Estimation and bootstrap for stochastically monotone Markov processes

Michael H. Neumann[1] ⬤

Abstract

The Markov property is shared by several popular models for time series such as autoregressive or integer-valued autoregressive processes as well as integer-valued ARCH processes. A natural assumption which is fulfilled by corresponding parametric versions of these models is that the random variable at time t gets stochastically greater conditioned on the past, as the value of the random variable at time $t-1$ increases. Then the associated family of conditional distribution functions has a certain monotonicity property which allows us to employ a nonparametric antitonic estimator. This estimator does not involve any tuning parameter which controls the degree of smoothing and is therefore easy to apply. Nevertheless, it is shown that it attains a rate of convergence which is known to be optimal in similar cases. This estimator forms the basis for a new method of bootstrapping Markov chains which inherits the properties of simplicity and consistency from the underlying estimator of the conditional distribution function.

1 Introduction

We consider a time-homogeneous Markov chain $(X_t)_{t \in \mathbb{N}_0}$ driven by a transition kernel which satisfies a certain monotonicity property: the conditional distribution of the random variable at time t gets stochastically greater as the value of the variable at time $t - 1$ increases. Such a condition is actually satisfied by several popular models for time series such as autoregressive or integer-valued autoregressive as well as

✉ Michael H. Neumann
michael.neumann@uni-jena.de

1 Institut für Mathematik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

integer-valued ARCH processes under natural assumptions on the involved parameters. To be specific, we assume that, for each fixed z, $F_x(z) := P(X_t \leq z \mid X_{t-1} = x)$ is antitonic (monotonically non-increasing) in x. This assumption allows us to employ a nonparametric antitonic estimator $\widehat{F}_x(z)$ of the function $x \mapsto F_x(z)$. Our estimator does not involve any tuning parameter which controls the degree of smoothing and is therefore easy to apply. Moreover, its consistency does not require smoothness properties of the function $x \mapsto F_x(z)$; the postulated monotonicity suffices. Theorem 2.1 states that the estimator $\widehat{F}_x(z)$ converges in L^1 norm, weighted by the stationary distribution of the Markov chain, with a rate of $n^{-1/3}$ which is believed to be the optimal one.

The estimator of $F_x(z)$ serves as a basis for a new bootstrap method for Markov chains. Among several other methods, those proposed by Rajarshi (1990) and Paparoditis and Politis (2002) are the closest ones to our proposal. While Rajarshi's bootstrap procedure is based on a nonparametric estimate of the one-step transition *density*, Paparoditis and Politis (2002) used in their so-called local bootstrap a *local resampling* of the original data set. In both papers, the proof of consistency of the respective bootstrap method is based on the assumption of a smooth transition density. In contrast, our approach does not require any smoothness assumption on the transition mechanism; it is merely based on the monotonicity assumption on the Markov kernel. We show its applicability for Markov chains with state space $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Consistency of bootstrap can be shown in a most transparent way by a so-called coupling of the original process and its bootstrap counterpart, i.e. we define versions $(\widetilde{X}_t)_{t \in \mathbb{N}_0}$ and $(\widetilde{X}_t^*)_{t \in \mathbb{N}_0}$ of these processes on a common probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ such that the corresponding random variables $\widetilde{X}_t$ and $\widetilde{X}_t^*$ are equal with a high probability. Somewhat surprisingly, this natural approach was rarely used in statistics. Using Mallows metric to measure the distance between variables from the original and the bootstrap process, it was implicitly employed in the context of independent random variables by Bickel and Freedman (1981) and Freedman (1981). A more explicit use of coupling was made, in the context of U- and V-statistics, but again in the independent case, by Dehling and Mikosch (1994) and Leucht and Neumann (2009). For dependent data, this approach was adopted by Leucht and Neumann (2013), Leucht et al. (2015), and Neumann (2021). Our second main result, Theorem 3.1, describes the results of our coupling approach. The stationary distribution $P_{X^*}^*$ of the bootstrap process converges in total variation norm and in probability to that of the original process. The coupled process is ϕ-mixing with coefficients decaying at an exponential rate and the corresponding values $\widetilde{X}_t^0$ and $\widetilde{X}_t^{*,0}$ of a stationary version of the coupled process coincide with a probability converging to 1. These general results can then be used to prove bootstrap consistency for specific statistics. The proofs of our main theorems and some auxiliary results a postponed to a final Sect. 4.

2 An estimator of a monotone family of distribution functions

Suppose that we observe random variables $X_0, X_1, \ldots, X_n$, where $\mathbf{X} = (X_t)_{t \in \mathbb{N}_0}$ is a strictly stationary Markov chain with state space $D \subseteq \mathbb{R}$, defined on a probability

space $(\Omega, \mathcal{A}, P)$. We denote the stationary distribution by P_X and the corresponding distribution function by F_X. Let $(F_x)_{x \in \mathbb{R}}$ defined as $F_x(z) = P(X_t \leq z \mid X_{t-1} = x)$ be the corresponding family of conditional distribution functions. We impose the following as our key assumption.

(A1) For each $z \in \mathbb{R}$, the function $x \mapsto F_x(z)$ is monotonically non-increasing, i.e. if $x_1 < x_2$, then $P(X_t \leq z \mid X_{t-1} = x_1) \geq P(X_t \leq z \mid X_{t-1} = x_2)$.
In addition we suppose that

(A2) $\mathbf{X} = (X_t)_{t \in \mathbb{N}_0}$ is strong mixing with exponentially decaying coefficients $\alpha_X(k)$, i.e.

$$\alpha_X(k) = O(\rho^k),$$

for some $\rho \in [0, 1)$.

Assumption (A1) may be paraphrased as follows. If $x_1 < x_2$ and if Y_1 and Y_2 are random variables following the respective conditional distributions $P^{X_t \mid X_{t-1} = x_1}$ and $P^{X_t \mid X_{t-1} = x_2}$, then Y_2 is stochastically not smaller than Y_1. It turns out that this assumption is actually satisfied by popular classes of Markov chain models under natural assumptions. Here is a list of models we have in mind:

(1) *Nonlinear autoregressive processes with non-decreasing link* The process $\mathbf{X} = (X_t)_{t \in \mathbb{N}_0}$ is assumed to obey the model equation

$$X_t = f(X_{t-1}) + \varepsilon_t \quad \forall t \in \mathbb{N},$$

where $(\varepsilon_t)_{t \in \mathbb{N}}$ is a sequence of i.i.d. random variables and ε_t is independent of $X_{t-1}, \ldots, X_0$. If the function $f : \mathbb{R} \to \mathbb{R}$ is monotonically non-decreasing, then, for $x_1 < x_2$,

$$\begin{aligned}
P(X_t \leq z \mid X_{t-1} = x_1) = P(\varepsilon_t \leq z - f(x_1)) &\geq P(\varepsilon_t \leq z - f(x_2)) \\
&= P(X_t \leq z \mid X_{t-1} = x_2).
\end{aligned}$$

Furthermore, if ε_t has an everywhere positive density and if

$$|f(x)| \leq \gamma |x| - \epsilon \quad \forall x \geq K,$$

for some $\gamma < 1, \epsilon > 0$, and $K < \infty$, then the process $\mathbf{X}$ has a unique stationary distribution and satisfies (A2); see e.g. Doukhan (1994).

(2) *Branching processes with immigration* Let $X_0, (Z_{t,k})_{t,k \in \mathbb{N}}$ and $(\varepsilon_t)_{t \in \mathbb{N}}$ be mutually independent random variables taking values in $\mathbb{N}_0$. We assume that $(Z_{t,k})_{t,k \in \mathbb{N}}$ as well as $(\varepsilon_t)_{t \in \mathbb{N}}$ are sequences of identically distributed random variables. Then the process $\mathbf{X} = (X_t)_{t \in \mathbb{N}_0}$ given by

$$X_t = \sum_{k=1}^{X_{t-1}} Z_{t,k} + \varepsilon_t \quad \forall t \in \mathbb{N}$$

is a branching process with immigration. In the special case of $Z_{t,k} \sim \text{Bin}(1, \alpha)$ we obtain a so-called first-order integer-valued autoregressive (INAR(1)) process which was proposed by McKenzie (1985) and Al-Osh and Alzaid (1987). Since the $Z_{t,k}$ are non-negative random variables, it is obvious that (A1) is fulfilled. If in addition $E\varepsilon_t < \infty$ and $EZ_{t,k} < 1$, then **X** has a unique stationary distribution and satisfies (A2); see Pakes (1971).

(3) *Poisson-INARCH processes* The process $\mathbf{X} = (X_t)_{t\in\mathbb{N}_0}$ is an integer-valued ARCH process of order 1 with Poisson innovations (Poisson-INARCH(1)) if

$$X_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}\big(f(X_{t-1})\big),$$

where $\mathcal{F}_s$ denotes the σ-algebra generated by $X_0, \ldots, X_s$. If f is monotonically non-decreasing, then we obtain, for $x_1 < x_2$ and $Y_1 \sim \text{Poisson}(f(x_1))$, $Y_2 \sim \text{Poisson}(f(x_2))$,

$$P(X_t \leq z \mid X_{t-1} = x_1) = P(Y_1 \leq z) \geq P(Y_2 \leq z) = P(X_t \leq z \mid X_{t-1} = x_1),$$

i.e., (A1) is fulfilled. Furthermore, if in addition

$$f(x) \leq \gamma x - \epsilon \qquad \forall x \geq K,$$

for some $\gamma < 1$, $\epsilon > 0$, and $K < \infty$, then **X** has a unique stationary distribution and satisfies (A2); see e.g. Theorem 2 in Doukhan (1994, Sec. 2.4, p. 90).

We consider an estimator of $F_x(z) = P\big(X_t \leq z \mid X_{t-1} = x\big)$ which takes into account that the function $x \mapsto F_x(z)$ is monotonically non-increasing under (A1). Nonparametric estimators of monotone functions have a long history and were proposed e.g. by Brunk (1955) and Ayer et al. (1955). Denote by $\mathbb{1}(\cdot)$ the indicator function. For $z \in D$ and $x \in \{X_0, \ldots, X_{n-1}\}$, we define

$$\widehat{F}_x^{(\max-\min)}(z) := \max_{v:\, v \geq x} \min_{u:\, u \leq x} \frac{\sum_{t=1}^n \mathbb{1}\big(X_t \leq z, X_{t-1} \in [u, v]\big)}{\#\{t \leq n:\ X_{t-1} \in [u, v]\}} \qquad (2.1a)$$

and

$$\widehat{F}_x^{(\min-\max)}(z) := \min_{u:\, u \leq x} \max_{v:\, v \geq x} \frac{\sum_{t=1}^n \mathbb{1}\big(X_t \leq z, X_{t-1} \in [u, v]\big)}{\#\{t \leq n:\ X_{t-1} \in [u, v]\}}. \qquad (2.1b)$$

It is well-known that $\widehat{F}_x^{(\max-\min)}(z) = \widehat{F}_x^{(\min-\max)}(z)$ for all $x \in \{X_0, \ldots, X_{n-1}\}$, see e.g. Theorem 1 in Brunk (1955) and Theorem 1.4.4 in Robertson, Wright, and Dykstra (1988, p. 23). As pointed out by Deng and Zhang (2020), (2.1a) and (2.1b) have to be modified for $x \notin \{X_0, \ldots, X_{n-1}\}$. Since it could well happen that an interval with $x \in [u, v]$ does not contain any point from the collection $\{X_0, \ldots, X_{n-1}\}$ we set $n_{u,v} = \#\{t \leq n:\ X_{t-1} \in [u, v]\}$, $n_{u,*} = \#\{t \leq n:\ u \leq X_{t-1}\}$, $n_{*,v} = \#\{t \leq$

$n\colon X_{t-1} \le v\}$, and define

$$\widehat{F}_x^{(\max-\min)}(z) := \max_{v\colon v\ge x,\, n_{*,v}>0}\ \min_{u\colon u\le x,\, n_{u,v}>0}\ \frac{\sum_{t=1}^n \mathbb{1}\big(X_t \le z,\, X_{t-1} \in [u,v]\big)}{\#\{t \le n\colon X_{t-1} \in [u,v]\}}$$
(2.2a)

and

$$\widehat{F}_x^{(\min-\max)}(z) := \min_{u\colon u\le x,\, n_{u,*}>0}\ \max_{v\colon v\ge x,\, n_{u,v}>0}\ \frac{\sum_{t=1}^n \mathbb{1}\big(X_t \le z,\, X_{t-1} \in [u,v]\big)}{\#\{t \le n\colon X_{t-1} \in [u,v]\}}.$$
(2.2b)

The estimators $\widehat{F}_x^{(\max-\min)}(z)$ and $\widehat{F}_x^{(\min-\max)}(z)$ are both non-increasing in x as the maxima are taken over non-increasing classes indexed by x and the minima over non-decreasing classes. Furthermore, for fixed $x \in D$, the mappings $z \mapsto \widehat{F}_x^{(\max-\min)}(z)$ and $z \mapsto \widehat{F}_x^{(\min-\max)}(z)$ are non-decreasing which follows from the isotonicity of the functions $z \mapsto \mathbb{1}(X_t \le z, X_{t-1} \in [u,v])$. Furthermore, if $X_{[1]}, \ldots, X_{[n]}$ is an enumeration of the values in $\{X_1, \ldots, X_n\}$ in non-decreasing order, then it follows that, again for fixed $x \in D$, the mappings $z \mapsto \widehat{F}_x^{(\max-\min)}(z)$ and $z \mapsto \widehat{F}_x^{(\min-\max)}(z)$ are constant on the half-open intervals $[X_{[k]}, X_{[k+1]})$ ($k = 1, \ldots, n-1$), and attain the respective values 0 and 1 on $(-\infty, X_{[1]})$ and $[X_{[n]}, \infty)$. Hence, these estimators are genuine probability distribution functions.

