Instability of ranks and inference under long memory

Shuyang (Ray) Bai*

joint work with Murad S. Taqqu** and Ting Zhang**

University of Georgia* & Boston University**

June 19, 2018

Gaussian Subordination Model

A popular class of long-memory models: **Gaussian subordination**.

$\{Z_k\}$: standardized stationary long-memory Gaussian with

$$\mathrm{Cov}[Z_k, Z_0] \sim k^{-\beta_0}, \qquad 0 < \beta_0 < 1.$$

Model:

$$X_k = G(Z_k), \qquad G(z)\text{: a function s.t. } \mathbb{E}G(Z_i)^2 < \infty.$$

Hermite polynomials $H_m(\cdot)$: orthogonal polynomials under Gaussian measure.

$$H_0(z) = 1, \quad H_1(z) = z, \quad H_2(z) = z^2 - 1, \quad H_3(z) = z^3 - 3z, \ldots$$

L^2-expansion:

$$G(z) = \mu + g_{\boldsymbol{m}} H_m(z) + g_{m+1} H_{m+1}(z) + \ldots, \qquad g_{\boldsymbol{m}} \neq 0$$

***m*: Hermite rank.**

Key Property:

$$\mathrm{Cov}[X_k, X_0] \sim k^{-\beta_0 \boldsymbol{m}}.$$

$\beta_0 m > 1$: (X_n) has short memory.
$\beta_0 m < 1$: (X_n) has long memory with new parameter $\beta = \beta_0 m$.

Limit Theorems for Gaussian Subordination

Recall: $X_k = G(Z_k)$, (Z_k) standardized Gaussian, $\mathrm{Cov}[Z_k, Z_0] \sim k^{-\beta_0}$, m: Hermite rank.

Central Limit Theorem Breuer Major (1983), Chambers Slud (1989).
If $\beta_0 m > 1$ (short memory), then

$$\frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu) \Rightarrow \sigma B(t), \quad B(t) : \text{Brownian motion}, \quad \sigma^2 := \sum_{k=-\infty}^{+\infty} \mathrm{Cov}[X_k, X_0],$$

Non-Central Limit Theorem Dobrushin & Major (1979), Taqqu (1979).
If $\beta_0 m < 1$ (long memory with $\beta = \beta_0 m$), then

$$\frac{1}{n^{1-\beta/2}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu) \Rightarrow \nu Z_{m,\beta_0}(t), \qquad \nu\text{: scale constant.}$$

Hermite process: $Z_{m,\beta_0}(t) =$ $\quad$ ($m = 1$: fractional Brownian motion)

$$\int_{x_1 < x_2 < \ldots < x_m} \left[\int_0^t \prod_{j=1}^{m} (s - x_j)_+^{-\beta_0/2 - 1/2} ds \right] dB(x_1) dB(x_2) \ldots dB(dx_m).$$

Summary: m controls both the **normalization order** and the **asymptotic distribution**.

Statistical Challenges with Gaussian Subordination model

Recall Gaussian subordination $X_k = G(Z_k)$.

(One can carry out a similar discussion for non-Gaussian linear process $Z_k = \sum_i a_i \epsilon_{k-i}$).

To apply previous limit theorems for inference, one needs to know the Hermite rank:

$$m = \inf \left\{ k \geq 1 : \int G(z) H_k(z) \phi(z) dz \neq 0 \right\}.$$

- Situation 1: G is unknown.
  $X_k = G(Z_k)$ with G unspecified to account for distributional flexibility.
  E.g., error in regression $u_k = \beta_0 + \beta_1 v_k + G(Z_k)$.

- Situation 2: G is known.
  G arises from statistical procedure.
  E.g., Z_k observed, $G(z) = z^2$ ($m = 2$) arises when estimating variance of Z_k.

Situation 1: difficult to estimate m. Often assume $m = 1$ (justification?).

Situation 2: seems no problem?

