半正規分布に就いて

河田, 龍夫
第一生命

出版情報：統計数理研究. 1 (2), pp.98-104, 1942-03-15. Research Association of Statistical Sciences
バージョン：
権利関係：

半正規分布に就いて

會員 河田龍夫（第一生命）

（昭和十六年九月廿七日受理）

1. X_1, X_2 が二つの確率變數なるとき，その比 X_1/X_2 なる變數の分布に關する Steffensen の研究色々と是に就ての筆者の考へを書いて見る．この問題は實際にも屢々遭遇するに違ひないと思はれるから，この種の研究は必要であらうと考へられる．例へば A 地方の人の身長を X_1 とし B 地方の人の身長を X_2 としたとき，その比 X_1/X_2 がどんな風に分布されるか，又その平均値・標準偏差はどの位かといふ如き問題が起らう．實際筆者も，筆者の勤めてゐる會社の實務で，この種の問題に出會したので色々考へて見た次第である．(1)

身長の場合でもそうであるが，實際問題で正規分布函數に從ふと假定される變數を考へる事が非常に多い．勿論嚴密に云へば，分布がいくら正規分布函數に近いといつても，それと全く一致する事は先づない．例へば身長の分布の場合にしても身長の値は負になる筈がないのであり，又ひどく大きい値もとらない筈である．この實際と理論との相違は併し大抵問題にならないものである．それは例へば平均値 m，標準偏差 σ の正規分布のとき，$m+3\sigma$ より大なる値をとる確率が非常に小である事からも類推される．そんな譯で身長の分布の場合でも是を正規分布に從ふと假定するのである．

所が X_1/X_2 の比を考へると狀態は全く變つて來る．即ち X_2 及 X_1 が夫々 x, y なる値をとる確率を

$$f(x,y)\,dxdy = Ke^{-(a^2x^2+2bxy+c^2y^2)}dxdy$$

として X_1/X_2 が v といふ値をとる確率を求めると

(1)
$$\int_{-\infty}^{\infty} f(u, uv)|u|\,du = 2K\int_0^{\infty} e^{-(a^2+2bv+c^2v^2)u^2} u\,du = \frac{K}{a^2+2bv+c^2v^2}$$

となる．(2) X_1 と X_2 とが互に獨立であるとすれば (1) で $b=0$ とおけばよい．とにかく (1) の形で確率密度が與へられる（$b^2 < a^2c^2$ ならば (1) の分母は決して 0 とならない）．この式からすぐ判る様にその平均値を出して見ると

$$\int_{-\infty}^{\infty} \frac{Kv}{a^2+2bv+c^2v^2}\,dv$$

となり有限にならない．勿論標準偏差も考へられない．

所が身長の比のときはその平均値はどこかにある筈だし，標準偏差だつて何等のか値になる筈である．この矛盾は身長の分布が正規分布に從ふと假定した爲である．

(1) は Cauchy の分布として知られてゐるものである．だから理論的な興味は以上の結果に對しても充分あるけれども，實際統計では平均値・標準偏差のない分布は避けらるべきであらう．

2. 以上述べた所により，正規分布に従ふ二つの變數の商の問題が相當困難である事が了解されよう．實際種々の學者によつて其が認められたのである．上述の困難を避けるために J. F. Steffensen が半正規分布函數を考へてゐる.[3] その研究の概略を述べて見やう．

變數 X が $(x, x+dx)$ の値をとる確率が

$$f(x)\,dx = Ax^{\lambda-1}e^{-\frac{x^2}{2M}}dx \quad x>0, \tag{2}$$
$$= 0 \qquad\qquad\qquad\quad x\leqq 0$$

なるとき X は半正規分布に従ふといふ．この分布は正規分布とよく似てゐるが，$x\leqq 0$ では 0 になるといふ點に於て本質的に違つてゐる．この分布函數は χ^2-test のときの χ^2 分布函數に他ならないのであるから，その重要である事は說明の必要がないであらう．

$\lambda>1$, $M>0$ とする．(2) の存在範圍は $x=0$ で $\lambda>2$ ならば $x=0$ で x 軸に $f(x)$ は接する．

$\lambda=1$ としても $f(x)$ は正規分布の確率密度にはならない．云ふ迄もなく $f(x)$ は $x<0$ では常に 0 だからである．併し次の樣に考へれば $f(x)$ の $\lambda\to\infty$ の極限が正規分布の確率密度になると考へる事は出來る．