We choose as our estimator of $F_x(z)$

$$\widehat{F}_x(z) := \big(\widehat{F}_x^{(\max-\min)}(z) + \widehat{F}_x^{(\min-\max)}(z)\big)/2.$$

It follows that all of the above properties of $\widehat{F}_x^{(\max-\min)}(z)$ and $\widehat{F}_x^{(\min-\max)}(z)$ are inherited by $\widehat{F}_x(z)$. Its performance is characterized by the following theorem.

Theorem 2.1 *Suppose that (A1) and (A2) are fulfilled. Then*

$$\sup_z \left\{ E\left[\int_D \big|\widehat{F}_x(z) - F_x(z)\big|\, dP_X(x) \right] \right\} = O\big(n^{-1/3}\big).$$

The rate of convergence $n^{-1/3}$ is known to be optimal in related problems of estimating a monotone function on the basis of independent random variables; see e.g. Durot (2002, Theorem 1) and Zhang (2002, Theorem 2.3). We believe that this rate cannot be improved in our more delicate case of time series data. Note that Mösching and Dümbgen (2020) considered a nonparametric antitonic estimator of F_x in a regression context where the dependent variables, conditional on the regressors, are independent. They derived under additional Hölder conditions rates of uniform and pointwise convergence for this estimator.

Our approach to prove this result can be most easily explained if the distribution function F_X is continuous. We split the domain D into $k_n = \lfloor n^{1/3} \rfloor$ intervals $I_k = [x_{k-1}, x_k)$, where $x_0 = -\infty$ if $D = \mathbb{R}$, $x_0 = 0$ if $D = \mathbb{N}_0$ and, in both cases,

$x_k = F_X^{-1}(k/k_n) = \sup\{x: F_X(x) \geq k/k_n\}$ for $k = 1, \ldots, k_n - 1$, $x_{k_n} = \infty$. (As usual, $\lfloor a \rfloor$ denotes the largest integer less than or equal to a.) We can expect a favorable behavior of $\widehat{F}_x(z)$ if $N_k(\omega) := \#\{t \leq n: X_{t-1}(\omega) \in I_k\}$ is sufficiently large for all k. Let

$$A_n = \{\omega: N_k(\omega) \geq n/(2k_n) \quad \text{for all } k = 1, \ldots, k_n\}.$$

It follows from Lemma 4.2 that $P(A_n^c) = O(n^{-1/3})$. Since $\int_D |\widehat{F}_x(z) - F_x(z)| \, dP_X(x) \leq 1$ holds with probability 1, we obtain that

$$E\left[\int_D |\widehat{F}_x(z) - F_x(z)| \, dP_X(x) \, \mathbb{1}_{A_n^c}\right] \leq P(A_n^c) = O(n^{-1/3}). \qquad (2.3)$$

To estimate $E\left[\int_D \left(\widehat{F}_x(z) - F_x(z)\right)_+ dP_X(x) \, \mathbb{1}_{A_n}\right]$ we proceed as follows. For $x \in I_k$, $k \in \{2, \ldots, k_n\}$, we use the estimate

$$\left(\widehat{F}_x(z) - F_x(z)\right)^+ \mathbb{1}_{A_n} \leq \left(\widehat{F}_{x_{k-1}}(z) - F_{x_k}(z)\right)^+ \mathbb{1}_{A_n}$$

$$\leq \max_{v:\, v \geq x_{k-1}} \left\{ \frac{\left|\sum_{t=1}^n \left[\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)\right] \mathbb{1}(X_{t-1} \in [x_{k-2}, v])\right|}{\#\{t \leq n:\, X_{t-1} \in [x_{k-2}, v]\}} \mathbb{1}_{A_n} \right\}$$

$$+ \left(F_{x_{k-2}}(z) - F_{x_k}(z)\right).$$

We obtain from Lemma 4.3 that

$$E\left[\max_{v:\, v \geq x_{k-1}} \left\{ \frac{\left|\sum_{t=1}^n \left[\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)\right] \mathbb{1}(X_{t-1} \in [x_{k-2}, v])\right|}{\#\{t \leq n:\, X_{t-1} \in [x_{k-2}, v]\}} \mathbb{1}_{A_n} \right\}\right]$$

$$= O(n^{-1/3}).$$

Since

$$\sum_{k=2}^{k_n} \left(F_{x_{k-2}}(z) - F_{x_k}(z)\right) = \sum_{k=2}^{k_n} \left(F_{x_{k-2}}(z) - F_{x_{k-1}}(z)\right) + \sum_{k=2}^{k_n} \left(F_{x_{k-1}}(z) - F_{x_k}(z)\right)$$

$$= \left(F_{x_0}(z) - F_{x_{k_n-1}}(z)\right) + \left(F_{x_1}(z) - F_{x_{k_n}}(z)\right) \leq 2.$$

we conclude that

$$\sum_{k=2}^{k_n} E\left[\int_{I_k} \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x) \, \mathbb{1}_{A_n}\right] = O(n^{-1/3}).$$

Furthermore, the rough estimate

$$E\left[\int_{I_1} \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x) \, \mathbb{1}_{A_n}\right] \leq P_X(I_1) \leq n^{-1/3}$$

is obviously true, which leads to

$$E\left[\int_D \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x) \, \mathbb{1}_{A_n}\right] = O\left(n^{-1/3}\right). \qquad (2.4)$$

We can prove

$$E\left[\int_D \left(\widehat{F}_x(z) - F_x(z)\right)^- dP_X(x) \, \mathbb{1}_{A_n}\right] = O\left(n^{-1/3}\right). \qquad (2.5)$$

in complete analogy to (2.4). The result stated in Theorem 2.1 follows from (2.3) to (2.5). In the general case we have to take into account that the distribution function F_X is not necessarily continuous. This leads to a technically more involved proof which is presented in full detail in Sect. 4.

The following pictures give an impression of how the functions $x \mapsto F_x(z)$ are approximated by $\widehat{F}_x(z)$ for different values of z. We simulated a Poisson-INARCH process of order 1, where $X_t \mid X_{t-1}, X_{t-2}, \ldots \sim \text{Poisson}\left(f(X_{t-1})\right)$ and $f(x) = \min\{\alpha_0 + \alpha_1 x, \beta\}$. The parameters α_0 and α_1 are chosen as 2.0 and 0.5, respectively, and the truncation constant β is set to 6.0. For a sample size $n = 1000$ and $z = 0, 1, \ldots, 11$, the following pictures show $F_x(z)$ (red lines) and a corresponding estimate $\widehat{F}_x(z)$ (blue lines). These results are quite encouraging except for large values of x. We conjecture that this deficiency is caused by data sparsity in this region.
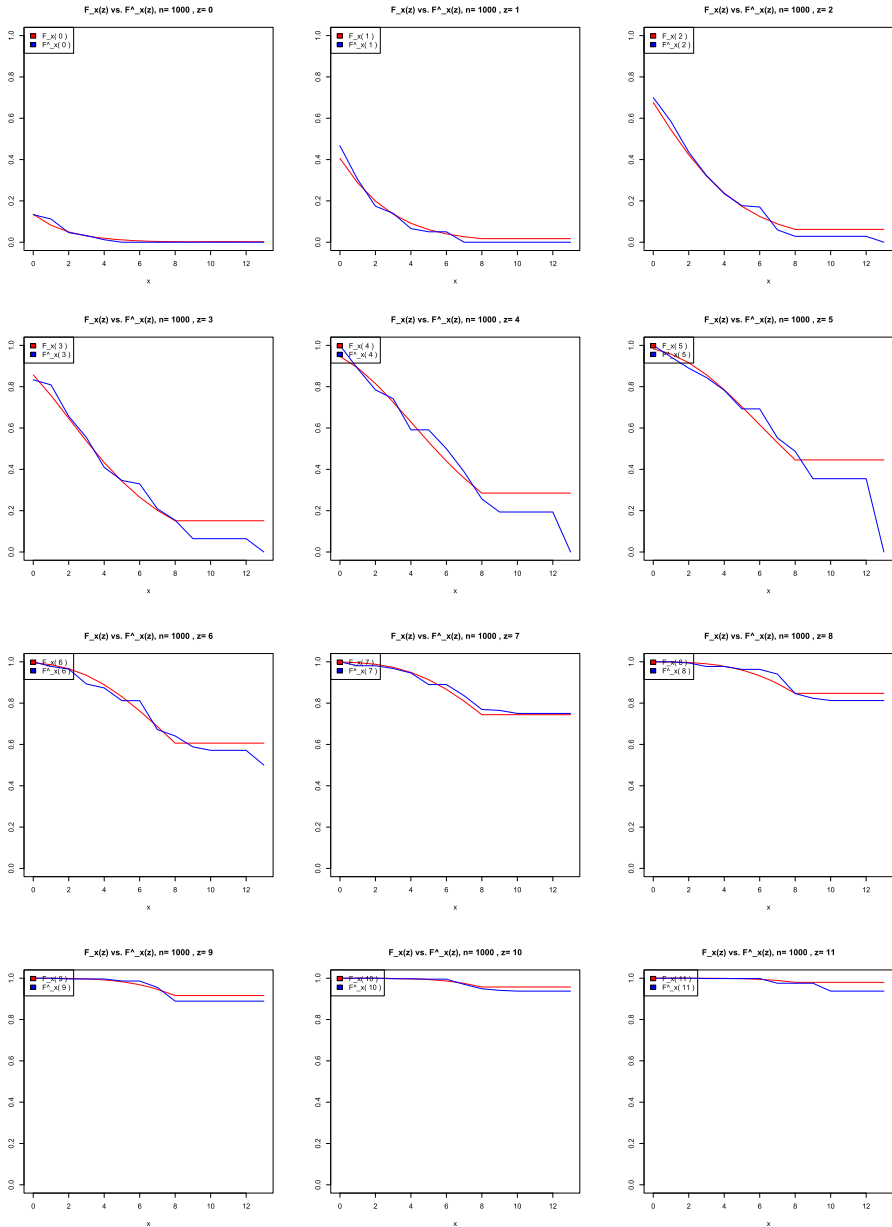
3 A new bootstrap method for Markov chains

Our estimator $\widehat{F}_x(z)$ can be used for bootstrapping Markov processes, and it is particularly suitable in case of Markov chains with a finite or countably infinite state space. In what follows we assume that $(X_t)_{t \in \mathbb{N}_0}$ is a stationary Markov chain which has a

state space $D \subseteq \mathbb{N}_0$. Bootstrap variates X_t^* are generated successively according to a slightly modified variant of our estimator $\widehat{F}_x(z)$. To prove consistency, we retain our monotonicity condition (A1), however, we replace (A2) by the following stronger condition which ensures that both the original and the bootstrap process satisfy a useful mixing condition and possess respective stationary distributions.

