

# 整数値自己回帰モデルの最近の発展

中嶋 雅彦<sup>1</sup>・酒折 文武<sup>2</sup>・川崎 能典<sup>3,4</sup>

(受付 2017 年 3 月 31 日；改訂 7 月 4 日；採択 7 月 4 日)

## 要 旨

整数値自己回帰モデル(Integer-valued Autoregressive Models, INAR モデル)は、自分自身の過去(整数値)に直接依存させて現時点での値(整数値)を説明するものであり、潜在過程に基づく動的な一般化線形モデルや一般化状態空間モデルとは、データへの接近法は大きく異なる。特に、自己回帰モデルの左辺と右辺において確率分布が整合的であるためには、周辺分布、自己回帰項、イノベーションのそれぞれの定式化が重要である。1980 年代後半から 90 年代初期にいくつかの研究が行われたあとやや停滞していたこの分野は、2000 年代後半から活性化してきている。本稿は、最近の整数値自己回帰モデルの発展を、ポアソン分布に基づく INAR モデルから出発して、定式化を変えた各種モデルを順に取り上げることで概観する。その過程で、必ずしも既存の文献で詳述ないし証明の与えられていない結果に関しては、著者なりに補った結果を付録に収めた。ポアソン分布の差から生成される分布に基づく INAR モデルについては、若干新しい提案と結果を加えることができた。その枠組みを利用した実データ解析例を、最後に紹介する。

キーワード：整数値時系列データ、間引き演算子、確率変数の分解、INAR(1) モデル、INAR( $p$ ) モデル、モーメント法。

## 1. はじめに

センサー技術など情報通信技術の発達に伴うデータ取得環境の変化の中で、様々な時系列データが収集・活用されるようになり、時系列データに対する統計的モデリングの重要性が高まっている。時系列解析においては、自己回帰移動平均(ARMA)モデルに代表される、イノベーションに正規分布を仮定した線形モデルが、現在でも最も頻繁に用いられるモデルであり、その対象は理学・工学のみならず経済学等の社会科学にも広がって久しい。

しかしながら、このような枠組みが、観測される時系列が整数値である場合でも妥当かどうかは注意を要する。例えば、マクロ経済時系列データの多くは、その取りうる値が非負で整数値とはいっても、概ね連続量と見なして差し支えないであろう。従って、イノベーションならびにデータに対する連続分布の仮定は近似として許される範囲として、あとは分布形とモデルの定式化の問題となる。

一方、観測データが計数時系列(count time series)で、とりわけ稀なイベントの計測から生じ

<sup>1</sup> 中央大学大学院 理工学研究科：〒112-8851 東京都文京区春日 1-13-27

<sup>2</sup> 中央大学 理工学部：〒112-8851 東京都文京区春日 1-13-27

<sup>3</sup> 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

<sup>4</sup> 総合研究大学院大学 統計科学専攻：〒190-8562 東京都立川市緑町 10-3

ているような場合、すなわち 0, 1, 2 からただか 10 程度の計数の連なりである場合はもはや正規近似は妥当性を欠き、データ分布に何らかの離散分布を仮定することが必要である。計数時系列に対するモデリング法として多くの成功事例を集めながら拡張されてきたのは、一般化線形モデル (generalized linear model, GLM; Nelder and Wedderburn, 1972) の時系列版である動的一般化線形モデル (Dynamic GLM, 以下 DGLM と略) であった。このような問題設定では、一般化された状態空間表現に基づく更新アルゴリズムを実現するための新たな計算法の提案が必要とされた。例としては Kitagawa (1987) による数値積分フィルタが挙げられる。

一般化状態空間モデルないし階層構造モデルで少数計測データ (small count data) を分析することの大きなメリットは、背後にある生起度 (intensity) が仮に非定常に近い振る舞いをするような場合でも、生起度を表す潜在時系列モデルの側で柔軟性に富んだ適切なモデルを置いてやれば、少なくとも現象の記述には支障がない点であろう。DGLM に関しては West and Harrison (1997), Fahrmeir and Tutz (2001) を、一般状態空間モデルの推定法に関しては Kitagawa (2010) の第 14 章と第 15 章を参照されたい。

さて、本稿で取り上げる整数値自己回帰モデル (Integer-valued Autoregressive Models, INAR モデル) は、整数値時系列のモデリングのための別種のアプローチである。DGLM や一般化状態空間モデルでは背後の潜在時系列に系列の履歴情報が縮約されるわけであるが、INAR モデルは整数値の履歴自体にモデルを依存させる。そうすると、例えばイノベーションに連続分布を仮定した線形 AR(1) モデル  $X_t = \alpha X_{t-1} + \epsilon_t$  からの類推で考えれば、自己回帰項  $\alpha X_{t-1}$  の部分が整数値となるよう定義を与え直し、かつ両辺の分布が整合する形でイノベーション  $\epsilon_t$  の分布を定式化しなければならない。INAR モデルにおける自己回帰項の定義は、2.1 節で説明する通り「間引き (thinning)」で与えられる。間引きの構造は、自己回帰項が 0 から  $X_{t-1}$  の値を取りうる形で定式化されており、それが 0 に近いのか  $X_{t-1}$  に近いのか、換言すれば時系列として持続性 (persistence) が希薄なのか強いのかは、間引き項を生成する確率分布のパラメータ設定次第となっている。

以下では、INAR モデルとその拡張に関する文献をサーベイしておこう。著者らが文献上確認できた INAR に関する最も古い論文は McKenzie (1985) である。INAR(1) モデルを  $p$  次に拡張した研究として、Al-Osh and Alzaid (1990), Du and Li (1991) がある。Du and Li (1991) では INAR( $p$ ) モデルのパラメータ  $\alpha_1, \alpha_2, \dots, \alpha_p$  をユールウォーカー法により推定できることを示しており、Al-Osh and Alzaid (1990) では、INAR( $p$ ) モデルの周辺分布がポアソン分布のとき、 $\sum_{i=1}^p \alpha_i \leq 1$  ならば 1 次のときと同様に確率変数が退化せずに分解可能であることが述べられている。

こうして 80 年代後半から 90 年代初頭に散発的に行われていた研究は、2000 年代後半に入っようやく活性化する。Leonenko et al. (2007) は INAR(1) モデルの周辺分布をポアソン分布ではなく負の二項分布とした NBD INAR(1) モデルを提案し、過分散なデータに対する INAR モデルを提案している。Savani and Zhigljavsky (2007a, 2007b) では、NBD INAR(1) モデルのパラメータのモーメント法による推定について言及しており、3 種類の推定量を提案し、比較している。

Fokianos (2011) は周辺分布をポアソン分布とした PINAR(1) モデルを中心に解説を加えているが、ハンドブックの一章ということもありサーベイ的な側面も強いことから、現時点でこの分野の研究を始める際にまず目を通しておくべき文献と言える。Rajarshi (2012) も INAR を含む離散確率過程のモノグラフであり、間引き項を二項分布とし、周辺分布を一般的によく知られた離散型分布としたモデルについて広く述べている点に特徴がある。

間引き項を二項分布ではなく幾何分布や超幾何分布とした研究もある。Rinstić et al. (2009) は、間引き演算子を幾何分布、周辺分布を幾何分布とした NGINAR(1) モデルを導入し、パラ

表 1. INAR モデルの概要.

