On Eigenvalues of the Transition Matrix of some Count-Data Markov Chains

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß

Department of Mathematics & Statistics,

Helmut Schmidt University, Hamburg

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Markov Chains and their Transition Matrices

Basic Facts

Markov Chains and Transition Matrix

Discrete-val. $(X_t)_{\mathbb{N}_0}$ is **Markov chain** iff "memory of length 1":

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}, \ldots) \;=\; P(X_t = x_t \mid X_{t-1} = x_{t-1}).$$

$(X_t)_{\mathbb{N}_0}$ even **homogeneous** MC iff

transition probabilities do not vary with time:

$$P(X_t = i \mid X_{t-1} = j) \;=\; p_{i|j} \qquad \text{for all } t \in \mathbb{N}.$$

Transition matrix $\mathbf{P} = (p_{i|j})_{i,j}$

of either finite or countably infinite dimension.

In this talk: count-data MC, i. e., range $\subseteq \mathbb{N}_0 = \{0, 1, \ldots\}$.

Important properties of MC can be recognized from **eigenvalues** of transition matrix $\mathbf{P}$.

Example 1: Pearson's GoF-Test.

Let $X_1, \ldots, X_T$ stem from reversible finite MC with range $\{0, \ldots, n\}$, let N_i denote number of X_t equal to i.

Tavaré & Altham (1983): Pearson's χ^2-statistic satisfies

$$\mathsf{X}^2 = \sum_{i=0}^{n} \frac{(N_i - Tp_i)^2}{Tp_i} \overset{D}{\to} \sum_{k=1}^{m} \frac{1+\lambda_k}{1-\lambda_k} \cdot Z_k^2 \qquad \text{for } T \to \infty,$$

where λ_k non-unit eigenvalues of $\mathbf{P}$, and Z_k i. i. d. $N(0,1)$.

Example 2: Forecasting.

Primitive finite MC is ergodic: $\mathbf{P}^h \;\to\; \boldsymbol{\pi}\,\mathbf{1}^\top$ for $h \to \infty$.

But how quick does $\mathbf{P}^h$ converge to $\boldsymbol{\pi}\,\mathbf{1}^\top$?

Denote distinct eigenvalues of $\mathbf{P}$ by $1 > |\lambda_2| \geq \ldots \geq |\lambda_r|$, where λ_2 "second largest e.v." with maximal multiplicity (say m_2).

Perron-Frobenius theorem implies (Seneta, 1983)

$$\mathbf{P}^h \;=\; \boldsymbol{\pi}\mathbf{1}^\top \;+\; O\big(h^{m_2-1}\cdot|\lambda_2|^h\big).$$

$\Rightarrow$ identify **forecasting horizon** with non-trivial forecasts (= differing from those w. r. t. stationary marginal distribution).

Derive relations for higher conditional (factorial) moments, where leading coefficient turns out to be required eigenvalue.

More precisely, for common types of count-data MCs, we have k^{th}-order polynomial structure like

$$E[X_t^k \mid X_{t-1}] = \gamma_{k,k}\, X_{t-1}^k + \ldots + \gamma_{k,1}\, X_{t-1} + \gamma_{k,0} \quad \text{with } \gamma_{k,k} \neq 0.$$

We use these to find coefficients $a_0^{(k)}, \ldots, a_k^{(k)}$ such that

$$E\big[\textstyle\sum_{r=0}^k a_r^{(k)}\, X_t^r \mid X_{t-1}\big] = \gamma_{k,k} \textstyle\sum_{j=0}^k a_j^{(k)}\, X_{t-1}^j.$$

Then we conclude that $\gamma_{k,k}$ is an eigenvalue of $\mathbf{P}$.

Eigenvalues of Count Data Markov Chains

CLAR(1) Processes

CLAR(1) Count Data Markov Chains

Grunwald et al. (2000):

Homogeneous MC $(X_t)_{\mathbb{N}_0}$ has **CLAR(1) structure** if

$$E[X_t \mid X_{t-1}] \;=\; \alpha_1 \cdot X_{t-1} + \alpha_0 \qquad \text{for some } \alpha_1, \alpha_0 \in \mathbb{R}.$$

Count data case: $\alpha_0 > 0$ to ensure that $X_t \geq 0$.

