研究ノート

所得分布 I

新井宏尚

1. 経済変量の分布の特性

ある経済変量の分布が正規分布に従うとき，その分布の構造を平均（算術平均）と標準偏差によって的確に把握することができる。しかし，所得のような経済変量の分布はその分布のもつ歪みにより，当然のことであるがもとのデータのままでは正規近似することはできない。そのため，Vilfred Pareto 以来いくつかの所得分布が提案されて来た。さまざまな分布の中でも代表的な分布としては，パレート分布，対数正規分布があげられるであろう。しかし，これらの分布はパラメーターの推定と解釈が容易であるという利点があるにもかかわらず，所得の全領域における分布の適合度を良好に示すことができない。この点，最近 Salem と Mount (1974) により，ガンマ分布が提案された。彼等によると，ガンマ分布は対数正規分布より適合度がまさることを 1960 年と 1969年の合衆国所得データを用いて立証し，パラメーターの解釈も与えた。しかし，なぜガンマ分布が発生するかということに関しては何ら説明していない。次に，Singh と Maddala (1976) は，対数正規，ガンマ分布と比較できうる程，適合度の良い分布を提案した。しかし，パラメーターについては全然解釈が与えられていない。本ノートではパラメーターの解釈と適合度の良好さは互いにトレード・オフな関係にあるのではないか点に注意して，ワイブル分

布の発生について Singh と Maddala との所得概念を使って説明し，次にガンマ分布の生成について考察する。

2. ワイブル分布とガンマ分布の発生について

信頼性理論から故障率（保険数学では死力とされている）なる概念を援用する。故障率とは(1)あるシステムが時刻 t まで稼動しつづけたという条件のもとでの次の時間間隔（t，$t+\Delta t$）の間に故障する条件は確率を求め，(2)これを Δt で割り，(3)この商の $\Delta t\to 0$ とおいた極限値をさすとされている。つまり，年齢 t に達した場合の瞬間故障率を示すと言える。Xを寿命時間を表わす確率変数とすれば故障率 $\lambda(t)$ は次式で与えられる〔4〕，〔5〕。

$$\lambda(t)=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\cdot\frac{P'[(t<X\leqq t+\Delta t)\cap(t<X)]}{P(t<X)}$$

$$=\lim_{\Delta t\to 0}\frac{\frac{1}{\Delta t}\int_t^{t+\Delta t}f(x)dx}{1-F(t)}=\frac{f(t)}{1-F(t)} \qquad (2\text{–}1)$$

(2–1) 式より明らかなように，$F(t)$ がわかれば $\lambda(t)$ は求めることができる。また，逆に，$\lambda(t)$ がわかれば以下明らかなように $F(t)$ も求まる。

$$\lambda(t)=\frac{F'(t)}{1-F(t)} \qquad \left(F(t)=\int_0^t f(t)dt\right)$$

より，$F(t)$ について解くために，次のように変形すると

$$\lambda(t)=-\frac{d}{dt}\{\log(1-F(t))\}$$

となり，さらに両辺を積分すると，

$$\int_0^T\lambda(t)dt=-\log(1-F(t)) \qquad (F(0)=0)$$

が得られる。したがって，

$$F(t)=1-e^{-\int_0^T\lambda(t)dt} \qquad (2\text{–}2)$$

が得られる。つまり，$\lambda(t)$ の形がわかれば，$F(t)$ は求められる。

次に，寿命 t の写像 $\phi(t)$ を所得と考えるとする〔3〕。いま，一番簡単な写像 $t=x$ の場合（xは所得とする），所得分布は (2-2) 式より

