



Title	Jensen-differenceに基づく多項分布の適合度検定統計量の漸近近似について
Author(s)	関谷, 祐里; Sekiya, Yuri; 佐藤, 義治 他
Citation	北海道大學工学部研究報告, 146, 63-73
Issue Date	1989-05-31
Doc URL	https://hdl.handle.net/2115/42189
Type	departmental bulletin paper
File Information	146_63-74.pdf



Jensen-difference に基づく多項分布の 適合度検定統計量の漸近近似について

関谷 祐里 佐藤 義治* 河口 至商
(昭和 63 年 12 月 20 日受理)

Asymptotic Approximations for the Distributions of Multinomial Goodness-of-fit Statistics Based on the Jensen-difference

Yuri SEKIYA, Yoshiharu SATO and Michiaki KAWAGUCHI
(Received December 20, 1988)

Abstract

The test statistic for a goodness-of-fit of multinomial distribution based on J-divergence, which is defined as the Jensen-difference of the entropy function, has been discussed. The asymptotic distribution of the test statistic J_α , defined as Jensen-difference of the entropy function of the order α , was obtained under null hypothesis.

In this paper, two asymptotic approximations of the distribution of J_1 under simple null hypothesis are proposed. One is an approximation of J_1 based on the asymptotic expansion with the discontinuous term for the null distribution. The other is an approximation of J_1 based on the asymptotic expansion of the moment under the null hypothesis. In the case of $\alpha \neq 1$ ($\alpha > 0$), the asymptotic approximations of the distribution of J_α under the equiprobable hypothesis are obtained in the same way as that of J_1 .

1. はじめに

多項分布の適合度検定統計量として、ピアソンの χ^2 統計量 (ここでは T_1 と表す)、対数尤度比統計量 T_0 、Freeman-Tukey 統計量 $T_{-1/2}$ をはじめとする

$$T_\lambda \equiv 2nI^\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k x_i \left\{ \left(\frac{x_i}{np_i} \right)^\lambda - 1 \right\}, \lambda \in R$$

という形の検定統計量がある^{2) 4)}。よく知られているように、これらの統計量はいずれも帰無仮説のもとでの極限分布が χ^2 分布に従い、標本数が十分大きい場合には χ^2 分布による近似を用いて検定を行っている。

等しい極限分布をもつこれらの検定統計量の比較を行うために、また、有限標本に基づく検定に対してより正確な棄却域を与えるために、帰無仮説のもとでの分布に対して漸近展開に基づく

近似について研究されている。Yarnold⁸⁾, Siotani and Fujikoshi⁵⁾, Read⁴⁾ はそれぞれ, T_1 , T_0 および $T_{-1/2}$, T_λ ($\lambda \in R$) に対して, 帰無仮説のもとでの各統計量の分布の漸近展開に基づいた離散修正項をもつ漸近近似式を求めている。また, Cressie and Read²⁾ は T_λ ($\lambda \in R$) の分布に対して, 帰無仮説のもとでの検定統計量のモーメントの漸近展開に基づいて極限分布を修正することによって得られる漸近近似式を与えている。

一方, $\{T_\lambda; \lambda \in R\}$ には含まれない検定統計量として, エントロピー関数の Jensen-difference として定義されている J-divergence¹⁾ による適合度検定統計量が考察されている。特に, オーダー α のエントロピー関数を用いた検定統計量 J_α は, $\alpha=1$ のとき, 帰無仮説のもとでの極限分布は χ^2 分布に従うことが知られている^{6) 7)}。

本論文では, 統計量の分布に対する漸近近似を与える前述の方法にしたがって, 単純帰無仮説のもとでの検定統計量 J_1 の分布に対して, 離散修正項をもつ漸近近似式 F_Y , およびモーメントに基づいて極限分布を修正して得られる漸近近似式 F_R を提案する。また, J_α ($\alpha > 0, \alpha \neq 1$) に対しては, 分布の一様性を検定するための帰無仮説の場合について, 同様の方法による漸近近似式 G_Y, G_R を求めている。

2. J-divergence による多項分布の適合度検定統計量

離散確率の全体を

$$F_k = \{p = (p_1, \dots, p_k) : p_i \geq 0, i=1, \dots, k, \sum_{i=1}^k p_i = 1\}$$

とする。 F_k の任意の要素 p, q に対して, オーダー α のエントロピー関数³⁾ $H_{k,\alpha}(p)$ に基づく J-divergence¹⁾ は, 次のように定義されている。

$$J_{k,\alpha}(p, q) \equiv H_{k,\alpha}\left(\frac{p+q}{2}\right) - \frac{1}{2}H_{k,\alpha}(p) - \frac{1}{2}H_{k,\alpha}(q)$$

ただし

$$H_{k,\alpha}(p) = \begin{cases} \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^k p_i^\alpha\right), & \alpha > 0, \alpha \neq 1, \\ -\sum_{i=1}^k p_i \log p_i, & \alpha = 1, \end{cases} \quad , p = (p_1, \dots, p_k)' \in F_k.$$

