

# Distributional Conformal Prediction for Markov Process

Kejin Wu

Department of Mathematics and Statistics  
Loyola University Chicago

Joint work with

Dehao Dai	Dimitris Politis
University of California San Diego	University of California San Diego

# Estimation vs. Prediction

---

	Estimation	Prediction
Target	Parameter $\theta$	Future value $Y_f$
Inference	Point Estimator $\hat{\theta}$ Confidence Interval (CI)	Point Predictor $\widehat{Y}_f$ Prediction Interval (PI)
Consistency	$\hat{\theta} \xrightarrow{P} \theta$	$\widehat{Y}_f   \mathbf{X} = \mathbf{x} \xrightarrow{P} \mathbb{E}(Y_f   \mathbf{X} = \mathbf{x})$
Asymptotic Validity	$\text{CVR} \rightarrow 1 - \alpha$ as $n \rightarrow \infty$	$\text{CVR} \rightarrow 1 - \alpha$ as $n \rightarrow \infty$
Finite Sample Validity	$\text{CVR} \geq 1 - \alpha$ for $\forall n$	$\text{CVR} \geq 1 - \alpha$ for $\forall n$

CVR: coverage of the interval.

# Naive Example

---

Classical linear regression problem with a fixed design:

- Assumptions:

- Linearity:  $\mathbb{E}(Y|X = \mathbf{x}) = \mathbf{x}^\top \beta^*$ .
- Independent Observations.
- Normality:  $\varepsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ .

- Output:

- $L_2$  point prediction of  $Y$  given  $x_f$ , i.e.,  $\widehat{Y}_f$ : consistent.
- Prediction interval (PI) for  $\widehat{Y}_f$ : valid with finite sample.
- Confidence interval (CI) for  $\widehat{Y}_f$ : valid with finite sample.

- Limits:

The coverage of the PI/CI may not be equal to  $1 - \alpha$  even asymptotically, if

- (i) Model is misspecified: nonlinear, etc.
- (ii)  $\varepsilon_i$  is not normal.
- (iii) Observations are dependent.

- Predictions for dependent data without constraint model assumptions.
  - Model-free Prediction Principle (Politis, 2015).
  - Distributional Conformal Prediction (Chernozhukov et al., 2021).
- Non-asymptotic coverage error bound of (unconditional) PI.
- Asymptotic validity of (conditional) PI.

# Conformal Prediction

---

Suppose a sequence of *i.i.d.*  $Y_t \in \mathbb{R}, t = 1, \dots, n$  without covariates.

- One-sided Quantile PI:  $(-\infty, \hat{q}_n]$ .
  - $\hat{q}_n$  is the  $1 - \alpha$  empirical quantile of  $\{Y_i\}_{i=1}^n$ .
  - Asymptotic validity: assuming the empirical quantile converges to the population quantile.
- Adjusted One-sided Conformal PI:  $(-\infty, \hat{q}_n]$ .
  - $\hat{q}_n = \lceil (1 - \alpha)(n + 1) \rceil$  smallest of  $Y_1, \dots, Y_n$ .
  - Finite sample validity,  $\mathbb{P}(Y_{n+1} \leq \hat{q}_n) \geq 1 - \alpha$ .
  - Only Exchangeability is needed.

# Full Conformal Prediction in Regression

---

- Augmented dataset  $\{(X_t, Y_t)\}_{t=1}^n \cup (\mathbf{x}, y)$ ;  $X_f = \mathbf{x}$ .
- Symmetric predictive model  $\hat{f}_y(\cdot)$ .
- Conformal scores: measure how well the future  $y$  conforms to the patterns of data

$$R_t^{(\mathbf{x}, y)} = |Y_t - \hat{f}_y(\mathbf{X}_t)|, \quad t = 1, \dots, n,$$

$$R_{n+1}^{(\mathbf{x}, y)} = |y - \hat{f}_y(\mathbf{x})|.$$

- Exchangeability:  $R_i^{(\mathbf{x}, y)}, i = 1, \dots, n + 1$  are exchangeable
- Conformal  $1 - \alpha$  PI:

$$\text{CPI}_n = \left\{ y : R_{n+1}^{(\mathbf{x}, y)} \leq \lceil (1 - \alpha)(n + 1) \rceil \text{ smallest of } R_1^{(\mathbf{x}, y)}, \dots, R_n^{(\mathbf{x}, y)} \right\}.$$

- Equivalently:

$$\text{CPI}_n = \{y : \underbrace{\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{R_i^{(x,y)} > R_{n+1}^{(x,y)}\}}_{p(y)} \geq \frac{\lfloor \alpha(n+1) \rfloor}{n}\}$$

- Split Conformal Prediction is another type of conformal prediction by separating the whole data as (1) training set; (2) calibration set to satisfy the exchangeability.
- Other conformal scores can be taken, e.g., scaled residual score, quantile score.

