統計数理（2004）
第 52 巻 第 2 号 367–380

特集「環境科学と統計科学の融合に向けて」
［原著論文］

測定の最小の刻み幅を考慮した一般化負の超幾何分布モデル

岩瀬 晃盛[1]・金藤 浩司[2,3]・岡田 光正[1]

（受付 2004 年 3 月 12 日；改訂 2004 年 9 月 21 日）

要　　旨

観測精度が無限に高い状況に対応する母集団分布として想定される連続分布においては，この連続分布に収束する離散分布は一つとは限らない．例えば従来は直感的に逆ガウス型分布が母集団分布として想定されてデータ解析がなされていた現象においても，データの記述の観点から本来は母集団分布としては離散分布が想定される場合がある．この場合，想定される離散分布としては一般化されたポアソン分布や一般化された負の超幾何分布がその候補となり得る．

本稿では，離散分布である一般化された負の超幾何分布を新たに提案した．さらに，その分布から連続分布である一般化されたベータ分布及び一般化されたべき逆ガウス型分布も新たに導出した．

キーワード：一般化されたベータ分布，一般化されたべき逆ガウス型分布，検出下限，変動係数．

1. はじめに

ダイオキシン類の例で述べれば或る物質の検出下限は $0.02\,\mathrm{pg/m^3}$ となっている．ここでの $0.02\,\mathrm{pg/m^3}$ はこれに達しない濃度は検出できない（またはその必要がない）ことを意味し，これを越える濃度の記述においてはこの濃度よりもさらに細かい記述は存在しない．つまりデータの記述の観点から言えば，データの最小の刻みの幅は $0.02\,\mathrm{pg/m^3}$ であることになる．検出下限と述べたときに，この量以下（または未満）は計測できないけれども，これ以上の量は望むだけの有効桁で計測できる場合を想定していることが多い．象徴的な例としては，液体の熱膨張を利用した普通見かける温度計などは，或る温度以下とか以上は計測できない．しかし上記のダイオキシン類の計測の場合は検出下限とデータの刻みの幅は一致している．この場合でのデータの最小の刻み幅は経済的理由または技術的理由によって定まる．

最小の刻み幅で記述されたデータは，離散量である．従ってこれらのデータが抽出されたであろう想定されるべき母集団の分布は離散分布である．

この考え方は一般的な考え方と異なる．一般的には母集団分布に連続分布を想定して，最小の刻み幅をもって記述された離散的なデータはその連続分布に従う確率変数の実現値であると見なす．しかしながら連続分布に従う確率変数の実現値であるならば無限桁でデータは表現さ

[1] 広島大学大学院 工学研究科：〒739–8527 広島県東広島市鏡山 1–4–1
[2] 統計数理研究所：〒106–8569 東京都港区南麻布 4–6–7
[3] 総合研究大学院大学 複合科学研究科統計科学専攻

れるはずである．もちろん無限桁で表現されたデータなど見られるはずもない．そこで本来は無限桁であるけれども(経済的または技術的な理由により)適当な最小の刻み幅で打ち切って「近似的」に表現していると解釈する．この考え方は最初に母集団分布があって，これに従属した形でデータが捉えられている．

これに反して，先に述べたような「最小の刻み幅で記述されたデータは離散量である．従ってこれらのデータが抽出されたであろう想定されるべき母集団分布は離散分布である」との考え方は，データが在ってそれに伴って母集団分布が想定されていることに注目されるべきである．

さて，データの記述の最小の刻み幅は量である．量とは数と単位との組である．先のダイオキシン類の場合では単位は $\mathrm{pg/m^3}$ である．従って，ここでの離散分布とは量の離散分布であることに注意されなければならない．

計測装置の検出能力が高まって且つその必要があればデータの最小の刻み幅は更に小さくなる．仮に無限に検出能力が高まってデータの最小の刻み幅が無限に小さくなっていく状況を想定すれば，これらのデータが抽出されるであろう想定されるべき母集団分布は連続分布となるべきである．

この理由によって，データの最小の刻み幅を限りなく小さくしていく「操作」に対応して，当初想定された離散分布が連続分布に何らかの意味で収束していくことを示さなければならない．このことが実現してはじめて収束した連続分布に従う確率変数の実現値として「近似的」な離散的なデータの記述が解釈されるものである．

データの最小の刻み幅を考慮した離散分布を想定し，この離散分布の(刻み幅を無限に小さくするという意味での)極限分布としての連続分布を母集団分布と仮定してデータを捉えるということの実際的な意味は，例えば母数の推定などにある．最小の刻み幅を持ったデータを用いて推定される母数はあくまでも離散分布の母数である．しかしこの推定値は仮に観測精度を高めていけば仮定された連続分布の母数の推定値になっているという関係をきちんと認識したことになっている．

2. 二項分布

単位を有する確率変数 X の確率関数 $p_X(x)$ が次式で与えられるとき，本稿では X は二項分布に従うと言う．

$$(2.1)\qquad p_X(x) = \begin{pmatrix} \frac{1-p}{c^2p} \\ \frac{(1-p)x}{c^2\mu} \end{pmatrix} p^{\frac{(1-p)x}{c^2\mu}} (1-p)^{\frac{1-p}{c^2p}-\frac{(1-p)x}{c^2\mu}}$$

$$x = 0\frac{c^2}{1-p}\mu, 1\frac{c^2}{1-p}\mu, 2\frac{c^2}{1-p}\mu\,, 3\frac{c^2}{1-p}\mu, \ldots, \frac{1-p}{c^2p}\frac{c^2}{1-p}\mu$$

ここで，μ は X と同じ単位を持つ量であり，c および p は無名数で，$\frac{1-p}{c^2p}=$ 自然数，$0<\mu<\infty, 0<c<\infty, 0<p<1$ である．明らかに $\frac{X}{\mu}$ の分布は μ に依存しないから，μ は X の尺度母数である．(2.1)式は通常見かける二項分布の確率関数の式よりも一見すると複雑に見えるけれども，観点を変えれば以下に述べる意味によって簡単な式であることがわかる．