多項分布 $M(n; \pi_1, \dots, \pi_k)$ からの観測値 $x = (x_1, \dots, x_k)'$ により, 単純帰無仮説 $H_0: \pi = p$, ただし $\pi = (\pi_1, \dots, \pi_k)'$, $p = (p_1, \dots, p_k)'$ を検定するための適合度検定統計量として

$$J_\alpha \equiv 8n J_{k,\alpha}\left(\frac{x}{n}, p\right)$$

を用いることができる。 J_α を具体的に表すと次のようになる。

$$J_1 = 4n \sum_{i=1}^k \left\{ p_i \log p_i + \frac{x_i}{n} \log \frac{x_i}{n} - \left(p_i + \frac{x_i}{n}\right) \log \left(\frac{1}{2} \left(p_i + \frac{x_i}{n}\right)\right) \right\}$$

$$J_\alpha = \frac{4n}{\alpha-1} \sum_{i=1}^k \left\{ p_i^\alpha + \left(\frac{x_i}{n}\right)^\alpha - 2 \left(\frac{1}{2} \left(p_i + \frac{x_i}{n}\right)\right)^\alpha \right\}, \quad \alpha > 0, \alpha \neq 1.$$

H_0 のもとでの J_α の極限分布については次の結果が知られている^{6) 7)}。

$$J_\alpha \xrightarrow{L} \sum_{i=1}^{k-1} \lambda_i G_{1i}, \quad n \rightarrow \infty.$$

ただし、 L は分布関数の収束を表し、 G_{1i} は互いに独立に自由度 1 の χ^2 分布に従う確率変数、 λ_i は $\Lambda\Psi$ の固有値、ここに $\Lambda = \text{diag}(\alpha p_1^{\alpha-2}, \dots, \alpha p_k^{\alpha-2})$ 、 $\Psi = \{(\delta_{ij} - p_i p_j); i, j = 1, \dots, k\}$ 、 $\text{rank } \Lambda\Psi = k-1$ 。特に、 J_1 は漸近的に自由度 $k-1$ の χ^2 分布に従う。

3. 分布の漸近展開を求めるための準備

検定統計量 J_1 の分布の漸近展開に基づく近似式を求めるための準備として、Siotani and Fujikoshi⁵⁾ の Lemma と Yarnold⁸⁾ の定理をここで述べておく。

$X = (X_1, \dots, X_k)'$ は多項分布 $M(n; p_1, \dots, p_k)$ に従う確率変数であるとする。変数変換

$$Y_i = (X_i - n p_i) / \sqrt{n}, \quad i = 1, \dots, k \quad (3.1)$$

を行い、 $Y = (Y_1, \dots, Y_r)'$ 、 $r = k-1$ とおくと、 Y は集合

$$M = \{y = (y_1, \dots, y_r)' : y = (\bar{x} - n \bar{p}) / \sqrt{n}, \bar{x} \in \tilde{M}\}$$

の要素を値としてとりうるような格子点分布に従う⁵⁾。ここで

$$\bar{p} = (p_1, \dots, p_r)', \quad \tilde{M} = \{\bar{x} = (x_1, \dots, x_r)' : x_i \text{ は非負整数}, i = 1, \dots, r, \sum_{i=1}^r x_i \leq n\}.$$

Siotani and Fujikoshi⁵⁾ の Lemma は次のように述べられている。

Lemma 1 (Siotani and Fujikoshi⁵⁾)

$y = (\bar{x} - n \bar{p}) / \sqrt{n}$ とする。 \tilde{M} の任意の要素 \bar{x} に対して

$$\Pr\{Y = y\} = n^{-r/2} \phi(y) \left\{ 1 + \frac{1}{\sqrt{n}} h_1(y) + \frac{1}{n} h_2(y) + O(n^{-3/2}) \right\} \quad (3.2)$$

となる。ここに

$$\phi(y) = (2\pi)^{-r/2} |\Omega|^{-1/2} \exp(-y' \Omega y / 2), \quad \Omega = \text{diag}(p_1, \dots, p_r) - \bar{p} \bar{p}' \quad (3.3)$$

$$h_1(y) = -\frac{1}{2} \sum_{i=1}^k \frac{y_i}{p_i} + \frac{1}{6} \sum_{i=1}^k \frac{y_i^3}{p_i^2},$$

$$h_2(y) = \frac{1}{2} (h_1(y))^2 + \frac{1}{12} \left(1 - \sum_{i=1}^k \frac{1}{p_i} \right) + \frac{1}{4} \sum_{i=1}^k \frac{y_i^2}{p_i^2} - \frac{1}{12} \sum_{i=1}^k \frac{y_i^4}{p_i^3},$$

$$y_k = -\sum_{i=1}^k y_i$$

である。さらに、変数変換

$$z = (z_1, \dots, z_r)' = Hy = A' D^{-1/2} \begin{bmatrix} I_r \\ -1 \dots -1 \end{bmatrix} y \quad (3.4)$$

を行う。ただし

$$D = \text{diag}(p_1, \dots, p_k), \quad \sqrt{p} = (\sqrt{p_1}, \dots, \sqrt{p_k})',$$