(A3) There exist a finite set $S = \{y \in D: \ y \leq \bar{s} \text{ and } P_X(\{y\}) > 0\}$, a probability measure Q on $(\mathbb{N}_0, 2^{\mathbb{N}_0})$, and constants $\delta > 0, \kappa > 0, \gamma > 0$, and $C < \infty$ such that

(i) $P(X_t \in S \mid X_{t-1} = x) \geq \delta > 0 \qquad \forall x \in \mathbb{N}_0$,
(ii) $P(X_t = y \mid X_{t-1} = x) \geq \kappa \cdot Q(\{y\}) \quad \forall x \in S, \ \forall y \in \mathbb{N}_0$,
(iii) $P(X_t \geq x) \leq C e^{-\gamma x} \qquad \forall x \in \mathbb{N}_0$.

(A3) (ii) means that the set S is a so-called small set and condition (A3) (i) ensures that this set can be reached from each point $x \in \Omega_X$ with a probability not smaller than δ. It follows from these conditions that

$$\inf_x P\big(X_{t+2} = y \mid X_t = x\big) \geq \delta \cdot \kappa \cdot Q\big(\{y\}\big) \quad \forall y \in \mathbb{N}_0.$$

Hence, Doeblin's minorization condition is satisfied and it follows that the process $(X_t)_{t \in \mathbb{N}_0}$ has a unique stationary distribution P_X, is geometrically ergodic, and is uniform (ϕ-) mixing with exponentially decaying coefficients.; see e.g. Theorem 1 in Doukhan (1994 Sec. 2.4, p. 88). In particular, a stationary version of the process satisfies (A2). Note that condition (A3) is satisfied e.g. by a Poisson-INARCH(1) process if the function f is bounded. While (i) and (ii) are obviously fulfilled, (iii) follows from the upper tail bound

$$P\big(Y \geq \lambda + x\big) \leq e^{-\frac{x^2}{2(\lambda+x)}} \qquad \forall x \geq 0$$

which holds for $Y \sim \text{Poisson}(\lambda)$; see Theorem 1 in Canonne (2017).

Before we fix the definition of our bootstrap process we check to what extent a process with transition distribution functions $\widehat{F}_x(\cdot)$ satisfies a suitable variant of condition (A3). It follows from Theorem 2.1 that

$$\widehat{F}_x(z) = F_x(z) + O_P\big(n^{-1/3}\big)$$

if $P_X(\{x\}) > 0$. This implies that

$$P\Big(\inf \big\{\widehat{F}_x(y) - \widehat{F}_x(y-1): \ x \in S, \ P_X(\{x\}) > 0\big\} \geq \frac{\kappa}{2} Q^*(\{y\}) \ \ \forall y \in \mathbb{N}_0\Big) \underset{n \to \infty}{\longrightarrow} 1,$$
$$(3.1)$$

where e.g.

$$Q^*(\{y\}) = \begin{cases} Q(\{y\}) & \text{if} \ \ y \leq \bar{y}, \\ 0 & \text{if} \ \ y > \bar{y} \end{cases}$$

and $\bar{y}$ such that $Q(\{0, 1, \ldots, \bar{y}\}) \geq 1/2$. Hence, a bootstrap process based on $\widehat{F}_x(\cdot)$ satisfies a variant of (A3) (ii) with a probability tending to 1.

For a variant of (A3) (i) to hold, it is important that $\inf_x\{\widehat{F}_x(\bar{s}) \colon x \in \mathbb{N}_0\} > 0$ is also satisfied with a probability tending to 1. This is not guaranteed to be true since the estimator $\widehat{F}_x(\bar{s})$ may get unreliable if x gets large. Indeed, the natural lower estimate of $\widehat{F}_x(\bar{s})$ is given by

$$\widehat{F}_x(\bar{s}) \geq \inf_{u \colon u \leq x} \frac{\sum_{t=1}^n \mathbb{1}(X_t \leq \bar{s}, \ X_{t-1} \in [u, \infty))}{\#\{t \leq n \colon \ X_{t-1} \in [u, \infty)\}}.$$

However, the right-hand side of this inequality can get arbitrarily close to 0 if x is large since then $P_X([x, \infty))$ gets small. It is actually a well-known shortcoming of nonparametric isotonic/antitonic estimators that they get unreliable near the ends of the domain of the explanatory variable. In view of this problem, we modify $\widehat{F}_x(z)$ for large x. Let

$$\widetilde{x} := \sup\left\{x \colon \#\{t \leq n \colon X_{t-1} \geq x\} \geq n^{2/3}\right\}.$$

Then $\#\{t \leq n \colon X_{t-1} \geq \widetilde{x}\} \geq n^{2/3} > \#\{t \leq n \colon X_{t-1} > \widetilde{x}\}$. We define

$$\widehat{\widehat{F}}_x(z) := \begin{cases} \widehat{F}_x(z) & \text{if} \quad x \leq \widetilde{x}, \\ \widehat{F}_{\widetilde{x}}(z) & \text{if} \quad x > \widetilde{x} \end{cases}.$$

In what follows we show that the modified estimator $\widehat{\widehat{F}}_x(\cdot)$ actually satisfies a suitable variant of (A3). To take advantage of Lemma 4.2 we embed the random truncation point $\widetilde{x}$ between two *nonrandom* points, $\widetilde{x}_l$ and $\widetilde{x}_u$. Let $\widetilde{x}_l := \sup\left\{x \colon P_X([x, \infty)) \geq 2n^{-1/3}\right\}$ and $\widetilde{x}_u := \sup\left\{x \colon P_X([x, \infty)) \geq (1/2)n^{-1/3}\right\}$. Then $P_X([\widetilde{x}_l, \infty)) \geq 2n^{-1/3} \geq P_X((\widetilde{x}_l, \infty))$ and $P_X([\widetilde{x}_u, \infty)) \geq (1/2)n^{-1/3} \geq P_X((\widetilde{x}_u, \infty))$. Since $\widetilde{x} > \widetilde{x}_u$ implies that $\#\{t \leq n \colon X_{t-1} > \widetilde{x}_u\} \geq n^{2/3}$ we obtain by Lemma 4.2 that

$$\begin{aligned} P\left(\widetilde{x} > \widetilde{x}_u\right) &\leq P\left(\#\{t \leq n \colon X_{t-1} > \widetilde{x}_u\} \geq n^{2/3}\right) \\ &\leq P\left(\#\{t \leq n \colon X_{t-1} > \widetilde{x}_u\} - nP_X((\widetilde{x}_u, \infty)) \geq n^{2/3}/2\right) \\ &= O(n^{-1}). \end{aligned}$$

On the other hand, if $\widetilde{x} \leq \widetilde{x}_u$, then

$$\inf_x \widehat{\widehat{F}}_x(\bar{s}) = \widehat{F}_{\widetilde{x}}(\bar{s}) \geq \widehat{F}_{\widetilde{x}_u}(\bar{s}).$$

Therefore,

$$P\left(\inf_x \widehat{\widehat{F}}_x(\bar{s}) \geq F_{\widetilde{x}_u}(\bar{s})/2\right) \xrightarrow[n \to \infty]{} 1. \tag{3.2}$$

Furthermore, since $\widetilde{x} < \widetilde{x}_l$ implies that $\#\{t \le n \colon X_{t-1} \ge \widetilde{x}_l\} < n^{2/3}$ we obtain by Lemma 4.2 that

$$
\begin{aligned}
P\big(\widetilde{x} < \widetilde{x}_l\big) &\le P\Big(\#\{t \le n \colon X_{t-1} \ge \widetilde{x}_l\} < n^{2/3}\Big) \\
&\le P\Big(\#\{t \le n \colon X_{t-1} \ge \widetilde{x}_l\} - n P_X\big([\widetilde{x}_l, \infty)\big) < -(n/2) P_X\big([\widetilde{x}_l, \infty)\big)\Big) \\
&= O\big(n^{-1}\big).
\end{aligned}
$$

On the other hand, $\widetilde{x} \ge \widetilde{x}_l$ yields that $\widehat{\widehat{F}}_x(z) = \widehat{F}_x(z)$ for all $x \le \widetilde{x}_l$. Hence, we obtain, for each $z \in \mathbb{N}_0$,

$$
\begin{aligned}
&E\left[\int \big|\widehat{\widehat{F}}_x(z) - F_x(z)\big|\, dP_X(x)\right] \\
&= E\left[\int_{\{x \colon x \le \widetilde{x}_l\}} \big|\widehat{F}_x(z) - F_x(z)\big|\, dP_X(x)\right] + P_X\big((\widetilde{x}_l, \infty)\big) + O\big(n^{-1}\big) \\
&= O\big(n^{-1/3}\big).
\end{aligned}
\tag{3.3}
$$

Now we are in a position to define our resampling algorithm generating the bootstrap variates:

1. Choose a starting value X_0^*.
2. For each $t \in \mathbb{N}_0$, suppose that $X_0^*, \dots, X_t^*$ have been generated already. Then X_{t+1}^* is generated such that is has, conditioned on $X_0^*, \dots, X_t^*$ and conditioned on the original sample $X_0, \dots, X_n$, a probability distribution function $\widehat{\widehat{F}}_{X_t^*}(\cdot)$.

In what follows, the symbol P^* refers to the distribution of the bootstrap variables conditioned on the original sample, e.g. $P^*\big(X_t^* \in A\big) = P\big(X_t^* \in A \mid X_0, \dots, X_n\big)$.

Let $K_n = \log n/(3\gamma)$. Then (A3)(iii) implies that $P(X_t > K_n) = O(n^{-1/3})$ and we obtain from (3.3) that

$$
\begin{aligned}
&\sum_{y=0}^{\infty} \sum_{x=0}^{\infty} \big|P^*\big(X_{t+1}^* = y \mid X_t^* = x\big) - P\big(X_{t+1} = y \mid X_t = x\big)\big|\, P_X(\{x\}) \\
&= \sum_{y=0}^{K_n} \sum_{x=0}^{\infty} \big|P^*\big(X_{t+1}^* = y \mid X_t^* = x\big) - P\big(X_{t+1} = y \mid X_t = x\big)\big|\, P_X(\{x\}) \\
&\quad + \sum_{x=0}^{\infty} \big(P^*\big(X_{t+1}^* > K_n \mid X_t^* = x\big) + P\big(X_{t+1} > K_n \mid X_t = x\big)\big)\, P_X(\{x\}) \\
&\le O\big(K_n\, n^{-1/3}\big) + \sum_{x=0}^{\infty} \big|P^*\big(X_{t+1}^* > K_n \mid X_t^* = x\big) - P\big(X_{t+1} > K_n \mid X_t = x\big)\big| \\
&\quad P_X(\{x\}) + 2\, P\big(X_t > K_n\big) \\
&= O_P\big(n^{-1/3} \log n\big).
\end{aligned}
\tag{3.4}
$$

Note that for all $t \in \mathbb{N}$ X_t^* takes values in $\{X_1, \ldots, X_n\}$, i.e. X_t^* lies in the collection of x with $P_X(\{x\}) > 0$. Hence, it follows from (3.1) and (3.2) that a process with transition distribution functions $\widehat{\widehat{F}}_x(\cdot)$ satisfies the following Doeblin-type condition with a probability tending to 1:

$$\inf_x P^*\big(X_{t+2}^* = y \mid X_t^* = x\big) \geq \inf_x \sum_{z \in S} P^*\big(X_{t+2}^* = y \mid X_{t+1}^* = z\big)$$
$$P^*\big(X_{t+1}^* = z \mid X_t^* = x\big)$$
$$\geq \frac{\kappa \, \delta}{4} \, Q^*(\{y\}) \quad \forall y \in \mathbb{N}_0.$$

This implies that, with a probability tending to 1, the bootstrap process is geometrically ergodic and has a unique stationary distribution $P_{X^*}^*$.

For a successful application of the bootstrap approximation the following properties are vitally important: With a probability tending to 1, conditioned on $X_0, \ldots, X_n$,

(a) the stationary distribution $P_{X^*}^*$ converges to P_X,
(b) the finite-dimensional distributions of $(X_t^*)_t$ converge to those of $(X_t)_t$.