- Residual-type conformal score is widely used, also for predicting dependent data.
- For the residual-type conformal score, the better the predictive model, the tighter the conformal prediction interval (Lei et al., 2018).

It is NOT Model-free.

# Model-free Prediction Principle

---

Main idea (Politis (2015)): For a data vector  $\underline{Y}_n := (Y_1, \dots, Y_n)'$  with possible covariates  $\underline{X}_n := (\mathbf{X}_1, \dots, \mathbf{X}_n)$ :

1. Find an invertible transformation function  $H_n$  which transforms  $\underline{Y}_n$  to *i.i.d.* vector  $(V_1, \dots, V_n) \stackrel{i.i.d.}{\sim} F_V$  with possible  $\underline{X}_n$ , s.t.,  $Y_n$  can be solved in terms of  $\underline{Y}_{n-1} := (Y_1, \dots, Y_{n-1})$ ,  $\mathbf{X}_n$  and  $V_n$ , i.e.,  $Y_n = H_n^{-1}(\underline{Y}_{n-1}, \mathbf{X}_n, V_n)$ ;
2. Determine the future response  $Y_f := H_n^{-1}(\underline{Y}_n, \mathbf{X}_f, V_f)$ , where  $V_f \sim F_V$  is independent with  $Y_f$ ,  $\mathbf{X}_f$  and  $(V_1, \dots, V_n)$ ;
3. Approximate the whole distribution of  $Y_f$  by Monte Carlo ( $F_V$  is known) or Bootstrap ( $F_V$  is estimated).

# Model-free Prediction with Markov Data

---

Assuming  $\underline{Y}_n$  is stationary Markov( $p$ ) and define  $\underline{X}_{t-1}^{(i)} = (Y_{t-1}, \dots, Y_{t-i})'$ :

1. Convert Markov( $p$ )  $\underline{Y}_n$  to *i.i.d*  $V_1, \dots, V_n$  by Probability Integral Transformation (PIT):

$$V_1 = F(Y_1); V_2 = F_1(Y_2|\underline{X}_1^{(1)}); \dots; V_p = F_{p-1}(Y_p|\underline{X}_{p-1}^{(p-1)})$$

$$\text{and } V_t = F(Y_t|\underline{X}_{t-1}^{(p)}) \text{ for } t = p+1, p+2, \dots, n.$$

Remark:

- We ignore  $V_1, \dots, V_p$  since they are kind of initial conditions.
- $V_{p+1}, \dots, V_n \stackrel{i.i.d}{\sim}$  Uniform(0, 1) which is a direct result of Rosenblatt's transformation with Markov( $p$ ) data (Rosenblatt, 1952).

2. The one-step ahead pseudo values  $Y_{n+1}^*$  can be generated by  $F^{-1}(V_{n+1}^* | \underline{X}_n^{(p)})$ , where  $V_{n+1}^*$  is simulated from  $\text{Uniform}(0, 1)$  or bootstrapped from  $\{V_{p+1}, \dots, V_n\}$ .
3. Repeat Step 2 enough times, then the conditional distribution of  $Y_{n+1}$  given  $\underline{X}_n^{(p)}$  can be approximated, resulting (1)  $L_2$  or  $L_1$  point predictions; (2) Quantile PI and (3) PI centered at a point predictor.

Remark:

- In Step 2, if  $V_{n+1}^*$  is simulated from  $\text{Uniform}(0, 1)$ , it is the so-called Limit Model-free prediction, closely related to the deep generative methods, e.g., (Zhou et al., 2023).