Short conclusion: better not trust a Hermite rank $m \geq 2$.

Perturbation of Gaussian Subordination

- A common statistical modeling principle:

A small perturbation of assumption should not drastically alter the conclusion.

$$\text{Model: } G(Z_k) \xrightarrow{\text{perturbation}} G \circ F(Z_k)$$

G: known or unknown, with Hermite rank m.

F: an uncontrollable perturbation transform.

Note: when G is known, F reflects the uncertainty prior to applying G.

Question: how likely does $G \circ F$ still have Hermite rank m?

Answer: if $m \geq 2$, very unlikely.
Indeed,

$$m \geq 2 \iff \int G \circ F(z) \cdot z \cdot \phi(z) \, dz = 0, \quad \text{where } H_1(z) = z.$$

The equality is very rigid. Departing from $F(z) = z$ easily breaks it down.

Shift Perturbation

As an example, let us consider a shift $F(z) = \theta_y(z) = z + y$.

$m(y)$: Hermite rank of $G \circ \theta_y$.

$$m(y) \geq 2 \iff H(y) = \int G(z+y) \cdot z \cdot \phi(z) dz = 0.$$

Bai & Taqqu (2018):

$$H \text{ is analytic } \Rightarrow H(y) = 0 \text{ occurs only for isolated } y's,$$

unless

$$H(y) \equiv 0 \quad \Rightarrow \quad G \text{ is a constant.}$$

In particular, if G is not constant,

$$\text{Hermite rank of } G \geq 2 \quad \Rightarrow \quad \text{Hermite rank of } G \circ \theta_y \text{ is 1 in a nbhd of } y = 0.$$

Similarly arguments apply to more general (parameterized) transforms F.

Interplay Between Shift Perturbation Size and Sample Size

When $m \geq 2$ and F is close to identity, would $G \circ F$ behave like $m \geq 2$?

For shift $F = \theta_y$, one can perform a "near-higher-order-rank" analysis as $y = y_n \to 0$ of

$$S_n(t) = \frac{1}{a_n} \sum_{n=1}^{\lfloor nt \rfloor} \left[G(Z_k + y_n) - \mathbb{E}G(Z_k + y_n) \right]$$

Recall $\mathrm{Cov}(Z_k, Z_0) \sim k^{-\beta_0}$, $\beta_0 \in (0,1)$.

- $\beta_0 > 1/m$, $m \geq 2$:

y_n	a_n	f.d.d. limit
$\ll n^{(\beta_0-1)/(2m-2)}$	$n^{1/2}$	$cB(t)$
$\approx n^{(\beta_0-1)/(2m-2)}$	$n^{1/2}$	$c_1 B(t) + c_2 Z_{1,\beta_0}(t)$
$\gg n^{(\beta_0-1)/(2m-2)}$	$n^{1-\beta_0/2} y_n^{1-m}$	$cZ_{1,\beta_0}(t)$

- $\beta_0 < 1/m$:

y_n	a_n	f.d.d. limit
$\ll n^{-\beta_0/2}$	$n^{1-\beta_0 m/2}$	cZ_{m,β_0}
$\approx n^{-\beta_0/2}$	$n^{1-\beta_0 m/2}$	$\sum_{k=1}^{m} c_k Z_{k,\beta_0}(t)$
$\gg n^{-\beta_0/2}$	$n^{1-\beta_0/2} y_n^{1-m}$	$cZ_{1,\beta_0}(t)$

Empirical Evidence

Observation:
(Z_k) Gaussian or linear, $\mathrm{Cov}(Z_k, Z_0) \sim k^{-\beta_0}$, $\beta_0 \in (0,1)$. Then "rank theory" predicts

$$\sum_{k=1}^{n} (Z_k - \bar{Z}_n)^2 \text{ has normalization } n^H,$$

where

$$H = H(\beta_0) := \min(1/2, 1 - \beta_0).$$

- Design of study:

Suppose we have a collection of long-memory time series data.
One of the series is $(Z_k, \ k = 1, \ldots, n)$.