確率変数 X の最小の刻み幅の量である $\frac{c^2}{1-p}\mu$ は任意に定められるものであり，先のダイオキシン類での例で言えば $\frac{c^2}{1-p}\mu = 0.02\,\mathrm{pg/m^3}$ となるものである．このように見かけ上は μ, c, p の 3 母数であるけれども，データの定められた最小の刻み幅での制限が加わるので実際は 2 母数である．

ここで X の母平均，母分散，母変動係数を計算すると

$$E[X] = \mu\,, \quad V[X] = c^2\mu^2\,, \quad CV[X] = \frac{\sqrt{V[X]}}{E[X]} = c$$

となる．

計測器の精度を上げて最小の刻み幅を小さくして計測しても，想定している母集団の中心とかばらつきとかは変化しないはずである．なぜならば，測定の最小の刻み幅は測定者側の都合で定まるものであり，対象としている現象とは無関係であるはずだからである．つまり今の場合についてこのことを述べれば，中心を母平均 $E[X]$，ばらつきを母変動係数 $CV[X]$ と定めれば，$E[X]$ も $CV[X]$ も変化させないで最小の刻み幅を小さくできる場合を考えれば，それは $p \to 0$ の場合だけである．

X の積率母関数は

$$\left((1-p) + p\exp\left(\frac{c^2\mu t}{1-p}\right)\right)^{\frac{1-p}{c^2 p}}$$

であるから，ここで $p \to 0$ とすると

$$\exp\left(\frac{1}{c^2}\left(\exp(c^2\mu t) - 1\right)\right)$$

となる．これは母平均および母変動係数がそれぞれ元の μ および c のままのポアソン分布の積率母関数である．つまり $p \to 0$ としたときの確率関数は

$$p_X(x) = \frac{1}{(\frac{x}{c^2\mu})!}\left(\frac{1}{c^2}\right)^{\frac{x}{c^2\mu}} \exp\left(-\frac{1}{c^2}\right)$$

$$x = 0c^2\mu, 1c^2\mu, 2c^2\mu, 3c^2\mu, \ldots, \quad 0 < c < \infty\,, \quad 0 < \mu < \infty$$

である．

このポアソン分布では $x = 0c^2\mu, 1c^2\mu, 2c^2\mu, 3c^2\mu, \ldots$ の上で定義されている．或る地点を車両が単位時間内に通過する台数を考えると，通常は何 [台/時間] というようにデータを記録する．しかしデータを記述するときに何 [10 台/時間] を単位にしてもかまわない．このときはデータは 0.1[10 台/時間] の刻みの幅で記述される．従って，この場合には

$$c^2\mu = 0.1[10\ \text{台/時間}]$$

という制約が母数に加わっている．従って見かけ上は 2 母数であっても実際は 1 母数である．例えば $\mu = 0.9[10$ 台/時間] であれば $c = \frac{1}{3}$ となり，一方で通常の [台/時間] を単位に採用すれば $\mu = 9$[台/時間] であるから $c = \frac{1}{3}$ となる．つまり c の値は μ の単位の取り方に依存しないことがわかる．この事実は母数 c が母変動係数であること，そして一般に母変動係数は無名数であることから自明であるけれども，直接的にデータの最小の刻み幅がどのように記述されているかに依存しないことを示したものになっている．

さて，ポアソン分布において分布の中心もばらつきも変化させないでデータの刻み幅を小さくできる場合は存在しない．つまり $c \to 0$ の場合や $\mu \to 0$ の場合は，どちらの場合も分布の中心もばらつきも変化させないという規則性に反する．

3. 一般化されたポアソン分布

先のポアソン分布は 1 母数であった．そこで 2 母数の一般化されたポアソン分布を考える．Consul (1989) が導入した generalized Poisson distribution model を変形して，無名数母数 λ を

導入することによって次のような一般化されたポアソン分布を考える．

$$\text{(3.1)}\qquad \frac{\frac{1}{(1-\lambda)c^2}\left(\frac{1}{(1-\lambda)c^2}+\frac{\lambda x}{(1-\lambda)^2c^2\mu}\right)^{\frac{x}{(1-\lambda)^2c^2\mu}-1}}{\left(\frac{x}{(1-\lambda)^2c^2\mu}\right)!}\exp\left(-\frac{1}{(1-\lambda)c^2}-\frac{\lambda x}{(1-\lambda)^2c^2\mu}\right)$$

$$x=0(1-\lambda)^2c^2\mu, 1(1-\lambda)^2c^2\mu, 2(1-\lambda)^2c^2\mu,\ldots,\quad 0<\mu<\infty,\quad 0<c<\infty,\quad 0\le\lambda<1\,.$$

このときの母平均，母分散，母変動係数は

$$E[X]=\mu\,,\quad V[X]=c^2\mu^2\,,\quad CV[X]=c$$

となる．分布の中心もばらつきも変えないで，X の刻み幅を変化できるのは λ のみであり，$\lambda\to 0$ の場合は刻み幅を小さくする場合に該当せず，先のポアソン分布に一致する．$\lambda\to 1$ の場合のみ刻み幅を小さくする場合に該当する．従って以下でこの場合を考察する．