We show these two properties by a coupling of the original process and its bootstrap counterpart, i.e. we define versions $(\widetilde{X}_t)_{t \in \mathbb{N}_0}$ and $(\widetilde{X}_t^*)_{t \in \mathbb{N}_0}$ on a common probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ such that the corresponding random variables $\widetilde{X}_t$ and $\widetilde{X}_t^*$ are equal with a high probability. We use the technique of *maximal coupling* [see e.g. Theorem 5.2 in Chapter I in Lindvall (1992)] and define the transition probabilities $\widetilde{\pi}$ driving the coupled process $\big((\widetilde{X}_t, \widetilde{X}_t^*)\big)_{t \in \mathbb{N}_0}$ as follows. For $x, y \in \mathbb{N}_0$, let $\pi(x, y) = P(X_{t+1} = y \mid X_t = x)$ and $\pi^*(x, y) = P^*(X_{t+1}^* = y \mid X_t^* = x)$. Then

$$\delta_{x,x^*} = \frac{1}{2} \sum_{y \in \mathbb{N}_0} \big|\pi(x, y) - \pi^*(x^*, y)\big|$$

is the total variation distance between the distributions with respective probability mass functions $\pi(x, \cdot)$ and $\pi^*(x^*, \cdot)$. Note that $\delta_{x,x^*} = \sum_y [\pi(x, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}] = \sum_y [\pi^*(x^*, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}]$. The transition probabilities of the coupled process are defined as

$$\widetilde{\pi}\big((x, x^*), (y, y)\big) = \min\big\{\pi(x, y), \pi^*(x^*, y)\big\} \quad \forall x, x^*, y \in \mathbb{N}_0, \quad (3.5a)$$

and, for $x, x^*, y, y^* \in \mathbb{N}_0$ such that $y \neq y^*$,

$$\widetilde{\pi}\big((x, x^*), (y, y^*)\big)$$
$$= \begin{cases} 0 & \text{if} \quad \delta_{x,x^*} = 0, \\ \frac{[\pi(x,y) - \min\{\pi(x,y), \pi^*(x^*,y)\}]\,[\pi^*(x^*,y^*) - \min\{\pi(x,y^*), \pi^*(x^*,y^*)\}]}{\delta_{x,x^*}} & \text{if} \quad \delta_{x,x^*} \neq 0. \end{cases}$$
$$(3.5b)$$

The corresponding Markov kernel $\widetilde{P}$ is defined as

$$\widetilde{P}\big((\widetilde{X}_{t+1}, \widetilde{X}_{t+1}^*) \in A \mid (\widetilde{X}_t, \widetilde{X}_t^*) = (x, x^*)\big)$$
$$= \sum_{(y, y^*) \in A} \widetilde{\pi}\big((x, x^*), (y, y^*)\big) \qquad \forall A \subseteq \mathbb{N}_0 \times \mathbb{N}_0.$$

Note that $[\pi(x, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}][\pi^*(x^*, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}]$
$= 0$ for all $x, x^*, y \in \mathbb{N}_0$, which implies in case of $\delta_{x,x^*} > 0$ that

$$\sum_{y^*:\ y^* \neq y} \frac{[\pi(x, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}][\pi^*(x^*, y^*) - \min\{\pi(x, y^*), \pi^*(x^*, y^*)\}]}{\delta_{x,x^*}}$$

$$= [\pi(x, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}] \sum_{y^* \in \mathbb{N}_0} \frac{\pi^*(x^*, y^*) - \min\{\pi(x, y^*), \pi^*(x^*, y^*)\}}{\delta_{x,x^*}}$$

$$= \pi(x, y) - \min\{\pi(x, y), \pi^*(x^*, y)\}.$$

Therefore we obtain

$$\widetilde{P}\big(\widetilde{X}_{t+1} = y \mid (\widetilde{X}_t, \widetilde{X}_t^*) = (x, x^*)\big)$$
$$= \sum_{y^* \in \mathbb{N}_0} \widetilde{\pi}\big((x, x^*), (y, y^*)\big)$$

$$= \begin{cases} \pi(x, y) & \text{if } \delta_{x,x^*} = 0, \\[2mm] \min\{\pi(x, y), \pi^*(x^*, y)\} + \sum_{y^*:\ y^* \neq y} & \\[1mm] \frac{[\pi(x,y) - \min\{\pi(x,y), \pi^*(x^*,y)\}][\pi^*(x^*,y^*) - \min\{\pi(x,y^*), \pi^*(x^*,y^*)\}]}{\delta_{x,x^*}} & \text{if } \delta_{x,x^*} \neq 0 \end{cases}$$

$$= \pi(x, y) = P\big(X_{t+1} = y \mid X_t = x\big)$$

and, likewise,

$$\widetilde{P}\big(\widetilde{X}_{t+1}^* = y^* \mid (\widetilde{X}_t, \widetilde{X}_t^*) = (x, x^*)\big) = \pi^*(x^*, y^*) = P^*\big(X_{t+1}^* = y^* \mid X_t^* = x^*\big).$$

Moreover, we have that

$$\widetilde{P}\big(\widetilde{X}_{t+1} = \widetilde{X}_{t+1}^* \mid (\widetilde{X}_t, \widetilde{X}_t^*) = (x, x^*)\big) = \sum_{y \in \mathbb{N}_0} \min\big\{\pi(x, y), \pi^*(x^*, y)\big\}$$
$$= 1 - \delta_{x,x^*}. \tag{3.6}$$

Hence, the conditional probability that the two random variables at time $t + 1$ are equal is maximized which explains the usage of the term maximal coupling.

The following theorem summarizes the results of our coupling approach. The stationary distribution $P_{X^*}^*$ of the bootstrap process converges in total variation norm and in probability to that of the original process. With a probability tending to 1, the

coupled process $\big((\widetilde{X}_t, \widetilde{X}_t^*)\big)_{t\in\mathbb{N}_0}$ is geometrically ϕ-mixing. And finally, the corresponding values $\widetilde{X}_t^0$ and $\widetilde{X}_t^{*,0}$ of a stationary version of the coupled process coincide with a probability converging to 1.

Theorem 3.1 *Suppose that (A1) and (A3) are fulfilled. Then*

(i) $d_{TV}\big(P_{X^*}^*, P_X\big) = O_{\widetilde{P}}\big(n^{-1/3}(\log n)^2\big)$.

(ii) *With a probability tending to 1, the process $\big((\widetilde{X}_t, \widetilde{X}_t^*)\big)_{t\in\mathbb{N}_0}$ is ϕ-mixing with coefficients $\phi_{\widetilde{X},\widetilde{X}*}(k)$ decaying at a geometric rate.*

(iii) *If $\big((\widetilde{X}_t^0, \widetilde{X}_t^{*,0})\big)_{t\in\mathbb{N}_0}$ is a stationary version of the coupled process, then*

$$\widetilde{P}\big(\widetilde{X}_t^0 \neq \widetilde{X}_t^{*,0}\big) = O_{\widetilde{P}}\big(n^{-1/3}(\log n)^2\big).$$

These general results can be used to prove bootstrap consistency for specific statistics. Suppose that $X_0, \ldots, X_n$ are observed and that (A1) and (A3) are fulfilled. To illustrate our advocated approach, we consider e.g. the parameter $\theta := P(X_{t-1} = x, X_t = y)$, which is consistently estimated by $\widehat{\theta}_n = n^{-1}\sum_{t=1}^n \mathbb{1}(X_{t-1} = x, X_t = y)$. It follows from a central limit theorem for ϕ-mixing processes (see e.g. Theorem 15.12 in Bradley (2007b)) that

$$S_n := \sqrt{n}\big(\widehat{\theta}_n - \theta\big) \xrightarrow{d} Y \sim \mathcal{N}\Big(0, \sigma_\infty^2\Big),$$

where $\sigma_\infty^2 = \sum_{k=-\infty}^\infty \mathrm{cov}\big(\mathbb{1}(X_0 = x, X_1 = y), \mathbb{1}(X_{|k|} = x, X_{|k|+1} = y)\big)$. The distribution of S_n can be approximated by that of its bootstrap counterpart,

$$S_n^* := \sqrt{n}\big(\widehat{\theta}_n^* - E^*\widehat{\theta}_n^*\big),$$

where $\widehat{\theta}_n^* = n^{-1}\sum_{t=1}^n \mathbb{1}(X_{t-1}^* = x, X_t^* = y)$ and $E^*\widehat{\theta}_n^* = E\big(\widehat{\theta}_n^* \,\big|\, X_0, \ldots, X_n\big)$. In order to prove that the distribution of S_n^* converges in probability to the same limit as that of S_n, we could use a central limit theorem for triangular arrays of ϕ-mixing random variables. Alternatively, we can use our coupling results and obtain bootstrap consistency almost for free. It follows from (iii) of Theorem 3.1 that

$$\widetilde{P}\Big((\widetilde{X}_{t-1}, \widetilde{X}_t) \neq (\widetilde{X}_{t-1}^*, \widetilde{X}_t^*)\Big) = O_{\widetilde{P}}\big(n^{-1/3}(\log n)^2\big).$$

Using this and a covariance inequality for ϕ-mixing random variables [see e.g. Theorem 1.2.2.3 in Doukhan (1994, p. 9)] we obtain

$$
\begin{aligned}
&\big|\mathrm{cov}\big(\mathbb{1}(\widetilde{X}_{s-1}, \widetilde{X}_s) - \mathbb{1}(\widetilde{X}_{s-1}^*, \widetilde{X}_s^*), \mathbb{1}(\widetilde{X}_{t-1}, \widetilde{X}_t) - \mathbb{1}(\widetilde{X}_{t-1}^*, \widetilde{X}_t^*)\big)\big| \\
&\quad \leq 2\,\phi\big(|s-t|-1\big)\,\big\|\mathbb{1}(\widetilde{X}_{s-1}, \widetilde{X}_s) - \mathbb{1}(\widetilde{X}_{s-1}^*, \widetilde{X}_s^*)\big\|_1 \big\|\mathbb{1}(\widetilde{X}_{t-1}, \widetilde{X}_t) - \mathbb{1}(\widetilde{X}_{t-1}^*, \widetilde{X}_t^*)\big\|_\infty \\
&\quad = O_{\widetilde{P}}\big(\phi_{\widetilde{X},\widetilde{X}*}(|s-t|-1)\,n^{-1/3}(\log n)^2\big),
\end{aligned}
$$

which implies that

$$
\widetilde{E}\left|\left(\widetilde{S}_n - \widetilde{S}_n^*\right)^2\right]
$$
$$
= \frac{1}{n} \sum_{s,t=1}^{n} \operatorname{cov}\left(\mathbb{1}(\widetilde{X}_{s-1}, \widetilde{X}_s) - \mathbb{1}(\widetilde{X}_{s-1}^*, \widetilde{X}_s^*), \mathbb{1}(\widetilde{X}_{t-1}, \widetilde{X}_t) - \mathbb{1}(\widetilde{X}_{t-1}^*, \widetilde{X}_t^*)\right)
$$
$$
= O_{\widetilde{P}}\left(\phi_{\widetilde{X},\widetilde{X}^*}(|s-t|-1)\, n^{-1/3}\, (\log n)^2\right).
$$

This implies that

$$
S_n^* \xrightarrow{d} Y \qquad \text{in probability.}
$$

If in addition $\sigma_\infty^2 > 0$, then we obtain by Lemma 2.11 of van der Vaart (1998) that

$$
\sup_x \left| P\left(S_n \le x\right) - P\left(S_n^* \le x \mid X_0, \ldots, X_n\right)\right| \xrightarrow{P} 0.
$$

Hence, we can use bootstrap quantiles to construct confidence intervals for θ such that their coverage probability converges to a prescribed level. Similar implications for other types of statistics are discussed in Leucht and Neumann (2013) and Neumann (2021).

Remark 1 In a similar context, Paparoditis and Politis (2002, Theorem 3.3) proved almost sure convergence of the bootstrap stationary distribution to the stationary distribution of the original process. Their method of proof is completely different from ours and employs classical tools from the theory of weak convergence such as Helly's theorem and the "uniqueness trick" which uses the fact that each subsequence contains a further subsequence converging to the same probability measure. We use a more direct approach based on a coupling of the original and the bootstrap process. The additional benefit is that we obtain a rate of convergence rather than consistency only.