Estimate $\hat{\beta}_0$ from (Z_k), plug in $H(\hat{\beta}_0)$.
Estimate $\widehat{H}$ directly from $(Z_k - \bar{Z}_n)^2$, $k = 1, \ldots, n$.

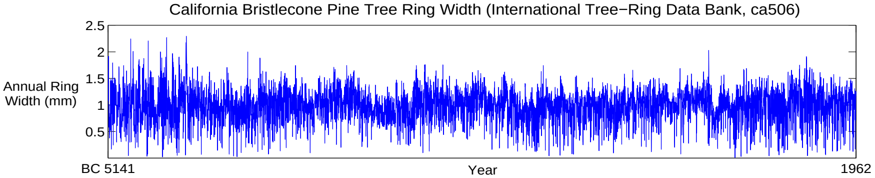
With "rank theory", one expects $H(\hat{\beta}_0) \approx \widehat{H}$ on average.

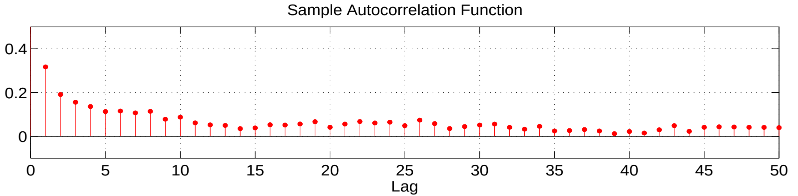
Data: Treering width sequence (The International Tree-Ring Data Bank).

Well-known to exhibit long memory since Mandelbrot & Wallis (1969).

Tree Ring Width

Figure: Tree Ring: Living Records of Climate





Failure of “rank theory”

Let $\delta = \widehat{H} - H(\hat{\beta}_0)$.

We contrast with fractional Gaussian noise (fGn), for which “rank theory” works.

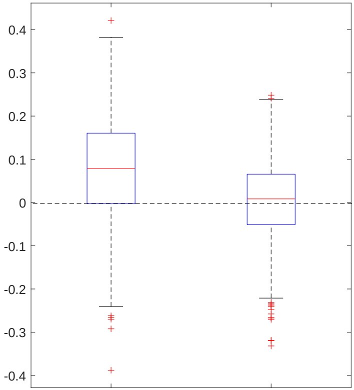


Figure: Box plot for δ's, Treering width (left) vs fGn (right). Aggregated Variance Method.

Beyond Sum: Instability of Whittle Estimator Asymptotics

(Z_k) centered long-memory Gaussian process with spectral density f_θ.

Whittle estimator:

$$\widehat{\theta}_n = \underset{\theta}{\operatorname{argmin}} \sum_{k,l=1}^{n} a_\theta(k-l) Z_k Z_l$$

where $a_\theta(n) = \int_{-\pi}^{\pi} \frac{e^{in\lambda}}{g_\theta(\lambda)} d\lambda$, $g_\theta(\lambda) \propto f_\theta(\lambda)$, $\int_{-\pi}^{\pi} \ln g_\theta(\lambda) d\lambda = 0$.

Fox & Taqqu (1986) and Giraitis & Surgailis (1990):

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2). \tag{1}$$

Achieving i.i.d. parametric rate $n^{-1/2}$ despite of having long memory.

Giraitis & Taqqu (1999):
Nice function G satisfying $\mathbb{E} G(Z_0) = 0$,

$$\rho_1 := \sum_{n \in \mathbb{Z}} \mathbb{E}\left[G'(Z_n) G(Z_0)\right] \frac{\partial}{\partial \theta} a_\theta(n).$$

If $G(x) = x$, then $\rho_1 = 0$. Departing from $G(x) = x$ likely yields $\rho_1 \neq 0$ (not by shift).