(2) の $f(x)$ に於てはモードは $x=\sqrt{(\lambda-1)M}$ になる．之は $f(x)$ の極大を捜せば容易に判る．今このモードの位置に原點を移すと $f(x)$ は

$$A(x+\sqrt{(\lambda-1)M})^{\lambda-1}e^{-\frac{1}{2M}(x+\sqrt{(\lambda-1)M})^2} \tag{2}$$

となる．是は

$$B_\lambda\left(1+\frac{x}{\sqrt{(\lambda-1)M}}\right)^{\lambda-1}e^{-x\sqrt{\frac{\lambda-1}{M}}}e^{-\frac{x^2}{2M}} \tag{4}$$

と書ける．茲に

$$B_\lambda = A[(\lambda-1)M]^{\frac{\lambda-1}{2}}\cdot e^{-\frac{\lambda-1}{2}}.$$

A の値は $\int_0^\infty f(x)\,dx=1$ なる如き値であるから

$$A\int_0^\infty x^{\lambda-1}e^{-\frac{x^2}{2M}}dx = 1$$

となる筈で，從て Γ 函數の定義に依てこの積分を Γ で表はして A を定めると

$$A=\frac{2}{(2M)^{\frac{\lambda}{2}}\Gamma\left(\frac{\lambda}{2}\right)} \tag{5}$$

となる．故に B_λ は

$$B_\lambda=\frac{2(\lambda-1)^{\frac{\lambda-1}{2}}e^{-\frac{\lambda-1}{2}}}{2^{\frac{\lambda}{2}}\Gamma\left(\frac{\lambda}{2}\right)\sqrt{M}} \tag{6}$$

となる．Stirling の公式によれば $\lambda\to\infty$ のとき

$$\Gamma\left(\frac{\lambda}{2}\right)\fallingdotseq\sqrt{2\pi}\left(\frac{\lambda}{2}-1\right)^{\frac{\lambda}{2}-1+\frac{1}{2}}e^{-\left(\frac{\lambda}{2}-1\right)}$$

であるから，之を (6) へ入れて整頓すると

$$B_\lambda \fallingdotseq \frac{1}{\sqrt{M\pi}}\left(\frac{\lambda-1}{\lambda-2}\right)^{\frac{\lambda-1}{2}}\frac{1}{\sqrt{e}} = \frac{1}{\sqrt{M\pi}}e^{-\frac{1}{2}}\left(\left(1+\frac{1}{\lambda-2}\right)^{\lambda-2}\right)^{\frac{\lambda-1}{2(\lambda-2)}}$$

となり，{} の中は e の定義により $\lambda\to\infty$ のとき e となる．故に $B_\lambda\to\frac{1}{\sqrt{\pi M}}$ となる．

又 $\left(1+\frac{x}{\sqrt{(\lambda-1)M}}\right)^{\lambda-1}e^{-x\sqrt{\frac{\lambda-1}{M}}}$ は

$$e^{(\lambda-1)\log\left(1+\frac{x}{\sqrt{(\lambda-1)M}}\right)-x\sqrt{\frac{\lambda-1}{M}}}$$

とかけ，この羃は $\lambda\to\infty$ のとき $\log\left(1+\frac{x}{\sqrt{(\lambda-1)M}}\right)\sim\frac{x}{\sqrt{(\lambda-1)M}}-\frac{x^2}{(\lambda-1)M}$ を使へば $\lambda\to\infty$ で $-\frac{x^2}{M}$ になる．故に上の極限は $e^{-\frac{x^2}{M}}$ になる．之を (4) へ入れると

$$\lim_{\lambda\to\infty}B_\lambda\left(1+\frac{x}{\sqrt{(\lambda-1)M}}\right)^{\lambda-1}e^{-x\sqrt{\frac{\lambda-1}{M}}}e^{-\frac{x^2}{2M}} = \frac{e}{\sqrt{M\pi}}^{-\frac{x^2}{M}}$$

となる．是は標準偏差 $\sqrt{\frac{M}{2}}$ で平均値 0 の正規分布の確率密度である．

以上により λ が大となれば半正規分布の確率密度は正規分布の確率密度に近づく．是に依つても半正規分布が正規分布に λ が大なれば非常によく似てゐる事が判るであらう．

3. 半正規分布の平均値・標準偏差はどうなるかといふと，これは積分の計算に依りすぐ求まる．即ち夫等を m,σ で表すと

$$m=\int_{-\infty}^{\infty}xf(x)\,dx = A\int_0^\infty x^\lambda e^{-\frac{x^2}{2M}}dx$$

$$= \frac{A}{2}(2M)^{\frac{\lambda+1}{2}}\Gamma\left(\frac{\lambda+1}{2}\right) = (2M)^{\frac{1}{2}}\frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)}$$

