

カイ 2 乗確率変数の重み付き和の分布について

鳥越 規央 (東海大・理)

1. はじめに

カイ 2 乗確率変数の重み付き和の形をした確率変数 $X = \sum_{i=1}^m c_i Y_i$ の分布のパーセント点の近似式による数値計算を考える. ここで $\{Y_i\}$ は自由度 d のカイ 2 乗分布に従う独立な確率変数の列, $\{c_i\}$ は $c_1 > c_2 > \cdots > c_m > 0$ を満たす定数の列である. このような形を持つ統計量はきわめて多い. 適合度検定に用いられる通常のカイ 2 乗統計量はもちろんのこと, [H86]においては一元配置モデルにおける $c_i = a/\{i(i+1)\}$, $m = a-1, d=1$ と $c_i = 2a/\{i(i+1)(i+2)(i+3)\}$, $m = a-1, d=1$ の累積カイ 2 乗統計量が議論され (ただし a は処理数), [SWT85]においてはグラフ表現による適合度検定における $c_i = 2/\{(4i^2-1)\pi\}$, $m = \infty, d=2$ となる統計量が提案されている. また経験分布関数に基づく適合度検定の統計量では様ざまの形の統計量が提案されている ([St74]). なお [Sr88]では自由度が 1 であるときの分布関数

$$\begin{aligned} & \int_0^1 \left[\frac{(-1)^{(m+1)/2} (e^{-x \tan^2(\frac{1}{2}\pi s) - \frac{x}{c_m}} - 1)}{\sqrt{c_m} (\sin^2(\frac{1}{2}\pi s) + \cos^2(\frac{1}{2}\pi s)/c_m)} \prod_{k=1}^{m-1} \{c_k \tan^2(\frac{1}{2}\pi s) - 1 + \frac{c_k}{c_m}\}^{-1/2} \right. \\ & + \frac{\sqrt{2}}{\pi \sqrt{1 - \frac{1}{2}t^2}} \sum_{j=1}^{m-1} \frac{(-1)^{(j+1)/2}}{\sqrt{c_j c_{j+1}}} \\ & \left. \left\{ \frac{e^{-h_j(\frac{1}{2}t^2)x} - 1}{h_j(\frac{1}{2}t^2)} \prod_j^- \frac{1}{\sqrt{|c_k h_j(\frac{1}{2}t^2) - 1|}} + \right. \right. \\ & \left. \left. \frac{e^{-h_j(1 - \frac{1}{2}t^2)x} - 1}{h_j(1 - \frac{1}{2}t^2)} \prod_j^- \frac{1}{\sqrt{|c_k h_j(1 - \frac{1}{2}t^2) - 1|}} \right\} \right] dt \end{aligned}$$

や自由度が偶数の場合の分布関数

$$\lim_{n \rightarrow \infty} -i \sum_{j=1}^n \frac{\lim_{z \rightarrow -1/c_j} \frac{d^{d-1}}{dz^{d-1}} \left(z + \frac{i}{c_j} \right) \frac{e^{-itz} - 1}{-it} \prod_{j=1}^n (1 - ic_j t)^{-d}}{(d-1)!}$$

が求められているが、数値計算の面においてかなり複雑な式となっており、それよりパーセント点を表計算ソフトの組み込み関数を用いて求めるようにすることは容易でない。ただし Π_j^- は $k = j$ と $j+1$ を除く k での積を表す。

本論では、一元配置モデルにおける累積カイ 2 乗統計量の形に絞り、キュムラントを等置した中心カイ 2 乗分布の定数倍の分布で近似し、カイ統計量やカイ 2 乗統計量の Cornish-Fisher 展開やカイ 2 乗統計量の Wilson-Hilferty 近似を用いて新しい近似式を求める。それを数値計算によって従来の近似式と比較検討し、新しい近似式の優位性を示す。

