

2重非心 F 分布の近似法についての研究

M2009MM012 前廣 芳孝

指導教員：松田 眞一

1 はじめに

2重非心 F 分布の確率密度関数はとても複雑な形をしており、直接パーセント点を求めることは困難である。このため、堀井 [2] は Tiku [5] や鳥越 [7] によって報告された2重非心 F 分布の近似法を用いてパーセント点近似によるパーセント点の導出を試みた。しかし、実際のデータをふまえた自由度と非心度を用いた場合、これら2つの近似手法でのパーセント点近似は自由度が小さい場合に大きな誤差が伴ってしまう。または、プログラミングでの計算途中にエラーとなり近似値を出せない場合が多く存在することがわかった。このため、これらの近似法は実用的ではなくモンテカルロ法で十分であることが報告された。しかし、モンテカルロ法は多くの計算と時間を必要とする。そこで、既存の近似法を元に自由度が小さい場合でも実用的な改良ができないかを研究の目的として取り上げた。

2 2重非心 F 分布とは

2重非心 F 分布 $F''(\nu_1, \nu_2, \lambda_1, \lambda_2)$ は以下で定義される確率変数である。(鳥越 [7] 参照)

$$F''(\nu_1, \nu_2, \lambda_1, \lambda_2) := \left\{ \frac{\chi_{\nu_1}^2(\lambda_1)}{\nu_1} \right\} / \left\{ \frac{\chi_{\nu_2}^2(\lambda_2)}{\nu_2} \right\} \quad (1)$$

このとき、分母と分子の χ はそれぞれパラメータ(自由度, 非心度)が (ν_1, λ_1) と (ν_2, λ_2) の独立した非心カイ2乗分布に従う確率変数である。2重非心 F 分布の確率密度関数は以下のように定義されている。

$$p_{F''}(x; \nu_1, \nu_2, \lambda_1, \lambda_2) := \sum_{r=0}^{\infty} \sum_{t=0}^{\infty} (-1)^{r+t} \frac{(\lambda_1/2)^r}{r!} \frac{(\lambda_2/2)^t}{t!} \cdot \left[\sum_{i=0}^r \sum_{j=0}^t (-1)^{i+j} \binom{r}{i} \binom{t}{j} p_F(x; \frac{\nu_1}{2} + i, \frac{\nu_2}{2} + j) \right] \quad (2)$$

($0 < x < \infty; \nu_1, \nu_2 = 1, 2, \dots; \lambda_1, \lambda_2 > 0$)

ここでの $\binom{r}{i}$ は r 個のものから i 個選ぶ組み合わせである。また、 $p_F(x; \frac{\nu_1}{2} + i, \frac{\nu_2}{2} + j)$ は自由度 $(\nu_1 + 2i, \nu_2 + 2j)$ の中心 F 分布の密度関数であるので、以下のような関数となる。

$$p_F(x; \frac{\nu_1}{2} + i, \frac{\nu_2}{2} + j) = \frac{(\nu_1/\nu_2)^{\nu_1/2+i} x^{\nu_1/2+i-1}}{B(\nu_1/2 + i, \nu_2/2 + j) (1 + (\nu_1/\nu_2)x)^{(\nu_1+\nu_2)/2+i+j}} \quad (3)$$

ここでの $B(\cdot, \cdot)$ はベータ関数である。

2重非心 F 分布は分散分析の交互作用や誤差因子における F 検定の検出力を求めるとき、タグチメソッドにおける SN 比を検定する問題を考えるときに用いられる。

3 2重非心 F 分布の近似法

堀井 [2] では、Tiku [5] と鳥越 [7] による近似法を元にした2つのパーセント点近似法の評価を行った (Tiku 法、Torigoe 法)。この近似法の詳しい導出方法は堀井 [2] を参照されたい。ここでは、Mudholkar et al. [3] を元に堀井 [2] では評価されなかった2つの近似方法について詳しく述べる。

