数理解析研究所講究録 1224 巻 2001 年 73-89

Approximations for a family of generalized hypergeometric distributions

筑波大・数学 飛田 英祐 (Eisuke Hida)
筑波大・数学 赤平 昌文 (Masafumi Akahira)

1. はじめに

離散型分布族の1つとして，一般超幾何分布族が考えられ，これはポアソン分布，2項分布，負の2項分布，超幾何分布，負の超幾何分布，対数級数分布などを含む一般的な分布族であることが知られている (Kemp[K68], Dacey[D72], 竹内 [Ta84]). この一般超幾何分布はすべての母数が大きくなるとき，正規分布で近似できることが知られ，さらに Edgeworth 型展開による近似式が Stirling の公式を用いて導かれる ([Ta84]). 本論では，一般超幾何分布の [Ta84] による Edgeworth 型近似式を改良し，数値的に比較してその精確性を確かめる ([HA00]). さらに，一般超幾何分布の下側確率の近似式を構成し，数値的に比較する．なお，関連する結果は [JKK92], [SO94], [M73], [SS81], [Tr83] 等に見られる．

2. 設定

本節においては [Ta84] と同じ設定で考える．まず，確率変数 X が確率関数

$$p_X(x) := P\{X = x\} = K\frac{\prod_{j=1}^{m} c_j(x)\prod_{j=1}^{n}\bar{d}_j[x]}{x!\prod_{j=1}^{k} a_j(x)\prod_{j=1}^{l}\bar{b}_j[x]}\theta^x \tag{2.1}$$

をもつ分布を一般超幾何分布 (generalized hypergeometric distribution) という．ただし，すべての a_j, b_j, c_j, d_j を非負値定数，$\theta > 0$, K はある定数とし，$a(x) = \Gamma(a+x)/\Gamma(a)$, $\bar{b}[x] = \Gamma(b)/\Gamma(b-x)$ とする．また，$M := \min\{b_1, \ldots, b_l, d_1, \ldots, d_n\} > 0$ とし，$x = 0, 1, \ldots, M$ とする．このような分布を，$\mathbf{a} = (a_1, \ldots, a_k)$, $\mathbf{b} = (b_1, \ldots, b_l)$, $\mathbf{c} = (c_1, \ldots, c_m)$, $\mathbf{d} = (d_1, \ldots, d_n)$ とし，

$$GHG(k, l, m, n; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$$

と表す．このとき，非負整数の組 (k, l, m, n) を一般超幾何分布のタイプといい，この分布は，特殊な場合として次のような分布を含む．

(k,l,m,n)	θ の範囲	分布
$(0,0,0,0)$	$(0,\infty)$	ポアソン分布
$(0,0,0,1)$	$(0,\infty)$	2項分布
$(0,0,1,0)$	$(0,1)$	負の2項分布
$(1,0,0,2)$	1	超幾何分布
$(0,1,1,1)$	1	負の超幾何分布
$(2,1,0,0)$	$(0,1)$	対数級数分布

3. 一般超幾何分布の近似

各 j について，a_j, b_j, c_j, d_j が大きいときの一般超幾何分布の Edgeworth 型近似は [Ta84] によって与えられているが，数値的な検討はなされていない．本節では，[Ta84] と同様にして近似を導出する際に，その途中で適切に係数を決めることによって [Ta84] の近似式を改良することを考慮した上で，数値検討も行う．

3.1 各点確率の近似

まず，各 j について $a_j = \alpha_j N + 1$, $b_j = \beta_j N + 1$, $c_j = \gamma_j N + 1$, $d_j = \delta_j N + 1$ とし，$\alpha_j > 0$, $\beta_j > 0$, $\gamma_j > 0$, $\delta_j > 0$ とする．ただし，$\alpha_0 := 0$ とし，$a_0 := 1$ とする．次に，$\theta = \theta_0 N^{k+l+1-m-n}$ $(\theta_0 > 0)$ として，$N \to \infty$ のときに (2.1) の近似を考える．ここで，Stirling の公式

$$\log\Gamma(x+1) = \log\sqrt{2\pi} + \left(x+\frac{1}{2}\right)\log x - x + \frac{1}{12x} + O\left(\frac{1}{x^3}\right) \tag{3.1}$$

を用いる．いま，$p_X(x)$ をモードの値 $x_0 = N\mu + O(N)$ を中心にして展開する．このとき，$p_X(x_0+1)/p_X(x_0) \approx 1$ より，μ は

$$\frac{\prod_{j=1}^{m}(\gamma_j+\mu)\prod_{j=1}^{n}(\delta_j-\mu)}{\prod_{j=0}^{k}(\alpha_j+\mu)\prod_{j=1}^{l}(\beta_j-\mu)}\theta_0 = 1 \tag{3.2}$$

を満たさなければならない．そこで，(3.2) の解 $\mu(>0)$ が存在するときに，μ の値を中心にした展開を考える．ここで，(3.2) の解は必ずしも一意的になるとは限らない．このとき，$z := (x-N\mu)/\sqrt{N}$ とおくと

$$\begin{aligned}\log p_X(x) = \log K &- \sum_{j=0}^{k}\log\Gamma(a_j+x) + \sum_{j=1}^{l}\log\Gamma(b_j-x)\\ &+ \sum_{j=1}^{m}\log\Gamma(c_j+x) - \sum_{j=1}^{n}\log\Gamma(d_j-x) + x\log\theta\end{aligned} \tag{3.3}$$