2. 累積カイ 2 乗統計量について

一元配置のモデル

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \dots, a; j = 1, \dots, r \quad \epsilon_{ij} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$$

で、水準に自然な順序がある場合、帰無仮説 $H_0: \mu_1 = \mu_2 = \dots = \mu_a$ の対立仮説を

$$H_1: \mu_1 \leq \mu_2 \leq \dots \leq \mu_a \quad (H'_1: \mu_1 \geq \mu_2 \geq \dots \geq \mu_a)$$

や

$$H_2: \mu_1 \leq \mu_2 \leq \dots \leq \mu_a \quad \text{or} \quad \mu_1 \geq \mu_2 \geq \dots \geq \mu_a$$

と限定することがある。この場合の検定方式で、実用的かつ比較的検出力の高いものとして累積カイ 2 乗法があげられる[H83]。 $\bar{y}_{i\cdot} = \sum_j y_{ij}/r$ とおいたとき $a-1$ 個の細分された統計量

$$t_i = \frac{1}{i}(\bar{y}_{1\cdot} + \cdots + \bar{y}_{i\cdot}) - \frac{1}{a-i}(\bar{y}_{i+1\cdot} + \cdots + \bar{y}_{a\cdot})$$

($i = 1, 2, \dots, a-1$)

は H_0 の細分

$$H_{i0} : \frac{1}{i}(\mu_1 + \cdots + \mu_i) = \frac{1}{a-i}(\mu_{i+1} + \cdots + \mu_a)$$

を H_1 (H'_1) の細分

$$H_{i1} : \frac{1}{i}(\mu_1 + \cdots + \mu_i) \leq \frac{1}{a-i}(\mu_{i+1} + \cdots + \mu_a)$$

$$(H'_{i1} : \frac{1}{i}(\mu_1 + \cdots + \mu_i) \geq \frac{1}{a-i}(\mu_{i+1} + \cdots + \mu_a))$$

に対して検定する統計量といえる.

$$V(t_i) = \frac{a}{i(a-i)} \frac{\sigma^2}{r}$$

より検定は

$$t_i^* = \frac{\sqrt{\frac{ri(a-i)}{a}}}{\hat{\sigma}} t_i$$

が非心t分布 $t(a(r-1), \gamma_i)$ に従うことを利用する. ただし

$$\hat{\sigma} = \frac{\sum \sum (y_{ij} - \bar{y}_{i\cdot})^2}{a(r-i)},$$

$$\gamma_i = \frac{1}{\hat{\sigma}} \sqrt{\frac{ri(a-i)}{a}} \left(\frac{\mu_1 + \cdots + \mu_i}{i} - \frac{\mu_{i+1} + \cdots + \mu_a}{a-i} \right)$$

である. ところで t_i は互いに独立にならない. そこで提案されている統計量として

$$\sum_{i=1}^{a-1} \frac{ri(a-i)}{a} t_i^2$$

という2乗和を考える. ここで H_0 の下で, t_i の適当な直交変換を施すことによって独立なカイ2乗分布の正係数の線形結合

$$\chi^{*2} = \frac{a}{1 \cdot 2} \chi_{(1)}^2 + \frac{a}{2 \cdot 3} \chi_{(2)}^2 + \cdots + \frac{a}{(a-1)a} \chi_{(a-1)}^2$$

を含む統計量を得る. ここで $\{\chi_{(i)}^2\}$ ($i = 1, \dots, a-1$) は互いに独立な自由度1のカイ2乗分布に従う確率変数となることが知られている ([H86]). この分布を自由度 ν の中心カイ2乗分布の定数倍 $c\chi^2(\nu)$ で近似することを考える. $c\chi^2(\nu)$ の2次までのキュムラントが $c\nu$, $2c^2\nu$ であるのに対し, χ^{*2} の1次, 2次キュムラントは $\kappa_1(\chi^{*2}) = a-1$, $\kappa_2(\chi^{*2}) = 4a^2 \sum_{i=1}^a i^{-2} - 2(3a^2 - 2a + 1)$ であるので

$$\nu = \frac{a-1}{c} \quad c = \frac{2a^2 \sum_{i=1}^a i^{-2} - (3a^2 - 2a + 1)}{a-1}$$