The following pictures give an impression of the effect of our coupling. As done for the pictures displayed in the previous section, we simulated a Poisson-INARCH process of order 1, where $X_t \mid X_{t-1}, X_{t-2}, \ldots \sim \mathrm{Poisson}\left(f(X_{t-1})\right)$ and $f(x) = \min\left\{\alpha_0 + \alpha_1 x, \beta\right\}$. The parameters α_0 and α_1 are chosen as 2.0 and 0.5, respectively, and the truncation constant β is set to 6.0. For respective sample sizes of $n = 200$ and $n = 1000$, Figs. 1 and 2 show one realization of independent and coupled versions of $X_1, \ldots, X_{50}$ and $X_1^*, \ldots, X_{50}^*$. While the pictures on the left of Figs. 1 and 2 let us at best hope for a similar behavior of the bootstrap and the original process, those on the right provide some evidence that the bootstrap process successfully mimics the behavior of the original process.
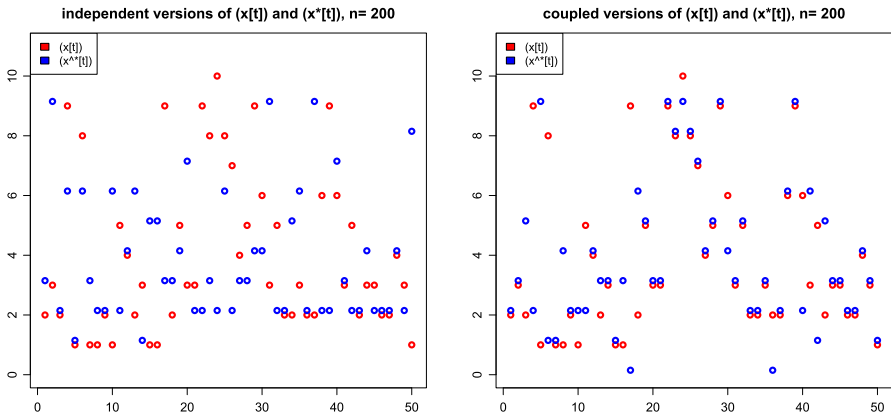
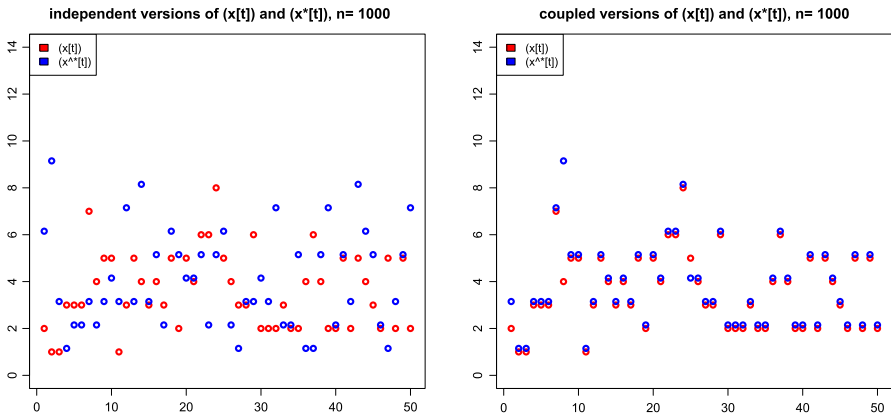**Fig. 1** Independent and coupled processes, n = 200

Fig. 2 Independent and coupled processes, n = 1000

4 Proofs

4.1 Proofs of the main results

Proof of Theorem 2.1 Our strategy to prove this result is already sketched in Sect. 2, in the special case where the distribution function F_X associated to P_X is continuous. In the general case with a possibly discontinuous function F_X, we have to take great care since we cannot split the domain D into intervals I_k such that $P_X(I_k) = 1/k_n$, where $k_n = \lfloor n^{1/3} \rfloor$. It could be the case that P_X has masses considerably larger than $1/k_n$ at single points which requires a modification of our previous approach.

To obtain an appropriate collection of intervals I_k, we define again suitable grid points $x_0, x_n, \ldots, x_{K_n}$. For technical reasons we choose them as a decreasing sequence. We set $x_0 := \infty$ and define recursively $x_k := \inf\{x : P_X((x, x_{k-1})) \le n^{-1/3}\}$ for $k \ge 1$. This procedure will terminate when $x_{K_n} = 0$ and $D = \mathbb{R}$ or when $x_{K_n} = -\infty$, for some K_n. In both cases we have that $D = [x_1, x_0) \cup \cdots \cup [x_{K_n}, x_{K_n-1})$.

For $k = 1, \ldots, K_n - 1$, i.e. with a possible exception of $k = K_n$, we have

$$P_X\big((x_k, x_{k-1})\big) \leq n^{-1/3} \leq \lim_{m \to \infty} P_X\big((x_k - 1/m, x_{k-1})\big) = P_X\big([x_k, x_{k-1})\big),$$

where the latter equality follows since the probability measure P_X is continuous from above. In the following we show that

$$E\left[\int_D \big(\widehat{F}_x(z) - F_x(z)\big)^+ \, dP_X(x) \right] = O\big(n^{-1/3}\big). \tag{4.1}$$

To this end, we consider the contributions by $E\big[\int_{[x_k, x_{k-1})} (\widehat{F}_x(z) - F_x(z))^+ \, dP_X(x) \big]$ separately. We distinguish between three possible cases.

Case 1 If $P_X\big([x_k, x_{k-1})\big) \leq 2n^{-1/3}$ and $k < K_n$, then we use for all $x \in [x_k, x_{k-1})$ in case of $N_{n,k} := \{t \leq n : X_{t-1} \in [x_k, x_{k-1})\} \neq 0$ the estimate

$$\begin{aligned}
\big(\widehat{F}_x(z) - F_x(z)\big)^+ &\leq \big(\widehat{F}_{x_k}(z) - F_{x_{k-1}}(z)\big)^+ \\
&\leq \max_{v:\, v \geq x_k} \left\{ \frac{\left| \sum_{t=1}^n [\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)] \, \mathbb{1}(X_{t-1} \in [x_{k+1}, v])\right|}{\#\{t \leq n : X_{t-1} \in [x_{k+1}, v]\} \vee 1} \right\} \\
&\quad + \big[F_{x_{k+1}}(z) - F_{x_{k-1}}(z)\big],
\end{aligned}$$

which leads to

$$\begin{aligned}
&E\left[\int_{[x_k, x_{k-1})} \big(\widehat{F}_x(z) - F_x(z)\big)^+ \, dP_X(x) \, \mathbb{1}_{\{N_{n,k} \neq 0\}} \right] \\
&\leq P_X\big([x_k, x_{k-1})\big) \\
&\quad \left\{ E\left[\max_{v:\, v \geq x_k} \left\{ \frac{\left| \sum_{t=1}^n [\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)] \mathbb{1}(X_{t-1} \in [x_{k+1}, v])\right|}{\#\{t \leq n : X_{t-1} \in [x_{k+1}, v]\} \vee 1} \right\} \right] \right. \\
&\quad \left. + \big[F_{x_{k+1}}(z) - F_{x_{k-1}}(z)\big] \right\} \\
&= O\Big(P_X\big([x_k, x_{k-1})\big) \big\{ n^{-1/3} + (F_{x_{k+1}}(z) - F_{x_{k-1}}(z)) \big\} \Big). \tag{4.2a}
\end{aligned}$$

Case 2 If $P_X\big([x_k, x_{k-1})\big) > 2n^{-1/3}$ then P_X has at x_k a point mass greater than $n^{-1/3}$ and we argue differently. In this case, we use for all $x \in (x_k, x_{k-1})$ in case of $N_{n,k} := \{t \leq n : X_{t-1} = x_k\} \neq 0$ the estimate

$$\begin{aligned}
\big(\widehat{F}_x(z) - F_x(z)\big)^+ &\leq \max_{v:\, v \geq x_k} \left\{ \frac{\left| \sum_{t=1}^n [\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)] \, \mathbb{1}(X_{t-1} \in [x_k, v])\right|}{\#\{t \leq n : X_{t-1} \in [x_k, v]\} \vee 1} \right\} \\
&\quad + \big[F_{x_k}(z) - F_{x_{k-1}}(z)\big],
\end{aligned}$$

which implies

$$
E\left[\int_{(x_k,x_{k-1})} \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x)\, \mathbb{1}_{\{N_{n,k}\neq 0\}}\right]
$$
$$
\leq P_X\big((x_k, x_{k-1})\big)
$$
$$
\left\{ E\left[\max_{v:\, v\geq x_k}\left\{\frac{\left|\sum_{t=1}^n [\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)]\, \mathbb{1}(X_{t-1} \in [x_k, v])\right|}{\#\{t \leq n:\, X_{t-1} \in [x_k, v]\} \vee 1}\right\}\right]\right.
$$
$$
\left. + \left[F_{x_k}(z) - F_{x_{k-1}}(z)\right]\right\}
$$
$$
= O\Big(P_X\big((x_k, x_{k-1})\big)\left\{n^{-1/3} + (F_{x_k}(z) - F_{x_{k-1}}(z))\right\}\Big). \tag{4.2b}
$$

For $x = x_k$, we use the simpler estimate

$$
\left(\widehat{F}_x(z) - F_x(z)\right)^+ \leq \max_{v:\, v>x_k}\left\{\frac{\left|\sum_{t=1}^n [\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)]\, \mathbb{1}(X_{t-1} \in [x_k, v])\right|}{\#\{t \leq n:\, X_{t-1} \in [x_k, v]\} \vee 1}\right\},
$$

and we obtain

$$
E\left[\int_{\{x_k\}} \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x)\, \mathbb{1}_{\{N_{n,k}\neq 0\}}\right]
$$
$$
\leq P_X(\{x_k\})\, E\left[\max_{v:\, v>x_k}\left\{\frac{\left|\sum_{t=1}^n [\mathbb{1}(X_t \leq z) - F_{X_{t-1}}(z)]\mathbb{1}(X_{t-1} \in [x_k, v])\right|}{\#\{t \leq n:\, X_{t-1} \in [x_k, v]\} \vee 1}\right\}\, \mathbb{1}_{A_n}\right]
$$
$$
= O\Big(P_X(\{x_k\})\, n^{-1/3}\Big). \tag{4.2c}
$$

Case 3   If $P_X([x_{K_n}, x_{K_n-1})) \leq 2n^{-1/3}$, then we can simply use the estimate

$$
E\left[\int_{[x_{K_n}, x_{K_n-1})} \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x)\right] \leq 2n^{-1/3}. \tag{4.2d}
$$

Finally, it follows from Lemma 4.2 that $P\big(\bigcup_k \{\omega:\, N_{n,k}(\omega) = 0\}\big) = O(n^{-1/3})$, which implies that

$$
E\left[\int_D \left(\widehat{F}_x(z) - F_x(z)\right)^+ dP_X(x)\, \mathbb{1}_{\bigcup_k\{N_{n,k}=0\}}\right]
$$
$$
\leq P\left(\bigcup_k \{\omega:\, N_{n,k}(\omega) = 0\}\right) = O(n^{-1/3}). \tag{4.2e}
$$

From (4.2a) to (4.2e) we obtain (4.1).