3.1 Tiku 法

$(F_{\nu_1, \nu_2, \lambda_1, \lambda_2} + \zeta)/\tau$ の分布の3次以下のモーメントを等置することによって自由度 (ν', ν_2) の中心 F 分布 $F(\nu', \nu_2)$ で近似する方法である。この手法を堀井 [2] では Tiku 法と呼んだ。最終的に以下の近似式が導き出される。

$$P\{F''_{\nu_1, \nu_2, \lambda_1, \lambda_2} < f\} \approx I_{\nu_0'}\left(\frac{1}{2}\nu_2, \frac{1}{2}\nu_0'\right) = 1 - \frac{1}{B(\frac{1}{2}\nu_2, \frac{1}{2}\nu_0')} \int_0^{\nu_0'} t^{\frac{1}{2}\nu_2-1} (1-t)^{\frac{1}{2}\nu_0'-1} dt \quad (4)$$

ただし $\nu_0' = 1 / \left(1 + \frac{\nu'}{\nu_2} \frac{f+\zeta}{\tau}\right)$ である。

3.2 Torigoe 法

近似式 (4) の左辺が F 分布で近似できること(カイ2乗分布でも表すことができる)を利用し、新たな統計量分布を得て、その分布を Cornish-Fisher 展開して新たな近似式を求めるものである。この手法を堀井 [2] では Torigoe 法と呼んだ。最終的に以下の近似式が導き出される。

$$-\frac{b_{\nu'} - \sqrt{f'_\alpha b_{\nu_2}}}{\sqrt{1 - b_{\nu'}^2 + f'_\alpha(1 - b_{\nu_2}^2)}} = u_\alpha + \frac{u_\alpha^2 - 1}{24\{1 - b_{\nu'}^2 + f'_\alpha(1 - b_{\nu_2}^2)\}} \cdot \left\{ \frac{1}{\nu'^2} + \frac{1}{\nu'^3} - f_\alpha'^{3/2} \left(\frac{1}{\nu_2^2} + \frac{1}{4\nu_2^3} \right) \right\} - \frac{2u_\alpha^3 - 5u_\alpha}{576\{1 - b_{\nu'}^2 + f'_\alpha(1 - b_{\nu_2}^2)\}^3} \cdot \left(\frac{1}{\nu'^2} - \frac{f_\alpha'^{3/2}}{\nu_2^2} \right)^2 \quad (5)$$

このとき u_α は $N(0, 1)$ の上側 100α パーセント点、 $f'_\alpha = (f_\alpha + \zeta)/\tau$ を示す。

3.3 MCL-E 法

Mudholkar et al. [3] の1つめの近似法は、エッジワース級数展開と Aty [1] によるキュムラント表現を組み合わせることによって標準正規変数の分布関数を用いた2重非心 F 分布の累積分布関数の近似式を導出するものである。

3.3.1 導出方法

エッジワース級数近似を展開するために2重非心 F 分布を以下のように置き換える。

$$P_r[F'' \leq c] = P_r[v = v_1 - c^{1/3}v_2 \leq 0] \quad (6)$$

このとき $v_1 = \{\chi_{\nu_1}^{\prime 2}(\lambda_1)/\nu_1\}^{1/3}$ 、 $v_2 = \{\chi_{\nu_2}^{\prime 2}(\lambda_2)/\nu_2\}^{1/3}$ を意味する。ここで Aty[1] の $\{\chi_{\nu}^{\prime 2}(\lambda)/(\nu + \lambda)\}$ のキムラントの表現を用いること、 v_1 と v_2 の独立性を用いることで以下のようなキムラント k_s ($v = v_1 - c^{1/3}v_2$ の $s = 1, 2, 3, 4$) の式を得ることができる。

$$\begin{aligned} k_1 &= (r_1/\nu_1)^{1/3}T_1(r_1, b_1) - (cr_2/\nu_2)^{1/3}T_1(r_2, b_2), \\ k_2 &= (r_1/\nu_1)^{2/3}T_2(r_1, b_1) + (cr_2/\nu_2)^{2/3}T_2(r_2, b_2), \\ k_3 &= (r_1/\nu_1)T_3(r_1, b_1) - (cr_2/\nu_2)T_3(r_2, b_2), \\ k_4 &= (r_1/\nu_1)^{4/3}T_4(r_1, b_1) - (cr_2/\nu_2)^{4/3}T_4(r_2, b_2) \end{aligned}$$