式(3.1)で示された一般化されたポアソン分布の積率母関数を対数変換した関数として記号 $\psi_X(t)$ を採用すれば

$$\psi_X(t)=\sum_{j=1}^{\infty}\frac{k_j(1-\lambda)^{2j}c^{2j}(\mu t)^j}{j!}$$

となる．ここで，$k_j, j=1,2,3,4,5,6$ は Consul (1989)の結果を用いることにより

$$\begin{aligned}
k_1&=(1-\lambda)^{-2}c^{-2}\\
k_2&=(1-\lambda)^{-4}c^{-2}\\
k_3&=(1-\lambda)^{-6}(1+2\lambda)c^{-2}\\
k_4&=(1-\lambda)^{-8}(1+8\lambda+6\lambda^2)c^{-2}\\
k_5&=(1-\lambda)^{-10}(1+22\lambda+58\lambda^2+24\lambda^3)c^{-2}\\
k_6&=(1-\lambda)^{-12}(1+55\lambda+328\lambda^2+444\lambda^3+120\lambda^4)c^{-2}
\end{aligned}$$

を得る．ここで刻み幅を小さくする場合に該当する極限操作として $\lambda\to 1$ とすれば

$$\begin{aligned}
\psi_X(t)&=\frac{1}{1!}\mu t+\frac{c^2}{2!}(\mu t)^2+\frac{3c^4}{3!}(\mu t)^3+\frac{15c^6}{4!}(\mu t)^4+\frac{105c^8}{5!}(\mu t)^5+\frac{945c^{10}}{6!}(\mu t)^6+\cdots\\
&=\frac{1}{c^2}\left(-\frac{1}{2}\frac{-2c^2\mu t}{1!}-\frac{1}{2}\left(-\frac{1}{2}\right)\frac{(-2c^2\mu t)^2}{2!}-\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{-(2c^2\mu t)^3}{3!}\right.\\
&\quad\left.-\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\frac{(-2c^2\mu t)^4}{4!}-\cdots\right)\\
&=\frac{1}{c^2}\left(1-\sqrt{1-2c^2\mu t}\right)
\end{aligned}$$

となる．これは母平均が μ，母変動係数が c の逆ガウス型分布 $IG(\mu,c^2)$ の積率母関数を対数変換した関数に一致する．つまり確率素分は

$$\frac{1}{\sqrt{2\pi c^2}}\left(\frac{x}{\mu}\right)^{-3/2}\exp\left(-\frac{1}{2c^2}\left(\sqrt{\frac{x}{\mu}}-\sqrt{\frac{\mu}{x}}\right)^2\right)\frac{dx}{\mu}\,,$$

$$0<x<\infty,\quad 0<\mu<\infty,\quad 0<c<\infty$$

である．これより(3.1)で与えられる一般化されたポアソン分布の刻みの幅を限りなく小さくしていったときの極限の分布として逆ガウス型分布が得られる．この事実そのものは既に Consul

(1989)が示したものであるけれども，極限の意味付けおよび尺度母数の導入などについて注目されたい．

或る最小の刻み幅を伴って記述されたデータが抽出されたであろう母集団分布としてここでの一般化されたポアソン分布を想定し，最小の刻み幅が仮に限りなく小さくなる状況を考えれば母集団分布は逆ガウス型分布に収束し，連続分布であるその逆ガウス型分布に従う母集団からのデータの「近似的」な記述としてその最小の刻み幅でのデータが解釈されるものである．従って，従来母集団分布として逆ガウス型分布が想定された下でのデータ解析の事例においては，データの在り方から考えれば本来的には母集団の分布モデルの一つして一般化されたポアソン分布(3.1)が想定されるべきものであって，上の極限操作の裏付けがあって初めて逆ガウス型分布が想定される母集団からの近似的な表現に基づく標本であると主張できることになる．

4. 負の二項分布

負の二項分布の確率関数は

$$p_X(x) = \binom{-\frac{1}{c^2(1-p)}}{\frac{x}{c^2p\mu}} p^{\frac{1}{c^2(1-p)}} (-1+p)^{\frac{x}{c^2p\mu}}\,, \tag{4.1}$$

$$x = 0c^2p\mu, 1c^2p\mu, 2c^2p\mu, 3c^2p\mu, \ldots, \quad 0<\mu<\infty\,, \quad 0<c<\infty\,, \quad 0<p<1$$

で与えられる．

ここで示された確率関数の表現は良く知られた「負の二項分布」の確率関数の表現と比較して煩雑である印象を与え，違和感を抱かせる可能性があるけれども，これは例えば正規分布 $N(\mu,\sigma^2)$ の確率素分の表現の方が $N(\mu,1)$ の確率素分の表現よりも煩雑であることに相当するものであり，違和感を持つ理由はない．

さて，この場合での母平均，母分散，母変動係数は

$$E[X]=\mu\,, \quad V[X]=c^2\mu^2\,, \quad CV[X]=c$$

となる．この場合にもデータの最初の刻み幅 $=c^2p\mu$ となる制約があるので見かけ上は 3 母数であるけれども実際は 2 母数である．この分布の中心もばらつきも変えないでデータの最小の刻み幅に相当する $c^2p\mu$ を小さくできるのは母数 p のみであり，極限 $p\to 0$ とすれば刻み幅は限りなく小さくなる．この分布の積率母関数は

$$m_X(t) = \left(\frac{1-(1-p)\exp(c^2p\mu t)}{p}\right)^{-\frac{1}{(1-p)c^2}}$$

となるから，$p\to 0$ を考えると

$$m_X(t) = (1-c^2\mu t)^{-\frac{1}{c^2}}$$

となる．これは変化させなかった元々の母平均と母変動係数を持つガンマ分布の積率母関数に他ならない．つまり確率素分は

$$\frac{(\frac{1}{c^2})^{\frac{1}{c^2}}}{\Gamma(\frac{1}{c^2})}\left(\frac{x}{\mu}\right)^{\frac{1}{c^2}-1}\exp\left(-\frac{1}{c^2}\frac{x}{\mu}\right)\frac{dx}{\mu}\,,$$

$$0<x<\infty\,, \quad 0<\mu<\infty\,, \quad 0<c<\infty$$

である．

負の二項分布においても，またガンマ分布においても母数 μ は同じ尺度母数の役割を担っていることに注意されたい．負の二項分布からのガンマ分布の導出は既に知られた事実であるけれども，分布の中心もばらつきも変えないでデータの最小の刻み幅を小さくしていったときの分布という観点からの一貫性に注意されたい．