The term $\int_D (\widehat{F}_x(z) - F_x(z))^- dP_X(x)$ can be analogously estimated which completes the proof of the theorem.   □

Proof of Theorem 3.1   (i) We construct a coupling of the original process and its bootstrap counterpart, where we use $\widetilde{\pi}\big((x, x^*), (y, y^*)\big)$ defined by (3.5a) and (3.5b) as transition probabilities and $\widetilde{P}$ as transition kernel. The initial values are chosen such that $\widetilde{X}_0 = \widetilde{X}_0^* \sim P_X$. Then, for each $t \in \mathbb{N}_0$, conditioned on $(\widetilde{X}_t, \widetilde{X}_t^*)$, the next pair $(\widetilde{X}_{t+1}, \widetilde{X}_{t+1}^*)$ is generated according to $\widetilde{P}$. It follows from (3.4) and (3.6) in particular that

$$
\begin{aligned}
& \widetilde{P}\big(\widetilde{X}_{t+1} \neq \widetilde{X}_{t+1}^*,\ \widetilde{X}_t = \widetilde{X}_t^*\big) \\
& = \sum_{x \in \mathbb{N}_0} \widetilde{P}\big(\widetilde{X}_{t+1} \neq \widetilde{X}_{t+1}^* \mid \widetilde{X}_t = \widetilde{X}_t^* = x\big)\, \widetilde{P}\big(\widetilde{X}_t = \widetilde{X}_t^* = x\big) \\
& = \sum_{x \in \mathbb{N}_0} \delta_{x,x}\, \widetilde{P}\big(\widetilde{X}_t = \widetilde{X}_t^* = x\big) \\
& \leq \frac{1}{2} \sum_x \sum_y \big|\pi(x, y) - \pi^*(x, y)\big|\, P_X(\{x\}) \\
& = O_{\widetilde{P}}\big(n^{-1/3} \log n\big).
\end{aligned}
$$

This implies first

$$
\widetilde{P}\big(\widetilde{X}_1 \neq \widetilde{X}_1^*\big) = O_{\widetilde{P}}\big(n^{-1/3} \log n\big),
$$

then

$$
\begin{aligned}
\widetilde{P}\big(\widetilde{X}_2 \neq \widetilde{X}_2^*\big) & \leq \widetilde{P}\big(\widetilde{X}_2 \neq \widetilde{X}_2^*,\ \widetilde{X}_1 = \widetilde{X}_1^*\big) + \widetilde{P}\big(\widetilde{X}_1 \neq \widetilde{X}_1^*\big) \\
& = O_{\widetilde{P}}\big(n^{-1/3} \log n\big),
\end{aligned}
$$

and after K_n such steps

$$
d_{TV}\big(P_X, \widetilde{P}^{\widetilde{X}_{K_n}^*}\big) \leq \widetilde{P}\big(\widetilde{X}_{K_n} \neq \widetilde{X}_{K_n}^*\big) = O_{\widetilde{P}}\big(n^{-1/3} \log n\, K_n\big).
$$

On the other hand, $(X_t^*)_{t \in \mathbb{N}_0}$, and therefore $(\widetilde{X}_t^*)_{t \in \mathbb{N}_0}$ as well, are geometrically ergodic. Hence, for $K_n = K \log n$ and K sufficiently large,

$$
d_{TV}\Big(\widetilde{P}^{\widetilde{X}_{K_n}^*}, P_{X^*}^*\Big) = O_{\widetilde{P}}\big(n^{-1/3}\big),
$$

which leads to

$$
\begin{aligned}
d_{TV}\big(P_X, P_{X^*}^*\big) & \leq d_{TV}\Big(P_X, \widetilde{P}^{\widetilde{X}_{K_n}^*}\Big) + d_{TV}\Big(\widetilde{P}^{\widetilde{X}_{K_n}^*}, P_{X^*}^*\Big) \\
& = O_{\widetilde{P}}\big(n^{-1/3} (\log n)^2\big).
\end{aligned}
$$

(ii) We couple the original and the bootstrap process according to (3.5a) and (3.5b) and show first that

$$
\widetilde{P}\big((\widetilde{X}_t, \widetilde{X}_t^*) \in S \times S \,\big|\, \widetilde{X}_{t-1} = x, \widetilde{X}_{t-1}^* = x^*\big)
$$
$$
\geq P\big(X_t \in S \,\big|\, X_{t-1} = x\big) \cdot P^*\big(X_t^* \in S \,\big|\, X_{t-1}^* = x^*\big) \qquad (4.3)
$$

holds for all $x, x^* \in \mathbb{N}_0$. Let $x, x^* \in \mathbb{N}_0$ be arbitrary. To simplify notation we set, for a generic set $B \subseteq \mathbb{N}_0$, $\pi(B) = \sum_{y \in B} \pi(x, y)$, $\pi^*(B) = \sum_{y \in B} \pi^*(x^*, y)$, and $\pi \wedge \pi^*(B) = \sum_{y \in B} \pi(x, y) \wedge \pi^*(x^*, y)$.
If $\pi \wedge \pi^*(S) \geq \pi(S) \cdot \pi^*(S)$, then (4.3) follows immediately. Suppose now the opposite, $\pi \wedge \pi^*(S) < \pi(S) \cdot \pi^*(S)$. Then $\delta_{x,x^*} > 0$, and it follows from (3.5a) and (3.5b)

$$
\widetilde{P}\big((\widetilde{X}_t, \widetilde{X}_t^*) \in S \times S \,\big|\, \widetilde{X}_{t-1} = x, \widetilde{X}_{t-1}^* = x^*\big)
$$
$$
= \sum_{y \in S} \pi(x, y) \wedge \pi^*(x^*, y),
$$
$$
+ \sum_{y, y^* \in S} \frac{\big(\pi(x, y) - \pi(x, y) \wedge \pi^*(x^*, y)\big)\big(\pi^*(x^*, y^*) - \pi(x, y^*) \wedge \pi^*(x^*, y^*)\big)}{\delta_{x,x^*}}
$$
$$
= \pi \wedge \pi^*(S) + \frac{1}{\delta_{x,x^*}}\big(\pi(S) - \pi \wedge \pi^*(S)\big)\big(\pi^*(S) - \pi \wedge \pi^*(S)\big).
$$
$$
= \pi(S) \cdot \pi^*(S) + \frac{1}{\delta_{x,x^*}}\Big\{\delta_{x,x^*}\big(\pi \wedge \pi^*(S) - \pi(S)\pi^*(S)\big)
$$
$$
+ \big(\pi(S) - \pi \wedge \pi^*(S)\big)\big(\pi^*(S) - \pi \wedge \pi^*(S)\big)\Big\}.
$$

Since $\delta_{x,x^*} = 1 - \pi \wedge \pi^*(\mathbb{N}_0) = \big(\pi(S) - \pi \wedge \pi^*(S)\big) + \big(\pi(S^c) - \pi \wedge \pi^*(S^c)\big)$ the term in curly braces is equal to

$$
\big(\pi(S) - \pi \wedge \pi^*(S)\big)\big(\pi \wedge \pi^*(S) - \pi(S)\pi^*(S)\big)
$$
$$
+ \big(\pi(S^c) - \pi \wedge \pi^*(S^c)\big)\big(\pi \wedge \pi^*(S) - \pi(S)\pi^*(S)\big)
$$
$$
+ \big(\pi(S) - \pi \wedge \pi^*(S)\big)\big(\pi^*(S) - \pi \wedge \pi^*(S)\big)
$$
$$
= \big(\pi(S) - \pi \wedge \pi^*(S)\big)\pi^*(S)\,\pi(S^c)
$$
$$
+ \big(\pi(S^c) - \pi \wedge \pi^*(S^c)\big)\big(\pi \wedge \pi^*(S) - \pi(S)\pi^*(S)\big)
$$
$$
= \pi(S^c)\big(\pi \wedge \pi^*(S) - \pi \wedge \pi^*(S)\pi^*(S)\big)
$$
$$
+ \pi \wedge \pi^*(S^c)\big(\pi(S)\pi^*(S) - \pi \wedge \pi^*(S)\big),
$$

and is therefore non-negative. This proves (4.3).
It follows from (3.4) that, for $y, y^* \in S$ such that $P_X(\{y^*\}) > 0$,

$$
\widetilde{P}\big(\widetilde{X}_{t+1} = \widetilde{X}_{t+1}^* = z \,\big|\, \widetilde{X}_t = y, \widetilde{X}_t^* = y^*\big) = \pi(y, z) \wedge \pi^*(y^*, z)
$$
$$
\geq \kappa\, Q(\{z\}) + O_P\big(n^{-1/3} \log n\big). \qquad (4.4)
$$

We obtain from (4.3) and (4.4) there exist some $\kappa^* > 0$ such that

$$
\begin{aligned}
&\widetilde{P}\big((\widetilde{X}_{t+1}, \widetilde{X}^*_{t+1}) = (z, z) \,\big|\, \widetilde{X}_{t-1} = x, \widetilde{X}^*_{t-1} = x^*\big) \\
&\geq \sum_{y, y^* \in S} \widetilde{P}\big((\widetilde{X}_{t+1}, \widetilde{X}^*_{t+1}) = (z, z) \,\big|\, \widetilde{X}_t = y, \widetilde{X}^*_t = y^*\big) \\
&\qquad \widetilde{P}\big((\widetilde{X}_t, \widetilde{X}^*_t) = (y, y^*) \,\big|\, \widetilde{X}_{t-1} = x, \widetilde{X}^*_{t-1} = x^*\big) \\
&\geq \kappa^*
\end{aligned}
\tag{4.5}
$$

holds with a probability tending to 1. Hence, with a probability tending to 1, the coupled process is ϕ-mixing with geometrically decaying coefficients.

(iii) According to (4.5), the coupled process $\big((\widetilde{X}_t, \widetilde{X}^*_t)\big)_{t \in \mathbb{N}_0}$ satisfies Doeblin's condition which implies in particular that this process has a unique stationary distribution. Let $\big((\widetilde{X}^0_t, \widetilde{X}^{*,0}_t)\big)_{t \in \mathbb{N}_0}$ be a stationary version of the coupled process. Since $\big((\widetilde{X}_t, \widetilde{X}^*_t)\big)_{t \in \mathbb{N}_0}$ is geometrically ergodic we obtain

$$
\begin{aligned}
\widetilde{P}\big(\widetilde{X}^0_t \neq \widetilde{X}^{*,0}_t\big) &\leq \widetilde{P}\big(\widetilde{X}_{K_n} \neq \widetilde{X}^*_{K_n}\big) + d_{TV}\big(\widetilde{P}^{(\widetilde{X}^0_{K_n}, \widetilde{X}^{*,0}_{K_n})}, \widetilde{P}^{(\widetilde{X}_{K_n}, \widetilde{X}^*_{K_n})}\big) \\
&= O_{\widetilde{P}}\big(n^{-1/3} (\log n)^2\big).
\end{aligned}
$$

$\square$

4.2 Some auxiliary lemmas

Lemma 4.1 *Suppose that $(X_t)_{t \in \mathbb{N}_0}$ is a Markov chain with state space $D \subseteq \mathbb{R}$ such that (A2) is fulfilled. For arbitrary $I \subseteq D$, let*

$$
\eta_t := \big[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1})\big]\, \mathbb{1}(X_{t-1} \in I),
$$

where $I \subseteq D$. Then, for arbitrary $\gamma < 1$,

$$
E\left[\left(\sum_{t=1}^n \eta_t\right)^4\right] = O\big((n\, p_I)^2 + n\, p_I^\gamma\big),
$$

where $p_I := P(X_0 \in I)$.

Proof In view of $E\big[\big(\sum_{t=1}^n \eta_t\big)^4\big] = \sum_{s,t,u,v=1}^n E[\eta_s \eta_t \eta_u \eta_v]$ we first consider the terms $E[\eta_s \eta_t \eta_u \eta_v]$. Let the indices be chronologically ordered, i.e. $1 \leq s \leq t \leq u \leq v \leq n$. Then it follows from the Markov property that

$$
E\big[\eta_s \eta_t \eta_u \eta_v\big] = 0 \quad \text{if} \quad u < v.
$$

Considering the remaining cases of $s \leq t \leq u = v$, we make use of the following equalities.

(a) $s = t = u = v$
   Then $E[\eta_s \eta_t \eta_u \eta_v] = E[\eta_s^4]$.
(b) $s = t < u = v$
   Then $E[\eta_s \eta_t \eta_u \eta_v] = \text{cov}(\eta_s^2, \eta_u^2) + E[\eta_s^2] E[\eta_u^2]$.
(c) $s < t \leq u = v$
   Then $E[\eta_s \eta_t \eta_u \eta_v] = \text{cov}(\eta_s, \eta_t \eta_u^2) = \text{cov}(\eta_s \eta_t, \eta_u^2)$.