$$F(x)=1-e^{-\int_0^M \lambda(x)dx} \tag{2-3}$$

で与えられる。したがって，$\lambda(t)$ の形が決定されれば良い。

次に $\lambda(t)$ の関数形について整理する。

(a) λ が一定の場合；

$$F(t)=1-e^{-\int_0^T \lambda(t)dt}=1-e^{-\lambda t} \tag{2-4}$$

となり指数分布が得られる。

(b) λ が時間の指数関数の場合（$\lambda(t)=t^{m-1}$）

$$F(t)=1-e^{-\int_0^T t^{m-1}\,dt}=1-e^{tm}$$

これを少し修正すると，

$$F(t)=1-e^{-\frac{t^m}{\alpha}} \tag{2-5}$$

がえられる。これに，t を $t-\gamma$ とおきかえると

$$F(t)=1-e^{-\frac{(t-\gamma)^m}{\alpha}} \qquad (t>\gamma) \tag{2-6}$$

というワイブル分布の一般形がえられる。(2-6) 式を微分すれば，

$$f(t)=\frac{m}{\alpha}(t-\gamma)^{m-1}\cdot e^{-\frac{(t-\gamma)^m}{\alpha}} \qquad (t>\gamma)$$

密度関数 $f(t)$ がえられる。さらに$t=x$（x；所得）とおくと，所得分布

$$f(x)=\frac{m}{\alpha}(x-x_0)^{m-1}\cdot e^{-\frac{(x-x_0)^m}{\alpha}} \tag{2-7}$$

がえられる。ただし，x_0 は最低所得である。

ここで，対数正規分布とパレート分布について $\lambda(t)$ について調べてみよう。対数正規分布の $p\cdot d\cdot f$ は

$$f(x)=\frac{M}{\sigma x\sqrt{2\pi}}\cdot e^{-\frac{(\log x-\log\eta)^2}{2\sigma^2}} \qquad (x\,;\ \text{寿命}) \tag{2-8}$$

ただし，η は分布の中央値，$M=\log_{10}e$

したがって，(2-1)，(2-8) 式より

$$\lambda(t)=\frac{f(t)}{\int_{\frac{\log t-1}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}dx}}\sim\frac{M(\log t-1)}{\sigma^2 t} \tag{2-9}$$

パレート分布の $p\cdot d\cdot f$, $c\cdot d\cdot f$ は $f(x)=\alpha x_0^d\cdot x^{-(\alpha+1)}$,

$F(x)=1-\left(\frac{x}{x_0}\right)^{-\alpha}$ で与えられるので，$\lambda(t)$ は

$$\lambda(t)=\alpha\cdot t^{-1} \tag{2-10}$$

で与えられる。パレート分布の $p\cdot d\cdot f$ とワイブル分布の $p\cdot d\cdot f$ は似てないように見える。しかし，$\lambda(t)$ を比較するとワイブル分布の場合の

$\lambda(t)=\frac{m}{\alpha}(t-\gamma)^{m-1}$ で $\gamma=0$ とおくと $\lambda(t)=\frac{m}{\alpha}t^{m-1}$ $(t>0)$ となり

(2-10) と良く似ている。

次に，ワイブル分布のパラメーターについて調べてみよう。ワイブル分布の平均と分散は次式で与えられる。

$$\mu=\int_0^\infty(\alpha X)^{\frac{1}{m}}e^{-X}dX=\alpha^{\frac{1}{m}}P(\frac{1}{m}+1) \qquad (2\text{-}11)$$

$$\sigma^2=\int_0^\infty[X-E(X)]^2 f(X)dX$$

$$=\alpha^{\frac{1}{m}}\left\{P(\frac{2}{m}+1)-P^2(\frac{1}{m}+1)\right\} \qquad (2\text{-}12)$$

ただし，　$X=\dfrac{x-\gamma}{\alpha^{\frac{1}{m}}}\geqq 0$　とおきかえる。

(2-11)，(2-12) から明らかなようにパラメーターに明解な解釈を与えることはむずかしい。ただし，m の値を $\widehat{CV}$（標本変動係数）で近似し（C•A•Cohen による。分布の真の変動係数は　$CV=\left\{\Gamma\left(\frac{2}{m}+1\right)-\Gamma^2\left(\frac{1}{m}+1\right)\right\}^{\frac{1}{2}}/\Gamma\left(\frac{1}{m}+1\right)$　で与えられる）mを既知とした場合，$\hat{\alpha}=\frac{1}{n}\sum_{i=1}^{n}x_i^m$　で与えられるので平均の1つとして解釈できるであろう（$m=1$　のときは指数分布，このとき　$\alpha=\mu$）。したがって，α とmの数値をつくり，mが形状を示すパラメーターを示すのでワイブル分布のモーメントを計算し，尖度を求め，mに何らかの解釈を与えられるかもしれない。