$A = (a_1, \dots, a_k)'$: (A, \sqrt{p}) が直行列になるような $k \times r$ 行列。

これを用いると (3.2) 式は次のように表すことができる⁵⁾。

$$\Pr\{Y = y\} = n^{-r/2} |\Omega|^{-1/2} \{f(z) + O(n^{-3/2})\} \quad (3.5)$$

ただし

$$f(z) = (2\pi)^{-r/2} \exp\left(-\frac{1}{2} z' z\right) \left\{ 1 + \frac{1}{\sqrt{n}} g_1(z) + \frac{1}{n} g_2(z) \right\} \quad (3.6)$$

$$g_1(z) = -\frac{1}{2} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i' z) + \frac{1}{6} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i' z)^3,$$

$$g_2(z) = \frac{1}{2} (g_1(z))^2 + \frac{1}{12} \left(1 - \sum_{i=1}^k \frac{1}{p_i}\right) + \frac{1}{4} \sum_{i=1}^k \frac{1}{p_i} (a_i' z)^2 - \frac{1}{12} \sum_{i=1}^k \frac{1}{p_i} (a_i' z)^4.$$

集合 B が extended convex set⁸⁾ であるとは、任意の $c \in \{1, \dots, r\}$ に対して B が

$$B = \{y = (y_1, \dots, y_r)' : \lambda_c(y^*) < y_c < \theta_c(y^*), y^* = (y_1, \dots, y_{c-1}, y_{c+1}, \dots, y_r)' \in B_c\}$$

と表されることである。ただし $B_c \subset R^{r-1}$, λ_c, θ_c は R^{r-1} 上の連続関数である。このとき次の定理が成り立つ。

定理 (Yarnold⁸⁾)

B が有界な extended convex set ならば

$$\Pr\{Y \in B\} = D_1 + D_2 + D_3 + O(n^{-3/2}) \quad (3.7)$$

となる。ただし

$$D_1 = \int \cdots \int_B \phi(y) \left\{ 1 + \frac{1}{\sqrt{n}} h_1(y) + \frac{1}{n} h_2(y) \right\} dy,$$

$$D_2 = -\frac{1}{\sqrt{n}} \sum_{c=1}^r n^{-(r-c)/2} \sum_{y_{c+1} \in M_{c+1}} \cdots \sum_{y_r \in M_r} \int \cdots \int_{B_c} [S_1(\sqrt{n} y_c + n p_c) \phi(y)]_{\lambda_c(y^*)}^{\theta_c(y^*)} dy_1 \cdots dy_{c-1},$$

$$D_3 = \frac{1}{n} \sum_{c=1}^r n^{-(r-c)/2} \sum_{y_{c+1} \in M_{c+1}} \cdots \sum_{y_r \in M_r} \int \cdots \int_{B_c} [-S_1(\sqrt{n} y_c + n p_c) h_1(y) \phi(y) + S_2(\sqrt{n} y_c + n p_c) \frac{\partial}{\partial y_c} \phi(y)]_{\lambda_c(y^*)}^{\theta_c(y^*)} dy_1 \cdots dy_{c-1},$$

$$S_1(y) = y - [y] - 1/2,$$

$S_2(y)$ は周期 1 の実数値周期関数で、 $S_2(y) = (y^2 - y + 1/6)/2$, ($0 \leq y \leq 1$),

$$[h(y)]_{\lambda_c(y^*)}^{\theta_c(y^*)} = h(y_1, \dots, y_{c-1}, \theta_c(y^*), y_{c+1}, \dots, y_r) - h(y_1, \dots, y_{c-1}, \lambda_c(y^*), y_{c+1}, \dots, y_r)$$

$$M_i = \{y_i : y_i = (x_i - n p_i) / \sqrt{n}, x_i \text{ は非負整数}\}$$

である。

上の定理において D_1 は、離散確率変数 Y を連続分布によって近似したときの Edgeworth 展開を与える項である。 D_2 と D_3 は離散修正項であり、 $D_2 = O(n^{-1/2})$, $D_3 = O(n^{-1})$ である。さらに、(3.4) 式で表される変換を用いて、 $\tilde{B} = \{z = (z_1, \dots, z_r)' : H^{-1} z \in B\}$ とおくならば、

$$D_1 = \int \cdots \int_{\tilde{B}} f(z) dz \quad (3.8)$$

と表すことができる。

4. J_1 の分布の漸近近似: F_Y

$X = (X_1, \dots, X_k)'$ は、多項分布 $M(n; p_1, \dots, p_k)$ に従う確率変数であるとして、帰仮説 H_0 のもとでの J_1 の分布を求める。

(3.1) 式と同様の変数変換を行うことによって、 J_1 は確率変数 Y の関数として次のように表すことができる。

$$J_1(Y) = 4n \sum_{i=1}^k \left\{ p_i \log p_i + \left(p_i + \frac{Y_i}{\sqrt{n}} \right) \log \left(p_i + \frac{Y_i}{\sqrt{n}} \right) - \left(2p_i + \frac{Y_i}{\sqrt{n}} \right) \log \left(p_i + \frac{Y_i}{2\sqrt{n}} \right) \right\}.$$

$$B_1 = \{y = (y_1, \dots, y_r) : J_1(y) < t\}$$

とおくと, B_1 は有界な extended convex set になる. また, B_1 をこのようにおくことによって J_1 の Exact な分布関数 $F_E(t)$ は $F_E(t) = \Pr\{J_1 < t\} = \Pr\{Y \in B_1\}$ となり, この確率の漸近展開は前に述べた Yarnold の定理を用いて求めることができる. それを $F_E(t) = D_1 + D_2 + D_3 + O(n^{-3/2})$ とおく.