を得る. よって x_α を χ^* の上側100 α パーセント点とすると,

$$\begin{aligned} 1 - \alpha &= P(\chi^{*2} < x_\alpha) \\ &= P\left(c\chi^2\left(\frac{a-1}{c}\right) < x_\alpha\right) \\ &= P\left(\chi^2\left(\frac{a-1}{c}\right) < \frac{x_\alpha}{c}\right) \end{aligned} \quad (2.1)$$

となる. これを利用して近似式を導出する.

3. カイ2乗確率変数の重み付き和の分布のパーセント点の近似式

3.1. カイ統計量を利用した近似式

(2.1) より

$$1 - \alpha = P\left(\chi^2\left(\frac{a-1}{c}\right) < \frac{x_\alpha}{c}\right)$$

$$\begin{aligned}
&= P\left(\frac{\chi^2\left(\frac{a-1}{c}\right)}{\frac{a-1}{c}} < \frac{x_\alpha/c}{(a-1)/c}\right) \\
&= P\left(\frac{c}{a-1} \chi^2\left(\frac{a-1}{c}\right) < \frac{x_\alpha}{a-1}\right) \\
&= P\left(\sqrt{\frac{c}{a-1}} \chi^2\left(\frac{a-1}{c}\right) < \sqrt{\frac{x_\alpha}{a-1}}\right)
\end{aligned}$$

となる. ここで $S_\nu := \sqrt{\chi^2(\nu)/\nu}$ と定義すると $\sqrt{\frac{c}{a-1}} \chi^2\left(\frac{a-1}{c}\right) = S_{(a-1)/c}$ となる. また S_ν のモーメントは

$$E(S_\nu) = \sqrt{\frac{\nu}{2}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} =: b_\nu \quad V(S_\nu) = 1 - b_\nu^2$$

より ([T96]), 規準化して

$$\begin{aligned}
1 - \alpha &= P\left(S_{\frac{a-1}{c}} < \sqrt{\frac{x_\alpha}{a-1}}\right) \\
&= P\left(\frac{S_{\frac{a-1}{c}} - b_{\frac{a-1}{c}}}{\sqrt{1 - b_{\frac{a-1}{c}}^2}} < \frac{\sqrt{x_\alpha/(a-1)} - b_{\frac{a-1}{c}}}{\sqrt{1 - b_{\frac{a-1}{c}}^2}}\right)
\end{aligned}$$

となる. よって

$$Y := \frac{S_{\frac{a-1}{c}} - b_{\frac{a-1}{c}}}{\sqrt{1 - b_{\frac{a-1}{c}}^2}}$$

の上側 100α パーセント点についての Cornish-Fisher 展開を得るため, Y の 3 次, 4 次キュムラントを求める.

補題 3.1. Y の 3 次, 4 次キュムラントをそれぞれ $\kappa_3(Y), \kappa_4(Y)$ とするとき, 次の近似式が成り立つ.

$$\begin{aligned}
\kappa_3(Y) &= -\frac{b_{\frac{a-1}{c}}}{(1 - b_{\frac{a-1}{c}}^2)^{3/2}} \left\{ 2\left(1 - b_{\frac{a-1}{c}}^2\right) - \frac{c}{a-1} \right\} \\
\kappa_4(Y) &= \frac{4(a+c-1)(1 - b_{\frac{a-1}{c}}^2) - 2c}{(a-1)(1 - b_{\frac{a-1}{c}}^2)^2} - 6
\end{aligned}$$