になる．いま，Stirling の公式 (3.1) を用いると

$$\begin{aligned}\log\Gamma(a_j+x) &= \log\sqrt{2\pi} + \{N(\alpha_j+\mu)+N^{1/2}z+\frac{1}{2}\}\log(N(\alpha_j+\mu)+N^{1/2}z)\\ &\quad - \{N(\alpha_j+\mu)+N^{1/2}z\} + \frac{1}{12\{N(\alpha_j+\mu)+N^{1/2}z\}} + O\left(\frac{1}{N^3}\right)\\ &= N^{1/2}\{\log N + \log(\alpha_j+\mu)\}z + \frac{1}{2(\alpha_j+\mu)}z^2\\ &\quad - \frac{1}{6\sqrt{N}}\left\{\frac{1}{(\alpha_j+\mu)^2}z^3 - \frac{3}{(\alpha_j+\mu)}z\right\}\\ &\quad + \frac{1}{12N}\left\{\frac{1}{(\alpha_j+\mu)^3}z^4 - \frac{3}{(\alpha_j+\mu)^2}z^2\right\}\\ &\quad + \mathrm{const} + o\left(\frac{1}{N}\right)\end{aligned} \tag{3.4}$$

となる．また，(3.2) より

$$\sum_{j=1}^{m}\log(\gamma_j+\mu)+\sum_{j=1}^{n}\log(\delta_j-\mu)$$
$$-\sum_{j=0}^{k}\log(\alpha_j+\mu)-\sum_{j=1}^{l}\log(\beta_j-\mu)+\log\theta_0=0$$

となるから，(3.4) と同様のものを用いると (3.3) は

$$\log p_X(x)=\log K-\frac{1}{2\sigma^2}z^2+\frac{1}{6\sqrt{N}}(A_2z^3-3A_1z)$$
$$-\frac{1}{24N}(2B_3z^4-6B_2z^2)+o\left(\frac{1}{N}\right) \tag{3.5}$$

となる．ただし，$z=(x-N\mu)/\sqrt{N}$,

$$A_i=\sum_{j=0}^{k}\frac{1}{(\alpha_j+\mu)^i}+\sum_{j=1}^{l}\frac{1}{(\beta_j-\mu)^i}-\sum_{j=1}^{m}\frac{1}{(\gamma_j+\mu)^i}-\sum_{j=1}^{n}\frac{1}{(\delta_j-\mu)^i}\quad(i=1,2),$$
$$B_i=\sum_{j=0}^{k}\frac{1}{(\alpha_j+\mu)^i}-\sum_{j=1}^{l}\frac{1}{(\beta_j-\mu)^i}-\sum_{j=1}^{m}\frac{1}{(\gamma_j+\mu)^i}+\sum_{j=1}^{n}\frac{1}{(\delta_j-\mu)^i}\quad(i=1,2,3)$$

とし，$B_1=1/\sigma^2$ とする．このとき，(3.5) において $A:=A_1/A_2$，$w:=z-(C/\sqrt{N})$ とすると，

$$\log p_X(x)=\log K-\frac{1}{2\sigma^2}\left(w+\frac{C}{\sqrt{N}}\right)^2+\frac{A_2}{6\sqrt{N}}\left\{\left(w+\frac{C}{\sqrt{N}}\right)^3\right.$$
$$\left.-3A\left(w+\frac{C}{\sqrt{N}}\right)\right\}-\frac{1}{24N}(2B_3w^4-6B_2w^2)+o\left(\frac{1}{N}\right)$$
$$=-\frac{1}{2\sigma^2}w^2+\frac{A_2}{6\sqrt{N}}\left\{w^3-3\left(A+\frac{2C}{A_2\sigma^2}\right)w\right\}$$
$$-\frac{1}{24N}\left\{2B_3w^4-6(B_2+2A_2C)w^2+12C\left(A_1+\frac{C}{\sigma^2}\right)\right\}$$
$$+o\left(\frac{1}{N}\right) \tag{3.6}$$

となる．ただし，$C=\frac{1}{2}(A_2-A_1)\sigma^2$ とする．ここで，通常の Edgeworth 展開を考慮に入れて

$$A+\frac{2C}{A_2\sigma^2}=1$$

となるように C を決めていることに注意．したがって (3.6) より

$$p_X(x)=Ke^{-\frac{w^2}{2\sigma^2}}\left[1+\frac{A_2}{6\sqrt{N}}(w^3-3w)-\frac{1}{24N}\{2B_3w^4\right.$$
$$-6(B_2+A_2(A_2-A_1)\sigma^2)w^2+3(A_2^2-A_1^2)\sigma^2\}$$
$$\left.+\frac{A_2^2}{72N}(w^3-3w)^2+o\left(\frac{1}{N}\right)\right]$$
$$=:Kf_N(w) \tag{3.7}$$

と表される．ただし，$w=(x-N\mu-C)/\sqrt{N}$ とする．また，(3.6) の定数 K は

$$\sum_w \frac{1}{\sqrt{N}} f_N(w) = \int_{-\infty}^{\infty} f_N(w)dw + o\left(\frac{1}{N}\right) \tag{3.8}$$

より，

$$K = \frac{1}{\sqrt{2\pi N}\sigma} + O\left(\frac{1}{N\sqrt{N}}\right) \tag{3.9}$$