$|\alpha_1| < 1$ guarantees finite mean $E[X_t] = \alpha_0/(1-\alpha_1)$.

If $|\alpha_1| < 1$ and $V[X_t] < \infty$, then ACF of AR(1)-type: $\rho(k) = \alpha_1^k$.

So linear coefficient α_1 of conditional mean equals $\rho(1)$.

Many famous instances: INAR(1), INARCH(1), (beta-)binomial AR(1), binomial INARCH(1), . . .

Theorem:

Let $(X_t)_{\mathbb{N}_0}$ stationary CLAR(1) count data MC with $|\alpha_1| < 1$.
Then α_1 is an eigenvalue of transition matrix **P**.

Theorem implies that second largest eigenvalue of **P** has modulus not smaller than $|\alpha_1|$, i. e., α_1 **lower bound** for second largest eigenvalue.

Proof of presents main idea for subsequent proofs:

Proof: $E[X_t \mid X_{t-1}] \;=\; \alpha_1 \cdot X_{t-1} + \alpha_0$
is polynomial in X_{t-1} with leading coefficient α_1.

Proof: (cont.) We find coefficients $a_0^{(1)}, a_1^{(1)}$ such that

$$E\big[a_1^{(1)}\,X_t + a_0^{(1)} \mid X_{t-1}\big] \overset{!}{=} \alpha_1\,\big(a_1^{(1)}\,X_{t-1} + a_0^{(1)}\big),$$

namely $a_1^{(1)} := 1$ and $a_0^{(1)} = -\alpha_0/(1-\alpha_1)$. It follows that

$$\big(a_1^{(1)}\cdot 0 + a_0^{(1)},\ a_1^{(1)}\cdot 1 + a_0^{(1)},\ \ldots,\ a_1^{(1)}\cdot i + a_0^{(1)},\ \ldots\big)\,\big(p_{i|j}\big)_{i,j=0,1,\ldots}$$

$$= \big(\ldots, \textstyle\sum_{i=0}^{\infty} (a_1^{(1)}\cdot i + a_0^{(1)})\, p_{i|j}, \ldots\big)$$

$$= \big(\ldots, E\big[a_1^{(1)}\,X_t + a_0^{(1)} \mid X_{t-1} = j\big], \ldots\big)$$

$$= \big(\ldots, \alpha_1\,(a_1^{(1)}\cdot j + a_0^{(1)}), \ldots\big) \quad = \alpha_1\,\big(\ldots, a_1^{(1)}\cdot j + a_0^{(1)}, \ldots\big),$$

i. e., $\big(\ldots, a_1^{(1)}\,i + a_0^{(1)}, \ldots\big)$ left eigenvector with e.v. α_1. #

Some basics for remaining results . . .

Falling factorials $x_{(r)} = x \cdots (x-r+1),\ x_{(0)} := 1.$

Factorial moments $\mu_{(r)} := E\big[(X_t)_{(r)}\big]$ related to **raw moments** $\mu_r := E[X_t^r]$ via

$$\mu_{(n)} = \textstyle\sum_{j=1}^n s_{n,j}^{(1)} \cdot \mu_j, \qquad \mu_n = \textstyle\sum_{j=1}^n s_{n,j}^{(2)} \cdot \mu_{(j)},$$

where $s_{n,j}^{(1)}, s_{n,j}^{(2)}$ **Stirling numbers** of first/second kind.

Both falling factorials $x_{(r)}$ and powers x^r are **polynomial sequences of binomial type**:

$$(a+b)^n = \sum_{j=0}^n \binom{n}{j} a^j\, b^{n-j}, \qquad (a+b)_{(n)} = \sum_{j=0}^n \binom{n}{j} a_{(j)}\, b_{(n-j)}.$$

Lemma:

Let $(X_t)_{\mathbb{N}_0}$ homogeneous count-data MC such that k^{th} conditional raw moments, for $k = 1, \ldots, n$ with $n \in \mathbb{N}$, have following k^{th}-order polynomial structure:

$$E[X_t^k \mid X_{t-1}] \;=\; \gamma_{k,k}\, X_{t-1}^k + \ldots + \gamma_{k,1}\, X_{t-1} + \gamma_{k,0} \quad \text{with } \gamma_{k,k} \neq 0.$$

We set $\gamma_{0,0} = 1$, and assume $\gamma_{k,k} \neq \gamma_{j,j}$ for $k \neq j$.