<証明の概略> $S_\nu - b_\nu$ の 3 次, 4 次キユムラントはそれぞれ

$$\begin{aligned}\kappa_3(S_\nu - b_\nu) &= E[(S_\nu - b_\nu)^3] \\ &= E[S_\nu^3] - 3b_\nu E[S_\nu^2] + 3b_\nu^2 E[S_\nu] - b_\nu^3 \\ &= \left(1 + \frac{1}{\nu}\right)b_\nu - 3b_\nu + 3b_\nu^2 b_\nu - b_\nu^3 \\ &= -b_\nu \left\{ 2(1 - b_\nu^2) - \frac{1}{\nu} \right\}\end{aligned}$$

$$\begin{aligned}\kappa_4(S_\nu - b_\nu) &= E[(S_\nu - b_\nu)^4] - 3(E[(S_\nu - b_\nu)^2])^2 \\ &= E[S_\nu^4] - 4b_\nu E[S_\nu^3] + 6b_\nu^2 E[S_\nu^2] - 4b_\nu^3 E[S_\nu] + b_\nu^4 \\ &\quad - 3(E[S_\nu^2] - 2b_\nu E[S_\nu] + b_\nu^2)^2 \\ &= 1 + \frac{2}{\nu} - 4\left(1 + \frac{1}{\nu}\right)b_\nu^2 + 6b_\nu^2 - 4b_\nu^3 b_\nu + b_\nu^4 - 3(1 - 2b_\nu^2 + b_\nu^2)^2 \\ &= \frac{2}{\nu}(1 - 2b_\nu^2) + (1 - b_\nu^2)(1 + 3b_\nu^2) - 3(1 - b_\nu^2)^2 \\ &= (1 - b_\nu^2)\{4 - 6(1 - b_\nu^2)\} + \frac{4}{\nu}(1 - b_\nu^2) - \frac{2}{\nu}\end{aligned}$$

であり,

$$\kappa_i(Y) = \frac{1}{(1 - b_{\frac{a-1}{c}}^2)^{i/2}} \kappa_i(S_{\frac{a-1}{c}} - b_{\frac{a-1}{c}}), \quad (i = 3, 4)$$

であることより得られる.

定理 3.1. 累積カイ 2 乗統計量が従う分布の上側 100α パーセント点を x_α とするとき、次が成り立つ.

$$x_\alpha = (a - 1) \left[c + u_\alpha \sqrt{1 - b_{\frac{a-1}{c}}^2} \right]$$

$$\begin{aligned}
& + \frac{1}{6} b_{\frac{a-1}{c}} \left\{ \frac{c}{(a-1)(1-b_{\frac{a-1}{c}}^2)} - 2 \right\} (u_\alpha^2 - 1) \\
& + \frac{u_\alpha^3 - 3u_\alpha}{24\sqrt{1-b_{\frac{a-1}{c}}^2}} \left\{ \frac{4(a+c-1)(1-b_{\frac{a-1}{c}}^2) - 2c}{(a-1)(1-b_{\frac{a-1}{c}}^2)^2} - 6 \right\} \\
& - \frac{b_{\frac{a-1}{c}}^2}{36(1-b_{\frac{a-1}{c}}^2)^{5/2}} \left\{ 2\left(1-b_{\frac{a-1}{c}}^2\right) - \frac{c}{a-1} \right\}^2 (2u_\alpha^3 - 5u_\alpha) \Bigg]^2 \quad (3.1)
\end{aligned}$$

ただし u_α は標準正規分布の上側 100α パーセント点である.

<証明の概略> 補題3.1 と Cornish-Fisher 展開より

$$\begin{aligned}
\frac{\sqrt{x_\alpha/(a-1)} - b_{\frac{a-1}{c}}}{\sqrt{1-b_{\frac{a-1}{c}}^2}} &= u_\alpha + \frac{\kappa_3(Y)}{6} (u_\alpha^2 - 1) + \frac{\kappa_4(Y)}{24} (u_\alpha^3 - 3u_\alpha) \\
&\quad - \frac{\{\kappa_3(Y)\}^2}{36} (2u_\alpha^3 - 5u_\alpha) + \cdots
\end{aligned}$$

上式を x_α について解くと (3.1) を得る.