となる．

$$A\int_0^\infty x^2x^{\lambda-1}e^{-\frac{x^2}{2M}}dx$$

$$= 2M\frac{\Gamma\left(\frac{\lambda+2}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)} = \mu_2$$

が同様に得られる．従て

$$\sigma^2=\mu_2-m^2$$

であるから σ^2 も容易に計算出きる．

又

$$\mu_3 = A\int_0^\infty x^3x^{\lambda-1}e^{-\frac{x^2}{M+1}}dx$$

を計算すると同様に Γ 函數の定義から

$$\mu_3=(2M)^{\frac{3}{2}}\frac{\Gamma\left(\frac{3+\lambda}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)}$$

となる．$\beta_1=\frac{\mu_3^2}{\mu_2^3}$ は上の μ_2,μ_3 の値からすぐ書き下せる．

さて實際の data が與へられてゐて是に半正規分布函數をあてはめるのには，勿論そのまゝ (2) を使ふべきでなく，原點を $x=-c$ に移した

(7) $$\varphi(x) = A(x+c)^{\lambda-1}e^{-\frac{(x+c)^2}{2M}}$$

を使用すべきである事は云ふ迄もない．A は λ の函数であるから (7) には三つの定めるべき常数 c, λ, M を含むわけで，これ等の値を決定する．便宜上 $\xi = \sqrt{2}\,\Gamma\left(\frac{\lambda+1}{2}\right)/\Gamma\left(\frac{\lambda}{2}\right)$ とおけば，平均値 m，二次及三次のモーメント m_2, m_3 は

(8) $$m_1 = \xi\sqrt{M} - c$$

(9) $$m_2 = (\lambda - \xi^2)M$$

(10) $$m_3 = \xi(2\xi^2 - 2\lambda + 1)M\sqrt{M}$$

となる．是は上述 m, μ_2, μ の計算と全く同じである．(9), (10) には c がないからこの二つから λ, M が定められる (data から各モーメント，平均値を求めておく)．そして最後に (8) から c を求めるのである．尤も理論的に云へば是だけの事であるが，實際 (9), (10) から λ, M を求めるのはそう簡單には行かない．詳しい事は省くが Steffensen のやり方は $\beta_1 = \frac{m_3^2}{m_2^3}$ を求めこの β_1 を使つて次の如くして λ を求めるのである．

(11) $\beta_1 \leqq 0.0675$ のときは $$\lambda = \frac{1}{2\beta_1} + \frac{5}{4} - \frac{19}{8}\beta_1$$

(12) $0.0675 < \beta_1 \leqq 0.2359$ のときは $$\lambda = \frac{1 + \sqrt{10\beta_1 + 1}}{4\beta_1}$$

β_1 がこの範圍以外のときは後者により λ を求め，之を出發として補間によつつて定める．

尤もそんなにまでして λ の精しい値を求めなくとも，吾々の問題が方程式をとくのが目的でなく curve fitting にあるのだから，λ の大體の値を求めておき (8) と (9) との二つから正しく c, M を求めればよい．正規分布を當はめるときには平均値と標準偏差の二つの常數だけで曲線をきめるのであるから，それに比べれば λ の正確を求めなくとも尚勝つてゐる．

どうして (11), (12) を求めたかといふと考へは

$$\beta_1 = \frac{4\xi^2(\xi^2 - \lambda + \frac{1}{2})^2}{(\lambda - \xi^2)^3}$$

を $\frac{1}{\lambda}$ の冪級數に展開し $\frac{1}{\lambda}$ の高い次數の所を捨てて出したものである．この議論はこゝでは省略させて頂く．

この半正規分布を適用する例もあるが，是に就ては別の機會に讓る事にして，吾々は確率變數の商の場合にこの確率函數を應用する．

4. X_1 の確率密度を

$$By^{\mu-1}e^{-\frac{y^2}{2N}}$$

とし，X_2 の確率密度を

$$Ax^{\lambda-1}e^{-\frac{x^2}{2M}}$$

とする．そして X_1 と X_2 が互に獨立とすると，X_2, X_1 が夫々 (x, y) なる値をとる確率は

$$f(x,y)\,dxdy = ABx^{\lambda-1}y^{\mu-1}e^{-\frac{x^2}{2M}-\frac{y^2}{2N}}dxdy$$