と表される．したがって (3.6) と (3.9) より

$$\begin{aligned} p_X(x) = \frac{1}{\sqrt{2\pi N}\sigma} e^{-\frac{w^2}{2\sigma^2}} \Bigg[& 1 + \frac{A_2}{6\sqrt{N}}(w^3-3w) - \frac{1}{24N}\{2B_2w^4 \\ & -6(B_2 + A_2(A_2-A_1)\sigma^2)w^2 + 3(A_2^2 - A_1^2)\sigma^2\} \\ & + \frac{A_2^2}{72N}(w^3-3w)^2 + o\left(\frac{1}{N}\right)\Bigg] \end{aligned} \tag{3.10}$$

となる．一方，(3.5) から直接

$$\begin{aligned} p_X(x) = \frac{1}{\sqrt{2\pi N}\sigma} e^{-\frac{z^2}{2\sigma^2}} \Bigg[& 1 + \frac{1}{6\sqrt{N}}(A_2z^3 - 3A_1z) - \frac{1}{24N}(2B_3z^4 - 6B_2z^2) \\ & + \frac{1}{72N}(A_2z^3-3A_1z)^2 + o\left(\frac{1}{N}\right)\Bigg] \end{aligned} \tag{3.11}$$

と表すこともできる ([Ta84]).

3.2 片側確率の近似

前節において，一般超幾何分布の確率関数が (3.10) より

$$p_X(x) = \frac{1}{\sqrt{2\pi N}\sigma} f_N(w)$$

と表せることから，その下側確率 $P\{X \le x\}$ の近似を [Ta84] と同様にして考えることもできる．そのため，$W=(X-N\mu-C)/\sqrt{N}$ の分布のキュムラントを求める．したがって，H_j を j 次のエルミート多項式とすると

$$\int e^{itw} w^j \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{w^2}{2\sigma^2}} dw = (i\sigma)^j H_j(\sigma t) e^{-\frac{\sigma^2 t^2}{2}}$$

となる．いま $\tilde{H}_j(t) := (i\sigma)^j H_j(\sigma t)$ とおくと，W の分布の特性関数は

$$\begin{aligned} \phi_N(t) &= E(e^{itW}) = \int \frac{1}{\sqrt{2\pi}\sigma} e^{itw} f_N(w)dw \\ &= e^{-\frac{\sigma^2t^2}{2}} \Bigg\{ 1 + \frac{A_2}{6\sqrt{N}}(\tilde{H}_3(t) - 3\tilde{H}_1(t)) - \frac{1}{24N}\{2B_3\tilde{H}_4(t) - 6(B_2 + \sigma^2 A_2(A_2-A_1))\tilde{H}_2(t) \\ &\qquad -3\sigma^2(A_2^2 - A_1^2)\} + \frac{A_2^2}{72N}(\tilde{H}_6(t) - 6\tilde{H}_4(t) + 9\tilde{H}_2(t))\Bigg\} + O\left(\frac{1}{N\sqrt{N}}\right) \end{aligned}$$

と展開することができる．これより W の分布のキュムラント母関数は

$$\log\phi_N(t) = -\frac{\sigma^2t^2}{2} + \frac{A_2}{6\sqrt{N}}(\tilde{H}_3(t) - 3\tilde{H}_1(t)) - \frac{1}{24N}\{2B_3\tilde{H}_4(t) - 6(B_2 + \sigma^2A_2(A_2 - A_1))\tilde{H}_2(t) - 3\sigma^2(A_2^2 - A_1^2)\} + \frac{A_2^2}{72N}(\tilde{H}_6(t) - 6\tilde{H}_4(t) + 9\tilde{H}_2(t)) + O\left(\frac{1}{N\sqrt{N}}\right)$$

$$= -\frac{\sigma^2t^2}{2} + \frac{A_2}{2\sqrt{N}}3(\sigma^4 - \sigma^2)(it) - \frac{1}{4N}[2B_3\sigma^6 - (B_2 + \sigma^2A_2(A_2 - A_1))\sigma^4 - A_2^2(2\sigma^8 - \sigma^6)](it)^2 + \frac{A_2}{6\sqrt{N}}\sigma^6(it)^3 - \frac{1}{24N}(2B_3\sigma^8 - 3A_2^2\sigma^{10})(it)^4 + O\left(\frac{1}{N\sqrt{N}}\right) \tag{3.12}$$

となる．したがって (3.12) より W のキュムラントはそれぞれ

$$\kappa_1(W) = E[W] = \frac{A_2}{2\sqrt{N}}(\sigma^4 - \sigma^2) + O\left(\frac{1}{N\sqrt{N}}\right)$$

$$\kappa_2(W) = \sigma^2 - \frac{1}{2N}(2B_3\sigma^6 - B_2\sigma^4) + \frac{1}{2N}(2A_2^2\sigma^8 - A_1A_2\sigma^6) + O\left(\frac{1}{N\sqrt{N}}\right)$$

$$\kappa_3(W) = \frac{A_2}{\sqrt{N}}\sigma^6 + O\left(\frac{1}{N\sqrt{N}}\right)$$

$$\kappa_4(W) = -\frac{2B_3}{N}\sigma^8 + \frac{3A_2^2}{N}\sigma^{10} + O\left(\frac{1}{N\sqrt{N}}\right)$$