(3.8) 式より, D_1 は Z を密度関数 $f(z)$ をもつ連続な確率変数であるとみなしたときの J_1 の分布関数を表していると考えることができる. これを求めるために J_1 の特性関数

$$C(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\{itJ_1(H^{-1}z)\} f(z) dz$$

を求める. $J_1(H^{-1}z)$ をテーラー展開することによって

$$J_1(H^{-1}z) = z'z - \frac{1}{2\sqrt{n}} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z)^3 + \frac{7}{24n} \sum_{i=1}^k \frac{1}{p_i} (a_i'z)^4 + O(n^{-3/2}) \quad (4.1)$$

を得る.

ここで注意しておくが, Read⁴⁾ が扱った適合度検定統計量 T_λ に対しては

$$T_\lambda = z'z + \frac{\lambda-1}{3\sqrt{n}} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z)^3 + \frac{(\lambda-1)(\lambda-2)}{12n} \sum_{i=1}^k \frac{1}{p_i} (a_i'z)^4 + O(n^{-3/2})$$

となる. これと (4.1) 式を比較することにより, J_1 は $\{T_\lambda; \lambda \in \mathbb{R}\}$ には含まれない統計量であることがわかる.

(3.6) 式と (4.1) 式より, $C(t)$ は次のようになる.

$$C(t) = (2\pi)^{-n/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2} z'z\right) A(z) dz + O(n^{-3/2}) \quad (4.2)$$

ただし, $\sigma^2 = 1/(1-2it)$,

$$A(z) = 1 + \frac{1}{\sqrt{n}} \left[-\frac{1}{2} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z) + \frac{1-3it}{6} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z)^3 \right]$$

$$+ \frac{1}{n} \left[\frac{1}{12} (1-S) + \frac{1}{4} \sum_{i=1}^k \frac{1}{p_i} (a_i'z)^2 + \frac{1}{8} \left\{ \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z) \right\}^2 - \frac{2-7it}{24} \sum_{i=1}^k \frac{1}{p_i} (a_i'z)^4 \right.$$

$$\left. - \frac{1-3it}{12} \left\{ \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z) \right\} \left\{ \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z)^3 \right\} + \frac{(1-3it)^2}{72} \left\{ \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'z)^3 \right\}^2 \right],$$

$$S = \sum_{i=1}^k \frac{1}{p_i}.$$

(4.2) 式の積分を計算することにより, $C(t)$ は次のように得られる.

$$C(t) = \sigma^r + \frac{\sigma^r}{n} \left\{ \frac{1}{12} (1-S) + \frac{\sigma^2}{32} (S - k^2 + 2k - 2) + \frac{\sigma^4}{16} (k^2 - 1) + \frac{\sigma^6}{96} (5S - 3k^2 - 6k + 4) \right\}$$

$$+ O(n^{-3/2}).$$

$\sigma^r = (1-2it)^{-n/2}$ は自由度 r の χ^2 分布の特性関数であるから

$$D_1 = F_{\chi(r)}(t) + \frac{1}{96n} \left\{ l_0 F_{\chi(r)}(t) + l_1 F_{\chi(r+2)}(t) + l_2 F_{\chi(r+4)}(t) + l_3 F_{\chi(r+6)}(t) \right\} + O(n^{-3/2}) \quad (4.3)$$

を得る. ただし

$F_{x(r)}(t)$: 自由度 r の χ^2 分布関数,

$$\begin{aligned} l_0 &= 8(1-S), \quad l_1 = 3(S-k^2+2k-2), \\ l_2 &= 6(k^2-1), \quad l_3 = 5S-3k^2-6k+4. \end{aligned}$$

次に D_2 を求める。Siotani and Fujikoshi⁵⁾ にならって, D_2 の中に現れる

$$[S_1(\sqrt{n} y_c + n p_c) \phi(y)]_{\lambda_c(y^*)}^{\theta_c(y^*)}$$

をその漸近近似である

$$(2\pi)^{-r/2} |\mathcal{Q}|^{-1/2} \exp(-t/2) [S_1(\sqrt{n} y_c + n p_c)]_{\lambda_c(y^*)}^{\theta_c(y^*)}$$

によって近似すると, D_2 の漸近近似である \hat{D}_2 が, Yarnold⁸⁾ の結果を用いて次のように得られる。

$$\hat{D}_2 = [(N_1(t) - n^{r/2} V_1(t)) \exp(-t/2)] / [((2\pi n)^r \prod_{i=1}^k p_i)^{1/2}],$$

ここで,

$$N_1(t) = \#\{y: y \in M \text{ and } y \in B_1\},$$

$$V_1(t) = \int \cdots \int_{B_1} dy = |\mathcal{Q}|^{1/2} \int \cdots \int_{\tilde{B}_1} dz, \quad \tilde{B}_1 = \{z = (z_1, \dots, z_r) : J_1(H^{-1}z) < t\}.$$