For $s < u$, there exist $\binom{4}{2} = 6$ quadruples (t_1, t_2, t_3, t_4) such that $t_i = t_j = s$ for some $i \neq j$, and $t_k = t_l = u$ for some $k \neq l$. For $s < t < u$, there exist $4 \cdot 3 = 12$ quadruples (t_1, t_2, t_3, t_4) such that $t_i = s$, $t_j = t$ and $t_k = t_l = u$ for some i, j, k, l, $k \neq l$. Finally, for $s < t = u$, there exist 4 quadruples (t_1, t_2, t_3, t_4) such that $t_i = s$ for some i and $t_j = u$ for $j \neq i$. Therefore we obtain

$$E\left[\left(\sum_{t=1}^{n} \eta_t\right)^4\right] \leq \sum_{t=1}^{n} E[\eta_t^4] + 6 \sum_{1 \leq s < u \leq n} E[\eta_s^2] E[\eta_u^2]$$

$$+ 12 \sum_{r=1}^{n-1} \sum_{(s,t,u) \in \mathcal{T}_{n,r}^{(1)}} \left| \text{cov}(\eta_s, \eta_t \eta_u^2) \right|$$

$$+ 12 \sum_{r=1}^{n-1} \sum_{(s,t,u) \in \mathcal{T}_{n,r}^{(2)}} \left| \text{cov}(\eta_s \eta_t, \eta_u^2) \right|, \qquad (4.6)$$

where

$$\mathcal{T}_{n,r}^{(1)} = \left\{ (s, t, u): \ 1 \leq s < t \leq u \leq n, \ r := t - s \geq u - t \right\}$$
$$\mathcal{T}_{n,r}^{(2)} = \left\{ (s, t, u): \ 1 \leq s \leq t < u \leq n, \ r := u - t > t - s \right\}.$$

To estimate the last two terms on the right-hand side of (4.6) we use a well-known covariance inequality for α-mixing random variables,

$$\left| \text{cov}(X, Y) \right| \leq 4 \left[\alpha(\sigma(X), \sigma(Y)) \right]^{1 - 1/\alpha - 1/\beta} \|X\|_\alpha \|Y\|_\beta,$$

where $\alpha, \beta \in (1, \infty)$ are such that $1/\alpha + 1/\beta < 1$ and $\|X\|_\alpha < \infty$, $\|Y\|_\beta < \infty$; see e.g. Bradley (2007a, Corollary 10.16). Choosing $\alpha = \beta = 2/\gamma$ and taking into account that $|\eta_s| \leq 1$ and $E|\eta_s| \leq p_I$ we obtain that

$$\left| \text{cov}(\eta_s, \eta_t \eta_u^2) \right| \leq \left[\alpha_X(t - s - 1) \right]^{1 - \gamma/2 - \gamma/2} \|\eta_s\|_{2/\gamma} \|\eta_t \eta_u^2\|_{2/\gamma}$$
$$\leq \left[\alpha_X(t - s - 1) \right]^{1 - \gamma} p_I^\gamma$$

as well as

$$\left| \text{cov}(\eta_s \eta_t, \eta_u^2) \right| \leq \left[\alpha_X(u - t - 1) \right]^{1 - \gamma} p_I^\gamma.$$

Using $\#\mathcal{T}_{n,r}^{(1)} \leq n(r+1)$ and $\#\mathcal{T}_{n,r}^{(2)} \leq nr$ we obtain from (4.6)

$$
E\left[\left(\sum_{t=1}^{n} \eta_t\right)^4\right] \leq n\, p_I + 6\,(n\, p_I)^2
$$

$$
+ 12\, n \sum_{r=1}^{n-1} (2r+1)\big[\alpha_X(r-1)\big]^{1-\gamma}\, p_I^{\gamma}
$$

$$
= O\Big((n\, p_I)^2 + n\, p_I^{\gamma}\Big),
$$

which completes the proof. $\qquad\square$

Lemma 4.2 *Suppose that* $(X_t)_{t\in\mathbb{N}_0}$ *is a Markov chain with state space* $D \subseteq \mathbb{R}$ *and stationary distribution* P_X *such that (A2) is fulfilled. For arbitrary* $I \subseteq D$, *let* $N_n(I) := \#\{t \leq n\colon X_{t-1} \in I\}$. *Then, for arbitrary* $\delta > 0$, $\kappa < \infty$, *and* $P_X(I) \geq n^{\delta-1}$,

$$
P\big(|N_n(I) - n\, P_X(I)| > n\, P_X(I)/2\big) = O\big(n^{-\kappa}\big).
$$

Proof Let $q \in 2\mathbb{N}$ and $\epsilon > 0$. Since

$$
\sum_{r=1}^{\infty} r^{q-2}\, [\alpha_X(r)]^{\epsilon/(q+\epsilon)} < \infty
$$

it follows from an extension of Rosenthal's inequality (see e.g. Theorem 2 in Section 1.4.1 in Doukhan (1994)) that

$$
E\Big[\big|N_n(I) - n\, P_X(I)\big|^q\Big] = E\left[\left|\sum_{t=1}^{n}\big(\mathbb{1}(X_{t-1} \in I) - P_X(I)\big)\right|^q\right]
$$

$$
\leq C_q\left\{n^{q/2}\, P_X(I)^{q/(2+\epsilon)} + n\, P_X(I)^{q/(q+\epsilon)}\right\}.
\tag{4.7}
$$

Choosing $\epsilon > 0$ small enough we have that $n\, P_X(I)^{2-2/(2+\epsilon)} \geq n^{\delta'}$ for some $\delta' > 0$. Therefore we obtain from Markov's inequality that

$$
P\big(|N_n(I) - n\, P_X(I)| > n\, P_X(I)/2\big)
$$

$$
\leq C_q\, \frac{n^{q/2}\, P_X(I)^{q/(2+\epsilon)} + n\, P_X(I)^{q/(q+\epsilon)}}{(n\, \mathbb{1}_X(I)/2)^q}
$$

$$
= O\Big(\big(n\, P_X(I)^{1-1/(2+\epsilon)}\big)^{-q/2} + n\,(n\, P_X(I))^{-q}\Big) = O\big(n^{-\kappa}\big),
$$

if q is chosen sufficiently large. $\qquad\square$

Lemma 4.3 *Suppose that $(X_t)_{t \in \mathbb{N}_0}$ is a Markov chain with state space $D \subseteq \mathbb{R}$ and stationary distribution P_X such that (A2) is fulfilled. Then there exist some $C < \infty$ such that, for arbitrary $\underline{x} \leq \overline{x}$ with $P_X([\underline{x}, \overline{x}]) \geq n^{-1/3}$,*

$$E\left[\sup_{v:\, v \geq \overline{x}} \frac{\left| \sum_{t=1}^n \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(X_{t-1} \in [\underline{x}, v]) \right|}{\#\{t \leq n:\, X_{t-1} \in [\underline{x}, v]\} \vee 1} \right] \leq C n^{-1/3}$$

(4.8a)

and

$$E\left[\sup_{u:\, u \leq \underline{x}} \frac{\left| \sum_{t=1}^n \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(X_{t-1} \in [u, \overline{x}]) \right|}{\#\{t \leq n:\, X_{t-1} \in [u, \overline{x}]\} \vee 1} \right] \leq C n^{-1/3}.$$

(4.8b)

Proof We prove only (4.8a) since the proof of (4.8b) is completely analogous. The proof is carried out in two steps. First we consider the technically simpler case where the distribution function F_X is continuous. This allows us to define a suitable dyadic family of intervals which leads to a readily comprehensible proof. Afterwards we extend the result to the general case.

Step 1 Suppose that F_X is continuous. First we prove that for arbitrary $\delta > 0$ and each $v \geq \overline{x}$ there exists some $C < \infty$ such that

$$E\left[\sup_{x:\, \underline{x} \leq x \leq v} \left| \sum_{t=1}^n \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(X_{t-1} \in [\underline{x}, x]) \right| \right]$$
$$\leq C \sqrt{n\, P_X([\underline{x}, v])} + n^\delta.$$

(4.9)

To deal with the supremum we define a suitable system of dyadic intervals. Let $J_n \in \mathbb{N}$ be such that $n^{\delta-1}/2 < 2^{-J_n} P_X([\underline{x}, v]) \leq n^{\delta-1}$. For $j = 1, 2, \ldots, J_n$ and $k = 1, 2, \ldots, 2^j$, we set

$$x_{j,k} = F_X^{-1}\left(F_X(\underline{x}) + k 2^{-j} P_X([\underline{x}, v]) \right)$$

(4.10a)

and, for $j = 1, \ldots, J_n$,

$$B_{j,k} = \begin{cases} [\underline{x}, x_{j,1}] & \text{if } k = 1, \\ (x_{j,k-1}, x_{j,k}] & \text{if } k = 2, \ldots, 2^j. \end{cases}$$

(4.10b)

We have that

$$P_X(B_{j,k}) = 2^{-j} P_X([\underline{x}, v]).$$

(4.11)

We define partial sume as

$$S_{j,k} = \sum_{t=1}^{n} \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(X_{t-1} \in B_{j,k}).$$

Choosing γ such that $(1 - 2\delta)/(1 - \delta) \leq \gamma < 1$ we have that $n(2^{-j} P_X([\underline{x}, v]))^\gamma = O\big((n2^{-j} P_X([\underline{x}, v]))^2\big)$ for all $j = 1, \ldots, J_n$. Hence, the first term in the bound given in Lemma 4.1 dominates the second and we obtain, for $j = 1, \ldots, J_n,\ k = 1, \ldots, 2^j,$

$$E\big[S_{j,k}^4\big] = O\Big((n2^{-j} P_X([\underline{x}, v]))^2\Big),$$

which implies that

$$E\left[|S_{j,k}| \, \mathbb{1}\big(|S_{j,k}| > \sqrt{n P_X([\underline{x}, v])}\, 2^{-j/4}\big) \right]$$
$$\leq \frac{E\big[S_{j,k}^4\big]}{\big(\sqrt{n P_X([\underline{x}, v])}\, 2^{-j/4}\big)^3} = O\Big(\sqrt{n P_X([\underline{x}, v])}\, 2^{-5j/4}\Big).$$

Therefore, we obtain that

$$E\left[\max_{1 \leq k \leq 2^j} |S_{j,k}| \right]$$
$$\leq \sqrt{n P_X([\underline{x}, v])}\, 2^{-j/4} + \sum_{k=1}^{2^j} E\left[|S_{j,k}| \, \mathbb{1}\big(|S_{j,k}| > \sqrt{n P_X([\underline{x}, v])}\, 2^{-j/4}\big) \right]$$
$$= O\Big(\sqrt{n P_X([\underline{x}, v])}\, 2^{-j/4}\Big). \tag{4.12}$$

At the finest scale J_n, we define for $k = 1, \ldots, J_n,$

$$N_{J_n,k} = \#\big\{t \leq n: \ X_{t-1} \in B_{J_n,k}\big\}.$$

Note that $E N_{J_n,k} = n2^{-J_n} P_X([\underline{x}, v]) \leq n^\delta$. We obtain from (4.7) that

$$E\big[N_{J_n,k}\, \mathbb{1}(N_{J_n,k} > 2n^\delta)\big]$$
$$\leq E\left[|N_{J_n,k} - E N_{J_n,k}| \, \mathbb{1}\big(|N_{J_n,k} - E N_{J_n,k}| > n^\delta\big) \right]$$
$$\leq \frac{C_q \big\{ n^{q/2} \, (2^{-J_n} P_X([\underline{x}, v]))^{q/(2+\epsilon)} + n \, (2^{-J_n} P_X([\underline{x}, v]))^{q/(q+\epsilon)} \big\}}{n^{\delta(q-1)}}$$
$$= O\big(n^{-\kappa}\big)$$

holds for arbitrary $\kappa < \infty$ if q is chosen large enough. Since $2^{J_n} < n^{1-\delta}/(2P_X([\underline{x}, v]))$ $\leq n^{4/3-\delta}/2$ we obtain

$$E\left[\max_{1\leq k\leq 2^{J_n}}\{N_{J_n,k}\}\right] = 2n^{\delta} + \sum_{k=1}^{2^{J_n}} E\left[N_{J_n,k}\,\mathbb{1}(N_{J_n,k} > 2n^{\delta})\right] = O\left(n^{\delta}\right). \quad (4.13)$$

After these preparatory steps we are in a position to estimate the expected value of the supremum. For arbitrary $x \in [\underline{x}, v]$, there exist p and $(j_1, k_1), \ldots, (j_p, k_p)$, k, $1 \leq j_1 < \cdots < j_p \leq J_n$, such that $B_{j_1,k_1}, \ldots, B_{j_p,k_p}, B_{J_n,k}$ are adjacent intervals and

$$B_{j_1,k_1} \cup \cdots \cup B_{j_p,k_p} \subseteq [\underline{x}, x] \subseteq B_{j_1,k_1} \cup \cdots \cup B_{j_p,k_p} \cup B_{J_n,k}.$$

This implies that

$$\left|\sum_{t=1}^{n}\left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1})\right]\mathbb{1}\left(X_{t-1} \in [\underline{x}, x]\right)\right| \leq \sum_{l=1}^{p}\left|S_{j_l,k_l}\right| + N_{J_n,k}$$

and, therefore,

$$E\left[\sup_{x:\ \underline{x}\leq x\leq v}\left|\sum_{t=1}^{n}\left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1})\right]\mathbb{1}\left(X_{t-1} \in [\underline{x}, x]\right)\right|\right]$$
$$= \sum_{j=1}^{J_n} E\left[\max_{1\leq k\leq 2^j}\{|S_{j,k}|\}\right] + E\left[\max_{1\leq k\leq 2^{J_n}}\{N_{J_n,k}\}\right].$$

It follows from (4.12) and (4.13) that (4.9) is fulfilled.