名称	間引き演算子	周辺分布	誤差分布	$p$ 次の拡張	負の場合の拡張
PINAR	二項	ポアソン分布	ポアソン分布	あり	なし
NBD INAR	二項	負の二項分布	負の二項幾何分布	なし	なし
BINAR	二項・超幾何	二項分布	二項分布	あり	なし
SINAR	二項	スケラム分布	スケラム分布	あり	あり
NGINAR	負の二項	幾何分布	混合幾何分布	なし	なし
STINAR	負の二項	SDL 分布	混合 SDL 分布	なし	あり

メータ推定にも言及している. Weiß (2008) は, 間引き演算子を超幾何分布, 周辺分布を二項分布とした BINAR(1) モデルを考え, 確率変数の取りうる値が有限な場合の INAR モデルを提案した.

整数値全体をとる時系列データに対する INAR モデルとして Freeland (2010) は, ポアソン分布の差が従う分布としての Skellam 分布 (以下スケラム分布と記す) に着目し, 2つの独立な PINAR(1) モデルの差による SINAR(1) モデル (Skellam INAR(1) モデル) を提案した. なお, Freeland (2010) では, 同じパラメータを持つポアソン分布の差を取った研究しか述べられていない. また, Baretto-Souza and Bourguignon (2013) は, 2つの独立な NGINAR(1) 過程の差を考えた非対称離散ラプラス分布による STINAR(1) モデルを提案し, さらに負の相関をもつ場合への拡張を与えた. ここまで名前を挙げたモデル群を, その特徴と拡張性の現状という観点から整理したのが, 表 1 である.

以上を踏まえ本稿では, まず PINAR モデルから始めて, NBD INAR モデル, BINAR モデル, SINAR モデルの定義と性質について紹介する. 文献上必ずしも詳述ないし証明が与えられていない結果に関しては, 著者らなりに補った結果を付録に収めている. 本稿は基本的には研究詳解であるが, SINAR(1) モデルの拡張とパラメータ推定法に関しては, 若干新しい提案を加えることができたので 4.2 節に記し, 5 節で実データへの応用を試みた結果を報告する.

## 2. 平均と分散が等しい場合: PINAR モデル

非負整数値全体をサポートとする基本的な分布であるポアソン分布を  $X_t$  の周辺分布とする PINAR モデル (Rajarshi, 2012) は, INAR モデルの中で最も代表的なモデルである.

### 2.1 PINAR(1) モデル

PINAR(1) モデルは以下で表される:

$$(2.1) \quad X_t = \alpha \circ X_{t-1} + \varepsilon_t, \quad X_t \sim Po(\lambda).$$

いま,  $X_t$  を平均が  $\lambda$  であるポアソン分布  $Po(\lambda)$  に従う確率変数とする. また,  $\alpha \in (0, 1]$ ,  $X$  を非負整数値とし,  $\{Y_i\}$  を独立にベルヌーイ分布  $B(1, \alpha)$  に従う確率変数とする. このとき,  $\alpha \circ X$  を確率変数として  $\alpha \circ X = \sum_{i=1}^X Y_i$  と定義する. すなわち,  $\alpha \circ X$  のとりうる値は 0 から  $X$  までの整数値であり,  $X$  を“間引く”という意味で  $\circ$  を間引き演算子 (thinning operator) と呼ぶ. このとき,  $X_{t-1} \sim Po(\lambda), \alpha \circ X_{t-1} | X_{t-1} \sim B(X_{t-1}, \alpha)$  より,  $\alpha \circ X_{t-1}$  が平均  $\lambda\alpha$  のポアソン分布に従うことが示せる. したがって, ポアソン分布の再生性より,  $\{\varepsilon_t\} \sim Po(\lambda(1-\alpha))$  が成り立つこともわかり, PINAR(1) のパラメータは  $\alpha$  と  $\lambda$  のみで表されることが確認できる.

PINAR(1) モデルに関する基本的な性質を見ておこう. 条件付き期待値, 条件付き分散は

$$E(X_t | X_{t-1}) = E(\alpha \circ X_{t-1} + \varepsilon_t | X_{t-1}) = \alpha X_{t-1} + (1-\alpha)\lambda,$$

$$V(X_t|X_{t-1}) = V(\alpha \circ X_{t-1} + \varepsilon_t|X_{t-1}) = \alpha(1-\alpha)X_{t-1} + (1-\alpha)\lambda$$

となる。また、 $E(X_t)$  は条件付き期待値の性質を用いると

$$E(X_t) = E(E(X_t|X_{t-1})) = E(\alpha X_{t-1} + (1-\alpha)\lambda) = \lambda,$$

$V(X_t)$  は条件付き分散の性質を用いると

$$V(X_t) = V(E(X_t|X_{t-1})) + E(V(X_t|X_{t-1})) = \lambda$$

となり、周辺分布であるポアソン分布の期待値と分散に一致していることが確認できる。PINAR(1) モデルは、1 時点前の値に依存しているため 1 次のマルコフ性を持ち、状態  $i$  から  $j$  に推移する確率は

$$p_{ij} = P(X_{t+1} = j|X_t = i) = \sum_{z=0}^{\min(i,j)} \binom{i}{z} \alpha^z (1-\alpha)^{i-z} \frac{e^{-\lambda} \lambda^{j-z}}{(j-z)!}$$

となる。推移確率が時間  $t$  に依存しないので、この時系列過程は斉時性が成り立つことが分かる。また、このマルコフ系列は任意の  $i, j$  に対して  $p_{ij} > 0$  となることから、既約で非周期的であることが分かる。

## 2.2 PINAR( $p$ ) モデル

PINAR(1) モデルは  $p$  次へ自然に拡張できる (Al-Osh and Alzaid, 1990 参照)。PINAR( $p$ ) モデルは

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + \varepsilon_t$$

で定義される。ここで、 $X_t \sim Po(\lambda)$ ,  $\sum_{i=1}^p \alpha_i \leq 1$ ,  $\alpha_i \circ X_{t-i} \sim B(X_{t-i}, \alpha_i)$  である。このとき、 $\alpha \circ X_{t-j} \sim Po(\alpha_j \lambda)$  となり、 $\varepsilon_t \sim Po((1 - \sum_{i=1}^p \alpha_i) \lambda)$  であることが分かる。