となる．ただし，5 次以上のキュムラントはすべて N^{-1} より小さい order となる．

そこで，$P\{X \le x\}$ を計算するために

$$u = \frac{x - N\mu - C - (A_2(\sigma^4 - \sigma^2) - 1)/2}{\sqrt{N}\sigma}$$

とおく．ただし，分子は W の平均の項と，連続補正を考慮したものとする．このとき離散修正を施した Edgeworth 展開は

$$P\{X \le x\} = \sum_{t=0}^{x} p_X(t)$$

$$= \Phi(u) - \phi(u)\left\{\frac{\kappa_3}{6\sigma^3}H_2(u) + \frac{\kappa_4}{24\sigma^4}H_3(u) + \frac{\kappa_3^2}{72\sigma^6}H_5(u) - \frac{1}{24N\sigma^2}H_1(u) + O\left(\frac{1}{N\sqrt{N}}\right)\right\}$$

$$= \Phi(u) - \phi(u)\left\{\frac{A_2\sigma^3}{6\sqrt{N}}H_2(u) - \frac{1}{24N}(2B_3\sigma^4 - 3A_2^2\sigma^6)H_3(u) + \frac{A_2^2\sigma^6}{72N}H_5(u) - \frac{1}{24N\sigma^2}H_1(u) - \frac{1}{4N}\{2B_3\sigma^4 - B_2\sigma^2 + A_1A_2\sigma^4 + 2A_2^2(\sigma^6 - \sigma^4)\} + O\left(\frac{1}{N\sqrt{N}}\right)\right\} \tag{3.13}$$

と表される．ただし，$\Phi(\cdot)$, $\phi(\cdot)$ はそれぞれ標準正規分布 $N(0,1)$ の c.d.f., p.d.f. とする．また，これより $P\{X \le x\} = \Phi(u')$ とおくと

$$u' = u - \frac{A_2\sigma^3}{6\sqrt{N}}H_2(u) + \frac{A_2^2\sigma^6}{72N}(H_2^2(u) - H_5(u)) + \frac{1}{24N}(2B_3\sigma^4 - 3A_2^2\sigma^6)H_3(u) + \frac{1}{24N\sigma^2}H_1(u) + \frac{1}{4N}\{2B_3\sigma^4 - B_2\sigma^2 + A_1A_2\sigma^4 + 2A_2^2(\sigma^6 - \sigma^4)\} + O\left(\frac{1}{N\sqrt{N}}\right)$$

と表すこともできる．

4. 数値的比較検討

前節の 2 つの Edgeworth 型近似式 (3.10), (3.11) と下側確率の近似式 (3.13) を具体的な分布において数値的に比較検討する.

例 1 (ポアソン分布の場合). ポアソン分布 $P_O(\lambda)$ は，タイプ (0,0,0,0), の一般超幾何分布であり，その確率関数は

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (x = 0, 1, 2, \ldots ; \lambda > 0)$$

である. このとき，$\theta = \lambda$, $\theta_0 = \lambda/N$ とおくと，(3.2) の解として $\mu = N/\lambda$ が一意的に定まる. これを用いて $p_X(x)$ の真値とその近似 (3.10), (3.11) と Edgeworth 近似，さらに下側確率の真値とその近似 (3.13) と各点確率の近似式 (3.10) の和をとったものと Edgeworth 近似を数値的に比較を行った (表 4.1.1, 4.1.2 参照). その結果，各点確率については近似式 (3.10) は (3.11) と Edgeworth 近似より精確であることが分かる. また，下側確率についても近似式 (3.13) の精確さを読みとることができる.

表 4.1.1 ポアソン分布 $P_O(\lambda)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)	Edgeworth
0	0.0045	—	—	—
1	0.0454	−0.1123	−0.1197	−0.1633
2	0.2270	0.0604	0.0516	0.0531
3	0.7567	0.0435	0.0349	0.0407
4	1.8917	0.0184	0.0099	0.0140
5	3.7833	0.0059	−0.0024	−0.0006
6	6.3056	0.0036	−0.0048	−0.0046
7	9.0079	0.0058	−0.0026	−0.0033
8	11.2599	0.0082	−0.0002	−0.0010
9	12.5110	0.0090	0.0006	0.0002
10	12.5110	0.0084	0.0000	0.0003
11	11.3736	0.0079	−0.0005	0.0003
12	9.4780	0.0089	0.0004	0.0013
13	7.2908	0.0109	0.0025	0.0028
14	5.2077	0.0120	0.0035	0.0030
15	3.4718	0.0095	0.0010	−0.0003
16	2.1699	0.0031	−0.0052	−0.0070
17	1.2764	−0.0032	−0.0116	−0.0132
18	0.7091	−0.0030	−0.0113	−0.0119
19	0.3732	0.0090	0.0006	0.0024
20	0.1866	0.0312	0.0227	0.0285

表 4.1.2 ポアソン分布 $P_O(10)$ の下側確率 $P\{X \le x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)	Edgeworth
0	0.0045	—	—	—
1	0.0499	−0.5019	−0.3169	−0.5019
2	0.2769	−0.0470	−0.0149	−0.0470
3	1.0336	0.0172	0.0215	0.0172
4	2.9253	0.0152	0.0140	0.0152
5	6.7086	0.0063	0.0047	0.0063
6	13.0141	0.0010	0.0001	0.0010
7	22.0221	−0.0008	−0.0010	−0.0008
8	33.2820	−0.0009	−0.0007	−0.0009
9	45.7930	−0.0006	−0.0004	−0.0006
10	58.3040	−0.0004	−0.0003	−0.0004