次に, $V_1(t)$ の展開を求めるために, $J_1(H^{-1}z) = u'u$ となるような変換 $z \rightarrow u$ を考える。このような変換に対して

$$z = u + \frac{1}{4\sqrt{n}} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'u)^2 a_i + \frac{1}{96n} \left\{ \sum_{i=1}^k \frac{1}{p_i} (a_i'u)^3 a_i - 15(u'u)u \right\} + O(n^{-3/2})$$

という展開を得る。これより,

$$\begin{aligned} V_1(t) &= |\mathcal{Q}|^{1/2} \int \cdots \int_{u'u < t} \left| \frac{\partial z}{\partial u} \right| du \\ &= |\mathcal{Q}|^{1/2} \int \cdots \int_{u'u < t} \left[l_r + \frac{1}{2\sqrt{n}} \sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'u) + \frac{1}{32n} \left\{ -3 \sum_{i=1}^k \frac{1}{p_i} (a_i'u)^2 \right. \right. \\ &\quad \left. \left. + (2-5k)(u'u) + 4 \left(\sum_{i=1}^k \frac{1}{\sqrt{p_i}} (a_i'u) \right)^2 \right\} + O(n^{-3/2}) \right] du \\ &= \frac{\{(\pi t)^r \prod_{i=1}^k p_i\}^{1/2}}{\Gamma(r/2+1)} \left\{ 1 + \frac{t}{32n(k+1)} (S-9k^2+10k-2) \right\} + O(n^{-3/2}) \end{aligned}$$

を得る。ただし, $\Gamma(t)$ はガンマ関数。

もう一つの離散修正項である D_3 の計算は大変複雑である。 $D_1 + D_2 + D_3$ によって誤差項が $O(n^{-3/2})$ まで近似できるが, $D_3 = O(n^{-1})$ であるから, ここでは, Siotani and Fujikoshi⁵⁾ と同様の考え方により J_1 の分布 $F_E(t)$ の漸近近似として

$$F_Y(t) = D_1 + \hat{D}_2$$

を提案する。ここに, D_1, \hat{D}_2 はそれぞれ (4.3) と (4.4) で与えられる式である。

5. J_1 の分布の漸近近似: F_R

検定統計量 T_k の分布に対して、Cressie and Read²⁾が与えた近似式と同様の方法によって、 J_1 の分布 $F_E(t)$ のもう一つの漸近近似 F_R を求める。

$$J_1(Y) = \sum_{i=1}^k \frac{Y_i^2}{p_i} - \frac{1}{2\sqrt{n}} \sum_{i=1}^k \frac{Y_i^3}{p_i^2} + \frac{7}{24n} \sum_{i=1}^k \frac{Y_i^4}{p_i^3} + O(n^{-3/2})$$

より J_1 の平均と分散は

$$E\{J_1\} = k - 1 + \frac{m}{n} + O(n^{-3/2}), \quad V\{J_1\} = 2(k-1) + \frac{v}{n} + O(n^{-3/2})$$

と展開できる。ただし、 $m = (3S - 2k + 1)/8$, $v = (11S - k^2 - 10k)/4$ である。

そこで

$$\delta = 1 + (v/2(k-1)n), \quad \gamma = (k-1)(1 - \sqrt{\delta}) + (m/n)$$

とおくと

$$E\{(J_1 - \gamma)/\sqrt{\delta}\} = k - 1 + O(n^{-3/2}), \quad V\{(J_1 - \gamma)/\sqrt{\delta}\} = 2(k-1) + O(n^{-3/2})$$

となる。これより、 J_1 の分布 $F_E(t)$ の漸近近似として

$$F_R(t) \equiv F_{\chi(r)}((t - \gamma)/\sqrt{\delta})$$

が得られる。

6. J_α の分布の漸近近似: G_Y, G_R

$J_\alpha(\alpha > 0, \alpha \neq 1)$ に対しては一般の場合について漸近展開を求めるのは困難であるため、分布の一様性を検定する場合について考える。すなわち、帰無仮説は

$$H_0: \pi = p = \left(\frac{1}{k}, \dots, \frac{1}{k}\right),$$

である。

$$\text{変数変換 } Y_i = (X_i - n/k)/\sqrt{n}, \quad i = 1, \dots, k$$

を行うと検定統計量は

$$J_\alpha(Y) = \frac{4n}{\alpha-1} \sum_{i=1}^k \left\{ \frac{1}{k^\alpha} + \left(\frac{1}{k} + \frac{Y_i}{\sqrt{n}}\right)^\alpha - 2\left(\frac{1}{k} + \frac{Y_i}{2\sqrt{n}}\right)^\alpha \right\}$$

となる。

$B_\alpha = \{y = (y_1, \dots, y_r) : J_\alpha(y) < t\}$ とおくと J_α の Exact な分布関数は

$G_E(t) = \Pr\{J_\alpha < t\} = \Pr\{Y \in B_\alpha\}$ となる。

$$1 < \alpha \leq 2, \quad \text{または} \quad 0 < \alpha \leq \left[\log \frac{2k^2}{(k+1)^2} \right] / \left[\log \frac{2k}{k+1} \right]$$