5. 離散一様分布

無名数の非負の整数 N と正の量である μ を用いて，離散一様分布の確率関数を表現すると

$$p_X(x) = \frac{1}{N+1}\,, \quad x = 0\frac{2\mu}{N}, 1\frac{2\mu}{N}, 2\frac{2\mu}{N}, \ldots, N\frac{2\mu}{N} \tag{5.1}$$

となる．このとき

$$E[X] = \mu\,, \quad V[X] = \frac{1}{3}\left(1 + \frac{2}{N}\right)\mu^2\,, \quad CV[X] = \sqrt{\frac{1}{3}\left(1 + \frac{2}{N}\right)}$$

となる．このとき分布の中心もばらつきも完全に変化させないで刻みの幅に相当する $\frac{2\mu}{N}$ が限りなく小さくなる母数は存在しない．しかし N を 1 と比較して十分大きくしていくに従い刻みの幅は限りなく小さくなり，母平均は変化せず，母変動係数は有限な一定値である $\frac{1}{\sqrt{3}}$ に収束する．

X の積率母関数は

$$m_X(t) = \frac{\sinh\left(\frac{(N+1)2\mu t}{2N}\right)}{(N+1)\sinh\left(\frac{2\mu t}{2N}\right)}\exp(\mu t)$$

となる．双曲線正弦関数の加法定理より

$$m_X(t) = \exp(\mu t)\frac{\sinh(\mu t)}{(N+1)\tanh\left(\frac{2\mu t}{2N}\right)}$$

となるから，双曲線正接関数を展開することにより，無名数である μt を任意に固定して N を 1 と比較して十分大きくすると

$$m_X(t) = \frac{1}{\mu t}\exp(\mu t)\sinh(\mu t) = \frac{\exp(2\mu t) - 1}{2\mu t}$$

が得られ，これは X の確率素分が $\frac{dx}{2\mu}, 0 < x < 2\mu$ の連続一様分布であることを示している．

二項分布，負の二項分布および一般化されたポアソン分布の場合には分布の中心とばらつきを変化させないで刻み幅を小さくしていく母数に着目したのであるけれども，一様分布の例ではこの考え方を更に拡げて考えるべきであることを示している．即ち，離散分布の中心とばらつきを厳密に一定にしつつ刻みの幅を限りなく小さくすることによっての連続分布の導出よりも条件を緩くすることが可能で，離散分布の中心とばらつきが刻みの幅を限りなく小さくしていくときに或る有限な量や数に収束すれば良いこととなる．さらに共通にいえることは，刻み幅を限りなく小さくしていくために操作される母数は無名数であるという共通性を持っている．これは 1 と比較して十分小さいとか十分大きいとかの解釈を可能にしている．

6. 一般化された負の超幾何分布

負の超幾何分布の一つの一般化された分布として，確率関数が

$$p(x) = \frac{\binom{-\alpha}{x}\binom{-\beta}{n-x}}{\binom{-\alpha-\beta}{n}} \cdot \frac{{}_2F_1\left(\alpha+x,\gamma;\alpha+\beta+n;-\frac{1}{a}\right)}{{}_2F_1\left(\alpha,\gamma;\alpha+\beta;-\frac{1}{a}\right)} \tag{6.1}$$

$$x=0,1,2,\ldots,n\,,\quad 0<\alpha<\infty\,,\quad 0<\beta<\infty\,,\quad -\infty<\gamma<\infty\,,\quad 0<a<\infty$$

で与えられる分布を新たに導出する．ここで x は無名数であり，${}_2F_1(\alpha,\beta;\gamma;x)$ はガウスの超幾何関数である．この分布は次のように導出される．

任意の実数 γ および任意の正の実数 α,β,a について Gradshteyn and Ryzhik（2000）より

$$\frac{p^{\alpha-1}(1-p)^{\beta-1}(p+a)^{-\gamma}}{a^{-\gamma}B(\alpha,\beta){}_2F_1(\alpha,\gamma;\alpha+\beta;-\frac{1}{a})}dp\,,\quad 0<p<1$$

は定義領域が $(0,1)$ の一つの連続分布の確率素分を与える．ここで二項分布の確率関数

$$\binom{n}{x}p^x(1-p)^{n-x}\,,\quad x=0,1,2,\ldots,n\,,\quad 0<p<1$$

において，この二項分布を先の連続分布で混合する．即ち

$$\begin{aligned}&\binom{n}{x}\int_0^1 \frac{p^{\alpha+x-1}(1-p)^{\beta+n-x-1}(p+a)^{-\gamma}}{a^{-\gamma}B(\alpha,\beta){}_2F_1(\alpha,\gamma;\alpha+\beta;-\frac{1}{a})}dp\\ &\quad=\binom{n}{x}\frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)}\cdot\frac{{}_2F_1(\alpha+x,\gamma;\alpha+\beta+n;-\frac{1}{a})}{{}_2F_1(\alpha,\gamma;\alpha+\beta;-\frac{1}{a})}\end{aligned}$$

を得る．特に a として負の実数 $a=-\alpha\ (\alpha>0)$ を考えると

$$\binom{-\alpha}{n} = (-1)^n\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(1+n)},\quad \alpha>0$$

となる．これを用いて変形することにより一般化された負の超幾何分布 $p(x)$ が得られる．実際，特に $\gamma=0$ とすればガウスの超幾何関数の比の項は 1 となって負の超幾何分布の確率関数を得る．

