

連続型円周分布の離散化

静岡県立大学 経営情報学部 井本智明
統計数理研究所 統計思考院 清水邦夫

風向や野菜の収穫時期のような周期性を持つデータを解析する際には、周期性を持つ確率密度関数 $f(\theta)$ により定義される円周分布によってモデルリングを行うことが好ましい。しかし、このようなデータは連続型データであることが多いが、その性質上、収集されたデータは離散型として近似されたものになってしまうことが多い。

本研究では、連続型円周分布の確率密度関数 $f(\theta)$ から、 $n \geq 2$ に対し、

$$p_n(j) = \frac{1}{C_n} f\left(\frac{2\pi}{n}j\right), \quad j = 1, 2, \dots, n-1, \quad \text{where} \quad C_n = \sum_{j=0}^{n-1} f\left(\frac{2\pi}{n}j\right)$$

という手法によって離散型円周分布の確率関数 $p_n(j)$ を定義する。この確率関数 $p_n(j)$ は、基となる確率密度関数 $f(\theta)$ の近似関数となっており、その基準化係数 C_n は少なくとも有限和で表されることから、確率密度関数 $f(\theta)$ の基準化係数が複雑で計算が困難な場合にも容易に利用できる利点がある。

また、基となる確率密度関数 $f(\theta)$ がフーリエ級数表現可能な場合、つまり

$$f(\theta) = \frac{1}{2\pi} \left[1 + 2 \sum_{p=1}^{\infty} \{ \alpha_p \cos(p\theta) + \beta_p \sin(p\theta) \} \right]$$

とできる場合、確率関数 $p_n(j)$ の基準化係数 C_n は

$$C_n = \frac{n}{2\pi} \left(1 + 2 \sum_{p=1}^{\infty} \alpha_{np} \right)$$

とコサインモーメントのみでの表現が可能となる。この性質により、カーディオイド分布や巻き込みコーシー分布、そしてそれらの正弦歪分布 (Abe and Pewsey, 2011) の離散化分布が閉じた関数で表現可能なことを示す。また、Fernández-Durán 型分布 (Fernández-Durán, 2004) や4パラメータ拡張巻き込みコーシー分布 (Kato and Jones, 2015) の離散化分布も閉じた関数で表現可能なことを示し、それらの分布を利用したデータ当てはめを見ていく。

その他にも、確率関数 $p_n(j)$ の逆関数法を利用した乱数生成により、基の確率密度関数 $f(\theta)$ の近似的な乱数生成を行う。そして、その乱数を利用したシミュレーション実験により、基の確率密度関数 $f(\theta)$ と離散化した確率関数 $p_n(j)$ による推定値の乖離度を見ていく。

参考文献

- [1] Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions, *Statistical Papers*, 52, 783-707.
- [2] Fernández-Durán, J. J. (2004). Circular distributions based on nonnegative trigonometric sums. *Biometrics*, 60, 499-503.
- [3] Kato, S. and Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle, *Biometrika*, 102, 181-190.