のとき、 B_α は常に有界な extended convex set になるので、 J_1 のときと同様に Yarnold⁸⁾ の定理を使って、帰無仮説のもとでの J_α の Exact な分布の近似 $D_1^f + D_2^g$ を求めることができる。

(3.4) 式と同様の変換を行い、テーラー展開すると

$$J_\alpha(H^{-1}z) = \frac{\alpha}{k^{\alpha-1}} z' z + \frac{\alpha(\alpha-2)}{2\sqrt{n} k^{\alpha-3/2}} \sum_{i=1}^k (a_i' z)^3 + \frac{7\alpha(\alpha-2)(\alpha-3)}{48nk^{\alpha-2}} \sum_{i=1}^k (a_i' z)^4 + O(n^{-3/2})$$

を得る。これより, J_α の特性関数を求め, $(1-iat)^{-p}$ はガンマ分布

$$\Gamma(p, a) = \frac{1}{a\Gamma(p)} \left(\frac{x}{a}\right)^{p-1} \exp\left(-\frac{x}{a}\right)$$

の特性関数であることに注意すると, Df は次のように得られる。

$$\begin{aligned} Df &= G_{\Gamma(r/2)}(t) + \frac{1}{96n} \{c_0 G_{\Gamma(r/2)}(t) + c_1 G_{\Gamma(r/2+1)}(t) + c_2 G_{\Gamma(r/2+2)}(t) + c_3 G_{\Gamma(r/2+3)}(t)\} \\ &\quad + O(n^{-3/2}) \end{aligned} \quad (6.1)$$

ただし

$G_{\Gamma(f)}(t)$: ガンマ分布 $\Gamma(f, a)$ の分布関数, $a=2\alpha/k^{\alpha-1}$,

$$\begin{aligned} c_0 &= 8(1-k^2), \quad c_1 = 3\{-k^2(\alpha^2-11\alpha+10) - 2k(2\alpha^2-\alpha-2) + (5\alpha^2-13\alpha+6)\}, \\ c_2 &= 3\{-k^2(5\alpha^2-5\alpha-2) + 2k(11\alpha^2-25\alpha+14) - (17\alpha^2-45\alpha+30)\}, \\ c_3 &= 2(k-1)(k-2)(9\alpha^2-24\alpha+16). \end{aligned}$$

次に, Dg を求める。 $J_\alpha(y)=t$ のとき

$$\phi(y) = (2\pi)^{-r/2} |\Omega|^{-1/2} \exp(-t/a) + O(n^{-1/2})$$

となるから

$$[S_1(\sqrt{n}y_c + np_c)\phi(y)]_{\lambda_c(y^*)}^{\theta_c(y^*)}$$

をその漸近近似である

$$(2\pi)^{-r/2} |\Omega|^{-1/2} \exp(-t/a) [S_1(\sqrt{n}y_c + np_c)]_{\lambda_c(y^*)}^{\theta_c(y^*)}$$

によって近似する。すると, J_1 のときと同様にして

$$\widehat{Dg} = \{[N_\alpha(t) - n^{r/2} V_\alpha(t)] \exp(-t/a)\} / \{[(2\pi n)^r k^{-k}]^{1/2}\}$$

を得る。ここで

$N_\alpha(t) = \#\{y: y \in M \text{ and } y \in B_\alpha\}$, $V_\alpha(t)$: B_α の体積である。

$V_\alpha(t)$ を求めるために $J_\alpha(H^{-1}z) = (\alpha/k^{\alpha-1})u'u$ となる変換 $u \rightarrow z$ を考えると

$$\begin{aligned} z &= u - \frac{\sqrt{k}(\alpha-2)}{4\sqrt{n}} \sum_{i=1}^k (a_i'u)^2 a_i + \frac{1}{96n} \{k(8\alpha^2-25\alpha+18) \sum_{i=1}^k (a_i'u)^3 a_i - 15(\alpha-2)^2 (u'u)u\} \\ &\quad + O(n^{-3/2}) \end{aligned}$$

と表される。これにより,

$$\begin{aligned} V_\alpha(t) &= |\Omega|^{1/2} \int \cdots \int_{u'u < \frac{2t}{\alpha}} \left| \frac{\partial z}{\partial u} \right| du \\ &= |\Omega|^{1/2} \int \cdots \int_{u'u < \frac{2t}{\alpha}} \left[1 - \frac{\alpha-2}{2\sqrt{n}} \left(\sqrt{k} - \frac{1}{\sqrt{k}} \right) \sum_{i=1}^k (a_i'u) + \frac{1}{32n} \{4(\alpha-2)^2 (k-2) \left(\sum_{i=1}^k (a_i'u) \right)^2 \right. \\ &\quad \left. + (\alpha-2)(\alpha-k-5) + 9k+3\} (u'u) \right] + O(n^{-3/2}) du \\ &= \frac{\{(2\pi t/a)^r k^{-k}\}^2}{\Gamma(r/2+1)} \left[1 + \frac{tk^{\alpha-1}(k-1)(\alpha-2)\{\alpha(-k-5) + 9k+3\}}{32n\alpha(k+1)} \right] + O(n^{-3/2}) \end{aligned}$$

(6.3)