PINAR( $p$ ) モデルにおいて、 $X_t$  の条件付き確率分布は

$$P(X_t = k|X_{t-1} = x_1, \dots, X_{t-p} = x_p) = \prod_{i=1}^p \sum_{z_i=0}^{\min(x_i, k)} \binom{x_i}{z_i} \alpha_i^{z_i} (1-\alpha_i)^{x_i-z_i} \frac{\lambda^{k-\sum_{i=1}^p z_i} e^{-\lambda}}{(j-\sum_{i=1}^p z_i)!},$$

条件付き期待値、条件付き分散は

$$(2.2) \quad E(X_t|X_{t-1}, X_{t-2}, \dots, X_{t-p}) = \sum_{i=1}^p \alpha_i X_{t-i} + \left(1 - \sum_{i=1}^p \alpha_i\right) \lambda,$$

$$(2.3) \quad V(X_t|X_{t-1}, X_{t-2}, \dots, X_{t-p}) = \sum_{i=1}^p \alpha_i (1-\alpha_i) X_{t-i} + \left(1 - \sum_{i=1}^p \alpha_i\right) \lambda$$

で与えられる。この証明は Al-Osh and Alzaid (1990) では示されていないので、付録 A.2 に記しておく。

## 2.3 PINAR モデルにおけるパラメータ推定

PINAR モデルにおけるパラメータ推定はモーメント法を用いるのが最も単純である。ここでは PINAR(1) モデルにおけるパラメータ  $\alpha, \lambda$  のモーメント法での推定量について述べておく。まず、1 次の自己共分散  $\text{Cov}(X_t, X_{t-1})$  は

$$(2.4) \quad \text{Cov}(X_t, X_{t-1}) = E(X_{t-1}(\alpha \circ X_{t-1} + \varepsilon_t)) - E(X_t)E(X_{t-1}) = \alpha V(X_{t-1})$$

である。定常性より  $V(X_t) = V(X_{t-1})$  であるので、1 次の自己相関  $\rho(1)$  は

$$\rho(1) = \frac{\text{Cov}(X_t, X_{t-1})}{\sqrt{V(X_t)V(X_{t-1})}} = \alpha$$

となることが分かる。よって、 $\alpha$  のモーメント法による推定量は、1 次の標本自己相関  $\hat{\rho}(1)$  を用いて  $\hat{\alpha} = \hat{\rho}(1)$  とすればよい。ただし、 $\alpha$  は間引きの確率を表すため非負であるが、標本自己相関は負の値をとることもありうるため注意が必要である。さらに、 $X_t \sim Po(\lambda)$  なので、 $E(X_t) = \lambda$  であることから、 $\lambda$  のモーメント法による推定量は標本平均  $\bar{X}$  を用いて  $\hat{\lambda} = \bar{X}$  とすればよい。なお、 $\hat{\lambda}$  は、 $E(\hat{\lambda}) = \lambda$ 、 $V(\hat{\lambda}) = \frac{1}{nT(\lambda)}$  であることから、不偏性と有効性をもつことがわかる。

もちろん最尤推定量などを考えることも可能であるが、推定量は陽関数としては得られず数値的に解くことになるため、最尤推定量のもつ漸近的な性質以上のことを言うことが困難となる。

### 3. その他の非負整数値自己回帰モデル

本節では、まず周辺分布を負の二項分布とした NBD INAR(1) モデルについて述べる。次に周辺分布を二項分布とした BINAR(1) モデルについて述べる。

#### 3.1 過分散の場合：NBD INAR(1) モデル

これまで、INAR(1) の周辺分布をポアソン分布とした PINAR モデルについて述べた。ポアソン分布には平均と分散が等しいという特徴があるので、データの平均と分散がほぼ等しければ PINAR(1) は妥当なモデルであると考えられる。しかしながら、平均よりも分散が大きい、すなわち過分散の場合には PINAR(1) モデルは適切でない。そこで、過分散なモデルとして、INAR(1) モデルの周辺分布を負の二項分布 (NBD) とした NBD INAR(1) モデルについて説明する。なお、これらにおけるパラメータ推定も 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る。詳しくは Leonenko et al. (2007), Savani and Zhigljavsky (2007a, 2007b) を参照されたい。

##### 3.1.1 NBD INAR(1) モデル

NBD INAR(1) モデルの式を次のように定義する：

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \quad \alpha \circ X_{t-1} | X_{t-1} \sim B(X_{t-1}, \alpha), \quad X_t \sim NBD(\gamma, \beta).$$

ここで  $NBD(\gamma, \beta)$  はパラメータ  $\gamma, \beta$  の負の二項分布であり、その確率関数は

$$f(x) = \binom{\gamma + x - 1}{x} \left( \frac{\beta}{\beta + 1} \right)^\gamma \left( \frac{1}{\beta + 1} \right)^x \quad (x = 0, 1, 2, \dots)$$

である。このとき、 $\alpha \circ X_{t-1} \sim NBD(\gamma, \frac{\beta}{\alpha})$  である。証明は Leonenko et al. (2007) を参照されたい。

この NBD INAR(1) モデルにおいては、 $X_t$  と  $\alpha \circ X_{t-1}$  の従う負の二項分布の第 2 パラメータが異なるため、 $\varepsilon_t$  の分布を再生性から求めることはできない。そこで、 $\varepsilon_t$  の確率母関数を求めると、

$$G_{\varepsilon_t}(s) = \frac{G_{X_t}(s)}{G_{\alpha \circ X_{t-1}}(s)} = \alpha^\gamma \left( 1 - (1 - \alpha) \frac{\beta}{\alpha + \beta} \left( 1 - \frac{\alpha s}{\alpha + \beta} \right)^{-1} \right)^{-\gamma}$$

となる。これは、負の二項幾何分布  $NBG(\gamma, \frac{\beta}{\alpha + \beta}, \alpha)$  の確率母関数なので、 $\varepsilon_t \sim NBG(\gamma, \frac{\beta}{\alpha + \beta}, \alpha)$

であることが分かる. なお, パラメータ  $m, p, \theta$  の負の二項幾何分布  $NBG(m, p, \theta)$  は, 以下の確率関数をもつ確率分布である:

$$f(x) = \sum_{k=0}^{\infty} \binom{k+x-1}{x} p^k (1-p)^x \binom{m+k-1}{k} \theta^m (1-\theta)^k \quad (x=0, 1, 2, \dots).$$

この確率母関数は

$$\begin{aligned} G(s) &= \theta^m (1 - (1-\theta)p(1-(1-p)s)^{-1})^{-m} \\ &= \left( \frac{\theta}{1 - (1-\theta)\frac{p}{1-(1-p)s}} \right)^m \end{aligned}$$

であり, 負の二項分布の確率母関数の中に幾何分布の確率母関数が含まれている. これが負の二項幾何分布とよばれる理由である.