例 2（2項分布の場合）． 2項分布分布 $B(n,p)$ は，タイプ $(0,0,0,1)$ の一般超幾何分布であり，その確率関数は

$$p_X(x) = \binom{n}{x} p^x q^{n-x} \qquad (x = 0, 1, \ldots, n;\ 0 < p < 1,\ q = 1-p)$$

である．このとき，$d = n+1$，$\theta = \theta_0 = p/q$ とおくと

$$p_X(x) = K \frac{\bar{d}[x]}{x!} \theta^x$$

と表される．ただし，K はある定数とする．いま (3.2) において，$\delta = n/N$ とおくと

$$\frac{\delta - \mu}{\mu} \cdot \frac{p}{q} = 1$$

となり，この解として $\mu = \delta p$ が一意的に定まる．これを用いて $p_X(x)$ の真値とその近似 (3.10), (3.11) と Edgeworth 近似，さらに下側確率の真値とその近似 (3.13) と各点確率の近似式 (3.10) の和をとったものと Edgeworth 近似を数値的に比較を行った (表 4.2.1, 4.2.2 参照)．その結果，各点確率については近似式 (3.10) は (3.11) と Edgeworth 近似より精確であることが分かる．また，下側確率についても近似式 (3.13) の精確さを読みとることができる．

表 4.2.1 2項分布 $B(20, 0.5)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)	Edgeworth
0	0.0001	—	—	—
1	0.0019	−0.0986	−0.1266	−0.2757
2	0.0181	0.0458	0.0300	0.0040
3	0.1087	0.0316	0.0195	0.0180
4	0.4621	0.0148	0.0044	0.0068
5	1.4786	0.0083	−0.0015	0.0002
6	3.6964	0.0078	−0.0018	−0.0013
7	7.3929	0.0091	−0.0006	−0.0008
8	12.0134	0.0099	0.0001	0.0000
9	16.0179	0.0100	0.0000	0.0002
10	17.6197	0.0099	0.0000	0.0003
11	16.0179	0.0100	0.0000	0.0002
12	12.0134	0.0099	0.0001	0.0000
13	7.3929	0.0091	−0.0006	−0.0008
14	3.6964	0.0078	−0.0018	−0.0013
15	1.4786	0.0083	−0.0015	0.0002
16	0.4621	0.0148	0.0044	0.0068
17	0.1087	0.0316	0.0195	0.0180
18	0.0181	0.0458	0.0300	0.0040
19	0.0019	−0.0986	−0.1266	−0.2757
20	0.0001	—	—	—

表 4.2.2 2項分布 $B(20, 0.5)$ の下側確率 $P\{X \leq x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)	Edgeworth
0	0.0181	−0.5459	0.4672	−0.5459
1	0.2133	0.0294	0.1113	0.0294
2	1.2118	0.0269	−0.0027	0.0269
3	4.4376	0.0072	−0.0282	0.0072
4	11.8197	−0.0003	−0.0207	−0.0003
5	24.5396	−0.0012	−0.0068	−0.0012
6	41.6625	−0.0006	0.0017	−0.0006
7	60.1027	−0.0003	0.0025	−0.0003

例 3 (負の 2 項分布の場合). 負の 2 項分布 $NB(n,p)$ は，タイプ $(0,0,1,0)$ の一般超幾何分布であり，その確率関数は

$$p_X(x) = \binom{x+n-1}{x} p^x q^{n-x} \qquad (x=0,1,\ldots,n;\ 0<p<1,\ q=1-p)$$

である．このとき，$c=n$，$\theta=\theta_0=q$ とすると

$$p_X(x) = K\frac{c(x)}{x!}\theta^x$$

と表される．ただし，K はある定数とする．いま，$\gamma=(n-1)/N$ とおくと，(3.2) の解として $\mu=\gamma q/p$ が一意的に定まる．これを用いて $p_X(x)$ の真値とその近似 (3.10), (3.11) と Edgeworth 近似，さらに下側確率の真値とその近似 (3.13) と各点確率の近似式 (3.10) の和をとったものと Edgeworth 近似を数値的に比較を行った (表 4.3.1, 4.3.2 参照)．その結果，各点確率については近似式 (3.10) は (3.11) と Edgeworth 近似より精確であることが分かる．また，下側確率についても近似式 (3.13) の精確さを読みとることができる．

表 4.3.1 負の 2 項分布 $NB(40,0.75)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)	Edgeworth
0	0.0010	6441.2400	——	319.3720
1	0.0100	769.3210	−0.3769	66.1710
2	0.0515	158.8540	−0.1304	25.6483
3	0.1804	41.6233	−0.1454	12.3797
4	0.4848	11.8389	−0.1743	6.7854
5	1.0666	3.15862	−0.1848	4.0082
6	1.9998	0.5862	−0.1763	2.4598
7	3.2854	−0.0515	−0.1532	1.5183
8	4.8254	−0.0934	−0.1208	0.9067
9	6.4339	−0.0067	−0.0845	0.4890
10	7.8815	0.0511	−0.0492	0.1926
11	8.9562	0.0499	−0.0194	−0.0238
12	9.5160	0.0167	0.0017	−0.1856
13	9.5160	−0.0031	0.0130	−0.3093
14	9.0062	0.0246	0.0163	−0.4063
15	8.1056	0.1000	0.0159	−0.4843
16	6.9657	0.1824	0.0182	−0.5486
17	5.7365	0.2110	0.0299	−0.6025
18	4.5414	0.1585	0.0554	−0.6475
19	3.4658	0.1069	0.0947	−0.6844
20	2.5560	0.3193	0.1427	−0.7136