このとき、

$$\begin{aligned} T_1(r, b) &= \left(1 - \frac{2(1+b)}{9r} - \frac{40b^2}{3^4r^2} + \frac{80(1+3b+33b^2-77b^3)}{3^7r^3} \right. \\ &\quad \left. + \frac{176(1+4b-210b^2-2380b^3-2975b^4)}{3^9r^4} \right), \\ T_2(r, b) &= \left(\frac{2(1+b)}{9r} - \frac{16b^2}{3^3r^2} + \frac{8(13+39b+405b^2-1025b^3)}{3^7r^3} \right. \\ &\quad \left. + \frac{160(1+4b-87b^2+1168b^3-1544b^4)}{3^8r^4} \right), \\ T_3(r, b) &= - \left(\frac{8b^2}{3^3r^2} - \frac{32(1+3b+21b^2-62b^3)}{3^6r^3} \right. \\ &\quad \left. - \frac{32(8+3b-177b^2+4550b^3-6625b^4)}{3^8r^4} \right), \\ T_4(r, b) &= \left(\frac{16(1+3b+12b^2-44b^3)}{3^6r^3} \right. \\ &\quad \left. - \frac{256(1+4b-6b^2+274b^3-458b^4)}{3^8r^4} \right), \end{aligned}$$

$r_i = \nu_i + \lambda_i$, $b_i = \lambda_i/r_i$, $i = 1, 2$ である。以上を用いてエッジワース級数展開を行うことにより、2重非心 F の累積分布関数のためのエッジワース級数近似を得ることができる。

$$\begin{aligned} P_r[F'' \leq c] &:= \Phi(d) \\ &- \left[\frac{\beta_1}{6}(d^2 - 1) + \frac{\beta_2}{24}(d^3 - 3d) + \frac{\beta_3^2}{72}(d^5 - 10d^3 + 15d) \right] \cdot \phi(d) \end{aligned} \quad (7)$$

このとき、 $d = -k_1/\sqrt{k_2}$ 、 $\beta_1 = k_3/k_2^{3/2}$ 、 $\beta_2 = k_4/k_2^2$ 、 $\Phi(\cdot)$ は標準正規変数の分布関数、 $\phi(\cdot)$ は確率密度関数を意味する。この近似法を導出者3人の頭文字 MCL とエッジワース級数展開の E を合わせ、本研究では MCL-E 法と呼ぶことにする。

3.4 MCL-M 法

Tiku[6] は非心 F 分布と合流型超幾何分布に対応するモーメントに関して2重非心 F の r 番目のモーメント μ_r' の表現を導き出した。これらの計算は困難なため、合流型超幾何分布を近似することによっての r 番目のモーメント近似法も Tiku は同時に導き出した (この3次の近似モーメントは2重非心 F 変数 $F''(\nu_1, \nu_2, \lambda_1, \lambda_2)$ の近似 $aF(\nu_1', \nu_2) + b$ のパラメータ (a, ν_1', b) を決定することに

使われる)。これらを利用し Mudholkar et al.[3] が非心 F 変数に基づくモーメント近似を導き出したものが2つ目の近似法である。

3.4.1 導出方法

2重非心 F の分母の非心カイ2乗は変数 $c\chi^{2\nu}$ によって近似できることを利用する。このときの c 、 ν (初めの2つのモーメントによって決まる) は $c = (\nu_2 + 2\lambda_2)/(\nu_2 + \lambda_2)$ と $\nu = (\nu_2 + \lambda_2)^2/(\nu_2 + 2\lambda_2)$ である。従って、2重非心 F 変数のモーメントは以下の変数とほぼ等しくなる。

$$aF'(\nu_1, \nu, \lambda_1) = a \frac{\chi_{\nu_1}^{\prime 2}/\nu_1}{\chi_{\nu}^{\prime 2}/\nu} \quad (8)$$

このとき、 $a = \nu_2/(\nu_2 + \lambda_2)$ 、 $F'(\nu_1, \nu, \lambda_1)$ はパラメータ (ν_1, ν, λ_1) での非心 F 変数である。このモーメント近似が2重非心 F 分布の確率とパーセント点近似に利用できることは明らかである。この近似方法を導出者3人の頭文字 MCL とモーメント近似の M を合わせ、本研究では MCL-M 法と呼ぶことにする。

4 近似法の比較

R を用いて近似法の精度の比較を行った。方法は自由度 (ν_1, ν_2) と非心度 (λ_1, λ_2) のパラメータを変更し、そのときのパーセント点近似値と真値との比較により精度を見る。

