

二重パラメータ化二重パレート対数正規分布の所得分布への当てはめ

総務省 岡本 政人

Reed [3]が提案した二重パレート対数正規分布 dPLN は、以下の確率密度関数(pdf)をもつ。

$$f_{dPLN}(x|\mu, \sigma, \alpha, \beta) = \frac{\alpha\beta}{\alpha+\beta} x^{\beta-1} e^{-\beta\mu+\beta^2\sigma^2/2} \Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right) + \frac{\alpha\beta}{\alpha+\beta} x^{-\alpha-1} e^{\alpha\mu+\alpha^2\sigma^2/2} \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right)$$

ここで、 $\sigma, \alpha, \beta > 0$, Φ は標準正規分布の累積分布関数、 $\Phi^c = 1 - \Phi$ 。dPLN は、多くの国の所得分布に関して第二種一般化ベータ分布 GB2 以上の適合度を達成しているほか([4]など)、多くの国における都市の人口規模分布への適合度も良いことが報告されている([1]など)。Reed and Wu [4]は dPLN を 5 パラメータ・モデル(GdPLN)に拡張したが、実データへの当てはめには特性関数の逆変換を数値計算で行う必要があり実用的でないことや、当てはめたモデルから推定した格差指標値の精度も考慮すると必ずしも適合度が向上するとは言えない。同様の問題は GB2 の 5 パラメータ拡張モデル GB についてもいえる。そこで、以下の pdf をもつ dPLN の 7 パラメータ拡張モデル dP²LN を検討した。

$$f_{dP^2LN}(x|\mu_L, \mu_R, \sigma_L, \sigma_R, \alpha, \beta, r) =$$

$$r \frac{1}{\beta} x^{\beta-1} e^{-\beta\mu_L+\beta^2\sigma_L^2/2} \Phi^c\left(\frac{\log x - \mu_L + \beta\sigma_L^2}{\sigma_L}\right) + (1-r) \frac{1}{\alpha} x^{-\alpha-1} e^{\alpha\mu_R+\alpha^2\sigma_R^2/2} \Phi\left(\frac{\log x - \mu_R - \alpha\sigma_R^2}{\sigma_R}\right)$$

ここで、 $\sigma_L, \sigma_R, \alpha, \beta > 0$, $0 < r < 1$ 。dPLN と同様に $h < \alpha$ ならば h 次モーメントが存在し、ジニ係数を解析的に表現でき、ローレンツ曲線がインプリシットに表される。ただし、dPLN とは異なり、 $\beta \geq 1$ のときの単峰性は一般には保証されない。また、dP²LN のフルモデルをパラメータに制約を課さずに最尤法で所得分布に当てはめると通常尤度は大幅に向上するものの、パラメータが不安定で格差指標の推定精度が悪化する傾向がみられるため、例えば $\sigma_L = \sigma_R$, $\mu_L < \mu_R$, $\beta = 1$ などといった制約条件の中から適切なものを選択し適用する必要がある。

dP²NL 及びその特殊ケースや既存の統計分布を全国消費実態調査の 2 人以上世帯の年収分布に最尤法で当てはめた結果を以下の図表に掲げる。制約 $\sigma_L = \sigma_R$ を課したモデル(dP²NL μ_r)は、pdf が単峰で最尤パラメータが安定し、既存モデルに比べて AIC が大幅に改善するとともに格差指標の推定精度を示すローレンツ曲線の RSSE 推定誤差(L-RSSE)も既存モデルの最小水準にほぼ匹敵している。なお、アメリカ及びイタリアの所得分布では、パラメータに適切な制約条件(dP²NL μ_r とは異なる)を課すことで AIC 及び L-RSSE とも既存モデルを大幅に上回る適合度を達成している。

表. 統計分布モデルの世帯年収分布への適合度

モデル	パラメータ数	2004 年		2009 年		1984~2009 年	
		AIC	L-RSSE	AIC	L-RSSE	AIC	L-RSSE
GG	3	135.2	0.028	116.4	0.022	220.5	0.032
SM	3	163.4	0.010	164.6	0.007	98.1	0.011
GB2	4	23.0	0.020	19.0	0.014	15.0	0.017
EκG1	4	73.0	0.014	70.2	0.008	47.0	0.015
EκG2	4	31.2	0.019	25.6	0.014	25.5	0.016
IEκG1	4	151.8	0.007	155.2	0.010	97.6	0.009
IEκG2	4	18.8	0.020	15.4	0.014	10.4	0.017
dPLN	4	0.0	0.019	0.0	0.013	0.0	0.017
GB	5	24.8	0.021	21.0	0.014	9.9	0.029
GdPLN	5	-55.4	0.030	-28.2	0.009	-20.8	0.041
dP ² LN μ_r	6	-140.8	0.009	-132.6	0.014	-61.7	0.011
dP ² LN	7	-170.2	0.028	-168.6	0.024	-77.3	0.192

注. AIC は dPLN との差を表章。dP²LN のフルモデルの pdf は 1989 年のみ双峰。

EκG1, EκG2 は第 1, 2 種拡張κ一般化分布, IEκG1, IEκG2 は各々 EκG1, EκG2 の逆分布([2])。

参考文献 (プレゼン資料は http://www.geocities.jp/stat_okamoto/jp/ に掲載)

- [1] Giesen, K. and Suedekum, J. (2012). The size distribution across all “cities”: a unifying approach, IZA discussion paper 6352, Institute for the Study of Labor (IZA).
- [2] Okamoto, M. (2013). Extension of the κ-generalized distribution: new four-parameter models for the size distribution of income and consumption, LIS working paper 600.
- [3] Reed, W.J. (2003). The Pareto law of incomes – an explanation and an extension, *Physica A* **319**, 579–597.
- [4] Reed, W.J. and Wu, F. (2008). New four- and five-parameter models for income distributions, in *Modeling Income Distributions and Lorenz Curves* by D. Chotikapanich (ed.), New York: Springer, 211–223.

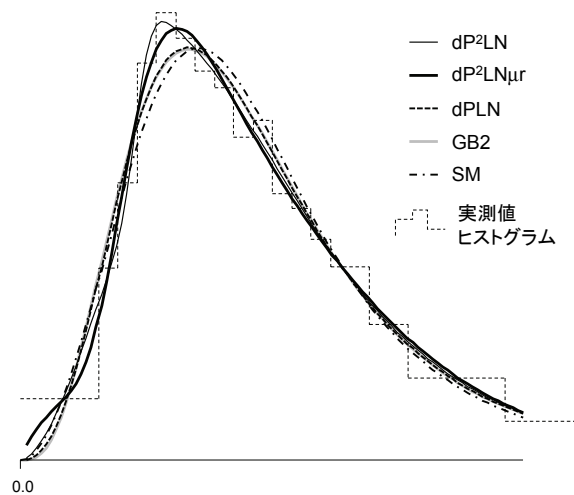


図. 世帯年収の推定密度分布(2009 年)