表 4.3.2 負の 2 項分布 $NB(40, 0.75)$ の下側確率 $P\{X \le x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)	Edgeworth
2	0.0626	0.4835	−0.2222	36.8765
3	0.2430	0.1533	−0.1652	18.6908
4	0.7278	0.0317	−0.1713	10.7603
5	1.7944	−0.0298	−0.1793	6.7469
6	3.7941	−0.0630	−0.1777	4.4872
7	7.0795	−0.0791	−0.1663	3.1094
8	11.9049	−0.0845	−0.1479	2.2166
9	18.3388	−0.0832	−0.1256	1.6105
10	26.2202	−0.0783	−0.1027	1.1843
11	35.1765	−0.0716	−0.0815	0.8767
12	44.6924	−0.0644	−0.0638	0.6505
13	54.2084	−0.0573	−0.0503	0.4820

例 4 (超幾何分布の場合). 超幾何分布 $H(M, n, L)$ は，タイプ $(1, 0, 0, 2)$，$\theta = 1$ の一般超幾何分布であり，その確率関数は

$$p_X(x) = \binom{M}{x}\binom{L-M}{n-x}\Big/\binom{L}{n} \quad (x = 0, 1, \ldots, \min(n, M))$$

である．このとき，$a = L - M - n + 1$，$d_1 = M + 1$，$d_2 = n + 1$ とすると

$$p_X(x) = K\frac{\prod_{j=1}^{2} \bar{d}_j[x]}{x!a(x)}.$$

と表される．ただし，K はある定数とする．いま，$\alpha = (L - M - n)/N$，$\delta_1 = M/N$，$\delta_2 = n/N$ とおくと，(3.2) は

$$\frac{(\delta_1 - \mu)(\delta_2 - \mu)}{\mu(\alpha + \mu)} = 1,$$

となり，この解として $\mu = \delta_1\delta_2/(\alpha + \delta_1 + \delta_2)$ が一意的に定まる．これを用いて $p_X(x)$ の真値とその近似 (3.10), (3.11) と Edgeworth 近似，さらに下側確率の真値とその近似 (3.13) と各点確率の近似式 (3.10) の和をとったものと Edgeworth 近似を数値的に比較を行った (表 4.4.1, 4.4.2 参照)．その結果，各点確率については近似式 (3.10) は (3.11) と Edgeworth 近似より精確であることが分かる．また，下側確率についても近似式 (3.13) の精確さを読みとることができる．

表 4.4.1 超幾何分布 $H(50, 25, 20)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)	Edgeworth
0	0.0000	—	—	—
1	0.0000	−0.4975	−0.4459	—
2	0.0003	−0.0152	−0.0032	−0.4434
3	0.0053	0.0175	0.0185	−0.0759
4	0.0548	0.0078	0.0042	−0.0082
5	0.3685	0.0035	−0.0028	0.0014
6	1.6750	0.0054	−0.0029	−0.0010
7	5.3041	0.0091	−0.0011	−0.0002
8	11.9342	0.0121	0.0000	0.0000
9	19.3221	0.0141	0.0000	−0.0001
10	22.6713	0.0160	−0.0001	−0.0001
11	19.3221	0.0181	0.0000	−0.0001
12	11.9342	0.0200	0.0000	0.0000
13	5.3041	0.0208	−0.0011	−0.0002
14	1.6750	0.0207	−0.0029	−0.0010
15	0.3685	0.0229	−0.0028	0.0014
16	0.0548	0.0328	0.0042	−0.0082
17	0.0053	0.0522	0.0185	−0.0759
18	0.0003	0.0410	−0.0032	−0.4434
19	0.0000	−0.3765	−0.4459	—
20	0.0000	−5.9376	—	—

表 4.4.2 超幾何分布 $H(40, 20, 15)$ の下側確率 $P\{X \leq x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)	Edgeworth
0	0.0000	−1.9175	−0.3489	−3.7685
1	0.0020	−0.1325	0.0458	−0.3679
2	0.0386	−0.0022	0.0185	−0.0414
3	0.3956	0.0017	−0.0014	−0.0018
4	2.4186	−0.0010	−0.0043	0.0010
5	9.5396	−0.0010	−0.0022	0.0004
6	25.7238	−0.0004	−0.0007	0.0001
7	50.0000	0.0000	−0.0004	0.0000

例 5 (負の超幾何分布の場合). 負の超幾何分布 $NH(n, M, L)$ は，タイプ $(0,1,1,1)$，$\theta = 1$ の一般超幾何分布であり，その確率関数は

$$p_X(x) = \frac{M!\ (L-M)!\ (n+x-1)!\ (L-n-x)!}{L!\ (n-1)!\ x!\ (M-n)!\ (L-M-x)!}$$
$$(x = 0, 1, \ldots, \min(L-n,\ L-M))$$

