2019年6月5日　統計数理研究所　オープンハウス

整数全体で値をとる自己回帰モデルの拡張と応用

川崎　能典　**モデリング研究系　教授**

1. SINAR(1) モデル

準備1：スケラム分布　X, Y が独立でそれぞれパラメータ λ_1, λ_2 のポアソン分布に従うとする．このとき，$Z = X - Y$ の従う分布を，パラメータ λ_1, λ_2 のスケラム(Skellam)分布という．ここでは，パラメータが λ_1, λ_2 のスケラム分布を $S(\lambda_1, \lambda_2)$ と記す．スケラム分布 $S(\lambda_1, \lambda_2)$ の確率関数は以下で与えられる．

$$f(z) = e^{-(\lambda_1+\lambda_2)}\lambda_1^z \sum_{x=\max(0,-z)}^{\infty} \frac{(\lambda_1\lambda_2)^x}{(x+z)!x!} \quad (x \in \mathbb{Z})$$

準備2：PINAR(1) モデル　X_t, Y_t をそれぞれ独立でパラメータの等しい PINAR(1) モデル(Poisson INteger-valued AutoRegressive model)とする．即ち，

$$\begin{aligned} X_t &= \alpha \circ X_{t-1} + \delta_t, \quad X_t \sim Po(\lambda), \quad \alpha \circ X_{t-1}|X_{t-1} \sim B(X_{t-1}, \alpha), \\ Y_t &= \alpha \circ Y_{t-1} + \eta_t, \quad Y_t \sim Po(\lambda), \quad \alpha \circ Y_{t-1}|Y_{t-1} \sim B(Y_{t-1}, \alpha) \end{aligned}$$

とする．ここで $\{B_i\}$ を独立にベルヌーイ分布 $B(1, \alpha)$ に従う確率変数とするとき，平均が λ であるポアソン分布 $Po(\lambda)$ に従う確率変数 X_{t-1} に対し，$\alpha \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} B_i$ で定義される．すなわち，$\alpha \circ X_{t-1}$ のとりうる値は 0 から X_{t-1} までの整数値であり，X_{t-1} を "間引く" という意味で $\circ$ を間引き演算子 (thinning operator) と呼ぶ．

SINAR(1) モデル　このとき, 2つのPINAR(1)モデルの差を取り，$Z_t = X_t - Y_t$, $\alpha \star Z_{t-1} = \alpha \circ X_{t-1} - \alpha \circ Y_{t-1}$, $\varepsilon_t = \delta_t - \eta_t$ とおくと

$$Z_t = \alpha \star Z_{t-1} + \varepsilon_t \qquad \{Z_t\} \sim S(\lambda, \lambda)$$

である．このモデルを SINAR(1) モデル(Skellam INteger-valued AutoRegressive model)という．周辺分布がスケラム分布であり，Z_t のとりうる値は整数全体 $\mathbb{Z}$ となっていることがわかる．

Z_t が正負の値をとりうるなると，その自己相関が正の場合のみならず負の場合についても積極的に検討が必要となる．Freeland (2010) では，奇数と偶数で X_t と Y_t を入れ替える方法が提案されている．すなわち，

$$Z_t = \begin{cases} X_t - Y_t & (t = 0, 2, 4, \dots) \\ Y_t - X_t & (t = 1, 3, 5, \dots) \end{cases}, \quad \varepsilon_t = \begin{cases} \delta_t - \eta_t & (t = 0, 2, 4, \dots) \\ \eta_t - \delta_t & (t = 1, 3, 5, \dots) \end{cases}$$

のように偶数と奇数で X_t と Y_t の差の順序を入れ替えると，$\rho(k) = (-\alpha)^k$ が成り立ち，1次の自己相関が負の場合を表現することができる．