3.2. カイ 2 乗統計量の Cornish-Fisher 展開を利用した近似式

(2.1) より

$$\begin{aligned}
1 - \alpha &= P\left(\chi^2\left(\frac{a-1}{c}\right) < \frac{x_\alpha}{c}\right) \\
&= P\left(\frac{\chi^2\left(\frac{a-1}{c}\right) - \frac{a-1}{c}}{\sqrt{\frac{2(a-1)}{c}}} < \frac{\frac{x_\alpha}{c} - \frac{a-1}{c}}{\sqrt{\frac{2(a-1)}{c}}}\right)
\end{aligned}$$

を得る. $\chi^2\left(\frac{a-1}{c}\right)$ を規準化することによって得られる統計量を

$$Z = \frac{\chi^2\left(\frac{a-1}{c}\right) - \frac{a-1}{c}}{\sqrt{\frac{2(a-1)}{c}}}$$

とおくと, Z の Cornish-Fisher展開を用いて新しい近似式を得る([Sb81])。

定理 3.2. 累積カイ 2 乗統計量が従う分布の上側 100α パーセント点を x_α とするとき、次の近似式が成り立つ。

$$\begin{aligned} x_\alpha = c \left\{ \frac{a-1}{c} + u_\alpha \sqrt{\frac{2(a-1)}{c}} + \frac{2}{3}(u_\alpha^2 - 1) \right. \\ \left. + \frac{u_\alpha^3 - 7u_\alpha}{9} \sqrt{\frac{c}{2(a-1)}} - \frac{c(3u_\alpha^4 + 7u_\alpha^2 - 16)}{405(a-1)} \right\} \quad (3.2) \end{aligned}$$

3.2. カイ 2 乗統計量の Wilson-Hilferty 近似式

$$W = \sqrt{\frac{9\nu}{2}} \left\{ \left(\frac{\chi^2(\nu)}{\nu} \right)^{\frac{1}{3}} - 1 + \frac{2}{9\nu} \right\}$$

の正規近似を用いることによって

$$\sqrt{\frac{9\nu}{2}} \left\{ \left(\frac{x_\alpha}{c\nu} \right)^{\frac{1}{3}} - 1 + \frac{2}{9\nu} \right\} \doteq u_\alpha$$

を得, x_α についてまとめると

$$x_\alpha \doteq c\nu \left(1 - \frac{2}{9\nu} + u_\alpha \sqrt{\frac{2}{9\nu}} \right)^3$$

となる([Sb81])。よって, 以下の新しい近似式を得る。

定理 3.3. 累積カイ 2 乗統計量が従う分布の上側 100α パーセント点を x_α とするとき、次の近似式が成り立つ.

$$x_\alpha \doteq (a-1) \left(1 - \frac{2c}{9(a-1)} + u_\alpha \sqrt{\frac{2c}{9(a-1)}} \right)^3 \quad (3.3)$$

4. 数値計算による比較検討

パーセント点の近似式(3.1), (3.2), (3.3) について処理数 a が 3 ~ 11 のときの上側 5 % 点, 10 % 点, そして 1 % 点の値を計算し [Sr88] との数値と比較した. その結果を図 4.1, 4.2, 4.3 に示した.