Then leading coefficients $\gamma_{0,0}, \gamma_{1,1}, \ldots, \gamma_{n,n}$ are eigenvalues of **P**.

Lemma holds in same way

if considering **falling factorials** instead of powers.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Eigenvalues of Count Data Markov Chains

Specific CLAR(1) Processes

CLAR(1) member of INARMA family:

INAR(1) model by McKenzie (1985),

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \epsilon_t.$$

Provided that innovations' mean μ_ϵ exists,

conditional mean linear in previous observation: $\mu_\epsilon + \alpha\, X_{t-1}$.

Theorem: Stationary INAR(1) process $(X_t)_{\mathbb{N}_0}$

with existing factorial moments $\mu_{(r),\epsilon}$ for $(\epsilon_t)_{\mathbb{N}}$.

Then $1, \alpha, \alpha^2, \alpha^3, \ldots$ are eigenvalues of transition matrix $\mathbf{P}$.

Specific CLAR(1) Count Data Markov Chains

CLAR(1) member of Poisson INGARCH family:

Poisson INARCH(1) model (Weiß, 2010),

$$X_t \mid X_{t-1}, \ldots \quad \sim \; \text{Poi}(\beta + \alpha\, X_{t-1}).$$

Conditional mean linear in previous observation: $\beta + \alpha\, X_{t-1}$.

Theorem: Stationary Poisson INARCH(1) process $(X_t)_{\mathbb{N}_0}$.

Then $1, \alpha, \alpha^2, \alpha^3, \ldots$ are eigenvalues of transition matrix $\mathbf{P}$.

Specific CLAR(1) Count Data Markov Chains

From now on **finite** case,

where Markov counts X_t have range $\{0, \ldots, n\}$.

Binomial AR(1) model by McKenzie (1985),

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

Conditional mean linear in previous observation: $\rho \cdot X_{t-1} + n\beta$,

where $\rho = \alpha - \beta$. So again CLAR(1)-type model.

Theorem: Stationary binomial AR(1) process $(X_t)_{\mathbb{N}_0}$.

Then $1, \rho, \rho^2, \ldots, \rho^n$ are the eigenvalues of transition matrix **P**.

Specific CLAR(1) Count Data Markov Chains

According to the previous Theorems:

- general CLAR(1): eigenvalue α_1;   but

- special case INAR(1): eigenvalues $1, \alpha, \alpha^2, \alpha^3, \ldots$
- special case INARCH(1): eigenvalues $1, \alpha, \alpha^2, \alpha^3, \ldots$
- special case binomial AR(1): eigenvalues $1, \rho, \rho^2, \ldots, \rho^n$

So one might conjecture that CLAR(1)-type count data model

with linear coefficient α_1 $(= \rho(1))$

always has eigenvalues of form $1, \alpha_1, \alpha_1^2, \ldots$

This conjecture, however, would not be true.

Binomial INARCH(1) model by Weiß & Pollett (2014),

$$X_t \mid X_{t-1}, X_{t-2}, \ldots \quad \sim \ \mathrm{Bin}\Big(n,\ \beta + \alpha\,\frac{X_{t-1}}{n}\Big).$$

Conditional mean linear in previous observation: $\alpha \cdot X_{t-1} + n\beta$.

So again CLAR(1)-type model.

Theorem: Stationary binomial INARCH(1) process $(X_t)_{\mathbb{N}_0}$.

Then the eigenvalues of transition matrix $\mathbf{P}$

are $\frac{n_{(k)}}{n^k}\,\alpha^k$ for $k = 0, \ldots, n$.

Eigenvalues decay even more quickly than in previous examples,

with (unique) second largest eigenvalue α.

Although we realized that CLAR(1)'s eigenvalues not necessarily of form $1, \alpha_1, \alpha_1^2, \ldots$, common feature of previous examples that second largest eigenvalue uniquely α_1.

But also this property not true for any CLAR(1)-type model.

Beta-binomial thinning: (Weiß & Kim, 2014)
Let thinning parameter α_ϕ follow $\text{BETA}\big(\frac{1-\phi}{\phi}\,\alpha,\ \frac{1-\phi}{\phi}\,(1-\alpha)\big)$, then conditional distribution of $\alpha_\phi \circ X$ given X is beta-binomial.