Now we are in a position to prove (4.8a). We define a dyadic sequence of growing intervals, $I_0 = [\underline{x}, \overline{x}]$ and, for $j \geq 1$

$$I_j = \left[\underline{x}, F_X^{-1}\left(F_X(\underline{x}) + 2^j P_X([\underline{x}, \overline{x}])\right)\right].$$

(There exists some $K_n \geq 0$ such that $P_X(I_j) = 2^j P_X([\underline{x}, \overline{x}])$ for $j = 0, \ldots, K_n$ and $P_X(I_{K_n+1}) < 2^{K_n+1} P_X(\underline{x}, \overline{x}))$. Then $I_{K_n+1} = I_{K_n+2} = \ldots$.) Define the event

$$A_n = \left\{\{\omega:\ \#\{t \leq n:\ X_{t-1}(\omega) \in I_j\} \geq n\,P_X(I_j)/2\ \text{ for all } j = 0, \ldots, K_n\}\right\}.$$

For $x \in I_{j+1} \setminus I_j$ we use the estimate

$$\frac{\left|\sum_{t=1}^{n}\left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1})\right]\mathbb{1}\left(X_{t-1} \in [\underline{x}, x]\right)\right|}{\#\{t \leq n:\ X_{t-1} \in [\underline{x}, x]\} \vee 1}$$
$$\leq \frac{\sup_{x\in I_{j+1}}\left|\sum_{t=1}^{n}\left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1})\right]\mathbb{1}\left(X_{t-1} \in [\underline{x}, x]\right)\right|}{\#\{t \leq n:\ X_{t-1} \in I_j\} \vee 1}.$$

It follows from Lemma 4.3 that $P\left(A_n^c\right) = O\left(n^{-1/3}\right)$, which implies that

$$
E\left[\sup_{x:\ \underline{x}\leq x\leq v}\frac{\left|\sum_{t=1}^n\left[\mathbb{1}(X_t\leq z)-P(X_t\leq z\mid X_{t-1})\right]\mathbb{1}\left(X_{t-1}\in[\underline{x},x]\right)\right|}{\#\{t\leq n\colon\ X_{t-1}\in[\underline{x},x]\}\vee 1}\right]
$$

$$
\leq E\left[\sup_{x:\ \underline{x}\leq x\leq v}\frac{\left|\sum_{t=1}^n\left[\mathbb{1}(X_t\leq z)-P(X_t\leq z\mid X_{t-1})\right]\mathbb{1}\left(X_{t-1}\in[\underline{x},x]\right)\right|}{\#\{t\leq n\colon\ X_{t-1}\in[\underline{x},x]\}}\ \mathbb{1}_{A_n}\right]
$$

$$
+ P\left(A_n^c\right)
$$

$$
= \sum_{j=0}^{K_n}\frac{E\left[\left|\sum_{t=1}^n\left[\mathbb{1}(X_t\leq z)-P(X_t\leq z\mid X_{t-1})\right]\mathbb{1}\left(X_{t-1}\in I_{j+1}\right)\right|\right]}{n\,P_X(I_{j-1})/2}+O\left(n^{-1/3}\right)
$$

$$
= O\left(\sum_{j=0}^{K_n}2^{-j/2}/\sqrt{n\,P_X([\underline{x},\overline{x}])}\right)
$$

$$
+ O\left(n^{-1/3}\right) = O\left(n^{-1/3}\right), \tag{4.14}
$$

i.e. (4.8a) is fulfilled.

Step 2 In case of a general stationary distribution P_X, the definition according to (4.10a) and (4.10b) does no longer guarantee that the convenient property of $P_X(B_{j,k}) = 2^{-j}P_X([\underline{x},v])$ holds true. In order to draw on the calculations in Step 1 we act as follows. Let $(V_t)_{t\in\mathbb{N}_0}$ be a sequence of independent random variables following a uniform distribution on $[0,1]$, which is independent of the process $(X_t)_{t\in\mathbb{N}_0}$. For the latter process we define an accompanying sequence $(U_t)_{t\in\mathbb{N}_0}$ of uniformly distributed random variables, where U_t depends on the pair (X_t, V_t) as follows. If F_X is continuous in the point X_t, then we simply set

$$
U_t := F_X(X_t).
$$

Otherwise, if F_X is discontinuous in X_t, then $P_X(\{X_t\}) = F_X(X_t) - F_X(X_t-0) > 0$ and we set

$$
U_t := F_X(X_t) - V_t/P_X(\{X_t\}).
$$

In both cases we have that

$$
X_t = F_X^{-1}(U_t),
$$

where $G^{-1}(t) = \inf\{x\colon G(x)\geq t\}$ denotes the generalized inverse of a generic distribution function G. Since F_X has at most countably many discontinuity points, it follows that the mapping $(X_t, V_t)\mapsto U_t$ is measurable. It also follows that U_t has a uniform distribution on $[0,1]$. Furthermore the process $\left((X_t, V_t)\right)_{t\in\mathbb{N}_0}$ has the same mixing properties as $(X_t)_{t\in\mathbb{N}_0}$, i.e.

$$
\alpha_{(X,V)}(r) = \alpha_X(r)\qquad \forall r\geq 1;
$$

see e.g. Lemma 8 in Bradley (1981). Now we obtain in complete analogy to the calculations leading to (4.9) in Step 1 that, for arbitrary $0 \leq \underline{u} < \overline{u} \leq 1$,

$$E\left[\sup_{u:\ \underline{u} \leq u \leq \overline{u}} \left| \sum_{t=1}^{n} \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(U_{t-1} \in [\underline{u}, u]) \right| \right]$$
$$\leq C\sqrt{n(\overline{u} - \underline{u})} + n^{\delta}. \tag{4.15}$$

It is easy to see that the following inclusions hold true for $\underline{x} \leq x$:

$$\left\{ F_X(\underline{x} - 0) < U_{t-1} \leq F_X(x) \right\} \subseteq \left\{ \underline{x} \leq X_{t-1} \leq x \right\}$$
$$\subseteq \left\{ F_X(\underline{x} - 0) \leq U_{t-1} \leq F_X(x) \right\}.$$

Indeed, the second inclusion follows immediately from the construction of U_{t-1}. Regarding the first one, note that it follows again from the construction of U_{t-1} that $F_X(\underline{x} - 0) < U_{t-1}$ implies $\underline{x} \leq X_{t-1}$. Furthermore, $U_{t-1} \leq F_X(x)$ implies $X_{t-1} = F_X^{-1}(U_{t-1}) \leq F_X^{-1}(F_X(x)) \leq x$. Since $P(U_{t-1} = F_X(\underline{x})) = 0$ we conclude that

$$\sup_{x:\ \underline{x} \leq x \leq \overline{x}} \left| \sum_{t=1}^{n} \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(X_{t-1} \in [\underline{x}, x]) \right|$$
$$\leq \sup_{u:\ F_X(\underline{x}-0) \leq u \leq F_X(\overline{x})} \left| \sum_{t=1}^{n} \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \right.$$
$$\left. \mathbb{1}(U_{t-1} \in [F_X(\underline{x} - 0), F_X(x)]) \right|$$

holds with probability one. Hence, we obtain from (4.13)

$$E\left[\sup_{x:\ \underline{x} \leq x \leq v} \left| \sum_{t=1}^{n} \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \mathbb{1}(X_{t-1} \in [\underline{x}, x]) \right| \right]$$
$$\leq E\left[\sup_{u:\ F_X(\underline{x}-0) \leq u \leq F_X(v)} \left| \sum_{t=1}^{n} \left[\mathbb{1}(X_t \leq z) - P(X_t \leq z \mid X_{t-1}) \right] \right. \right.$$
$$\left. \left. \mathbb{1}(U_{t-1} \in [F_X(\underline{x} - 0), F_X(x)]) \right| \right]$$
$$\leq C\sqrt{n\left(F_X(v) - F_X(\underline{x} - 0)\right)} + n^{\delta}$$
$$= C\sqrt{n\, P_X([\underline{x}, v])} + n^{\delta},$$

i.e. (4.9) holds true. $\qquad\square$

References

Al-Osh M, Alzaid A (1987) First-order integer-valued autoregressive (INAR(1)) processes. J Time Ser Anal 8(3):261–275

Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E (1955) An empirical distribution function for sampling with incomplete information. Ann Math Stat 26(4):641–647

Bickel PJ, Freedman DA (1981) Some asymptotic theory for the bootstrap. Ann Stat 9(6):1196–1217

Bradley RC (1981) Central limit theorems under weak dependence. J Multivar Anal 11:1–16

Bradley RC (2007a) Introduction to strong mixing conditions, vol I. Kendrick Press, Heber City

Bradley RC (2007b) Introduction to strong mixing conditions, vol II. Kendrick Press, Heber City

Brunk HD (1955) Maximum likelihood estimates of monotone parameters. Ann Math Stat 26(4):607–616

Canonne CL (2017) A short note on Poisson tail bounds. http://www.cs.columbia.edu/ccanonne/files/misc/2017-poissonconcentration.pdf. Accessed 20 April 2022

Dehling H, Mikosch T (1994) Random quadratic forms and the bootstrap for U-statistics. J Multivar Anal 51:392–413

Deng H, Zhang C-H (2020) Isotonic regression in multi-dimensional spaces and graphs. Ann Stat 48(6):3672–3698

Doukhan P (1994) Mixing: properties and examples. Lecture notes in statistics, vol 85. Springer, Berlin

Durot C (2002) Sharp asymptotics for isotonic regression. Probab Theory Relat Fields 122:222–240

Freedman DA (1981) Bootstrapping regression models. Ann Stat 9(6):1218–1228

Leucht A, Neumann MH (2009) Consistency of general bootstrap methods for degenerate U-and V-type statistics. J Multivar Anal 100:1622–1633

Leucht A, Neumann MH (2013) Dependent wild bootstrap for degenerate U- and V-statistics. J Multivar Anal 117:257–280

Leucht A, Neumann MH, Kreiss J-P (2015) A model specification test for GARCH(1,1) processes. Scand J Stat 42:1167–1193

Lindvall T (1992) Lectures on the coupling method. Wiley, New York

McKenzie E (1985) Some simple models for discrete variate time series. Water Resour Bull 21(4):645–650

Mösching A, Dümbgen L (2020) Monotone least squares and isotonic quantiles. Electron J Stat 14:24–49

Neumann MH (2021) Bootstrap for integer-valued GARCH(p,q) processes. Stat Neerl 75(3):343–363

Pakes AG (1971) Branching processes with immigration. J Appl Probab 8(1):32–42

Paparoditis E, Politis DN (2002) The local bootstrap for Markov processes. J Stat Plan Inference 108:301–328

Rajarshi MB (1990) Bootstrap in Markov-sequences based on estimates of transition density. Ann Inst Stat Math 42:253–268

Robertson T, Wright FT, Dykstra RL (1988) Order restricted statistical inference. Wiley, New York

van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge

Zhang C-H (2002) Risk bounds in isotonic regression. Ann Stat 30(2):528–555