ワイブル分布には，パレート，対数正規分布と同様，簡便な点があるので以下その手順を示す〔6〕。

ワイブル分布の　$c\cdot d\cdot f$　は　$F(x)=1-e^{-\frac{(x-\gamma)^m}{\alpha}}$　で与えられる。$x-\gamma$ をxでおきかえると，　$F(x)=1-e^{-\frac{x^m}{\alpha}}$　これを変形し両辺の対数を2回とると，

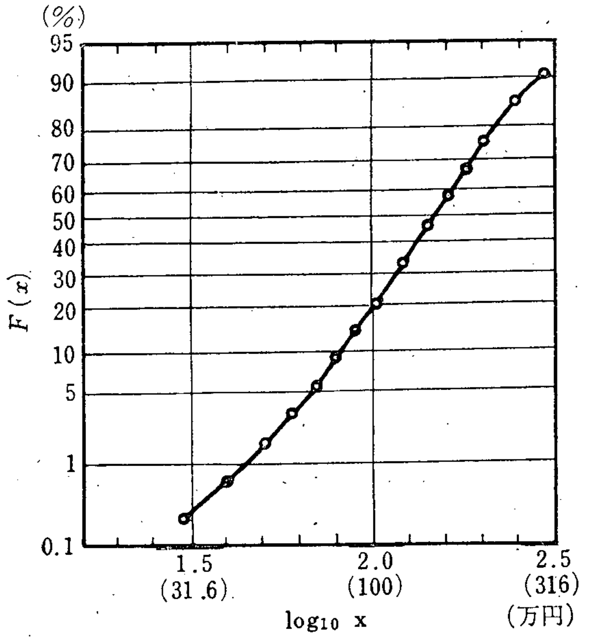

図 1　1972年全世帯年間収入の分布
(両対数)

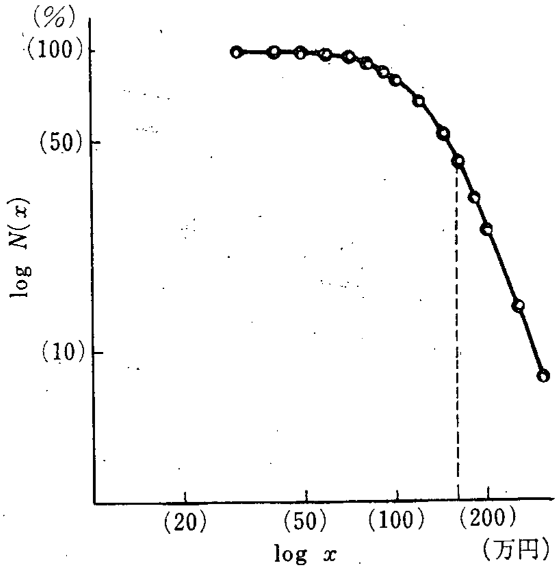

図 2　1972年全世帯年間収入の分布
(正規確率紙)