4.1 真値の定義

堀井 [2] によって定義された真値を用いる。この方法は R に元々入っているカイ2乗乱数 (rchisq) を用いたものである。非心カイ2乗乱数を10000組発生させて式(1)を計算し、それを10000回繰り返したパーセント点の平均の値を真値と定義する。このモンテカルロ法での真値は相対標準誤差が0.0003であり、非常に信頼の高い値が出ることが堀井 [2] によってわかっている。しかし、1つのパーセント点を求める計算に1.40GHzのノートPCにおいて約6分程度かかってしまう。また、乱数を用いての計算のため、同じ自由度と非心度の組み合わせであっても最終的な結果が若干異なってしまうという欠点がある。

4.2 プログラム

Tiku 法、Torigoe 法の近似式 (4)、(5) についてのプログラムも堀井 [2] によって作成されているのでそのプログラムを使用した。Tiku 法 (4) の積分部分は中点則を使用しパーセント点を求め、Torigoe 法 (5) はニュートン・ラプソン法を使用してパーセント点近似を求めるものである。MCL-E 法 (7) と MCL-M 法 (8) の近似式についてのプログラムは作成した。堀井 [2] と同様にソフトウェアは R を使用する。式 (7) は確率を求める形であるので、パーセント点を求める考え方として、 $(\nu_1, \nu_2, \lambda_1, \lambda_2, x)$ を決定する (x は内部の関数で自由度 (ν_1, ν_2) の中心 F 分布のパーセント点を求めることに使われる)。この x を変更していき、結果として帰ってくる2重非心 F 分布の下側確率の値が95パーセント点を求めたいならば0.95、5パーセント点ならば0.05になるまで繰り返す。このときの x

の中心 F 分布のパーセント点 x が 2 重非心 F 分布のパーセント点近似である。式 (8) のパーセント点を求める考え方として、2 重非心 F 分布の確率を次のように変換する。

$$Pr\{F'' \leq x\} = Pr\{aF' \leq x\} = Pr\left\{F' \leq \frac{x}{a}\right\} \quad (9)$$

よって、 $(\nu_1, \nu_2, \lambda_1, \lambda_2, x)$ を決定して “ x/a ” を用いての非心 F 分布の分布関数より求める。このときの x の値を変更していき、95 パーセント点を求めたいならば 0.95 に、5 パーセント点を求めたいならば 0.05 になるように x を定める。このときの x の値が 2 重非心 F 分布のパーセント点近似となる。また、プログラム内での x の探索方法として MCL-E 法、MCL-M 法ともに 2 分法を使用した。

4.3 自由度の大きい場合

自由度 1 を 5、10、20、自由度 2 を 10、20、30、非心度 1 と 2 を 5、10 に変化させパーセント点近似値を求める。このときの結果の一部を表 1 と表 2 に示す。表中では Tiku 法、Torigoe 法、MCL-E 法、MCL-M 法を TI、TO、E 法、M 法と略して記載する。また、計算途中で符号の中が負となり値が出せないものは “-” で示す。

表 1 $\alpha = 95\%$

ν_1	ν_2	λ_1	λ_2	真値	TI	TO	E 法	M 法
5	10	5	5	4.054	4.083	4.085	4.173	4.016
5	20	5	5	4.029	4.043	4.043	4.115	4.025
5	30	5	5	4.028	4.041	4.040	4.114	4.028
10	10	10	10	2.585	3.261	-	2.646	2.544
10	20	10	10	2.910	3.031	3.033	2.941	2.901
10	30	10	10	3.046	3.065	3.065	3.073	3.043

表 2 $\alpha = 5\%$

ν_1	ν_2	λ_1	λ_2	真値	TI	TO	E 法	M 法
5	10	5	5	0.338	0.347	0.347	0.317	0.338
5	20	5	5	0.422	0.396	0.396	0.395	0.422
5	30	5	5	0.459	0.435	0.435	0.429	0.459
10	10	10	10	0.387	-	-	0.378	0.387
10	20	10	10	0.543	0.730	-	0.533	0.543
10	30	10	10	0.627	0.635	0.635	0.615	0.627