## Kernel-based smoothed conditional CDF

---

Create the pair data  $\{(\underline{X}_{t-1}^{(p)}, Y_t)\}_{t=p+1}^n$ , the conditional CDF of  $Y$  given  $\underline{x}$  can be estimated by:

$$\widehat{F}(y|\underline{x}) = \frac{\frac{1}{n-p} \sum_{t=p+1}^n W_h(\underline{X}_{t-1}^{(p)}, \underline{x}) K\left(\frac{y-Y_t}{h_0}\right)}{\overline{W}_h(\underline{x})};$$

where

- $W_h(\underline{X}_{t-1}^{(p)}, \underline{x}) = \prod_{s=1}^p \frac{1}{h_s} w\left(\frac{X_{t,s} - x_s}{h_s}\right)$ .
- $\overline{W}_h(\underline{x}) = \frac{1}{n-p} \sum_{t=p+1}^n W_h(\underline{X}_{t-1}^{(p)}, \underline{x})$ .
- $w(\cdot)$ : univariate, symmetric kernel density function .
- $X_{t,s}$  and  $x_s$  are the  $s$ -th coordinate of  $\underline{X}_{t-1}^{(p)}$  and  $\underline{x}$ .
- $K$  is a smooth CDF that is strictly increasing.
- $h_1, \dots, h_p$  and  $h_0$  are bandwidths.

Define

$$\widehat{V}_t = \widehat{F}(Y_t | \underline{X}_{t-1}^{(p)}) \text{ for } t = p + 1, p + 2, \dots, n;$$

$\{\widehat{V}_t\}_{t=p+1}^n$  are approximately in asymptotic sense i.i.d. Uniform(0, 1). Thus,

- Model-free Prediction can work with  $\widehat{F}(\cdot|\cdot)$  asymptotically.
- Model-free Prediction can work with leave-one-out kernel estimator  $\widehat{F}_t(\cdot|\cdot)$  asymptotically.
- Pertinent Prediction Interval can further be made to capture the estimation variability in  $\widehat{F}$  (Pan and Politis, 2016; Wang and Politis, 2026).
- Conformal prediction is feasible with  $\{\widehat{V}_t\}_{t=p+1}^n$ , resulting so-called *Distributional Conformal Prediction* (Chernozhukov et al., 2021).

# Distributional CP for Markov process (MDCP)

---

- Determine a  $\mathcal{Y}_{\text{trail}}$  to be the candidate set of future values.
- Conformal scores:

$$\widehat{V}_t^{(y)} = \begin{cases} \widehat{F}^{(y)}(Y_t | \underline{X}_{t-1}^{(p)}) & \text{if } p+1 \leq t \leq n; \\ \widehat{F}^{(y)}(y | \underline{X}_{t-1}^{(p)}) & \text{if } t = n+1 \end{cases};$$

where  $\widehat{F}^{(y)}$  is kernel-based smoothed CDF on  $\{(\underline{X}_{t-1}^{(p)}, Y_t)\}_{t=p+1}^n \cup (\underline{X}_n^{(p)}, y)$  and  $y$  is a candidate from  $\mathcal{Y}_{\text{trail}}$ .

- Ranks ( $p$ -value)

$$\hat{p}(y) = \frac{1}{n-p+1} \sum_{t=p+1}^{n+1} \mathbb{1}\{\widehat{U}_t^{(y)} \geq \widehat{U}_{n+1}^{(y)}\};$$

where  $\widehat{U}_t^{(y)} := |\widehat{V}_t^{(y)} - 1/2|$ .

- Distributional Conformal  $(1 - \alpha)$ -PI:

$$\widehat{C}_{1-\alpha}^{\text{MDCP}}(\underline{X}_n^{(p)}) = \{y \in \mathcal{Y}_{\text{trial}} : \widehat{p}(y) > \alpha\}.$$

Remark:

- $\mathcal{Y}_{\text{trial}}$  is chosen to be a fine grid between  $-\max_{1 \leq t \leq n} |Y_t|$  and  $\max_{1 \leq t \leq n} |Y_t|$ .
- Conformal scores can be calculated by leave-one-out kernel estimator  $\widehat{F}_t^{(y)}(\cdot|\cdot)$  :

$$\widetilde{U}_t^{(y)} := \widehat{F}_t^{(Y_{n+1})}(Y_t | \underline{X}_{t-1}^{(p)}) = \frac{\frac{1}{n-p+1} \sum_{i=p+1, i \neq t}^{n+1} W_h(\underline{X}_{i-1}^{(p)}, \underline{X}_{t-1}^{(p)}) K\left(\frac{Y_t - Y_i}{h_0}\right)}{\frac{1}{n-p+1} \sum_{i=p+1, i \neq t}^{n+1} W_h(\underline{X}_{i-1}^{(p)}, \underline{X}_{t-1}^{(p)})}$$

resulting in the so-called Predictive Distributional Conformal prediction for Markov (PMDCP).