となり，(1) の所の式により $\frac{X_1}{X_2}$ が $v=\frac{y}{x}$ なる値をとる確率密度は

$$\psi(v)=ABv^{\mu-1}\int_0^{\infty}u^{\lambda+\mu-1}e^{-\frac{u^2}{2}\left(\frac{1}{M}-\frac{v^2}{N}\right)}du$$

となる．是は Γ なる函數の定義から

$$\psi(v)=K\frac{v^{\mu-1}}{\left(1+\frac{Mv^2}{N}\right)^{\frac{\lambda+\mu}{2}}},\quad v\geqq 0 \tag{13}$$

$$K=2\left(\frac{M}{N}\right)^{\frac{\mu}{2}}\frac{\Gamma\left(\frac{\lambda+\mu}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{\mu}{2}\right)} \tag{14}$$

となる．この分布函數 $\psi(v)$ は $v\to\infty$ のとき $\frac{1}{v^{\lambda+1}}$ の order であるから，λ が相當大きければ平均値・標準偏差，更に高次のモーメントが存在する．實際一般に r 次のモーメントを計算する．但し $r<\lambda$ としておく．

$$\begin{aligned}m_r&=K\int_0^{\infty}v^r\frac{v^{\mu-1}}{\left(1+\frac{Mv^2}{N}\right)^{\frac{\lambda+\mu}{2}}}dv\\&=\frac{K}{2}\left(\frac{N}{M}\right)^{\frac{\mu+r}{2}}\int_0^{\infty}t^{\frac{\mu+r}{2}-1}(1+t)^{-\frac{\lambda+\mu}{2}}dt\quad\left(v=\left(\frac{Nt}{M}\right)^{\frac{1}{2}}\text{ とおいた}\right)\\&=\frac{K}{2}\left(\frac{N}{M}\right)^{\frac{\mu+r}{2}}\int_0^{1}\theta^{\frac{\lambda-r}{2}-1}(1-\theta)^{\frac{\mu+r}{2}-1}d\theta\quad\left(t=\frac{1}{\theta}-1\text{ とおいた}\right).\end{aligned}$$

是は β 函數の定義により

$$=\frac{K}{2}\left(\frac{N}{M}\right)^{\frac{\mu+r}{2}}\frac{\Gamma\left(\frac{\lambda-r}{2}\right)\Gamma\left(\frac{\mu+r}{2}\right)}{\Gamma\left(\frac{\lambda+\mu}{2}\right)}$$

となり，是に (14) を入れると，結局

$$m_r=\left(\frac{N}{M}\right)^{\frac{r}{2}}\frac{\Gamma\left(\frac{\lambda-r}{2}\right)\Gamma\left(\frac{\mu+r}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}\quad(r<\lambda) \tag{15}$$

となる．

5. 以上で吾々の商の問題は理論的には解決がついたわけで，以上が Steffensen の研究の大體である．更に Cramér の書物[4]の中では，X_1, X_3 が共に Pearson の type III の法則に従ふとき X_1 が正規分布に従ひ X_2 が半正規分布に従ふときが論じてある（是は Steffensen の論文の中にも書いてある）．このときは X_1/X_2 は Pearson の type VII の法則に従ふ．

6. 理論的には以上の通りであるけれど，ある物の長さをくりかへして測つたときの分布の様に標準偏差が平均値に比べて小さいと云ふときを考へる．X_1 も X_2 も其の様な分布に従ふとし，X_1 X_2 の平均値を夫々 m_1, m_2 とすると X_1/X_2 の平均値も大體 m_1/m_2 だと考へられる．假に X_1, X_2 が半正規分布に従ふとして (15) に依て平均値を求めて見る．3. に依れば

$$m_1=(2N)^{\frac{1}{2}}\frac{\Gamma\left(\frac{\mu+1}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right)},\quad m_2=(2M)^{\frac{1}{2}}\frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)}$$

であるから，商の平均値は (15) から

$$\left(\frac{N}{M}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{\lambda-1}{2}\right)\Gamma\left(\frac{\mu+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}$$

$$=\frac{m_1}{m_2}\frac{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}{\Gamma\left(\frac{\mu+1}{2}\right)\Gamma\left(\frac{\lambda+1}{2}\right)}\cdot\frac{\Gamma\left(\frac{\lambda-1}{2}\right)\Gamma\left(\frac{\mu+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{\mu}{2}\right)}$$

$$=\frac{m_1}{m_2}\frac{\Gamma\left(\frac{\lambda-1}{2}\right)}{\Gamma\left(\frac{\lambda+1}{2}\right)}.$$