Tiku 法、Torigoe 法ともに求められない点が多く存在するが、MCL-E 法、MCL-M 法はすべての組み合わせで近似値を求めることができる。比較的自由度の大きい場合のシミュレーションでの 4 つの近似法の最大誤差と誤差平均の比較を表 3 を示す。MCL-M 法は 95、90 パーセント点近似とともに真値に対する誤差の割合が大体 1% 程度である。90 パーセント点近似では Torigoe 法が最も良い結果となっているが、近似値を求められない点が存在することを考えると最良とはいえない。10、5 パーセント点近似を見ると、MCL-M 法は最大誤差と誤差平均が非常に小さく真値に近い近似値を求めることができることがわかる。よって、自由度が比較的大きい場合は MCL-M 法が最も有効であるといえる。

表 3 誤差の比較

パーセント点	—	TI	TO	E 法	M 法
95%	最大誤差	0.811	0.178	0.133	0.059
	誤差平均	0.156	0.043	0.047	0.015
90%	最大誤差	0.565	0.019	0.339	0.025
	誤差平均	0.087	0.007	0.043	0.007
10%	最大誤差	0.150	0.014	0.036	0.001
	誤差平均	0.034	0.005	0.010	0.000
5%	最大誤差	0.256	0.026	0.040	0.001
	誤差平均	0.058	0.010	0.012	0.000

4.4 自由度の小さい場合

使用した自由度と非心度は堀井 [2] によって採取された静特性における望目特性の実データ (あるサーキットでのカーレースにおける 1 周のタイムのシミュレーションデータ) を元にしたものである。詳しい静特性の説明、シミュレーションの結果、導出法は堀井 [2] を参照されたい。最終的に導き出される分布は $F''(1, 5, 564588.16, 14.29)$ 、 $F''(1, 5, 77894.98, 14.83)$ 、 $F''(1, 5, 15386.60, 4.83)$ である (堀井 [2] では計算途中に誤りがあるため、最終的な分布は高橋 [4] による結果を参照)。実データの結果をふまえて、自由度 1 は 1 に自由度 2 は 5 に固定し、非心度 1 は 0、10、100、1000、10000、100000 に、非心度 2 は 0、2、4、6、8、10 に変化させシミュレーションを行った結果の一部を表 4 と表 5 に示す。Tiku 法と Torigoe 法については堀井 [2] によって実用的でないとわかっているため、表への記載はスペースの都合上省略する。

表 4 静特性 (95%)

ν_1	ν_2	λ_1	λ_2	真値	E 法	M 法
1	5	10	0	54.852	56.749	54.876
1	5	100	0	448.616	446.632	448.803
1	5	10000	0	43641.280	43396.180	43662.430
1	5	10	4	26.929	29.531	25.701
1	5	100	4	216.943	232.033	203.998
1	5	10000	4	21046.960	22537.120	19733.150
1	5	10	10	13.012	14.176	12.472
1	5	100	10	100.249	106.279	94.317
1	5	10000	10	9636.978	10234.980	9026.467

表 5 静特性 (5%)

ν_1	ν_2	λ_1	λ_2	真値	E 法	M 法
1	5	10	0	2.038	1.749	2.039
1	5	100	0	42.032	41.931	42.045
1	5	10000	0	4511.976	4508.188	4513.316
1	5	10	4	1.155	0.975	1.156
1	5	100	4	24.799	24.284	24.759
1	5	10000	4	2682.411	2628.578	2673.896
1	5	10	10	0.712	0.600	0.713
1	5	100	10	16.277	15.911	16.219
1	5	10000	10	1782.268	1742.777	1771.576

95、90パーセント点の場合は非心度2が0のときに理論上MCL-M法が有効であり、10、5パーセント点の場合はその非心度でもMCL-M法が有効であるとわかった。しかし、非心度2が0以外の場合は非心度1の値によってMCL-E法、MCL-M法ともに真値に対する誤差の割合が大きくなり、MCL-E法では最大で12%、MCL-M法では最大で7%ほどになる。また、MCL-E法、MCL-M法ともに非心度2の影響は小さく、非心度2を変化させても同じような誤差の周期を繰り返すことがわかった。