# Assumptions

---

- A1  $\{Y_t\}_{t \geq 1}$  forms a strictly stationary and geometrically ergodic Markov process of order  $p$ .
- A2 (i) The marginal density  $f_{\underline{x}}$  is positive for all  $\underline{x}$  and has continuous second-order derivatives;  
(ii)  $F(y|\underline{x})$  have continuous second-order derivatives and  $f(y|\underline{x})$  is positive for all  $y$  and  $\underline{x}$ ;  
(iii)  $F(y|\underline{x})$  is Lipschitz continuous in  $y$  for all  $\underline{x}$ .
- A3 let  $N = n - p + 1$ ,  $h_0 \asymp h \rightarrow 0$ ,  $Nh^p \rightarrow \infty$  and  $ph^2 \rightarrow 0$  when  $n \rightarrow \infty$ .
- A4 The domains of  $Y_t$  and  $\underline{X}_{t-1}^{(p)}$  are bounded, i.e.,  $(Y_t, \underline{X}_{t-1}^{(p)}) \in \mathcal{X} := [-M, M]^{p+1}$  with  $M$ , an arbitrarily large constant.
- A5  $w(\cdot)$  and  $K(\cdot)$  are positive, differentiable, and bounded on their whole domains with  
(i)  $\int w(v)dv = 1$ ,  $\int vw(v)dv = 0$  and  $\int v^2w(v)dv = C_w < \infty$ ;  
(ii)  $\int K'(z)dz = 1$ ,  $\int zK'(z)dz = 0$  and  $\int z^2K'(z)dz = C_K < \infty$ .

# Unconditional Coverage

---

Theorem (non-asymptotic unconditional coverage error)

Given  $\alpha \in (0, 1)$ , under assumptions A1-A5, taking  $C_\delta = O\left(\frac{\log N}{\sqrt{N}h^p} + h^2\right)$ , we have

$$|\mathbb{P}(Y_{n+1} \in \widehat{C}_{1-\alpha}^{MDCP}(\underline{X}_n^{(p)})) - (1 - \alpha)| \leq 24C_\delta + 4 \exp(-8NC_\delta^2) + 2S_N;$$

where  $S_N$  is the appropriate sequence which converges to 0 as  $n \rightarrow \infty$ .

Remark:

- $C_\delta$  measures the approximation error between  $\widehat{F}(\cdot|\cdot)$  and  $F(\cdot|\cdot)$ .
- $S_N$  term is related to the high probability condition of the convergence of  $\widehat{F}(\cdot|\cdot)$  to  $F(\cdot|\cdot)$ .

## Theorem (Conditional coverage guarantees I)

Given  $\alpha \in (0, 1)$ , if assumptions A1-A5 hold, the MDCP prediction interval is asymptotically valid, i.e.,

$$|\mathbb{P}(Y_{n+1} \in \widehat{C}_{1-\alpha}^{MDCP}(\underline{X}_n^{(p)}) | \underline{X}_n^{(p)}) - (1 - \alpha)| \xrightarrow{P} 0 \text{ for any } \underline{X}_n^{(p)}.$$

- Geometrically ergodic (Geometric  $\beta$ -mixing)  $\{Y_t\}_{t \geq 1}$  does not cover all types of Markov processes, e.g., the AR(1) process  $Y_t = \rho Y_{t-1} + \varepsilon_t$ ;  $0 \leq \rho \leq 1/2$ , with independent Bernoulli innovations, does not have the  $\alpha$ -mixing property (Andrews, 1984).
- $L^q$ - $m$ -approximable series  $Y_t$ , s.t.,  $\exists L^q$ - $m$ -approximator  $Y_t^{(m)}$ : (Hörmann and Kokoszka, 2010)
  - $Y_t = f(\varepsilon_t, \varepsilon_{t-1}, \dots)$ , for each  $t$ ; where the  $\varepsilon_i$  are i.i.d. in a real space.
  - $Y_t^{(m)} = f(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-m+1}, \varepsilon_{t-m}^{(t)}, \varepsilon_{t-m-1}^{(t)}, \dots)$ ; where  $\{\varepsilon_k^{(t)}\}$  of  $\{\varepsilon_k\}$  for each  $t$ .
  - $\sum_{m=1}^{\infty} (\mathbb{E} |Y_m - Y_m^{(m)}|^q)^{1/q} < \infty$ .