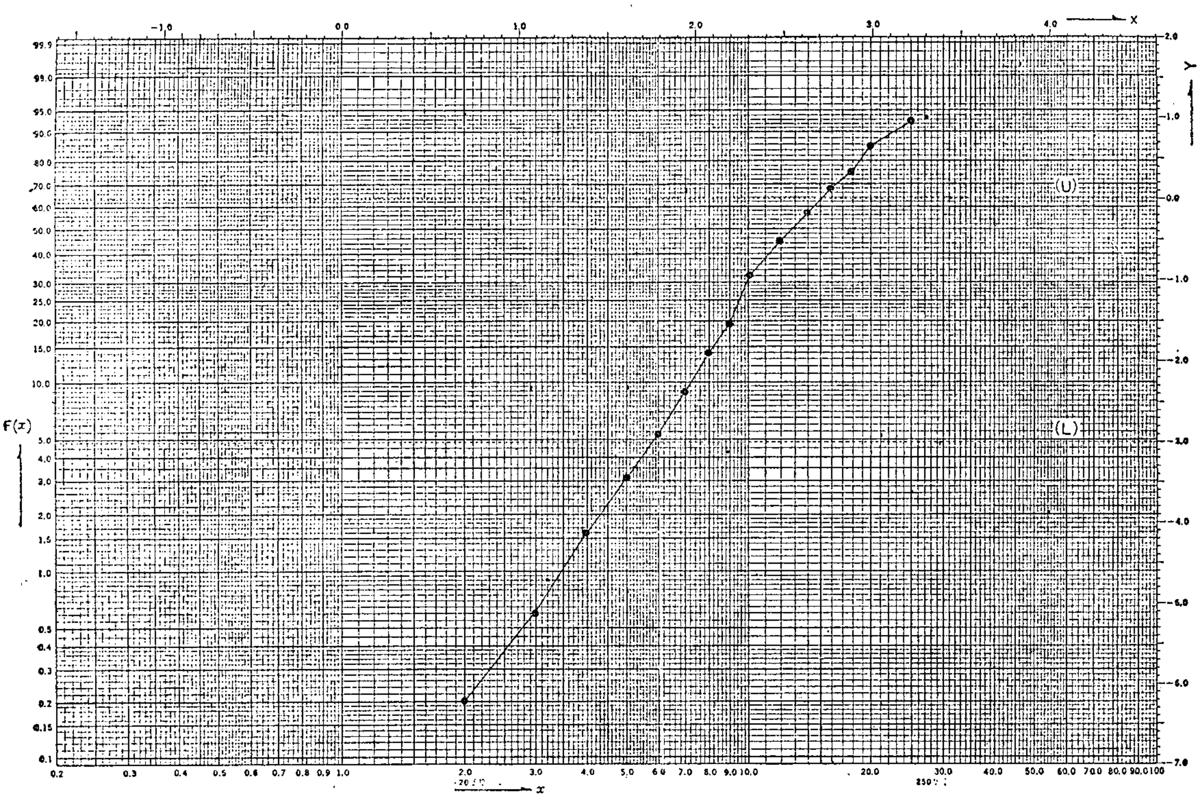

図 3　1972年全世帯年間収入の分布 (ワイブル確率紙)

$$\log\log\frac{1}{1-F(x)}=m\log x-\log\alpha$$

$$Y=mX-B$$

ただし，$Y=\log\log\{1/(1-F(x)\}$，$\log x=X$，$-B=-\log\alpha$ とする。したがって，x がワイブル分布に従っているか否かは，$Y=mX-B$ がワイブル確率紙で直線になっているかどうかを確かめれば良い。グラフで示すと図(2)のようになった。図(1)，(3)は〔6〕による。

次に，Mandelbrot (1960) によれば「所得とは，所得の構成要素を $X_i(i=1, 2, \cdots\cdots n)$ とするとき，その和として定義する。」としており，さらに「U_i がパレート分布」としている。ここでは，U_i が $\lambda(t)$=一定の分布，つまり X_i が指数分布に従うとするわけである。従って，所得分布は $X=\sum X_i$ の分布を求めることになる。当然のことであるが，X はガンマ分布 $g(x)=\dfrac{\alpha^n x^{n-1}}{\Gamma(n)}\cdot e^{-\alpha x}$ に従う。一応知られていることであるが証明しておく。

証明　帰納法で証明する。

$n=1$ のとき $g(x)=\alpha\cdot e^{-\alpha x}$ となり自明

次に，n のとき成立するとして，$n+1$ のとき成立することを示す。

$$g(x)=\frac{\alpha^n x^{n-1}}{\Gamma(n)}\cdot e^{-\alpha x}=\frac{\alpha^n x^{n-1}}{(n-1)!}e^{-\alpha x}=\alpha\frac{(\alpha x)^{n-1}}{(n-1)!}e^{-\alpha x} \qquad (2\text{-}13)$$