### 3.1.2 Stochastic thinning による NBD INAR(1) モデル

先の NBD INAR(1) モデルでは,  $X_t$  の周辺分布および間引き項  $\alpha \circ X_{t-1}$  は負の二項分布であったが, 誤差項  $\varepsilon_t$  は NBG 分布となった. 誤差分布も同じ負の二項分布とするために, Leonenko et al. (2007) では, NBD INAR(1) モデルにおけるパラメータ  $\alpha$  を定数ではなくベータ分布に従って変動する Stochastic thinning (確率の間引き) とするモデルを提案している.

いま,

$$X_t = A_t \circ X_{t-1} + \varepsilon_t, \quad X_t \sim NBD^{(\gamma, \beta)}$$

とし, 間引きの確率を表すパラメータ  $A_t$  が定数ではなく, 独立にベータ分布に従うとする. すなわち,  $A_t \circ X_{t-1} | X_{t-1} \sim B(X_{t-1}, A_t)$ ,  $A_t \sim \text{Beta}(a, \gamma - a)$ ,  $X_t \sim NBD^{(\gamma, \beta)}$  である. このとき,  $A_t \circ X_{t-1} \sim NBD^{(a, \beta)}$  が成り立つことが示される. このことより,  $X_t$  と  $A_t \circ X_{t-1}$  の第 2 パラメータがともに  $\beta$  のため,  $\varepsilon_t$  の分布が負の二項分布の再生性より  $\varepsilon_t \sim NBD^{(\gamma-a, \beta)}$  となることがわかる. このモデルを, stochastic thinning による NBD INAR(1) モデルという.

### 3.1.3 Mixed thinning による NBD INAR(1) モデル

3.1.1 節の間引きパラメータが定数  $\alpha$  のモデルと, 3.1.2 節の確率変数  $A_t$  のモデルを混合し,  $\alpha A_t$  を間引きパラメータとする一般的なモデル

$$X_t = \alpha A_t \circ X_{t-1} + \varepsilon_t$$

も Leonenko et al. (2007) で提案されている. ここで,  $\alpha A_t \circ X_{t-1} | X_{t-1} \sim B(X_{t-1}, \alpha A_t)$ ,  $A_t \sim \text{Beta}(a, \gamma - a)$ ,  $X_t \sim NBD^{(\gamma, \beta)}$  である. このとき,  $\alpha A_t \circ X_{t-1} \sim NBD^{(a, \frac{\beta}{\alpha})}$ ,  $\varepsilon_t \sim NBD^{(\gamma, \frac{\beta}{\alpha})} \cdot NBG(\gamma, \frac{\beta}{\alpha+\beta}, \alpha)$  となることが示される. このモデルを, mixed thinning (混合間引き) による NBD INAR(1) モデルという.

## 3.2 サポートが有限の場合: BINAR(1) モデル

ここまで,  $X_t$  のとりうる値が非負の整数全体のモデルについて考えてきた. しかし実際には, ある有限の値  $n$  までしか値をとらない場合も考えられる. ここでは, 周辺分布を最も代表的な二項分布としたモデルについて説明する.

単純に INAR(1) モデルの周辺分布を二項分布とした場合, 誤差分布が一般的に知られた分布にならない. その改善策として 2 つの方法が提案されている. 誤差分布を用いず, 2 つの二項演算子を用いる方法と, 超幾何演算子を用いる方法である. 詳細は, Weiß (2008, 2009), Rajarshi (2012) を参照されたい. 原論文では両方とも BINAR(1) モデルとして言及されているが, ここでは区別のためにそれぞれ  $\text{BINAR}_b(1)$  と  $\text{BINAR}_h(1)$  と呼ぶこととする.

**3.2.1 BINAR<sub>b</sub>(1) : 2つの二項演算子を用いたモデル**

まず, 2つの二項演算子を用いたモデルを説明する.  $X_t \sim B(n, p)$ ,  $\rho \in [\max(-\frac{1-p}{p}, -\frac{p}{1-p}), 1]$ ,  $\alpha \circ X_{t-1} | X_{t-1} \sim B(X_{t-1}, \alpha)$  とする. また,  $\beta = p(1 - \rho)$ ,  $\alpha = \beta + \rho$  とする. このとき,

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}) \quad (t \geq 1)$$

を BINAR<sub>b</sub>(1) モデルという.

右辺により与えられる分布が  $X_t$  の分布, すなわち二項分布  $B(n, p)$  であることを確認しておこう. 右辺の確率母関数は

$$\begin{aligned} E(s^{\alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})}) &= E((1 - \alpha + \alpha s)^{X_{t-1}}) E((1 - \beta + \beta s)^{n - X_{t-1}}) \\ &= (1 - \beta + \beta s)^n \left( 1 - p + p \left( \frac{1 - \alpha + \alpha s}{1 - \beta + \beta s} \right) \right)^n \end{aligned}$$

となる. ここで,  $\beta = p(1 - \rho)$ ,  $\alpha = \beta + \rho$  を代入すると,

$$E(s^{\alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})}) = (1 - p + ps)^n$$

となることが示される.

BINAR<sub>b</sub>(1) モデルにおける  $X_t$  の 1 次の自己相関  $\rho(1)$  は  $\rho$  である. 証明は付録 A.3 に記した.

**3.2.2 BINAR<sub>h</sub>(1) : 超幾何演算子を用いたモデル**

次に, 間引き演算子に超幾何演算子を用いたモデルについて説明する. いま,  $X_t \sim B(N, p)$  とし,  $n/N \diamond X_{t-1} | X_{t-1} \sim HG(N, X_{t-1}, n)$  とする. ここで  $HG(N, m, n)$  は超幾何分布であり, その確率関数は

$$f(x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}} \quad (x = 0, 1, 2, \dots, \min(m, n))$$

である. このとき, 次のモデルを BINAR<sub>h</sub>(1) モデルという:

$$X_t = n/N \diamond X_{t-1} + \varepsilon_t \quad (t \geq 1).$$

このモデルにおいて,  $n/N \diamond X_{t-1} \sim B(n, p)$  となることが示される(付録 A.4 参照). したがって, 二項分布の再生性より  $\{\varepsilon_t\} \sim B(N - n, p)$  となることが分かる.

また, このモデルは PINAR( $p$ ) モデルと同様, 二項分布の再生性より  $p$  次への拡張が可能であるが, Weiß(2008, 2009)では提案されていない.