Beta-binomial AR(1) model by Weiß & Kim (2014),

$$X_t \;=\; \alpha_\phi \circ X_{t-1} \;+\; \beta_\phi \circ (n - X_{t-1}).$$

Conditional mean linear in previous observation: $\rho \cdot X_{t-1} + n\beta$,

where $\rho = \alpha - \beta$. So again CLAR(1)-type model.

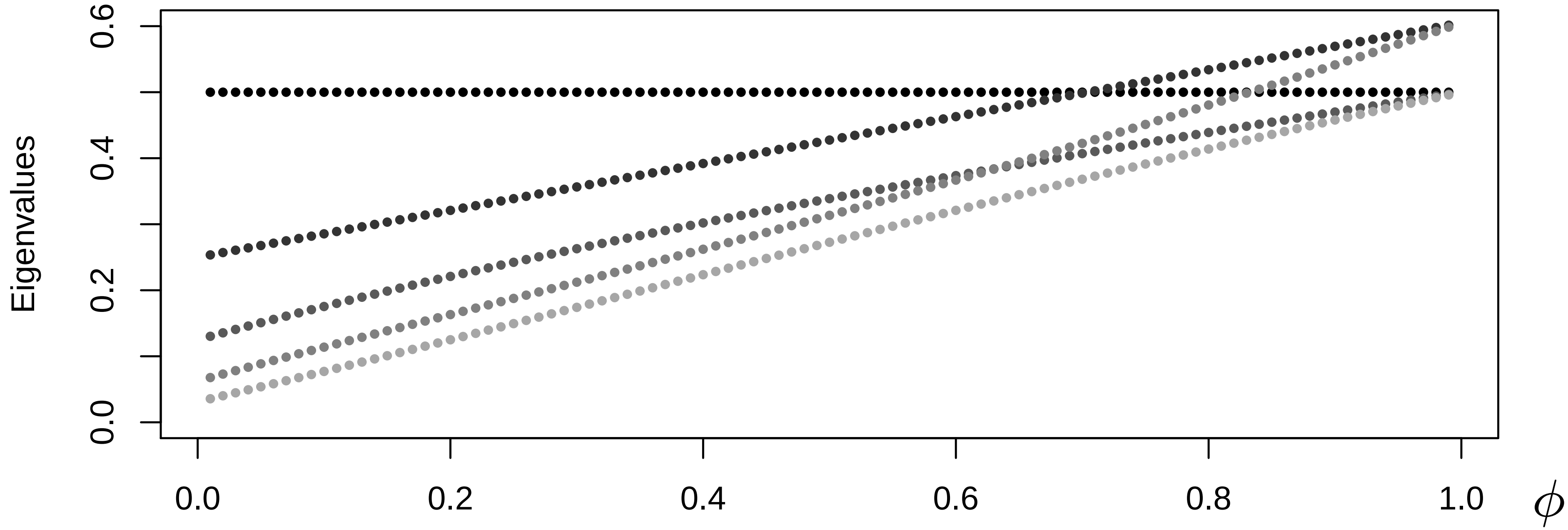
Theorem: Stationary beta-binomial AR(1) process $(X_t)_{\mathbb{N}_0}$.

Then eigenvalues of transition matrix $\mathbf{P}$ are $(k = 0, \ldots, n)$

$$\sum_{r=0}^{k} (-1)^{k-r} \binom{k}{r} \frac{(\frac{1-\phi}{\phi}\alpha + r - 1)_{(r)}}{(\frac{1-\phi}{\phi} + r - 1)_{(r)}} \frac{(\frac{1-\phi}{\phi}\beta + k - r - 1)_{(k-r)}}{(\frac{1-\phi}{\phi} + k - r - 1)_{(k-r)}}.$$

Note that values of eigenvalues do not depend on n.

Beta-binomial AR(1)'s eigenvalues for $\pi = 0.3$, $\rho = 0.5$, graphs for $k = 1$ (black) and $k = 2, \ldots, 5$ (dark gray to light gray):



$\Rightarrow$ Eigenvalue ρ not always second largest one,

and second largest eigenvalue may have multiplicity > 1.

Specific CLAR(1) Count Data Markov Chains

In a nutshell:

Linear coefficient of conditional mean of CLAR(1) model
(which equals $\rho(1)$ if variance exists)

always an eigenvalue of transition matrix,

but not necessarily second largest one
(only lower bound for second largest eigenvalue),
and its multiplicity does not need to be 1.