7. 一般化された負の超幾何分布からの一般化されたベータ分布の導出

（6.1）式において，以降，簡単の為に $p(x)=g(x;\alpha,\beta,n)\cdot h(x;\alpha,\beta,\gamma,a,n)$ とする．
但し

$$g(x;\alpha,\beta,n) = \frac{\binom{-\alpha}{x}\binom{-\beta}{n-x}}{\binom{-\alpha-\beta}{n}}\text{（}=\text{負の超幾何分布の確率関数）},$$

$$h(x;\alpha,\beta,\gamma,a,n) = \frac{{}_2F_1(\alpha+x,\gamma;\alpha+\beta+n;-\frac{1}{a})}{{}_2F_1(\alpha,\gamma;\alpha+\beta;-\frac{1}{a})}$$

と置く．ここで $p = g \cdot h$ での g と h とを分けて考察する．

最初は g に関しての考察を行う．この負の超幾何分布に従う確率変数 X を次のように変換する．

$$Y = \frac{\mu}{n\lambda}X\,, \quad \text{ただし} \quad \alpha = \frac{1}{c^2}, \quad \beta = \frac{1}{\lambda c^2}$$

とする．ここで $0 < \mu < \infty, 0 < \lambda < \infty, 0 < c < \infty$ である．このとき負の超幾何分布の確率関数である $g(x; \alpha, \beta, n)$ は次の表現となる．

$$\frac{\begin{pmatrix} -\frac{1}{c^2} \\ n\lambda\frac{y}{\mu} \end{pmatrix} \begin{pmatrix} -\frac{1}{\lambda c^2} \\ n - n\lambda\frac{y}{\mu} \end{pmatrix}}{\begin{pmatrix} -\frac{1}{c^2}\left(1 + \frac{1}{\lambda}\right) \\ n \end{pmatrix}}\,, \quad y = 0\frac{\mu}{n\lambda}, 1\frac{\mu}{n\lambda}, \ldots, n\frac{\mu}{n\lambda}\,.$$

このときの母平均，母変動係数は

$$E[Y] = \frac{\mu}{1+\lambda}\,, \quad CV[Y] = \frac{c}{\sqrt{1+\lambda+\lambda c^2}}\sqrt{1 + \frac{1+\lambda}{n\lambda c^2}}$$

である．

ここで $n \to \infty$ を考えたとき，Y の母平均は n に依存せず $\frac{\mu}{1+\lambda}$ のまま変わらず，Y の母変動係数は有限な値である $\frac{c}{\sqrt{1+\lambda+\lambda c^2}}$ に収束するし，最小の刻み幅に相当する $\frac{\mu}{n\lambda}$ はゼロに収束する．つまり $g(x; \alpha, \beta, n)$ は $n \to \infty$ とすれば

$$\frac{1}{B(\frac{1}{c^2}, \frac{1}{\lambda c^2})}\left(\lambda\frac{y}{\mu}\right)^{\frac{1}{c^2}-1}\left(1 - \lambda\frac{y}{\mu}\right)^{\frac{1}{\lambda c^2}-1}\lambda\frac{dy}{\mu}\,, \quad 0 < \frac{y}{\mu} < \frac{1}{\lambda}$$

へ収束する．これは母平均が $\frac{\mu}{1+\lambda}$，母変動係数が $\frac{c}{\sqrt{1+\lambda+\lambda c^2}}$ のベータ分布の確率素分に他ならない．つまり負の超幾何分布を連続化した分布がベータ分布であることを示している．更にここで $\lambda \to 0$ とすると

$$\frac{(\frac{1}{c^2})^{\frac{1}{c^2}}}{\Gamma(\frac{1}{c^2})}\left(\frac{y}{\mu}\right)^{\frac{1}{c^2}-1}\exp\left(-\frac{1}{c^2}\frac{y}{\mu}\right)\frac{dy}{\mu}\,, \quad 0 < y < \infty$$

となり，これは $\lim_{\lambda \to 0}\frac{\mu}{1+\lambda} = \mu$，母変動係数が $\lim_{\lambda \to 0}\frac{c}{\sqrt{1+\lambda+\lambda c^2}} = c$ のガンマ分布であることがわかる．つまりベータ分布の定義領域を $(0, \infty)$ としたときの分布としてガンマ分布が得られるという既知の事実が再確認できる．

次に重み関数として解釈される $h(x; \alpha, \beta, \gamma, a)$ の考察を行う．変数変換は $\frac{\mu x}{n\lambda}$ であったから，$h(x; \alpha, \beta, \gamma, a)$ の分子は

$$\begin{aligned}
&\lim_{n \to \infty} {}_2F_1\left(\alpha + n\lambda\frac{y}{\mu}, \gamma; \alpha + \beta + n; -\frac{1}{a}\right) \\
&\quad = 1 + \gamma\left(\lambda\frac{y}{\mu}\right)\frac{1}{1!}\left(-\frac{1}{a}\right) + \gamma(\gamma+1)\left(\lambda\frac{y}{\mu}\right)^2\frac{1}{2!}\left(-\frac{1}{a}\right)^2 + \cdots \\
&\quad = \sum_{k=0}^{\infty}(\gamma)_k\left(-\frac{1}{a}\lambda\frac{y}{\mu}\right)^k\frac{1}{k!} \\
&\quad = {}_1F_0\left(\gamma; -\frac{1}{a}\lambda\frac{y}{\mu}\right)
\end{aligned}$$

$$= \left(1+\frac{1}{a}\lambda\frac{y}{\mu}\right)^{-\gamma}, \quad \text{ただし}\ 0<\lambda\frac{y}{\mu}<1,\ (a)_n = a(a+1)\cdots(a+n-1)$$

となって，結果的に

$$h(x;\alpha,\beta,\gamma,a,n) = \frac{\left(1+\frac{1}{a}\lambda\frac{y}{\mu}\right)^{-\gamma}}{{}_2F_1\left(\frac{1}{c^2},\gamma;\frac{1}{c^2}+\frac{1}{\lambda c^2};-\frac{1}{a}\right)}$$