**4.  $\mathbb{Z}$  上の整数値自己回帰モデル**

これまででは,  $X_t$  が非負整数値をとる場合のモデルについて言及してきた. 本節では, 負の整数値も含む整数全体  $\mathbb{Z}$  に拡張したモデルとして, 周辺分布をスケラム分布とするモデルについて述べる.

**4.1 SINAR(1) モデル**

ここでは, Freeland (2010)で提案されている 1 次の SINAR モデル(Skellam INAR モデル)について述べる.

まずはスケラム分布について確認しておく.  $X, Y$  が独立でそれぞれパラメータ  $\lambda_1, \lambda_2$  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  $\lambda_1, \lambda_2$  のスケラム分布という. ここでは, パラメータが  $\lambda_1, \lambda_2$  のスケラム分布を  $S(\lambda_1, \lambda_2)$  と表記する. スケ

ラム分布  $S(\lambda_1, \lambda_2)$  の確率関数は

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

である。スケラム分布の性質については、例えば Alzaid and Omair (2010) を参照されたい。

スケラム分布を用いて、SINAR(1) モデルは次のように定義される。  $X_t, Y_t$  をそれぞれ独立でパラメータの等しい PINAR(1) モデルとする。つまり、

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

とする。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

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

である。このモデルを SINAR(1) モデルという。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となっていることがわかる。

$Z_t$  が正負の値をとりうるとなると、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、先ほどと同様に

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

とし、

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

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。

#### 4.2 SINAR(1) モデルの拡張

さて、Freeland (2010) では  $X_t, Y_t$  の2つの PINAR(1) モデルのパラメータが等しい場合のみを論じている。ここでは、  $\lambda$  が異なるモデルへの拡張を提案する。  $X_t, Y_t$  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

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

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

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

を考える。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 4.1 節で述べた 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。以降では、この拡張したモデルのことを改めて SINAR(1) モデルと呼ぶことにする。



この(拡張した)SINAR モデルの性質を述べる.  $\varepsilon_t$  のモーメント母関数, 期待値, 分散はそれぞれ

$$(4.1) \quad \begin{aligned} M_{\varepsilon_t}(s) &= \exp \left\{ -((1-\alpha)\lambda_1 + (1-\alpha)\lambda_2) + (1-\alpha)\lambda_1 e^s - (1-\alpha)\lambda_2 e^{-s} \right\} \\ E(\varepsilon_t) &= (1-\alpha)(\lambda_1 - \lambda_2) \\ V(\varepsilon_t) &= (1-\alpha)(\lambda_1 + \lambda_2) \end{aligned}$$

であり(付録 A.5 を参照), このモーメント母関数より  $\varepsilon_t \sim S((1-\alpha)\lambda_1, (1-\alpha)\lambda_2)$  であることがわかる. そして,

$$M_{\alpha * Z_{t-1}}(s) = \exp \left\{ -(\alpha\lambda_1 + \alpha\lambda_2) + \alpha\lambda_1 e^s + \alpha\lambda_2 e^{-s} \right\}$$

であり,  $\alpha * Z_{t-1} \sim S(\alpha\lambda_1, \alpha\lambda_2)$  であることもわかる. 条件付き期待値, 条件付きモーメント母関数はそれぞれ

$$(4.2) \quad \begin{aligned} E(Z_t | Z_{t-1} = z) &= \alpha z + (1-\alpha)(\lambda_1 - \lambda_2) \\ E(\exp \{s Z_t\} | Z_{t-1} = z) &= \frac{M_{\varepsilon_t}(s)}{P(Z_{t-1} = z)} \sum_{y=0}^{\infty} \frac{(\lambda_1(1-\alpha + \alpha e^s))^{z+y}}{(z+y)!} \frac{(\lambda_2(1-\alpha + \alpha e^{-s}))^y}{y!} \end{aligned}$$

である(付録 A.6 を参照).

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

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

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  $\lambda_1, \lambda_2$  の不偏推定量である.

### 5. 実例

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を, 拡張した SINAR(1) モデルにあてはめた例を紹介する. データとしては, 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を用いる. データは J リーグウェブサイトより取得した. 図 1 は全 34 試合(節)における得失点差の分布である. 標本平均は 0.47, 標本分散は 1.5 であった. また, 図 2 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, 比較的上昇と下降を交互に繰り返している傾向があり, 負の相関がありそうに見える. この時系列データのコレログラムは図 3 のようになり, 1 次の自己相関係数は  $-0.16$  と, 想定されたように負の値である. すなわち,  $-0.16$  と値こそ大きくはないが相関があり, 各試合での得失点差は独立ではないといえる.

そこで, SINAR(1) モデル

$$Z_t = \alpha * Z_{t-1} + \varepsilon_t$$

によりモデル化することを考える. 1 次の自己相関が負であることから,

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

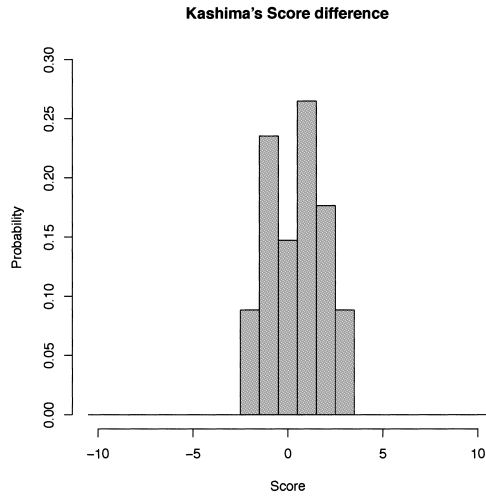


図 1. 2015 年鹿島アントラーズの得失点差のヒストグラム.

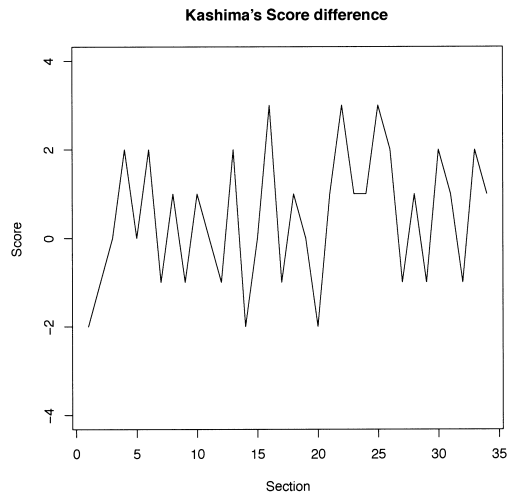


図 2. 2015 年鹿島アントラーズの得失点差のグラフ.