である．このとき，$b = L - n + 1, c = n$，$d = L - M + 1$ とすると

$$p_X(x) = K\frac{c(x)\ \bar{d}[x]}{x!\ \bar{b}[x]}$$

と表される．ただし，K はある定数とする．いま，$\beta = (L-n)/N, \gamma = (n-1)/N$，$\delta = (L-M)/N$ とおくと，(3.2) は

$$\frac{(\gamma + \mu)(\delta - \mu)}{\mu(\beta - \mu)} = 1$$

となり，この解として $\mu = \gamma\delta/(\beta + \gamma - \delta)$ が一意的に定まる．これを用いて $p_X(x)$ の真値とその近似 (3.10), (3.11) と Edgeworth 近似，さらに下側確率の真値とその近似 (3.13) と各点確率の近似式 (3.10) の和をとったものと Edgeworth 近似を数値的に比較を行った (表 4.5.1, 4.5.2 参照)．その結果，各点確率については近似式 (3.10) は (3.11) と Edgeworth 近似より精確であることが分かる．また，下側確率についても近似式 (3.13) の精確さを読みとることができる．

表 4.5.1 負の超幾何分布 $NH(40, 20, 10)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)	Edgeworth
0	0.0218	——	−0.6406	−0.2567
1	0.1453	0.0102	0.0591	0.2651
2	0.5236	0.0400	0.0418	0.0791
3	1.3464	0.0128	0.0078	0.0062
4	2.7552	−0.0031	−0.0057	−0.0158
5	4.7474	−0.0080	−0.0061	−0.0158
6	7.1210	−0.0079	−0.0019	−0.0086
7	9.4947	−0.0068	0.0019	−0.0015
8	11.4040	−0.0064	0.0030	0.0027
9	12.4407	−0.0066	0.0020	0.0044
10	12.3814	−0.0067	0.0003	0.0045
11	11.2559	−0.0065	−0.0007	0.0037
12	9.3305	−0.0064	−0.0002	0.0016
13	7.0178	−0.0070	0.0017	−0.0029
14	4.7474	−0.0082	0.0047	−0.0096
15	2.8484	−0.0081	0.0090	−0.0155
16	1.4836	−0.0023	0.0154	−0.0128
17	0.6483	0.0151	0.0213	0.0136
18	0.2244	0.0419	−0.0024	0.0923
19	0.0551	−0.0068	−0.2648	0.2810
20	0.0073	——	——	0.6311

表 4.5.2 負の超幾何分布 $NH(40, 20, 10)$ の下側確率 $P\{X \leq x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)	Edgeworth
0	0.0218	——	−0.6406	−0.2567
1	0.1671	−0.5636	−0.0322	0.1971
2	0.6907	−0.0913	0.0239	0.1077
3	2.0371	−0.0148	0.0133	0.0406
4	4.7923	−0.0033	0.0023	0.0082
5	9.5396	−0.0022	−0.0019	−0.0038
6	16.6607	−0.0019	−0.0019	−0.0058
7	26.1554	−0.0013	−0.0005	−0.0043
8	37.5593	−0.0008	0.0006	−0.0021
9	50.0000	−0.0005	0.0009	−0.0005
10	62.3814	−0.0004	0.0008	0.0005

例 6 (一般超幾何分布の場合). 一般超幾何分布で, (i) タイプ $(1,0,0,2)$, $\theta=1.5$, 2 と (ii) タイプ $(0,1,1,1)$ $\theta=2$ の場合について考える. まず (i) の場合, 確率関数は

$$p_X(x)=K\frac{\prod_{j=1}^{2}\bar{d}_j[x]}{x!\,a(x)}\theta^x \qquad (x=0,1,\ldots,\min\{d_1,\ d_2\}),$$

と表される. ただし, K はある定数とする. また, (3.2) より方程式

$$\frac{(\delta_1-\mu)(\delta_2-\mu)}{\mu(\alpha-\mu)}\theta_0=1$$

の μ の解が存在する. また, (ii) においても同様に μ を求めることができる. これを用いて $p_X(x)$ の真値とその近似 (3.10), (3.11) と Edgeworth 近似, さらに下側確率の真値とその近似 (3.13) と各点確率の近似式 (3.10) の和をとったものと Edgeworth 近似を数値的に比較を行った (表 4.6.1 ~ 4.6.6 参照). その結果, 各点確率については近似式 (3.10) は (3.11) と Edgeworth 近似より精確であることが分かる. また, 下側確率についても近似式 (3.13) の精確さを読みとることができる.

表 4.6.1 タイプ $(1,0,0,2)$, $\theta=1.5$ の一般超幾何分布 $(a,b,c,d_1,d_2)=(11,0,0,31,21)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)
0	0.0000	201.1810	—
1	0.0000	36.3066	—
2	0.0000	10.5663	−0.6873
3	0.0005	3.6016	−0.2168
4	0.0063	1.2169	−0.0527
5	0.0526	0.3490	0.0100
6	0.3081	0.0669	0.0277
7	1.3051	0.0090	0.0230
8	4.0647	0.0153	0.0100
9	9.4131	0.0217	−0.0010
10	16.3082	0.0165	−0.0045
11	21.1794	0.0103	−0.0012
12	20.5777	0.0119	0.0035
13	14.8655	0.0178	0.0033
14	7.8973	0.0170	−0.0049
15	3.0326	0.0070	−0.0177
16	0.8201	0.0185	−0.0247
17	0.1501	0.1725	−0.0066
18	0.0174	0.9184	0.0688
19	0.0011	4.5290	0.2552
20	0.0000	32.2268	0.4710