2. 拡張SINAR(1) モデル

ここからは，X_t, Y_t をそれぞれ独立な PINAR(1) モデルとし，そのポアソン分布の平均は異なるとする．すなわち

$$\begin{aligned} X_t &= \alpha \circ X_{t-1} + \delta_t \quad \{X_t\} \sim Po(\lambda_1),\ \alpha \circ X_{t-1}|X_{t-1} \sim B(X_{t-1}, \alpha), \\ Y_t &= \alpha \circ Y_{t-1} + \eta_t \quad \{Y_t\} \sim Po(\lambda_2),\ \alpha \circ Y_{t-1}|Y_{t-1} \sim B(Y_{t-1}, \alpha) \end{aligned}$$

とする．このとき，この2つのPINAR(1)モデルの差を取ったモデル

$$Z_t = \alpha \star Z_{t-1} + \varepsilon_t \quad \{Z_t\} \sim S(\lambda_1, \lambda_2) \tag{1}$$

を考える．ただしここで，$Z_t = X_t - Y_t$, $\alpha \star Z_{t-1} = \alpha \circ X_{t-1} - \alpha \circ Y_{t-1}$, $\varepsilon_t = \delta_t - \eta_t$ である．$\lambda_1 = \lambda_2$ のときはSINAR(1)モデルであり，このモデルはそれを拡張したものになっている．

ε_t のモーメント母関数, 期待値, 分散は，中嶋・酒折・川崎 (2017) に与えられている．ここではパラメータ推定法だけ紹介する．パラメータ推定は，PINAR(1)モデルと同様にモーメント法などで行うことができ，推定量

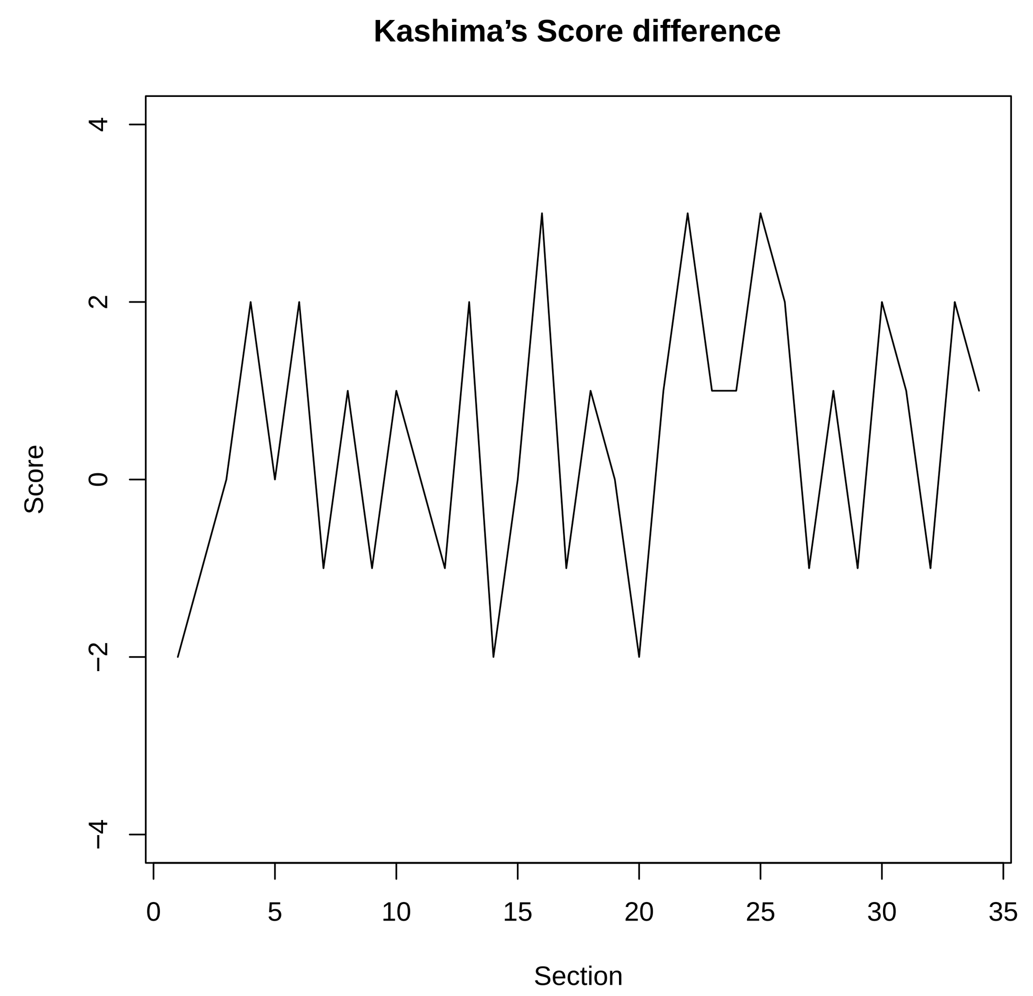
$$\begin{cases} \hat{\alpha} = \hat{\rho}(1) \\ \hat{\lambda}_1 = \frac{1}{2}\left(S_Z^2 + \bar{z}\right) \\ \hat{\lambda}_2 = \frac{1}{2}\left(S_Z^2 - \bar{z}\right) \end{cases}$$