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

のように偶数と奇数で得点数  $X_t$  と失点数  $Y_t$  の差の順序を入れ替える, すなわち得失点差  $Z_t$  の正負を交互に入れ替える. パラメータをモーメント法で推定すると

$$\begin{cases} \hat{\alpha} & = -\hat{\rho}(1) = 0.16 \\ \hat{\lambda}_1 & = 1.33 \\ \hat{\lambda}_2 & = 0.86 \end{cases}$$

が得られる. しかし, この結果自体の解釈は難しい.

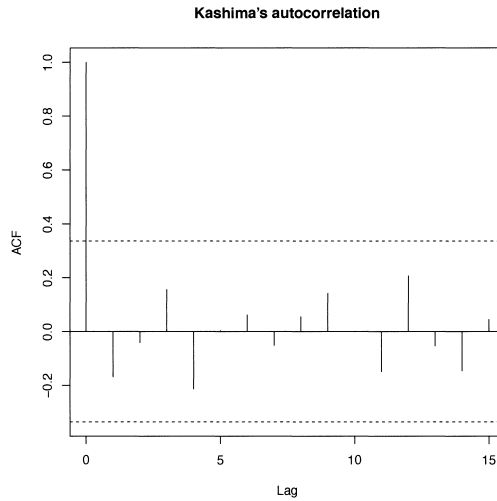


図 3. 2015 年鹿島アントラーズの得失点差のコレログラム.

1 次の自己相関が小さく、このモデルを用いての予測の精度はあまり高くはないが、1 期先予測などを行うことは可能である。例えば、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してみよう。例えば  $Z_{35} = 0$  となる確率は、 $Z_{34} = 1$  である条件の下で、

$$\begin{aligned}
 P(Z_{35} = 0|Z_{34} = 1) &= \dots + P(\hat{\alpha} * Z_{34} = -3|Z_{34} = 1)P(\varepsilon_{35} = 3) \\
 &\quad + P(\hat{\alpha} * Z_{34} = -2|Z_{34} = 1)P(\varepsilon_{35} = 2) \\
 &\quad + \dots \\
 &\quad + P(\hat{\alpha} * Z_{34} = 2|Z_{34} = 1)P(\varepsilon_{35} = -2) \\
 &\quad + P(\hat{\alpha} * Z_{34} = 3|Z_{34} = 1)P(\varepsilon_{35} = -3) + \dots
 \end{aligned}$$

のように計算することができる。ここで、 $\hat{\alpha} = -0.16 < 0$  であり、 $t = 35$  が奇数であることから、

$$\begin{aligned}
 P(\alpha * Z_{34} = v|Z_{34} = z_{34}) &= \frac{P(\alpha \circ X_{34} - \alpha \circ Y_{34} = v, Z_{34} = z_{34})}{P(Z_{34} = z_{34})} \\
 &= \frac{1}{P(Z_{34} = z_{34})} \sum_x \sum_w P(\alpha \circ X_{34} = w|X_{34} = x)P(Y_{34} = x - z_{34}) \\
 &\quad \cdot P(X_{34} = x)P(\alpha \circ Y_{34} = w + v|Y_{34} = x - z_{34})
 \end{aligned}$$

により求まる。この結果をまとめたものが表 2 である。表 2 には、得失点差が独立なスケラム分布  $S(1.33, 0.86)$  に従うとした場合の確率も併記している。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、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。実際、得失点差の推移をみると、得失点差が 1 であった場合の次の節においては、ほとんどのケースで得失点差が負(とくに -1)になっており、SINAR(1) モデルではこの結果を反映していると解釈することができる。

表 2. SINAR モデルによる推定.

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
確率 (DGLM)	0.021	0.065	0.151	0.246	0.238	0.157	0.076

ところで、同じデータに対して DGLM で分析することは可能だろうか。少なくとも、得失点差系列自体に DGLM をあてはめるために、ポアソン過程の生起度(intensity)として単一の系列をどうモデル化して与えるかは自明ではないように思われる。この点に鑑みれば、正負双方を取り得る整数値時系列データに対しては、DGLM のようなパラメータ駆動型のモデルよりは、INAR モデルのような観測値駆動型のモデルのほうが構築しやすさからメリットがあると言えよう。

ここでは、得点系列と失点系列にそれぞれ別々の非定常ポアソンモデルをあてはめた結果とその含意を述べる。まずポアソンパラメータは時変( $\lambda_t$ )とし、非負制約を課して  $\log \lambda_t$  に対して  $\log \lambda_t = \log \lambda_{t-1} + v_t$  というランダムウォーク型のシステム方程式を仮定する。システムノイズには正規分布を仮定し、 $v_t \sim N(0, \tau^2)$  とする。観測時系列  $Y_t$  は、ポアソン分布  $P(Y_t | \lambda_t) = e^{-\lambda_t} (\lambda_t)^{Y_t} / Y_t!$  に従うものとする。ここでは状態変数の分布を短冊状に近似して Kitagawa (1987) の数値積分フィルタ・平滑化で、時変生起度  $\lambda_{t|T}$  ( $t = 1, \dots, T$ ) を求める。状態の初期分布は、各系列の標本平均を  $\bar{\lambda}$  と書くとする、 $N(\bar{\lambda}, \bar{\lambda}^2)$  で与えた。

システムノイズの分散( $\tau^2$ )は、得点系列に関しては  $1.12 \times 10^{-4}$  と、失点系列に関しては  $4.01 \times 10^{-3}$  と推定された。このとき最終の第 34 節でのフィルタ値はそれぞれ 1.657 と 1.067 と得られたので、ランダムウォーク予測で仮想的第 35 節もこの生起度で得失点が発生すると仮定する。10 万回のシミュレーションから得られた得失点差の相対頻度を、表 2 の最終行に示している。傾向としては、独立スケラム分布の結果に近いが、今回の DGLM では得失点差系列に見られる負の相関(高周波成分の強い平均回帰性)がモデルに取り込めていないことを考えれば自然である。2 本の対数生起度系列(潜在時系列)に関する多変量モデリングが現象説明能力の改善に資するかどうかは、今後の研究課題である。

独立であっても、得点と失点に DGLM をあてはめることの効用はある。時変生起度の平滑化分布を描画してみれば、これらの整数値時系列が定常という仮定が妥当であったかどうかのチェックを、少なくとも視覚的に行うことが可能である。図 4 が今回の推定結果である。ポアソンの生起度に非定常な振る舞いを許す枠組みで分析したわけだが、平滑化分布の中央値に多少の凹凸は見られるものの、得点、失点ともに定常ポアソン過程とみなして問題ないであろう。

## 6. おわりに

本研究では、整数値をとる時系列データに対するモデルとして近年研究が進んでいる PINAR モデル、NBD INAR モデル、BINAR モデル、SINAR モデルについて、その定義や性質、そしてパラメータ推定法を説明した。