表 4.6.2 タイプ $(1,0,0,2)$, $\theta = 1.5$ の一般超幾何分布 $(a,b,c,d_1,d_2) = (11,0,0,31,21)$ の下側確率 $P\{X \leq x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)
3	0.0005	0.0974	−0.2434
4	0.0069	0.0506	−0.0678
5	0.0595	0.0320	0.0010
6	0.3676	0.0249	0.0234
7	1.6727	0.0204	0.0231
8	5.7374	0.0151	0.0138
9	15.1505	0.0090	0.0046
10	31.4586	0.0035	−0.0001
11	52.6381	0.0000	−0.0006

表 4.6.3 タイプ $(1,0,0,2)$, $\theta = 2$ の一般超幾何分布 $(a,b,c,d_1,d_2) = (11,0,0,31,21)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)
2	0.0000	150.6610	—
3	0.0000	61.5599	−0.9488
4	0.0007	26.0501	−0.3560
5	0.0073	10.2095	−0.0528
6	0.0573	3.2430	0.0792
7	0.3238	0.6355	0.1022
8	1.3447	0.0174	0.0674
9	4.1520	0.0517	0.0197
10	9.5910	0.1104	−0.0105
11	16.6079	0.0614	−0.0139
12	21.5147	0.0032	−0.0004
13	20.7232	0.0221	0.0117
14	14.6789	0.0839	0.0080
15	7.5156	0.0843	−0.0136
16	2.7100	0.0031	−0.0423
17	0.6613	0.1775	−0.0550
18	0.1023	2.2546	−0.0124
19	0.0089	15.2547	0.1742
20	0.0003	127.1020	0.8425

表 4.6.4 タイプ $(1,0,0,2)$, $\theta = 2$ の一般超幾何分布 $(a,b,c,d_1,d_2) = (11,0,0,31,21)$ の下側確率 $P\{X \le x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)
5	0.0080	0.0327	−0.0826
6	0.0654	−0.0139	0.0593
7	0.3892	−0.0211	0.0950
8	1.7339	−0.0152	0.0736
9	5.8858	−0.0079	0.0356
10	15.4769	−0.0030	0.0070
11	32.0847	−0.0004	−0.0038
12	53.5994	0.0007	−0.0025

表 4.6.5 タイプ $(0,1,1,1)$, $\theta = 2$ の一般超幾何分布 $(a,b,c,d) = (0,47,9,21)$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.11)	GHG approx. (3.10)
0	0.0067	4.2192	—
1	0.0524	0.8174	−0.0727
2	0.2213	0.2038	0.0452
3	0.6638	0.0434	0.0408
4	1.5747	0.0068	0.0261
5	3.1193	0.0070	0.0139
6	5.3256	0.0128	0.0048
7	7.9885	0.0137	−0.0012
8	10.6513	0.0094	−0.0036
9	12.7068	0.0041	−0.0024
10	13.5997	0.0020	0.0008
11	13.0502	0.0043	0.0037
12	11.1859	0.0088	0.0036
13	8.5033	0.0116	−0.0009
14	5.6689	0.0098	−0.0086
15	3.2596	0.0049	−0.0150
16	1.5772	0.0052	−0.0130
17	0.6185	0.0310	0.0058
18	0.1848	0.1364	0.0377
19	0.0375	0.5316	−0.0368
20	0.0039	2.9176	—

表 4.6.6 タイプ $(0,1,1,1)$, $\theta=2$ の一般超幾何分布 $(a,b,c,d)=(0,47,9,21)$ の下側確率 $P\{X \le x\}$ の真値と近似式との相対誤差

x	True value (%)	GHG approx. (3.13)	Sum of GHG (3.10)
0	0.0067	−0.6935	—
1	0.0591	0.1509	−0.1173
2	0.2804	0.1233	0.0151
3	0.9442	0.0687	0.0278
4	2.5189	0.0357	0.0219
5	5.6382	0.0189	0.0150
6	10.9638	0.0107	0.0095
7	18.9523	0.0064	0.0055
8	29.6036	0.0038	0.0028
9	42.3104	0.0020	0.0015
10	55.9101	0.0008	0.0013

上記より，ここで求めた近似式 (3.10) が，[Ta84] での近似式 (3.11) と Edgeworth 展開より比較的精確な結果を与えていることから，(3.11) を改良しているといえるであろう.

参考文献

[D72] Dacey, M. F. (1972). A family of discrete probability distributions defined by the generalized hypergeometric series. *Sankhyā* **B 34**, 234–250

[HA00] Hida, E. and Akahira, M. (2000). *An approximation to the generalized hypergeometric distribution.* Submitted for publication.

[JKK92] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions* (2nd ed.). Wiley, New York.

[K68] Kemp, A. W. (1968). A wide class of discrete distributions and the associated differential equations. *Sankhyā* **A 30**, 401–410.

[M73] Molenaar, W. (1973). *Approximations to the Poisson, Binomial and Hyper-Geometric Distribution Functions.* Mathematical Centre, Amsterdam.

[SO94] Stuart, A. and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics Volume 1: Distribution Theory* (6th ed.). Edward Arnold, London.

[SS81] Sibuya, M. and Shimizu, R. (1981). The generalized hypergeometric distributions. *Ann. Inst. Statist. Math.*, **33 A**, 177–190.

[Ta84] 竹内 啓 (1984). 一般超幾何分布. 応用統計学 **13**, 83–101.

[Tr83] Tripathi, R. C. (1983). *Kemp families of distributions.* Encyclopedia of Statistical Sciences Vol. 4.