## Conditional Coverage for $L^2$ - $m$ -approximable $\{Y_t\}_{t \geq 1}$

A1'  $\{Y_t\}$  is an  $L^2$ - $m$ -approximable series with  $m$  is larger than the Markov order  $p$ .

Also, we denote  $\sum_{m=1}^{\infty} \left( \mathbb{E} \left| Y_m - Y_m^{(m)} \right|^2 \right)^{1/2} = \sum_{m=1}^{\infty} \delta(m) < \infty$ . We require that  $\delta(m)$  converges to 0 as  $m \rightarrow \infty$  s.t.,  $\delta(m) = o(h^{p+1})$  and  $m = o(n)$ .

### Theorem (Asymptotically conditional coverage guarantees II)

*Under A1' and A2 to A5, given  $\alpha \in (0, 1)$ , taking  $\{Y_t^{(m)}\}$  as the  $L^2$ - $m$ -approximator of the original series with the latest  $p$  values replaced with the values from the original series, we have*

$$\left| \mathbb{P} \left( Y_{n+1} \in \widehat{C}_{1-\alpha}^{MDCP}(\mathbf{X}_n) \mid \mathbf{X}_n \right) - (1 - \alpha) \right| \xrightarrow{P} 0.$$

# Simulation studies

---

Setup:

- $R = 1000$  datasets.
- $p = 1$ .
- error  $\epsilon_t \sim N(0, 1)$  or Laplace with unit variance.
- $S = 5000$  pseudo values of  $Y_{n+1}$  used to evaluate PIs.
- 

$$CVR = \frac{1}{R} \sum_{i=1}^R CVR_i \text{ and } LEN = \frac{1}{R} \sum_{i=1}^R LEN_i;$$

where

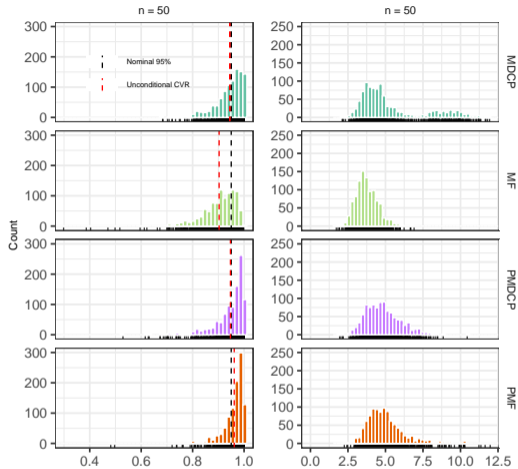
$$CVR_i = \frac{1}{S} \sum_{j=1}^S \mathbb{1}\{Y_{n+1,j} \in [L_i, U_i]\} \text{ and } LEN_i = U_i - L_i.$$

# Simulation Studies

- Model 2:  $Y_{t+1} = 0.8 \log(3Y_t^2 + 1) + \epsilon_{t+1}$ ;  $\epsilon_{t+1}$  are normal.

	Nominal coverage 90%				Nominal coverage 95%			
	CVR	LEN	CVR Sd	LEN Sd	CVR	LEN	CVR Sd	LEN Sd
$n = 50$								
MDCP	0.881	3.822	0.074	1.624	0.945	5.358	0.053	2.200
PMDCP	0.886	3.607	0.087	0.705	0.947	4.822	0.061	1.141
MF	0.852	3.234	0.091	0.639	0.903	3.765	0.076	0.733
PMF	0.922	4.023	0.065	0.797	0.962	5.066	0.046	1.396
$n = 250$								
MDCP	0.887	3.466	0.042	1.240	0.945	4.360	0.029	1.593
PMDCP	0.887	3.319	0.056	0.396	0.945	4.100	0.048	0.520
MF	0.873	3.197	0.055	0.397	0.923	3.742	0.046	0.455
PMF	0.900	3.466	0.051	0.435	0.948	4.192	0.042	0.579