さらに本論文では、SINAR(1) モデルに関する拡張を提案し、それに伴いいくつかの性質を示すことができた。これらの結果により、今までよりもさらに複雑な構造をもつ整数値時系列データに対するモデリングの可能性が広がったといえよう。しかしながら、拡張したモデルにおけるパラメータ構造の問題が残されており、これは今後の課題である。また、他のモデルに関する  $p$  次への拡張への可能性、そしてそのパラメータ推定についての研究も残されている。最後に、パラメータ推定法が確立されていないモデルでの推定法の提案も今後の研究課題といえる。

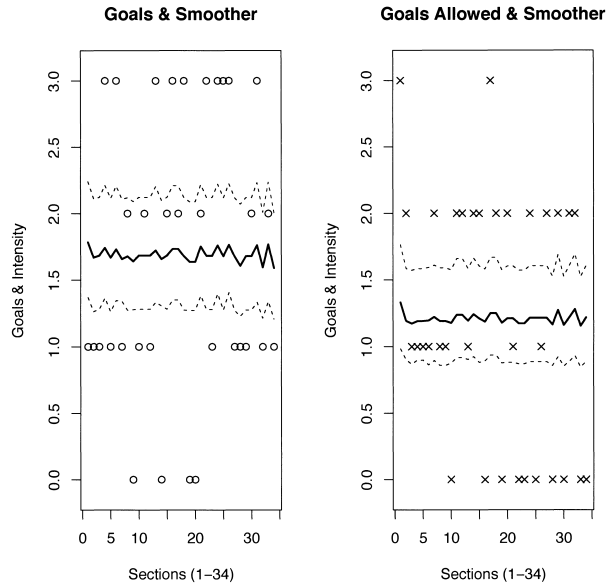


図 4. 得点(左パネル中 o)失点(右パネル中 x)の時系列と, DGLM で推定した時変生起度の平滑化系列. 平滑化系列は太実線が 50% 点, 細破線は下が 3% 点, 上が 97% 点. 横軸は試合の開催節.

謝 辞

改稿につながるコメントを寄せてくださった匿名の査読者にこの場を借りて感謝申し上げます。本研究を進めるにあたり、中嶋は統計数理研究所の特別共同利用研究員制度(平成 28 年度)を利用した。

付録. 証明

A.1 PINAR( $p$ ) モデルの誤差分布の導出

まず,  $\alpha_i \circ X_{t-i}$  は

$$M_{\alpha_i \circ X_{t-i}}(s) = \exp \{ \lambda \alpha_i (s - 1) \}$$

である。ここで,  $\sum_{i=1}^p \alpha_i \circ X_{t-i}$  の確率母関数は、ポアソン分布の再生性より

$$(A.1) \quad G_{\sum_{i=1}^p \alpha_i \circ X_{t-i}}(s) = \exp \left\{ \sum_{i=1}^p \alpha_i \lambda (s - 1) \right\}$$

となる。したがって,  $\varepsilon_t$  の確率母関数は,

$$G_{\varepsilon_t}(s) = \frac{G_{X_t}(s)}{G_{\sum_{i=1}^p \alpha_i \circ X_{t-i}}(s)} = \exp \left\{ \left( 1 - \sum_{i=1}^p \alpha_i \right) \lambda (s - 1) \right\}$$

となる。□

**A.2 (2.2), (2.3) の証明**

まず, 条件付き期待値は,

$$\begin{aligned} E(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-p}) &= E(\alpha_1 \circ X_{t-1} + \dots + \alpha_p \circ X_{t-p} + \varepsilon_t | X_{t-1}, X_{t-2}, \dots, X_{t-p}) \\ &= E(\alpha_1 \circ X_{t-1} | X_{t-1}) + \dots + E(\alpha_p \circ X_{t-p} | X_{t-p}) + E(\varepsilon_t) \\ &= \sum_{i=1}^p \alpha_i X_{t-i} + \left(1 - \sum_{i=1}^p \alpha_i\right) \lambda \end{aligned}$$

となる. 続いて条件付き分散は

$$\begin{aligned} V(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-p}) &= V(\alpha_1 \circ X_{t-1} + \dots + \alpha_p \circ X_{t-p} + \varepsilon_t | X_{t-1}, X_{t-2}, \dots, X_{t-p}) \\ &= V(\alpha_1 \circ X_{t-1} | X_{t-1}) + \dots + V(\alpha_p \circ X_{t-p} | X_{t-p}) + V(\varepsilon_t) \\ &= \sum_{i=1}^p \alpha_i (1 - \alpha_i) X_{t-i} + \left(1 - \sum_{i=1}^p \alpha_i\right) \lambda \end{aligned}$$

となる.  $\square$

**A.3 BINAR<sub>b</sub>(1) モデルにおける  $X_t$  の 1 次の自己相関  $\rho(1)$  の導出**

1 次の自己共分散は

$$\text{Cov}(X_t, X_{t-1}) = \text{Cov}(\alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})) = \rho V(X_{t-1})$$

となるので, 1 次の自己相関は  $\rho(1) = \rho$  となることが分かる.  $\square$

**A.4  $n/N \diamond X_{t-1} \sim B(n, p)$  の証明**

$X$  と  $n/N \diamond X$  の同時分布を求め, そのあと  $n/N \diamond X$  の周辺分布を求める. まず,  $X$  と  $n/N \diamond X$  の同時分布は

$$\begin{aligned} P(n/N \diamond X = k, X = x) &= P(n/N \diamond X = k | X = x) P(X = x) \\ &= \frac{\binom{x}{k} \binom{N-x}{n-k}}{\binom{N}{n}} \binom{N}{x} p^x (1-p)^{N-x} \\ &= \binom{n}{k} p^k (1-p)^{n-k} \binom{N-n}{x-k} p^{x-k} (1-p)^{N-n-(x-k)} \end{aligned}$$

である. ここで,  $x$  の範囲が  $k \leq x \leq N - n + k$  であることに注意を払うと,  $n/N \diamond X$  の確率関数は

$$\begin{aligned} P(n/N \diamond X = k) &= \sum_{x=k}^{N-n+k} \binom{n}{k} p^k (1-p)^{n-k} \binom{N-n}{x-k} p^{x-k} (1-p)^{N-n-(x-k)} \\ &= \binom{n}{k} p^k (1-p)^{n-k} \sum_{x'=0}^{N-n} \binom{N-n}{x'} p^{x'} (1-p)^{N-n-x'} \quad (x' = x - k \text{ とおく}) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

となるので,  $n/N \diamond X \sim B(n, p)$  となることが分かる.  $\square$

**A.5 (4.1)の証明**

まず,

$$(A.2) \quad \begin{aligned} M_{\alpha \star Z_{t-1}}(s) &= E(e^{s(\alpha \circ X_{t-1} - \alpha \circ Y_{t-1})}) \\ &= \exp \{ -(\alpha \lambda_1 + \alpha \lambda_2) + \alpha \lambda_1 e^s + \alpha \lambda_2 e^{-s} \} \end{aligned}$$

