

Pólya-gamma 分布によるデータ拡大と応用

入江 薫（東京大学経済学部）

Polson, Scott and Windle (2013) で提案された Pólya-gamma (PG) 拡大は、二項尤度に関する積分表現にもとづくデータ拡大の一種である。潜在変数を導入することで尤度関数が条件付きで正規尤度となり、完全条件付き事後分布の解析的計算が可能になるのが特徴で、ロジット回帰を含む多くの非正規モデルのベイズ分析に活用されている。潜在変数の条件付き分布は PG 分布と呼ばれる確率分布になるが、当該論文により効率的な乱数生成法が考案され、ギブスサンプラーにより容易に事後分布を計算することが可能である。手法の普及とともに、ロジット回帰以外のモデルへの拡張や、データ拡大の効率性に関する理論的考察などの研究が多くなされている。

本講演では PG 拡大の基礎事項と先行研究を概観したのち、具体的なデータとモデルへの応用研究を三つ紹介する。第一の研究では、動的なロジット回帰に PG 拡大を適用し、フィルタリング・スムージングを活用して事後計算を行う。このモデルおよび計算手法は財・サービスのカテゴリ別に計算されたインフレ率のデータに現れる過剰なゼロを説明するのに用いられ、実際に消費者物価指数のデータ分析に応用される。第二の研究では、カウントデータの分析に用いられる負の二項分布に対して PG 拡大を適用し、カウント値の時系列データの分析を行う。タクシーの需要データをリアルタイムで分析および予測するために、PG 拡大を利用した逐次モンテカルロ法を考案する。第三の研究では、回帰分析において回帰係数が重みパラメータである（単体上に値をとる）場合の事後分析を扱い、ロジスティック正規分布を事前分布に用いた場合に PG 拡大が利用可能であることを示す。このような制約付き回帰問題は、単調性や凸性などの性質（形状制約）を満たす関数のノンパラメトリック推定に見られる。単調増加、凸、かつ端点条件を持つ関数の具体例としてローレンツ曲線を取り上げ、所得データに基づく日本の経済的不平等の変化を PG 拡大を用いて分析する。

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American statistical Association*, 108(504), 1339-1349.

Han, G., & Irie, K. (2024). Zero-inflated stochastic volatility model for disaggregated inflation data with exact zeros. *arXiv preprint arXiv:2403.10945*.

Masuda, R., Irie, K., & Aktekin, T. (202X). Sequential analysis of count-valued multivariate time series with seasonal factors: real-time monitoring and forecasting of demand in ride-sharing, mimeo.

Hiraki, D., Hamura, Y., Irie, K., & Sugawara, S. (2024). State-space modeling of shape-constrained functional time series. *arXiv preprint arXiv:2404.07586*.