が得られる。従って、 J_α の分布 $G_E(t)$ の一つの漸近近似として

$$G_Y(t) = D_f^g + \bar{D}_g^g$$

が得られる。ここに D_f^g , \bar{D}_g^g はそれぞれ (6.1), (6.2) で与えられた式である。

次に

$$J_\alpha(Y) = \frac{\alpha}{k^{\alpha-2}} \sum_{i=1}^k Y_i^2 + \frac{\alpha(\alpha-2)}{2\sqrt{n}k^{\alpha-3}} \sum_{i=1}^k Y_i^3 + \frac{7\alpha(\alpha-2)(\alpha-3)}{48nk^{\alpha-4}} \sum_{i=1}^k Y_i^4 + O(n^{-3/2})$$

より

$$E\{J_\alpha\} = \frac{\alpha(k-1)}{k^{\alpha-1}} + \frac{m_\alpha}{n} + O(n^{-3/2}), \quad V\{J_\alpha\} = \frac{2\alpha^2(k-1)}{k^{2\alpha-2}} + \frac{v_\alpha}{n} + O(n^{-3/2})$$

を得る。ただし

$$m_\alpha = \frac{\alpha(k-1)}{16k^{\alpha-1}} (\alpha-2) \{k(7\alpha-13) - 7\alpha + 5\},$$

$$v_\alpha = \frac{\alpha^2(k-1)}{2k^{2\alpha-2}} \{5k(\alpha-2)(2\alpha-3) - 13\alpha^2 + 35\alpha - 22\}.$$

そこで

$$\delta_\alpha = 1 + v_\alpha / \left[n \cdot \frac{2\alpha^2(k-1)}{k^{2\alpha-2}} \right], \quad \gamma_\alpha = \frac{\alpha(k-1)}{k^{\alpha-1}} (1 - \sqrt{\delta_\alpha}) + \frac{m_\alpha}{n}$$

とおくと J_α の分布の漸近近似として

$$G_R(t) \equiv G_{\Gamma(r/2)}\{(t - \gamma_\alpha) / \sqrt{\delta_\alpha}\}$$

が得られる。

7. 各近似の比較

これまでに求めた近似式が検定統計量の Exact な分布をどの程度よく近似しているかということ調べるために、3項分布の場合について数値計算による比較を行った。

図1と図2は、 $p=(0.2, 0.4, 0.4)'$ とした帰無仮説のもとでの検定統計量 J_1 についてのグラフである。 J_1 の Exact な分布 F_E に対する近似として、極限分布である χ^2 分布 $F_{\chi(r)}$ 、離散修正項をもつ近似 F_Y 、モーメントに基づいて極限分布を修正した近似 F_R の3つの近似法を比較している。図1は $n=10$ の場合の分布関数とその近似を表している。なお、補助線として、確率 0.95 を示す線と、 χ^2 分布における5%点、 $\chi_{0.05}^2(r)$ を示す線がかかっている。グラフからわかるように $F_Y(t)$ は必ずしも単調増加関数ではないが、全体として、 $F_E(t)$ をよく近似しているといえる。図2は n の変化にともなう各近似の変化の様子を見るために、 $t = \chi_{0.05}^2(r)$ における各関数の値を表したグラフである。 n の値が大きくなるに従って、 F_Y による近似がかなりよくなっていく様子がわかる。 p をいろいろに変えても、図1、図2とほぼ同様の結果が得られる。また、 $\min_{1 \leq j \leq 3} p_j$ の値が大きいときほど F_Y は良い近似を与えている。

図3は、 $\alpha=1.5$ とした検定統計量 $J_{1.5}$ の Exact な分布関数 G_E と、3つの近似 G_Γ , G_Y , G_R を表した $n=10$ の場合のグラフである。補助線として確率 0.95 を示す線と、漸近分布である G_Γ の5%点を示す線もかかっている。

α の値を一つ決めると検定統計量 J_α が一つ定まる。そこで、 n の値を固定したときの各統計量に対する近似の様子を表したのが図4である。各 α に対応する漸近ガンマ分布の5%点における

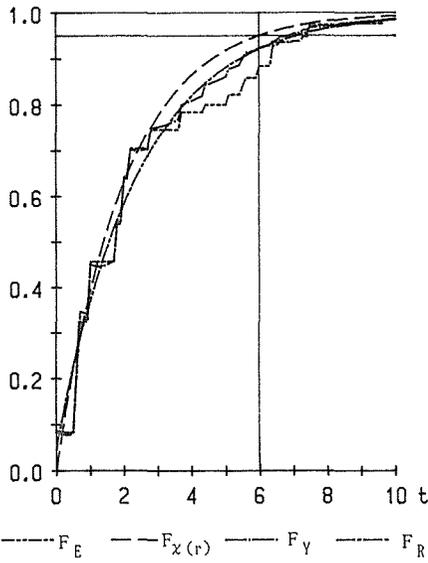


図1 J_1 の Exact な分布関数 $F_E(t)$ と、その近似 $F_{X(r)}(t)$, $F_Y(t)$, $F_R(t)$.
 X の分布は $p = (0.2, 0.4, 0.4)'$, $n = 10$ の 3 項分布.