5 新たな近似法の提案

実データをふまえたシミュレーションにより、95と90パーセント点近似でMCL-E法は真値よりも大きい近似値を示し、MCL-M法は真値よりも小さい近似値を示すことがわかった。また、非心度1の値が大きい場合はMCL-E法とMCL-M法の誤差の絶対値がほぼ等しくなることが多い。このため、2つの近似値の中間点を取れば真値に近づくと考えられる。これをME法と呼ぶことにすると

$$\text{ME法の近似値} = \frac{(\text{MCL-M法の近似値} + \text{MCL-E法の近似値})}{2} \quad (10)$$

である。堀井 [2] による実データ (静特性、動特性) を元にした自由度と非心度でシミュレーションを行い、近似法の場合分け表を作成した。これを表6に示す。

表6 近似法の場合分け

95、90パーセント点近似	使用する近似法
非心度2が0	M法
自由度2が5以下	ME法
自由度2が6から7で非心度1が10以下	M法
自由度2が6から7で非心度1が11以上	ME法
自由度2が8から10で非心度1が25以下	M法
自由度2が8から10で非心度1が26以上	ME法
自由度2が11から14で非心度1が50以下	M法
自由度2が11から14で非心度1が51以上	ME法
10、5パーセント点近似	M法

表6を元にME法とM法をうまく使いわけることで、表4での近似は表7のように改善できる。

表7 場合分けを用いた近似結果

ν_1	ν_2	λ_1	λ_2	真値	近似値	近似法
1	5	10	0	54.852	54.876	M法
1	5	100	0	448.616	448.803	M法
1	5	10000	0	43641.280	43662.430	M法
1	5	10	4	26.929	27.616	ME法
1	5	100	4	216.943	218.015	ME法
1	5	10000	4	21046.960	21135.135	ME法
1	5	10	10	13.012	13.324	ME法
1	5	100	10	100.249	100.298	ME法
1	5	10000	10	9636.978	9630.724	ME法

6 まとめ

10、5パーセント点近似の場合はMCL-M法は真値に対する誤差の割合が最大でも0.6%であり実用であると結論付けた。95、90パーセント点近似の場合はME法を提案し、自由度と非心度の値による場合分けの表6を導出した。表6をもとにMCL-M法とME法の場合分けを行うことで、真値に対する誤差の割合を最大でも2.5%程度までにすることが可能となる。また、この2.5%程度の誤差がでる場合の非心度1は10の場合である。実データをもとにした非心度1は非常に大きい場合が多く、ME法を用いると真値に対する誤差の割合が1%以下にできる。また、非心度2が0の場合はMCL-M法を用いることでほぼ正確な値を出すことができることがわかった。このため、95、90パーセント点近似でも近似法は有効であると結論付けた。以上により、近似法によるパーセント点近似は自由度の小さい実データにおいても十分実用であると結論付ける。

7 おわりに

自由度の値をさらに増やしての場合分け、非心度1が10以下の場合の検討など今後の課題がまだまだ残っている。しかし、モンテカルロ法では1つの値を導出するのに約6分程度かかるのに対し、近似法を用いると一瞬でパーセント点近似値を求めることができる。これらの近似法うまく使いこなすことで作業効率を格段に上げることができるだろう。

参考文献

- [1] Aty, A.S.H. (1954): Approximate formula for the percentage points and the probability integral of the non-central χ^2 -distribution, *Biometrika* 41, 538-40.
- [2] 堀井 里佳子 (2010): タグチメソッドにおけるSN比の統計的分布について, 2009年度南山大学大学院数理情報研究科修士論文.
- [3] Mudholkar, G.S., Chaubey, Y.P. and Lin, C. (1976): APPROXIMATIONS FOR THE DOUBLY NON-CENTRAL-F DISTRIBUTION, *Communications in Statistics-Theory and Methods*, Vol.5, 49-63.
- [4] 高橋 知也 (2011): タグチメソッドのSN比における信頼区間の適用方法の研究, 2010年度南山大学大学院数理情報研究科修士論文.
- [5] Tiku, M.L. (1965): Series expansion for the doubly noncentral F-distribution, *Austral. J. Statist.* 7, 78-89.
- [6] Tiku, M.L. (1972): A note on the distribution of the doubly non-central F-distribution, *Austral. J. Statist.* 14, 37-40.
- [7] 鳥越 規夫 (1997): 2重非心F分布のパーセント点の近似について,
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/0916-4.pdf>.