となる．

さて，$p(x)=g(x;\alpha,\beta,n)\cdot h(x;\alpha,\beta,\gamma,a,n)$ であったから，一般化された負の超幾何分布を $n\to\infty$ とすることにより（つまり最小の刻みの幅を限りなく小さくしていくことにより）連続化した分布として次式を得る．

$$\frac{\left(\lambda\frac{y}{\mu}\right)^{\frac{1}{c^2}-1}\left(1-\lambda\frac{y}{\mu}\right)^{\frac{1}{\lambda c^2}-1}}{B\left(\frac{1}{c^2},\frac{1}{\lambda c^2}\right)}\cdot\frac{\left(1+\frac{1}{a}\lambda\frac{y}{\mu}\right)^{-\gamma}}{{}_2F_1\left(\frac{1}{c^2},\gamma;\frac{1}{c^2}+\frac{1}{\lambda c^2};-\frac{1}{a}\right)}\lambda\frac{dy}{\mu},\ 0<\frac{y}{\mu}<\frac{1}{\lambda} \tag{7.1}$$

ちなみに，無名数化された確率変数として，更に見易くするために

$$x=\lambda\frac{y}{\mu}, \alpha=\frac{1}{c^2}, \beta=\frac{1}{\lambda c^2}$$

と変換した確率素分を示せば，一般化された負の超幾何分布を連続化した分布として，新たに一般化されたベータ分布の確率素分が次式のように得られる．

$$f(x;\alpha,\beta,\gamma,a)dx = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\cdot\frac{(a+x)^{-\gamma}}{a^{-\gamma}\ {}_2F_1(\alpha,\gamma;\alpha+\beta;-\frac{1}{a})}dx, \quad 0<x<1 \tag{7.2}$$

$n\to\infty$ として導出されたこの一般化されたベータ分布が確率素分の条件を満たしていることは直接積分を実行することによっても確認される．

8. 一般化されたベータ分布からの一般化されたべき逆ガウス型分布の導出

前節で尺度母数を入れた一般化された負の超幾何分布から尺度母数の入った一般化されたベータ分布を導出し，最後に無名数化された一般化されたベータ分布の確率素分を示した．この定義領域は $(0,1)$ である．定義域を $(0,b)$（ただし b は無名数）とすれば

$$\frac{x^{\alpha-1}(b-x)^{\beta-1}}{b^{\alpha+\beta-1}B(\alpha,\beta)}\cdot\frac{(x+a)^{-\gamma}}{a^{-\gamma}\ {}_2F_1(\alpha,\gamma;\alpha+\beta;-\frac{b}{a})}dx, \quad 0<x<b, \quad 0<b<\infty \tag{8.1}$$

となり，更に定義領域を (A,B) とした一般化されたベータ分布は

$$y=(B-A)x+A, \quad 0<x<1, \quad a=\frac{A}{B-A}$$

とすることにより

$$\frac{(y-A)^{\alpha-1}(B-y)^{\beta-1}}{(B-A)^{\alpha+\beta-1}B(\alpha,\beta)}\cdot\frac{y^{-\gamma}}{A^{-\gamma}\ {}_2F_1(\alpha,\beta;\alpha+\beta;1-\frac{B}{A})}dy, \quad A<y<B, \quad 0<A<B<\infty$$

となる．これに更にべき変換を加えて，$\xi>0$ の条件の下で

$$t^{\xi}=(B-A)x+A, \quad 0<x<1, \quad a=\frac{A}{B-A}, \quad 0<A<B<\infty$$

と変換すれば

$$\frac{(t^{\xi}-A)^{\alpha-1}(B-t^{\xi})^{\beta-1}}{(B-A)^{\alpha+\beta-1}B(\alpha,\beta)}\cdot\frac{\xi t^{\xi(-\gamma+1)-1}}{(B-A)^{\alpha+\beta-1}A^{-\gamma}\ {}_2F_1(\alpha,\gamma;\alpha+\beta;1-\frac{B}{A})}dt, \quad A^{\frac{1}{\xi}}<x<B^{\frac{1}{\xi}}$$

となる．

この最後のべき変換を加えた一般化されたベータ分布の確率素分において，特に制約式

$$\alpha=\beta=\frac{1}{2\lambda\xi^2c^2}\,,\quad \gamma=\beta+\frac{1}{2}\,,\quad AB=1\,,\quad \sqrt{B}-\sqrt{A}=\frac{1}{\sqrt{\lambda}}$$

を与える．この制約の下では Prudnikov et al. (1990)により

$$\begin{aligned}{}_2F_1\left(\alpha,\gamma;\alpha+\beta;1-\frac{A}{B}\right)&={}_2F_1\left(\beta,\beta+\frac{1}{2};2\beta;1-B^2\right)\\&={}_2F_1\left(\beta,\beta+\frac{1}{2};\beta+\left(\beta+\frac{1}{2}\right)-\frac{1}{2};1-B^2\right)\\&=(B^2)^{-\frac{1}{2}}\,{}_2F_1\left(2\beta-1,2\beta;2\beta;\frac{1-B}{2}\right)\\&=\frac{1}{B}\,{}_1F_0\left(2\beta-1;\frac{1-B}{2}\right)=\frac{1}{B}\left(\frac{1+B}{2}\right)^{1-2\beta}\end{aligned}$$

となり，また，

$$A=\frac{1+2\lambda-\sqrt{1+4\lambda}}{2\lambda}>0\,,\quad B=\frac{1+2\lambda+\sqrt{1+4\lambda}}{2\lambda}>0\,,\quad B-A=\frac{\sqrt{1+4\lambda}}{\lambda}>0$$