参考文献

- [H83] 広津千尋.(1983). 統計的データ解析. 日本規格協会
- [H86] Hirotsu, C.(1986). Cumulative chi-squared statistic as a tool for testing goodness of fit. *Biometrika*, **73**, 165-173.
- [Sb81] 柴田義貞 (1981). 正規分布. 東京大学出版会.
- [Sr88] 白旗慎吾 (1988). カイ二乗確率変数の重み付き和の分布関数の計算. 計算機統計学, **73**, 165-173.
- [St74] Stephens, M. A.(1974). EDF statistics for goodness of fit and some comparisons. *J. Amer. Statist. Assoc.*, **69**, 730-737.
- [SWT85] Shirahata, S., Wakimoto, K. and Tarumi, T.(1985). A goodness of fit test based on the linked line chart. *Aust. J. Statist.*, **27**, 163-171.
- [To96] Torigoe, N.(1996). Approximations to non-central χ^2 and F distributions. *J. Japan Statist. Soc.*, **26**, 145-159.

図4.1 χ^2 の上側5%点の値における白旗の近似式と他の近似式との誤差

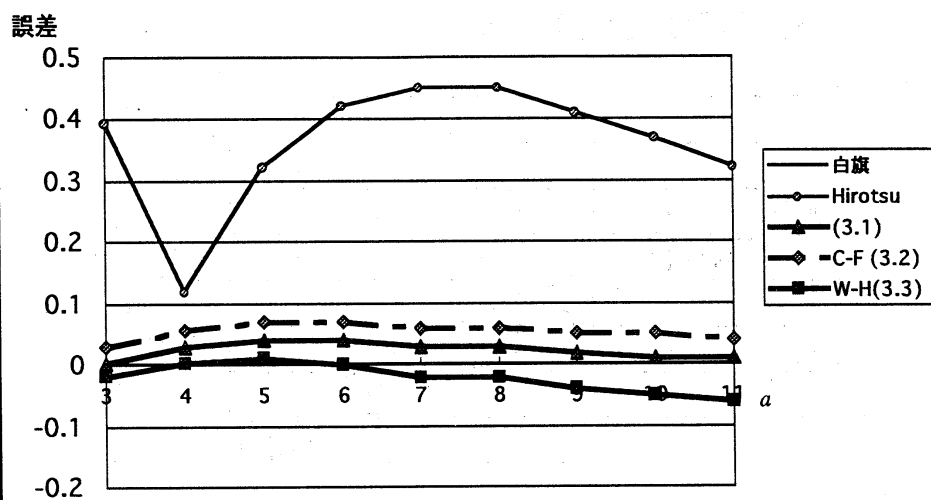


図4.2 χ^2 の上側10%点の値における白旗の近似式と他の近似式との誤差

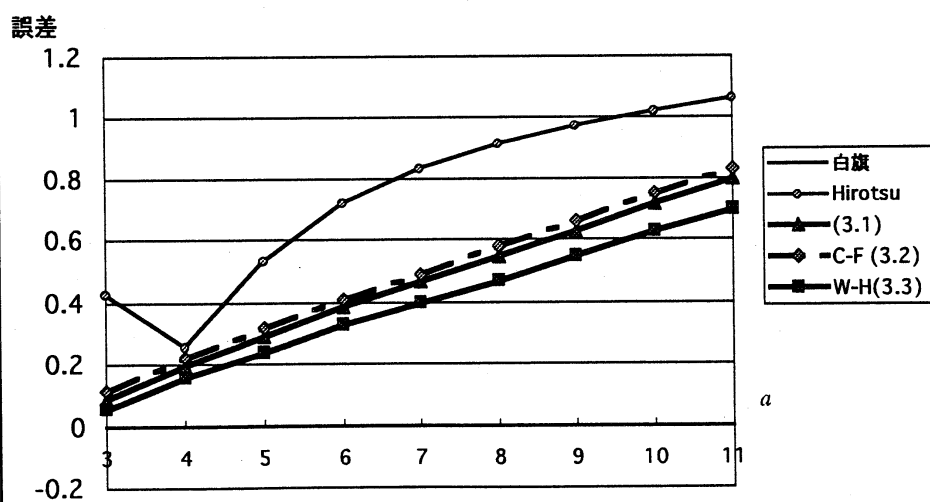


図4.3 χ^2 の上側1%点の値における白旗の近似式と他の近似式との誤差