Perhaps $\rho(1)$ also eigenvalue for non-CLAR(1) models?

Binomial AR(1) model with density-dependent colonization by Weiß & Pollett (2014),

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \beta_t \circ (n - X_{t-1}),$$

with α, a, b satisfying $\alpha, a, a+b \in (0;1)$ and $\beta_t := \alpha\,(a+b\,X_{t-1}/n)$.

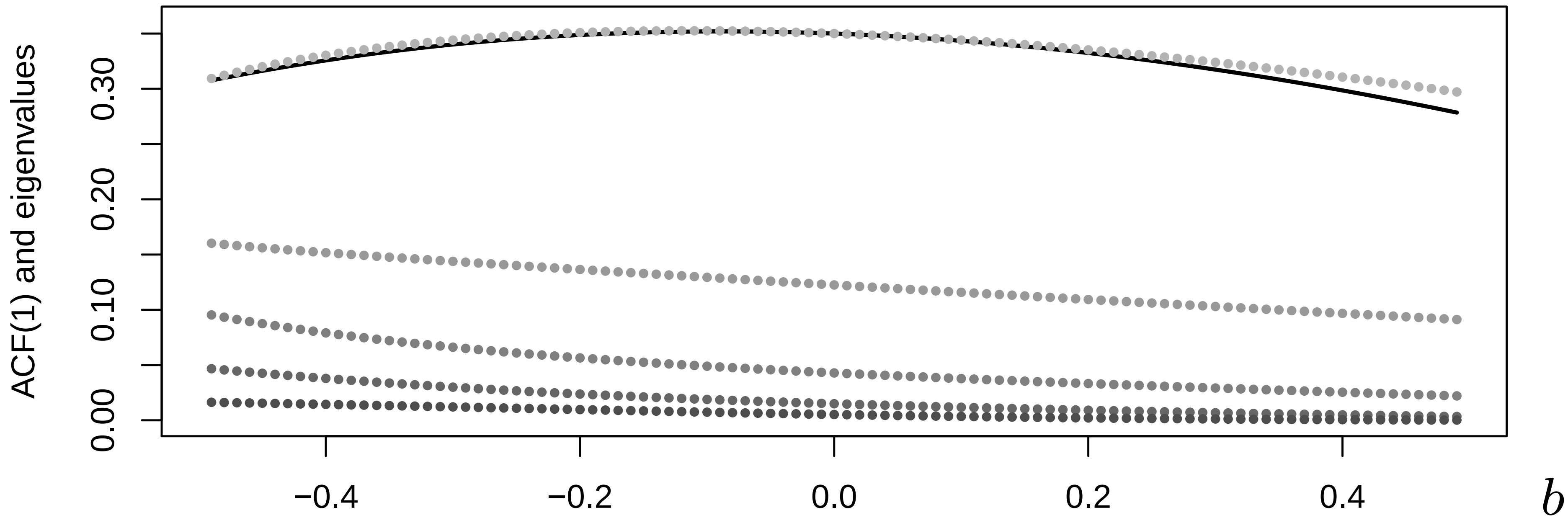
Conditional mean quadratic function in X_{t-1}:

$$E[X_t \mid X_{t-1}] \;=\; -\tfrac{1}{n}\alpha\, b \cdot X_{t-1}^2 \;+\; \alpha\,(1-a+b)\cdot X_{t-1} \;+\; n\,\alpha\, a.$$

So no CLAR(1)-type model.