となって，これらを代入することによって次の結果を得る．

結果． 一般化されたベータ分布において

$$t^\xi=(B-A)x+A\,,\quad 0<x<1\,,\quad 0<\xi<\infty\,,\quad a=\frac{A}{B-A}\,,\quad 0<A<B<\infty\,,$$

$$\alpha=\beta=\frac{1}{2\lambda\xi^2c^2}\,,\quad \gamma=\beta+\frac{1}{2}\,,\quad AB=1\,,\quad \sqrt{B}-\sqrt{A}=\frac{1}{\lambda}$$

なる変換および母数への制約を加えることによって，新たに次の 3 母数の一般化されたべき逆ガウス型分布の確率素分 $f(t;c,\xi,\lambda)dt$ を得る．

$$\xi\sqrt{\lambda}\frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\frac{1}{2\lambda\xi^2c^2}+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2\lambda\xi^2c^2}\right)}t^{-1-\frac{\xi}{2}}\left(1-\lambda(t^{\frac{\xi}{2}}-t^{-\frac{\xi}{2}})^2\right)^{\frac{1}{2\lambda\xi^2c^2}-1}dt\,,\quad A^{\frac{1}{\xi}}<t<B^{\frac{1}{\xi}}\,. \tag{8.2}$$

ただし，

$$A=\frac{1+2\lambda-\sqrt{1+4\lambda}}{2\lambda}\,,\quad B=\frac{1+2\lambda+\sqrt{1+4\lambda}}{2\lambda}$$

である．尚，尺度母数を導入すれば 4 母数となるものである．

図 1，2，3 に一般化されたべき逆ガウス型分布を示している．

特に，この結果において，$\lim_{\lambda\to 0}A=0,\lim_{\lambda\to 0}B=\infty$ であることに注意すれば，$\lambda\to 0$ のときは，上式で与えられた確率素分は，岩瀬・平野(1990)が提案したべき逆ガウス型分布の確率素分

$$\frac{1}{\sqrt{2\pi}c}t^{-1-\frac{\xi}{2}}\exp\left(-\frac{1}{2(\xi c)^2}(t^{\frac{\xi}{2}}-t^{-\frac{\xi}{2}})^2\right)dt\,,\quad 0<t<\infty$$

である．ただし，$0<c<\infty,-\infty<\xi<\infty$ である．

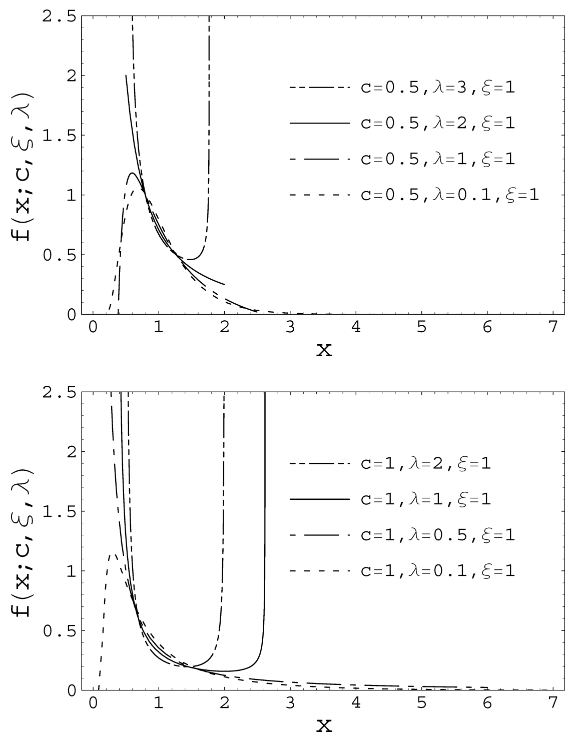

図 1. 一般化されたべき逆ガウス型分布の密度関数(1).

既知の結果として上式でさらに特に $\xi = 1$ としたときには逆ガウス型分布 $IG(1, c^2)$ の確率素分

$$\frac{1}{\sqrt{2\pi}c} t^{-\frac{3}{2}} \exp\left(-\frac{1}{2c^2}\left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)^2\right) dt, \quad 0 < t < \infty$$

を与え，特に $\xi \to 0$ のときには対数正規分布 $LN(1, c^2)$ の確率素分

$$\frac{1}{\sqrt{2\pi}c} t^{-1} \exp\left(-\frac{1}{2c^2}(\log t)^2\right) dt, \quad 0 < t < \infty$$

を与える.

9. 考察

本稿では，離散分布である一般化された負の超幾何分布を提案し，これから連続分布である一般化されたベータ分布及び一般化されたべき逆ガウス型分布を新たに導出した.

観測精度が無限に高い状況に対応する母集団分布として想定される連続分布においても，この連続分布に収束する離散分布は一つとは限らない．例えば従来は直感的に逆ガウス型分布が母集団分布として想定されてデータ解析がなされていた現象について，データの記述の観点から本来的には離散分布が想定されるべきであるが，その想定される離散分布としては一般化されたポアソン分布や一般化された負の超幾何分布が候補となり得ることを意味している.

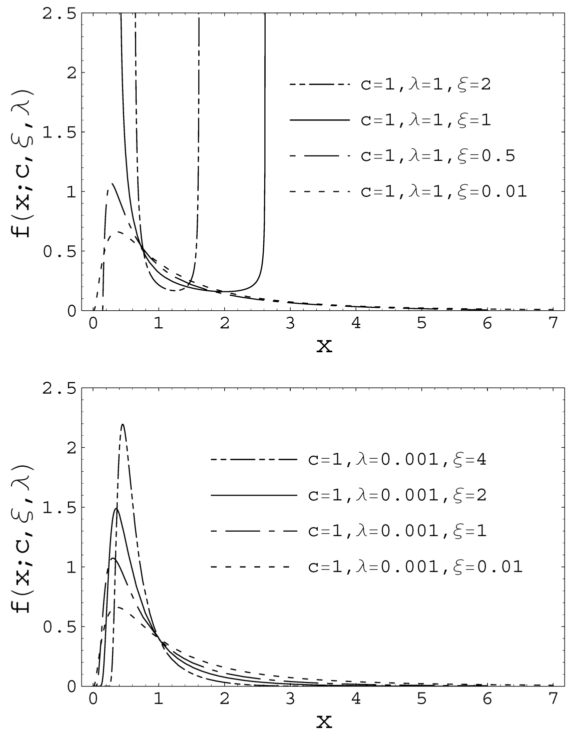

図 2. 一般化されたべき逆ガウス型分布の密度関数(2).