Model-2 with normal errors,  $n = 50$



Model-2 with normal errors,  $n = 250$

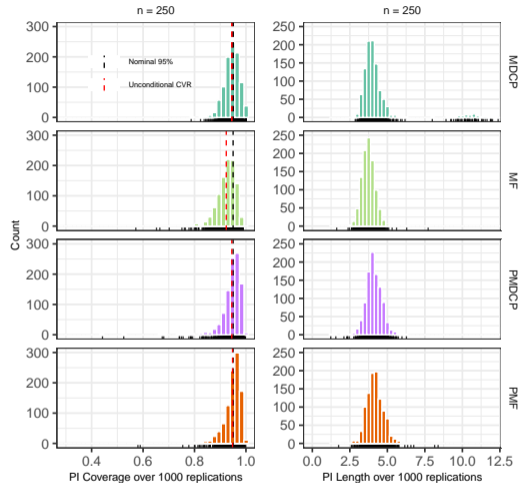


Figure 1: Conditional coverage and PI length for Model-2 simulation with normal error.

- We take the log returns of the weekly S&P 500 price index from 1 January 1988 to 31 December 1997, i.e., in total 521 observations. Revealed by Chen and Hong (2012), this returns series can be seen as a Markov(1) data.
- Rolling-window predictions are taken with a window size of  $\lceil 0.7n \rceil$ .
- The average coverage rate and average length of PI for rolling prediction are calculated throughout the whole series.

Method	Nominal coverage 90%			Nominal coverage 95%		
	CVR	LEN	LEN Sd	CVR	LEN	LEN Sd
MF	0.8344	0.0480	0.0054	0.9045	0.0578	0.0087
PMF	0.8535	0.0513	0.0061	0.9427	0.0655	0.0102
CP	0.8726	0.0512	0.0012	0.9172	0.0616	0.0022
MDCP	0.8280	0.0557	0.0239	0.9363	0.0690	0.0243
PMDCP	0.8535	0.0499	0.0058	0.9236	0.0643	0.0109

**Table 1:** The out-of-sample prediction performance of different PIs on log-returns of S&P500 price index with  $[0.7n]$  as rolling window size.

Thank you!

## References

---

- Andrews, D. W. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4):930–934.
- Chen, B. and Hong, Y. (2012). Testing for the markov property in timeseries. *Econometric Theory*, 28(1):130–178.
- Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48):e2107794118.
- Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics*, 38(3).
- Lei, J., G’Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111.
- Pan, L. and Politis, D. N. (2016). Bootstrap prediction intervals for markov processes. *Computational Statistics & Data Analysis*, 100:467–494.
- Politis, D. N. (2015). *Model-free prediction and regression: a transformation-based approach to inference*. Springer.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472.
- Wang, Y. and Politis, D. N. (2026). Model-free bootstrap and conformal prediction in regression: conditionality, conjecture testing, and pertinent prediction intervals. *Journal of Nonparametric Statistics*, 38(1):311–345.
- Zhou, X., Jiao, Y., Liu, J., and Huang, J. (2023). A deep generative approach to conditional sampling. *Journal of the American Statistical Association*, 118(543):1837–1848.

# Proof of Theorem (Conditional coverage guarantees I)

---

The proof hinges on the definition of  $\beta$ -mixing directly:

$$\begin{aligned} \beta_X(k) &= E \sup_{B \in \sigma(X_s, s \geq k)} |\mathbb{P}(B \mid \sigma(X_s, s \leq 0)) - \mathbb{P}(B)| \\ &= \frac{1}{2} \sup \sum_{i=1}^I \sum_{j=1}^J \left| \mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i) \mathbb{P}(B_j) \right|; \end{aligned}$$

where the last sup is taken among all the pairs of partitions  $\{A_1, \dots, A_I\}$  and  $\{B_1, \dots, B_J\}$  s.t.  $A_i \in \sigma(X_s, s \leq 0)$  for all  $i$  and  $B_j \in \sigma(X_s, s \geq k)$  for all  $j$ .