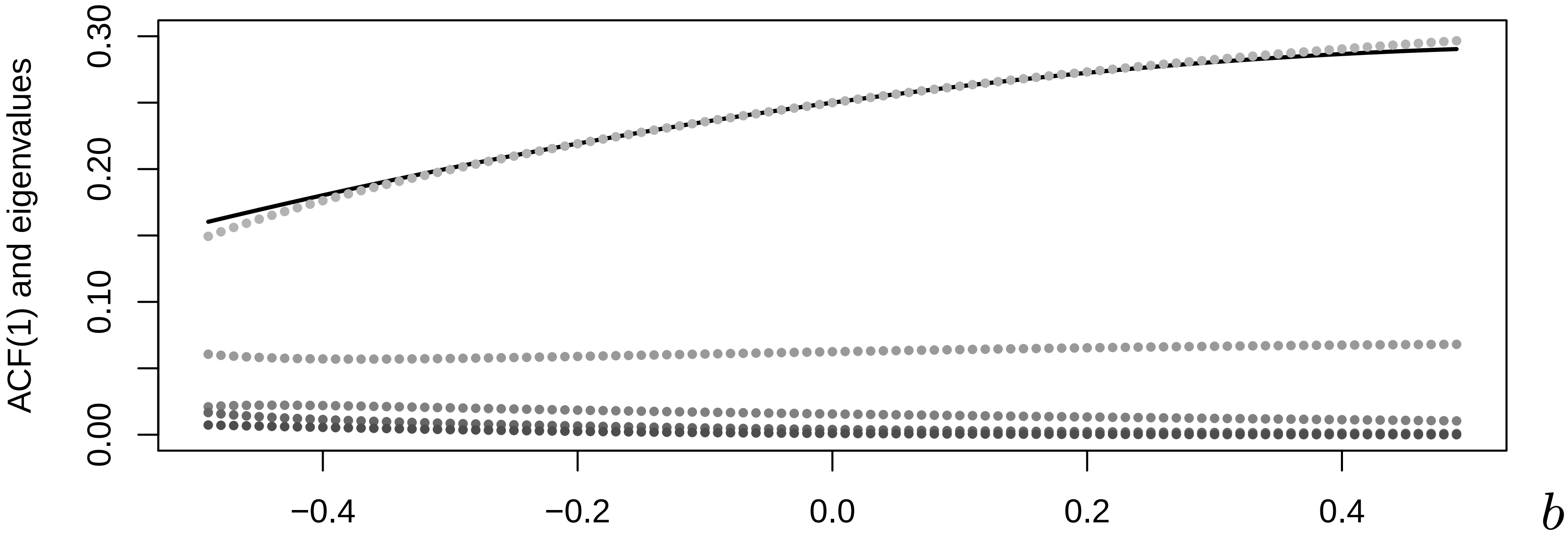
Specific CLAR(1) Count Data Markov Chains

DDC-binomial AR(1)'s eigenvalues for $n = 5$, $\alpha = 0.7$, $a = 0.5$ (gray dots), corresponding value of $\rho(1)$ as black line:



Second largest eigenvalue (lightest gray) close to $\rho(1)$, but usually different.

DDC-binomial AR(1)'s eigenvalues for $n = 5$, $\alpha = 0.5$, $a = 0.5$ (gray dots), corresponding value of $\rho(1)$ as black line:



So second largest eigenvalue may also be smaller than $\rho(1)$.

$\Rightarrow$ For non-CLAR(1), $\rho(1)$ limited insight into eigenstructure.

Conclusions

- General approach based on conditional moments to derive eigenstructure of count-data MCs. Especially second-largest eigenvalue relevant for applications.
- For CLAR(1) models, linear coefficient of conditional mean ($= \rho(1)$) always eigenvalue and hence lower bound for second-largest eigenvalue.

  Although $\rho(1)$ often equals second-largest eigenvalue, of multiplicity one, and remaining eigenvalues are powers of it, all these rules *do not generally hold*.
- For non-CLAR(1) models, $\rho(1)$ generally not an eigenvalue.

Thank You

for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

weissc@hsu-hh.de

Literature

Weiß (2017): *On eigenvalues of the transition matrix ...*
Methodology and Computing in Applied Probability 19, 997–1007.

Grunwald et al. (2000): *Non-Gauss. ... AR(1) models*. Austral. N. Z. J. Stat. 42, 479–495.

McKenzie (1985): *S. s. models for discr. v. time series*. Water Resources Bull. 21, 645–650.

Seneta (1983): *Non-negative matrices and Markov chains*. Springer Verlag.

Tavaré & Altham (1983): *Serial dependence of observations ...* Biometrika 70, 139–144.

Weiß (2010): *The INARCH(1) model for ...* Comm. Stat. Simul. Comp. 39, 1269–1291.

Weiß & Kim (2014): *Diagnosing ... extra-binom. ...* Appl. Stoch. Mod. ... 30, 588–608.

Weiß & Pollett (2014): *Binom. autoregressive processes ...* J. T. Ser. Anal. 35, 115–132.