

二重パレート対数正規分布及び第二種一般化ベータ分布の所得分布への当てはめ

総務省 岡本 政人

所得分布に適合がよい統計分布として第二種一般化ベータ分布(GB2)^[2]とその特殊ケースが知られている。GB2 は 4 つのパラメータをもち、その確率密度関数(pdf)は以下のように表される。

$$gb2(x; a, b, p, q) := \frac{ax^{ap-1}}{b^{ap}B(p, q)[1+(x/b)^a]^{p+q}} \quad (1)$$

$p = 1, q = 1$ の場合が各々 Singh-Maddala 分布(SM), Dagum 分布(Da)になる。しかし、適合度がより高いモデルの探索が続けられている。Reed^[3]は、各個人(あるいは各世帯)の所得の対数値がブラウン運動に従って推移し(i.e. $d \log X_t = \mu dt + \sigma dB_t$, ここで B_t は標準ブラウン運動), 個人の出現時点(労働市場参入時点)から観測時点までの期間が指数分布に従う場合, 観測時点の所得分布は以下の pdf をもつ 4 パラメータの二重パレート対数正規分布(dPLN)になり, この分布が多くの国において GB2 よりも(尤度等の頻度ベースの評価基準に関して)適合度が高いことを示した。

$$dpln(x; \mu, \sigma^2, \alpha, \beta) := \frac{\alpha\beta}{\alpha+\beta} \left[x^{\beta-1} e^{-\beta\mu+\beta^2\sigma^2/2} \Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right) + x^{-\alpha-1} e^{\alpha\mu+\alpha^2\sigma^2/2} \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right) \right] \quad (2)$$

ここで, $\alpha, \beta, \sigma > 0$, Φ は標準正規分布の累積分布関数(cdf), $\Phi^c = 1 - \Phi$ 。Clementi, *et al.*^[1]は、一般化統計力学の κ -指数関数を応用した以下の pdf をもつ 3 パラメータの κ -一般化分布(κG)を米独のパネルデータ PSID, GSOEP の等価可処分所得分布に当てはめ, SM 及び Da よりも(尤度に関して)適合度が高いことを示した。

$$\kappa g(x; a, b, \kappa) := \frac{ax^{a-1} \exp_{\kappa}(-(x/b)^a)}{b^a \sqrt{1+\kappa^2(x/b)^{2a}}}, \quad \exp_{\kappa}(x) := (\sqrt{1+\kappa^2 x^2} + \kappa x)^{1/\kappa} \quad (3)$$

ここで, $a, b > 0, 0 \leq \kappa < 1, \lim_{\kappa \rightarrow 0} \exp_{\kappa}(x) = e^x$ 。 κG の cdf は $1 - \exp_{\kappa}(-(x/b)^a)$ である。なお、一般化統計力学における別種の deformed 指数関数 $\exp^{(q)}(x) := [1 + (1 - q)x]^{1/(1-q)}$ を応用すると SM 分布になる。

全国消費実態調査の年収階級別集計値にこれらの統計分布と 2 パラメータの対数正規分布(LN)を当てはめた結果を表に掲げる。最尤法を適用すると, GB2 及び dPLN は, SM, Da 及び κG よりもパラメータ数を考慮しても有意に適合度が高く, しかも両者間の尤度比は拡大傾向にある。これは, 高齢層で GB2 及び dPLN の適合度が相対的に高いことと, 高齢化が影響している。dPLN は GB2 に比べて尤度が高い傾向にあるが, ジニ係数の差は小さい。両分布の比較では年収区分に留意が必要で, 十分位階級のように高所得層が粗い区分を用いると, dPLN の適合度は GB2 を下回る場合がある。

尤度等の頻度ベースで優っていても金額シェアベースや格差指標の推定精度でみると必ずしも優っているとは限らない。実際, GB2 及び dPLN の推定 Lorenz 曲線の二乗誤差の平方根(L-RSE)は, 1989 年以降 SM よりも大きい。ジニ係数の推定精度も SM に劣っている。そこで, 金額シェアベース等でも適合度が良い当てはめを行うため(他の統計分布の探索は今後の課題として)2 つの方法を試みた。一つは, Lorenz 曲線の二乗誤差の最小化法(L-OLS)で, 1999 年以降, 尤度, L-RSE とも SM に優る。ただし, L-RSE が有意に小さいのは 2009 年のみで, しかも最尤法に比べると尤度等は犠牲にすることになる。もう一つは, 世帯主の年齢階級別年収分布に最尤法で統計分布を当てはめ, 全年齢の年収分布を混合分布として扱う方法で, dPLN, GB2 とも(合成)尤度はパラメータ数の違いを考慮しても 1999 年以降, L-RSE は 2004 年以降 SM に優り, 全年齢に最尤法を適用した場合よりも両評価基準で適合度が改善する。