が得られる．ここで，S_Z^2 は Z の不偏標本分散である．また，$\hat{\lambda}_1, \hat{\lambda}_2$ はそれぞれパラメータ λ_1, λ_2 の不偏推定量である．

3. 実例：サッカーにおける得失点差時系列

整数値自己回帰モデルの実データへの適用として，サッカーにおける各試合の得失点差のデータに拡張SINAR(1)モデルにあてはめた例を紹介する．データは2015年サッカーJ1リーグ戦における，鹿島アントラーズの各試合の得失点差である．

全34試合(節)における得失点差の標本平均は0.47，標本分散は1.5であった．下の図はその推移を表す折れ線グラフである．正負の値をとりながら，上昇と下降を交互に繰り返している傾向から負の相関が予想されるが，実際1次の自己相関係数は -0.16 と負の値(有意)であり，得失点差時系列には従属性が認められる．



そこで，(1)式で与えられる拡張SINAR(1) モデルによる分析を行う．その際，1次の自己相関が負であることを考慮し，前述の要領で偶数時点と奇数時点で得点数 X_t と失点数 Y_t の差の順序を入れ替えて Z_t を構成する．パラメータをモーメント法で推定すると

$$\hat{\alpha} = -\hat{\rho}(1) = 0.16, \quad \hat{\lambda}_1 = 1.33, \quad \hat{\lambda}_2 = 0.86$$

が得られる．

1次の自己相関が整数値時系列の1期先予測にどのような差異をもたらすかを検証しよう．第34節(最終節)の得失点差は1であり，これ以上の試合はなかったが，もし次の試合があったとした場合の得失点差を予測することは可能である．例えば $Z_{34} = 1$ である条件の下で $Z_{35} = 0$ となる確率を評価，というような作業を繰り返せば良い．計算の詳細は中嶋・酒折・川崎 (2017) に譲るが，時系列構造を仮定しない独立なスケラム分布に基づく分析に照らして予測結果をまとめたのが下の表である．

得失点差	−3	−2	−1	0	1	2	3
確率(独立スケラム)	0.016	0.060	0.163	0.282	0.252	0.143	0.058
確率(SINAR(1))	0.021	0.076	0.191	0.295	0.235	0.118	0.043

いずれのモデルを用いた場合においても得失点差が0の確率が最も大きく，続いて得失点差1，得失点差 -1 の確率が大きいことがわかる．しかし, その確率自体を比較してみると，時系列相関を考慮したSINAR(1)モデルのほうが，全体的に得失点差が負の方向に分布が寄っていることがわかる．

謝辞 本研究は中嶋雅彦氏(中央大学大学院：当時)，酒折文武准教授(中央大学)との共同研究であり，この研究を進めるにあたって中嶋氏は統計数理研究所の特別共同利用研究員制度を利用した．

参考文献 中嶋雅彦，酒折文武，川崎能典 (2017). 整数値自己回帰モデルの最近の発展，統計数理，第65巻2号，323–339.