ここで

$$g_n(x)=\alpha\frac{(\alpha x)^{n-1}}{(n-1)!}e^{-\alpha x}$$

とすると，$g^n(x)$ と $g(x)$ の合成積 $g_{n+1}(x)$ となるから $g_n * g_1=g_{n+1}$ を示せば良い。

$$g_{n+1}(t)=\int_0^t g_n(t-x)\cdot g_1(x)dx=\alpha\cdot\left(\frac{\alpha t}{n!}\right)^n\cdot e^{-dt}$$

で全ての t について (2-13) は成立する。

また，X_i が $\lambda(t)=\frac{m}{\alpha}\cdot t^{m-1}$ の形をしている分布，つまりワイブル分布に従う変数の場合には，その和 $X=\sum_{i=1}^{n}X_i^m$（各 X_i は独立とする）はカイ2乗分布に従う。次に，ガンマ分布のパラメーター解釈についてみよう。ガンマ分布の $p\cdot d\cdot f$ は周知のように次の形で与えられる。

$$f(x\ ;\ \alpha,\ \lambda)=\lambda^{\alpha}\cdot x^{\alpha-1}\cdot e^{-\lambda x}/\Gamma(\alpha) \quad ただし \quad \Gamma(\alpha)=\int_0^{\infty}e^{-u}u^{\alpha-1}du$$

$(u>0)$

これより

(a)算術平均は α/λ

(b) モードは $(\alpha-1)/\lambda$

(c) メディアンは $(3\alpha-1)/3\lambda$

(d) 標本標準偏差は $\sqrt{\alpha}/\lambda$

(e) 積率表示の歪度は $1/\sqrt{\alpha}$

(f) 集中度係数は $C=2\cdot B_{(0.5)}(\alpha,\ \alpha+1)-1$

ただし，$B_{(0.5)}$ は不完全ベータ関数の50%点までの値。

(g) タイトルのエントロピー表示の不平等度は次式で与えられる。

$$I=\int_0^{\infty}\left[\frac{x}{E(x)}\right]\log\left[\frac{x}{E(x)}\right]f(x\ ;\ \alpha,\ \lambda)dx=\frac{1}{\alpha}-\phi(\alpha)-\log\alpha$$

ただし $E(x)=\alpha/\lambda$。$\phi(\alpha)/\Gamma'(\alpha)=d\log\Gamma(\alpha)/d\alpha$(注) はディガンマ関数である。

(h)ジニ平均差係数は $2\alpha c/\lambda$

となる。以上の中で単一パラメーターで与えられるのは (e), (f), (g) のみである。なお，指数分布，ガンマ分布，ワイブル分布を含む分布としては拡張ガン

マ分布　$\Gamma(x)=\frac{\lambda^k m}{\Gamma(k/m)}x^{k-1}e^{-(\lambda x)^m}$　$\lambda>0,\quad k>0,\quad m>0,\quad x\geqq 0$ がある

がこれら分布についてもデータ解析が必要であろう。また，ワイブル，ガンマ，対数正規分布の比較検討，推定問題については稿を改める予定である。

注　ディガンマ関数 $\psi(\alpha)$ は $\frac{d\log \Pi(\lambda)}{dx}=\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}$ と定義されることもある。

参考文献

(1) Mandelbrot. B. "The Pareto-Lévy Law and distribution of income," *International Economic Review*, Vol. 1, 1960, pp. 79–106.

(2) A. B. Z Salem, and T. D. Mount, "A convenient destriaive model of Income distribution," *Econometrica*, Vol. 42, 1974, pp. 1115–1127.

(3) S. K Singh, and and G. S Maddala, "A function for size distribution of Incomes," *Econometrica*, Vol. 44, 1976, pp. 963–969.

(4) 岩田暁一「経済現象における分布」Vol. 1, p. 36『行動計量学』1975　参照

(5) 中山一郎編『現代統計学大辞典』pp. 709–711　有斐閣　参照

(6) 真壁　肇『ワイブル確率紙の使い方』日本規格協会　1966　参照

(7) 森口繁一他『数学公式』p. 9　岩波全書　1970　参照

(8) 依田　浩『信頼性理論入門』朝倉書房　1972　参照