是は λ が大きくなれば $\frac{\Gamma\left(\frac{\lambda-1}{2}\right)}{\Gamma\left(\frac{\lambda+1}{2}\right)}$ は段々 0 に近づく値であつて，$\frac{m_1}{m_2}$ とは相當異つた値になる．標準偏差にしても同様の事が云へる．この矛盾は半正規分布のときはその存在範圍が $x\geqq 0$ であるといふ點にあると考へられる．この意味で實際問題に應用する場合，半正規分布を考へただけでは不充分で更に研究の必要が認められる．即ち X_1, X_2 の平均値に較べて其の標準偏差が小さいとき X_1/X_2 の平均値と平均値の商とがどの程度に異ふかといふ様な問題が必要と思ふ．商の標準偏差にしても同様に想像されるものと實際の標準偏差とどの位異ふかといふ事である．

以上の如き問題に對しては，確率函數が正規とか半正規とかいふ風に特別なものにとらないでも極く一般に取扱へる．但し商の分布其者ははつきりした式で出せない．それでこの場合，X_2 の標準偏差が次第に小さくなつたときどんな分布函數に近づくか，又そのときの誤差はどの位か，又標準偏差が ∞ になればどうなるか，といふ様に收斂定理の形に於て考へて見れば面白いと考へる．

以上の方針でやつて見た結果をかいて見る．詳しい事はアクチュアリー會報に載せる心算である．

7. X_1 の分布は全く任意ととする．但し標準偏差は有限とする．X_2 の分布はその平均値 m_2 に關して對稱とし，X_2 は (m_2-a, m_2+a) 以外の値をとらぬと假定する (但し $m_2>a$)．且 X_1 と X_2 は互に獨立としておく．是だけを假定して 6. の問題の結果だけを書く．先づ X_1 の平均値・標準偏差を夫々 m_1, σ_1 とし X_2 の標準偏差を σ_2 とする．結果は次の如くなる．

X_1/X_2 の平均値は

$$m=\frac{m_1}{m_2}(1+\varepsilon)$$

となり，標準偏差は

$$\sigma=\sqrt{\frac{m_1^2}{m_2^2}\left(\frac{\sigma_1^2}{m_1^2}+\frac{\sigma_2^2}{m_2^2}\right)+\gamma}$$

となる．茲に

$$\varepsilon=\frac{\sigma_2^2}{m_2^2}+\frac{\gamma_4}{m_2^4}+\cdots\cdots$$

$$\eta = \frac{3\sigma_1^2\sigma_2^2}{m_2^4} + \frac{5\sigma_1^2\gamma_4}{m_2^6} + \frac{m_1^2(3\gamma_4 - \sigma_2^4)}{m_2^6} + \cdots\cdots$$

但し γ_4 は X_2 の平均値の周りの四次のモーメントである.

この結果によれば σ_2 が m_2 に比べて小さければ X_1/X_2 の平均値は殆ど $m_2 < m_1$ に等しくなり, その標準偏差は $\frac{m_1}{m_2}\sqrt{\frac{\sigma_1^2}{m_1^2} + \frac{\sigma_2^2}{m_2^2}}$ に略等しくなる.

實際問題で α も m_2 に比べて小さい場合が多いから, 正規分布の當はまる様なものでは $\alpha = 4\sigma_2$ 位としてよく, 從て σ_2/m_2 が小であれば $|\varepsilon_1|$ は非常に小さい.

次に $\alpha \to 0$ のときは X_1/X_2 の分布が $\frac{X_1}{m_2}$ の分布に近づく事は殆ど明らかであらう. 嚴密な證明はよい練習問題と思ふ. そのときの近づき方がどんなものかを計算する事も出きる.

それで X_2 の分布に於て α が小さいときは X_1/X_2 の分布が大體 X_1 の分布と考へてよい場合が多く, そのときの誤差を計算して正しい判斷をすればよい. この誤差がどの程度かといふ事を出すのは割合に簡單であるから讀者に委せやう.

註

(1) この問題は第一生命取締役龜田博士に頂いたので最後の結果も御指示を仰ぐ所多かつた. 博士に感謝の意を表す次第です.

(2) 龜田博士, 確率論及其應用, p. 106 參照.

(3) J. F. Steffensen, On the semi-normal distributions, Skand. Akt. 1937.

(4) H. Cramér, Random variables and probability distributions, camb. Tracts 1937).