データの記述の在り方から，つまり離散量であるデータの在り方から本来的には離散分布が想定されるべきであり，その極限分布である連続分布を想定した母集団からの「近似的」な表現として離散データを認識することが可能であることを支えるのは，離散分布と連続分布との間の整合性である．どのような離散分布を想定するかは実質科学的な知見に基づいてなされることが望ましい．データの最小の刻み幅をどのくらいの量とするかはその時点での技術水準，資金，利用目的などによって決まるものである．その意味で任意の最小の刻み幅に対応できるように離散分布が定義されていることは大切である．

本稿で示した離散分布の確率関数は通常の離散分布の確率関数の表現と比較して煩雑な印象を与え，何故わざわざそのような理解しにくい表現を採用するのかとの批判を受けるけれども，離散分布の通常の確率関数の表現では尺度母数を入れていない表現であるから当然簡単な表現となるものである．そもそも数式表現としての見かけ上の簡単さが優先されるべきではなく，離散分布と連続分布との整合性が優先されるべきであると考える．離散の世界も連続の世界も共に区別なく扱う分布モデルの構築の一つの方法論の展開事例を示すことが本稿の目的である．

本稿において「分布の中心」や「分布のばらつき」などの曖昧な表現をしていることには理由がある．例えば本稿においては「分布の中心」を表現するものとして母算術平均，即ち母平均(期待値)を採用したけれども，これ以外の「分布の中心」を表現する定義を排除しないことが含まれている．例えば確率関数が $\frac{6}{\pi^2}\frac{1}{k^2}, k=1,2,\ldots$ で与えられる確率変数の母平均は存在しないから，自動的に本稿での方法で扱う離散分布の範疇からは除外されてしまっている(ここでは話を簡単にするために母数を入れていない)．これらの分布を含めて議論しようとすれ

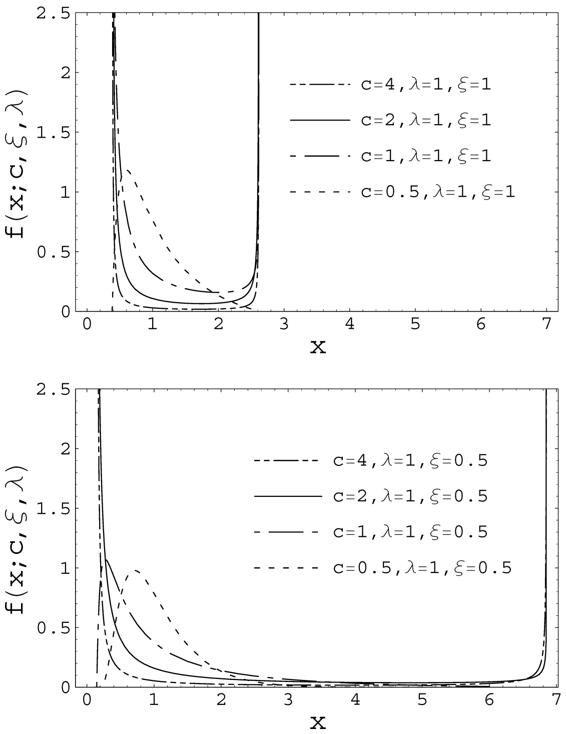

図 3.　**一般化されたべき逆ガウス型分布の密度関数**(3).

ば他の意味での「分布の中心」を採用して議論され得ることを容認している．分布の中心やばらつきなどがどのように定義されたものを採用すべきであるかについては恣意性があるわけであり，このことを考慮して意識的に曖昧な表現をしているものである．

謝　辞

2 人の査読者の有益な助言に対して感謝いたします．

本研究は，統計数理研究所共同研究プログラム(15-共研-3001)「水環境問題に於ける統計的認識手法の確立と実務での検証」，および日本学術振興会科学研究費補助金(15300097)「水質モニタリングとモデリングにおける統計的問題の解決と実証的研究」の研究成果に基づくものである．

参　考　文　献

Consul, P. C. (1989). *Generalized Poisson Distributions*, Marcel Dekker, New York.

Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products*, 6th ed., Academic Press, New York.

岩瀬晃盛，平野勝臣(1990)．べき逆ガウス型分布とその応用，応用統計学, **19**, 163–176.

Prudnikov, A. P., Brychkov, Yu. A. and Marichev, O. I. (1990). *Integrals and Series*, Volume 3, Gordon and Breach Science Publishers, New York.

Note on a Generalized Negative Hypergeometric Distribution

Kōsei Iwase[1], Koji Kanefuji[2] and Mitsumasa Okada[1]

[1]Graduate School of Engineering, Hiroshima University
[2]The Institute of Statistical Mathematics; The Graduate University for Advanced Studies

A discrete distribution converging on a certain continuous distribution is not necessarily unique. For example, a discrete distribution should be essentially assumed to be a population distribution from the viewpoint of description of data even if the inverse Gaussian distribution is assumed as a population distribution for convenience and data analysis is conducted. As assumed for discrete distribution, the generalized Poisson distribution and the generalized negative hypergeometric distribution can serve as a candidate in this case.

A generalized negative hypergeometric distribution is newly proposed. The generalized beta distribution and the generalized power inverse Gaussian distribution are obtained as limiting distributions in this distribution.

Key words: Generalized negative hypergeometric distribution, generalized power inverse Gaussian distribution, coefficient of variation.