# Lemma on the Property of $\widehat{F}(\cdot|\cdot)$

## Lemma

Under A1 to A5, we have

$$\sup_{\mathbf{x}, y} |\widehat{F}(y|\mathbf{x}) - F(y|\mathbf{x})| = O\left(\frac{\log N}{\sqrt{N}h^p} + h^2\right),$$

with probability at least  $1 - S_n$ , where  $S_n$  is one appropriate sequence converges to 0; see below remark for the discussion about this sequence.

Remark:

Denote  $h^p = N^{-\tau}$  for  $0 < \tau < 1$ . From the proof of Lemma 4,  $S_n$  is taken as  $1 - 2J_n N^{-C_9} - 2J_n N^{-C'_9}$ , where  $C_9 > 0$  and  $\tau < \frac{p+1}{2} + \tau \frac{(p+2)(p+1)}{2p} - C_9 < 0$ . Similarly,  $C'_9 > 0$  is taken as another appropriate constant s.t.,  $\frac{p}{2} + \frac{p\tau}{2} - C'_9 < 0$ . This is possible when  $Nh^p$  is large enough.

# Algorithm of Model-Free (MF) Bootstrap Method

---

**Algorithm 1** Model-Free (MF) Bootstrap Method for Markov( $p$ )

---

**Require:** Data  $\{(\mathbf{X}_{t-1}, Y_t)\}_{t=1}^n$ . Some large positive integer  $M$ . Bootstrap size  $B$ . Significant level  $\alpha$ .

- 1: Use (1) to obtain the  $V_t = \widehat{F}(Y_t|\mathbf{X}_{t-1})$  for  $t = p + 1, \dots, n$ .
- 2: Calculate  $\widehat{Y}_{n+1}$ , the predictor of  $Y_{n+1}$  by the sample mean

$$\widehat{Y}_{n+1} = \frac{1}{n-p} \sum_{t=p+1}^n \widehat{F}^{-1}(V_t|\mathbf{X}_n).$$

- 3: **for** step  $i \in \{1, \dots, B\}$  **do**
- 4: Resample randomly with replacement the transformed data  $V_{p+1}, \dots, V_n$  to create the data  $V_{-M}^*, V_{-M+1}^*, \dots, V_0^*, V_1^*, \dots, V_n^*, V_{n+1}^*$ .
- 5: Draw  $Y_{-M}^*, \dots, Y_{-M+p-1}^*$  from any consecutive  $p$  values of the dataset  $(Y_1, \dots, Y_n)$ ; Let  $\mathbf{X}_{-M+p-1}^* = (Y_{-M+p-1}^*, \dots, Y_{-M}^*)$ . Especially, for  $p = 1$ ,  $\mathbf{X}_{-M}^* = Y_{-M}^*$  is drawn from  $(Y_1, \dots, Y_n)$ .
- 6: Generate  $Y_t^* = \widehat{F}^{-1}(V_t^*|\mathbf{X}_{t-1}^*)$  for  $t = -M + p, \dots, n$ .
- 7: Calculate the bootstrap future value  $Y_{n+1}^* = \widehat{F}^{-1}(V_{n+1}^*|\mathbf{X}_n)$ .
- 8: Calculate the bootstrap mean  $\widehat{Y}_{n+1}^* = \frac{1}{n-p} \sum_{t=p+1}^n \widehat{F}^{-1}(V_t^*|\mathbf{X}_n)$  where

$$\widehat{F}^*(Y|\mathbf{X}) = \frac{\frac{1}{n-p} \sum_{t=p+1}^n W_h(\mathbf{X}_{t-1}^*, \mathbf{X}) K\left(\frac{Y-Y_t^*}{h_0}\right)}{\overline{W}_h(\mathbf{X})}.$$

- 9: Calculate the bootstrap root  $Y_{n+1}^* - \widehat{Y}_{n+1}^*$ .
- 10: **end for**
- 11: The  $B$  bootstrap root replicates are collected in the form of an empirical distribution whose  $\alpha$ -quantile is denoted  $q(\alpha)$ .

**Output:** The  $(1 - \alpha) \times 100\%$  equal-tailed predictive interval for  $Y_{n+1}$  is given by

$$\widehat{C}_{1-\alpha}^{\text{MF}}(\mathbf{X}_n) = \left[ \widehat{Y}_{n+1} + q(\alpha/2), \widehat{Y}_{n+1} + q(1 - \alpha/2) \right].$$