となる. ここで, (A.2)の結果を用いると,  $M_{\varepsilon_t}(s)$  は

$$\begin{aligned} M_{\varepsilon_t}(s) &= \frac{M_{Z_t}(s)}{M_{\alpha \star Z_{t-1}}(s)} \\ &= \exp \{ -((1-\alpha)\lambda_1 + (1-\alpha)\lambda_2) + (1-\alpha)\lambda_1 e^s - (1-\alpha)\lambda_2 e^{-s} \} \end{aligned}$$

となる. これは  $S((1-\alpha)\lambda_1, (1-\alpha)\lambda_2)$  のモーメント母関数であることから,  $\varepsilon_t \sim S((1-\alpha)\lambda_1, (1-\alpha)\lambda_2)$  である. このことから, 期待値, 分散はそれぞれ,

$$\begin{cases} E(\varepsilon_t) = (1-\alpha)(\lambda_1 - \lambda_2) \\ V(\varepsilon_t) = (1-\alpha)(\lambda_1 + \lambda_2) \end{cases}$$

であることが分かる. □

**A.6 (4.2)の証明**

まず, 条件付き期待値は

$$\begin{aligned} E(Z_t | Z_{t-1} = z) &= E(E(\alpha \star Z_{t-1} | Z_{t-1} = z)) + E(\varepsilon_t) \\ &= \alpha z + (1-\alpha)(\lambda_1 - \lambda_2) \end{aligned}$$

である. 一方, 条件付きモーメント母関数は

$$\begin{aligned} E(\exp \{s Z_t\} | Z_{t-1} = z) &= E(\exp \{s(\alpha \star Z_{t-1} + \varepsilon_t)\} | Z_{t-1} = z) \\ &= \frac{M_{\varepsilon_t}(s)}{P(Z_{t-1} = z)} \sum_{y=0}^{\infty} P(X_{t-1} = z+y, Y_{t-1} = y) M_{\alpha \circ X_{t-1}}(s) M_{\alpha \circ Y_{t-1}}(-s) \\ &= \frac{M_{\varepsilon_t}(s)}{P(Z_{t-1} = z)} \sum_{y=0}^{\infty} \frac{(\lambda_1(1-\alpha + \alpha e^s))^{z+y}}{(z+y)!} \frac{(\lambda_2(1-\alpha + \alpha e^{-s}))^y}{y!} \end{aligned}$$

となる. これ以上は解析的に解くことができない. □

参 考 文 献

Al-Osh, M. A. and Alzaid, A. A. (1990). An integer-valued  $p$ th-order autoregressive structure (INAR( $p$ )) process, *Journal of Applied Probability*, **27**(2), 314–324.

Alzaid, A. A. and Omair, M. A. (2010). On the poisson difference distribution inference and applications, *Bulletin of the Malaysian Mathematical Sciences Society*, **33**(1), 17–45.

Baretto-Souza, W. and Bourguignon, M. (2013). A skew true INAR(1) process with application, arXiv: 1306.0156 [stat.ME].

Du, J. G. and Li, Y. (1991). The integer-valued autoregressive (INAR( $p$ )) model, *Journal of Time Series Analysis*, **12**, 129–142.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear*

- Models*, 2nd ed., Springer-Verlag, New York.
- Fokianos, K. (2011). Count time series models, *Handbook of Statistics Vol.30* (eds. T. S. Rao, S. S. Rao and C. R. Rao), 315–347, North-Holland, Oxford.
- Freeland, R. K. (2010). True integer value time series, *ASTA Advances in Statistical Analysis*, **94**, 217–229.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion), *Journal of the American Statistical Association*, **76**, 1032–1064.
- Kitagawa, G. (2010). *Introduction to Time Series Modeling*, Chapman & Hall/CRC, Boca Raton.
- Leonenko, N., Savani, V. and Zhigljasky, A. (2007). Autoregressive negative binomial processes, *Annales de l'Institut de Statistique de l'Université de Paris*, **51**(1), 25–47.
- McKenzie, E. (1985). Some simple models for discrete variate time series, *Journal of the American Water Resources Association*, **21**(4), 645–650.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society Series A*, **135**, 370–384.
- Rajarshi, M. B. (2012). *Statistical Inference for Discrete Time Stochastic Processes*, Springer, New Delhi.
- Rinstić, M. M., Bakouch, H. S. and Nastić, A. S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process, *Journal of Statistical Planning and Inference*, **139**, 2218–2226.
- Savani, V. and Zhigljavsky, A. (2007a). Efficient parameter estimation for independent and INAR(1) negative binomial samples, *Metrika*, **65**(2), 207–225.
- Savani, V. and Zhigljavsky, A. (2007b). Efficient estimation of parameters of the negative binomial distribution, *Communications in Statistics — Theory and Methods*, **35**(5), 1–17.
- Weiß, C. H. (2008). Thinning operations for modeling time series of counts — A survey, *ASTA Advances in Statistical Analysis*, **92**(3), 319–341.
- Weiß, C. H. (2009). A new class of autoregressive models for time series of binomial counts, *Communications in Statistics — Theory and Methods*, **38**(4), 447–460.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed., Springer-Verlag, New York.



## Recent Development of Integer-valued Autoregressive Models

Masahiko Nakajima<sup>1</sup>, Fumitake Sakaori<sup>2</sup> and Yoshinori Kawasaki<sup>3,4</sup>

<sup>1</sup>Graduate School of Science and Engineering, Chuo University

<sup>2</sup>Faculty of Science and Engineering, Chuo University

<sup>3</sup>The Institute of Statistical Mathematics

<sup>4</sup>Department of Statistical Science, SOKENDAI

Integer-valued Autoregressive Models (INAR models) express current observations, which are integers that depend on past integer-value observations. INAR models completely differ from Dynamic Generalized Linear Models and General State-Space Models, which employ an unobservable/latent process to model integer-valued time series. In INAR models, the choice of the marginal and/or innovation distribution and the definition of the ‘autoregressive part’ are very important to ensure compatibility (in the distributional sense) of both sides of an autoregressive model. Although publications sporadically appeared from the mid 1980s to the early 1990s, notable contributions were not reported for more than a decade. However in the late 2000s, new results began appearing on a regular basis. In this paper, we introduce recent developments in INAR models beginning with Poisson INAR. For some theorems and propositions without proofs in existing literature, we compiled our notes in the appendix. Additionally, we include new results on an INAR model based on the difference of two Poisson distributions and use real data analysis as an illustration.