}^k \frac{1}{p_j}(a_j'z)^4 + O(n^{-3/2}).$$

The J_1 term can be obtained by using the formula (5.3) and the same method as in the case of T_2. The final result is given by

(5.4)
$$J_1 = \Pr(\chi_r^2 < c) + \frac{1}{n}\sum_{j=0}^3 g_j \Pr(\chi_{r+2j}^2 < c) + O(n^{-3/2})$$

where

$$g_0 = \frac{1}{12}\left(1 - \sum_{j=1}^k \frac{1}{p_j}\right),\quad g_1 = \frac{1}{32}(-k^2 + 4k - 3),$$

$$g_2 = \frac{1}{32}\left(2k^2 - 2k - 1 + \sum_{j=1}^{k} \frac{1}{p_j}\right),$$

$$g_3 = \frac{1}{96}\left(- 3k^2 - 6k + 4 + 5 \sum_{j=1}^{k} \frac{1}{p_j}\right).$$

Applying an argument similar to that in the case of T_2, we can obtain an asymptotic approximation for the J_2 term given by

$$\hat{J}_2 = (N_3 - n^{r/2} V_3) \exp\left(- \frac{1}{2} c\right) \Big/ \{(2n)^r \textstyle\prod_{j=1}^{k} p_j\}^{1/2} \tag{5.5}$$

where N_3 is the number of lattice points in B_3, i.e.,

$$N_3 = \#\{x : x \in L \quad \text{and} \quad T_3(x) < c\} \tag{5.6}$$

and

$$V_3 = \int \cdots \int_{B_3} dx. \tag{5.7}$$

Similarly we can derive an asymptotic expansion for V_3. For this, we consider the transformation $z \to u$ such that $T_3(H^{-1}z) = u'u$. From (5.3) we can write the transformation as

$$z = u + \frac{1}{4\sqrt{n}} \sum_{j=1}^{k} \frac{1}{\sqrt{p_j}} (a_j' u)^2 a_j - \frac{5}{32n} (u'u)u + O(n^{-3/2}).$$

Therefore we have

$$\begin{aligned} V_3 &= |\Omega|^{1/2} \int \cdots \int_{u'u<c} \left| \frac{\partial z}{\partial u} \right| du \\ &= |\Omega|^{1/2} \int \cdots \int_{u'u<c} \Bigg[1 + \frac{1}{2\sqrt{n}} \sum_{j=1}^{k} \frac{1}{\sqrt{p_j}} a_j' u \\ &\quad + \frac{1}{32n} \Bigg\{ - (2+5r) u'u + 4\Bigg(\sum_{j=1}^{k} \frac{1}{\sqrt{p_j}} a_j' u \Bigg)^2 \\ &\qquad - 4 \sum_{j=1}^{k} \frac{1}{p_j} (a_j' u)^2 \Bigg\} + O(n^{-3/2}) \Bigg] du \\ &= V_1 \left\{ 1 - \frac{3c}{32n} (k-1)(3k-1)(k+1)^{-1} + O(n^{-3/2}) \right\} \end{aligned} \tag{5.8}$$

where V_1 is given by (4.13).

The formulas (5.4) and (5.5) will be useful in getting closer approximations to $\Pr(T_3 < c)$.

References

[1] R. N. Bhattacharya and R. Ranga Rao, Normal Approximation and Asymptotic Expansions, Wiely, New York (1975).
[2] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland, Discrete Multivariate Analysis: Theory and Practice, The MIT Press, Cambridge (1975).
[3] C. G. Esseén, Fourier analysis of distribution functions: A mathematical study of the Laplace-Gaussian law, Acta Math., **77** (1945), 1–125.
[4] P. G. Hoel, On the chi-square distribution for small samples, Ann. Math. Statist., **9** (1938), 158–165.
[5] R. Ranga Rao, On the central limit theorem in R_k, Bull. Amer. Math. Soc., **67** (1961), 359–361.
[6] J. K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set, Ann. Math. Statist., **43** (1972), 1566–1580.

Department of Information and Behavioral Sciences,
Faculty of Integrated Arts and Sciences,
Hiroshima University
and
Department of Mathematics,
Faculty of Science,
Hiroshima University