なお, κG は, 適合がよいとは言いがたいが, L-OLS 法で推定される変動係数の水準や変化傾向が実際に近い。米国の Survey of Consumer Finances の 2004 年以前の等価所得分布では, κG は, 尤度, L-RSE・格差指標とも適合度が高く, 他の国では κG は有力な候補になる可能性がある。

表. 全国消費実態調査の年収分布への適用結果(2人以上の世帯)

		1994 (公式 Gini 係数 = 0.297)						2009 (公式 Gini 係数 = 0.317)					
		LN	SM	Da	κG	GB2	dPLN	LN	SM	Da	κG	GB2	dPLN
最尤法	max ℓ	$\Delta 297.9$	0.0	$\Delta 59.3$	$\Delta 122.9$	19.6	20.9	$\Delta 17.2$	0.0	$\Delta 63.9$	$\Delta 208.5$	73.8	83.3
	L-RSE	0.0128	0.0125	0.0066	0.0074	0.0176	0.0174	0.0155	0.0071	0.0244	0.0170	0.0138	0.0131
	Gini	0.2986	0.2940	0.2993	0.2982	0.2919	0.2919	0.3089	0.3116	0.3183	0.3166	0.3070	0.3073
L-OLS 法	ℓ	$\Delta 298.1$	$\Delta 13.9$	$\Delta 78.2$	$\Delta 130.4$	$\Delta 20.0$	$\Delta 16.5$	$\Delta 21.2$	$\Delta 11.2$	$\Delta 149.7$	$\Delta 236.0$	50.6	66.0
	L-RSE	0.0127	0.0033	0.0046	0.0061	0.0033	0.0031	0.0100	0.0051	0.0089	0.0106	0.0038	0.0035
	Gini	0.2987	0.2993	0.2997	0.2999	0.2994	0.2994	0.3113	0.3131	0.3136	0.3141	0.3125	0.3125
最尤法 (粗区分)	max ℓ	$\Delta 306.5$	0.0	$\Delta 39.8$	$\Delta 93.2$	16.4	19.0	$\Delta 18.6$	0.0	$\Delta 51.5$	$\Delta 182.1$	72.8	81.2
	L-RSE	0.0761	0.0682	0.0781	0.0776	0.0636	0.0628	0.0083	0.0057	0.0216	0.0200	0.0139	0.0143
	Gini	0.2987	0.2944	0.3011	0.3020	0.2903	0.2895	0.3087	0.3125	0.3198	0.3193	0.3052	0.3051
年齢別 最尤法	合成 ℓ^*	$\Delta 98.0$	9.2	8.7	$\Delta 23.2$	7.2	9.9	$\Delta 28.7$	89.1	81.2	78.4	137.4	141.5
	L-RSE	0.0056	0.0024	0.0046	0.0029	0.0022	0.0031	0.0064	0.0085	0.0046	0.0036	0.0014	0.0016
	Gini	0.2966	0.2983	0.3000	0.2985	0.2977	0.2993	0.3094	0.3076	0.3094	0.3102	???	0.3122

* 合成対数尤度は, 年齢階級別年収分布に当てはめて得られる全年齢年収(混合)分布の対数尤度

各分布の(合成)対数尤度 ℓ は, 3 パラメータ分布の中で最も適合がよい SM 分布の(全年齢)最尤法による対数尤度との差を表章

Reed は, dPLN のジニ係数に解析的な算式が存在しないと述べているが, 実際には存在し, さらに混合分布のジニ係数も解析的に算出可能であることが分かった。これは LN と共通し, 他の統計分布には無い特徴である。

参考文献

- [1] Clementi, F., Gallegati, M., and Kaniadakis, G. (2007). κ -generalized statistics in personal income distribution, *European Physical Journal B*, Vol. **52**, 187-193.
- [2] McDonald, J. B. (1984). Some generalized functions for the size distribution of income, *Econometrica*, Vol. **52**, 647-663.
- [3] Reed, W. J. (2003). The Pareto law of incomes – an explanation and an extension, *Physica A*, Vol. **319**, 579-597.