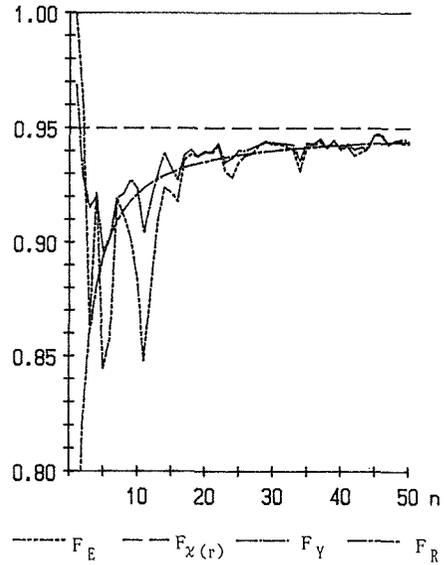


図2 n の変化にともなう J_1 の各近似の変化の様子.
 $t = \chi_{0.05}^2(r)$ における各関数の値の変化.
 $p = ((0.2, 0.4, 0.4)'$.

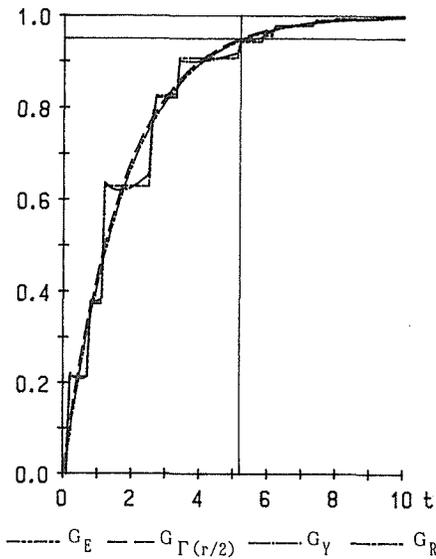


図3 $J_{1.5}$ の Exact な分布関数 $G_E(t)$ と、その近似 $G_{\Gamma(r/2)}(t)$, $G_Y(t)$, $G_R(t)$.
 X の分布は $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})'$, $n = 10$ の 3 項分布.

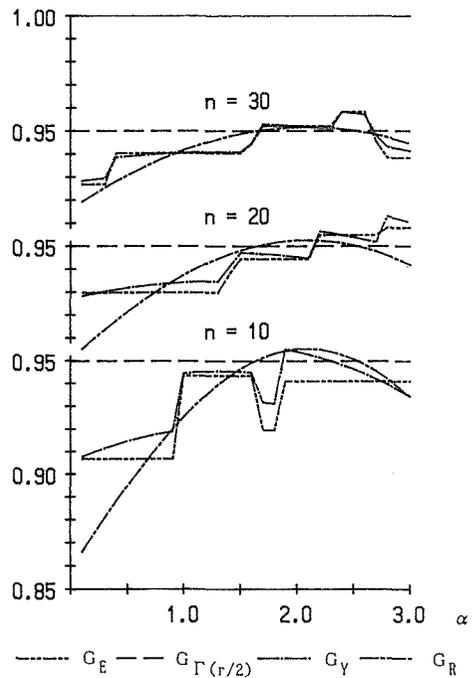


図4 $t = \Gamma_{0.05}(r/2, a)$ における J_α の各近似の様子.
 $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})'$.

J_α の Exact な確率と、その近似値を表したグラフである。下から順に $n=10, 20, 30$ の場合のグラフが示されている。 n の値が大きくなるにしたがって、離散修正項の効果がよく現れ、 G_Y による近似は G_E に対して、この図で示された α については、ほぼ一様によい近似となっている。また、 $n=30$ になると、 $\alpha=2$ の付近では G_L や G_R もかなりよい近似を示している。

F_Y と G_Y は Yarnold の定理における一つの離散修正項 D_2 の漸近近似である \hat{D}_2 を用いた $D_1 + \hat{D}_2$ の形の近似式であるにすぎないが、検定統計量 J_α ($\alpha=1$ も含めて) の Exact な分布の漸近近似として有効であると思われる。

また、統計量のモーメントに基づいて極限分布を修正することによって得られた近似式 F_R は、 F_Y に比べて計算量が少ないという利点があり、検定の棄却限界点の付近ではよい近似を与えられ、と思われる。

参 考 文 献

- 1) Burbea, J. and Rao, C. R. : IEEE Trans. Information Theory, **28** (1982), p. 489-495
- 2) Cressie, N. and Read, T. R. C. : J. R. Statist. Soc. B, **46** (1984), p. 440-464
- 3) Havrda, J. and Charvát, F. : Kybernetika, **3** (1967), p. 30-35
- 4) Read, T. R. C. : Ann. Inst. Statist. Math., **36** (1984), p. 59-69
- 5) Siotani, M. and Fujikoshi, Y. : Hiroshima Math. J., **14** (1984), p. 115-124
- 6) 種市信裕 : 第 14 回日本行動計量学会大会発表論文抄録集 (1986), p. 85-86
- 7) 種市信裕 : 第 55 回日本統計学会講演報告集 (1987), p. 49-50
- 8) Yarnold, J. K. : Ann. Math. Statist., **43** (1972), p. 1566-1580