If Gaussian Z_k is replaced by $G(Z_k)$, and $\rho_1 \neq 0$, then

$$n^{\beta_0/2}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma_2^2).$$

What Should One Do?

- Issue: asymptotics developed based on "rank theory" may not be reliable.

An ad hoc solution: assume "rank $=1$" always:

Stick to convergence rate $n^{-\beta_0/2}$ and asymptotic normality no matter what.

Problem: may not approximate well the situation of "near-higher-order-rank".

- Reformulate issue: uncertainty in **normalization order** and in **asymptotic distribution**.

- Prescription: **Resampling** (self-adaptive to normalization/self-normalization).

Basic Notation and Setup for Inference

Sample block: $\mathbf{X}_p^q = (X_p, \ldots, X_q)$.

Unknown parameter of interest: θ.

$T_n(\,\cdot\,;\theta) : \mathbb{R}^n \to \mathbb{R}$ a function of n samples designed for inference of θ, which satisfies:

$$T_n(\mathbf{X}_1^n;\theta) \xrightarrow{\mathcal{L}} T \qquad \text{as } n \to \infty,$$

for some non-degenerate T.

If distribution of T is known (no nuisance parameter), can use it for inference directly.

Example: Inference of Mean

$$\theta = \mu, \quad T_n(X_1^n; \theta) = \frac{\bar{X}_n - \mu}{D_n}, \quad D_n = D_n(\mathbf{X}_1^n)\text{: a normalizer to ensure } T_n \xrightarrow{\mathcal{L}} T.$$

- When (X_n) is i.i.d. with finite variance σ^2, use sample standard deviation:
$D_n = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2} \xrightarrow{p} \sigma$. $T \overset{\mathcal{L}}{=} N(0,1)$.

- When (X_n) has short memory, use consistent estimate of long-run standard deviation:
$D_n = \sqrt{\sum_k w(k/h)\widehat{\gamma}(k)} \xrightarrow{p} \sqrt{\sum_k \gamma(k)}$, $\gamma(k) := \mathrm{Cov}[X_k, X_0]$. $T \overset{\mathcal{L}}{=} N(0,1)$.
w : window function, $h = h_n$: bandwidth parameter.

- A self-adaptive normalizer for short/long memory, light/heavy tails, Shao (2010):
$D_n = \sqrt{\frac{1}{n^3}\sum_{k=1}^n \left[\sum_{i=1}^k X_i - k\bar{X}_n\right]^2}$.

$$\text{If } \frac{1}{n^H}\sum_{i=1}^{[ns]}(X_i - \mu) \Rightarrow \nu Z(s), \quad \text{then} \quad \frac{\bar{X}_n - \mu}{D_n} \xrightarrow{\mathcal{L}} T = \frac{Z(1)}{\sqrt{\int_0^1 [Z(s) - sZ(1)]^2 ds}}.$$

E.g. $Z(s) =$ Brownian motion, Hermite process, stable process, etc.

Resampling Under Dependence

- **Block Bootstrap** (Kunsch 1989).
  1. Estimate θ by a consistent estimator $\widehat{\theta}_n = \widehat{\theta}_n(\mathbf{X}_1^n)$.
  2. Choose a block size b. Form $n - b + 1$ successive blocks (with overlap)
  $$\mathbf{X}_1^b, \mathbf{X}_2^{b+1}, \dots, \mathbf{X}_{n-b+1}^n.$$
  3. Sample randomly with replacement $[n/b]$ blocks.
  Paste them into $\mathbf{X}^*$ of length $b \times [n/b] \approx n$.
  Obtain $T^* := T_{b[n/b]}(\mathbf{X}^*; \widehat{\theta}_n)$ on the bootstrapped sample $\mathbf{X}^*$.
  4. Repeat the last step N times getting bootstrapped copies: $T_1^*, \dots, T_N^*$.
  5. Use the empirical distribution of $\{\bar{T}_i^*\}$ to approximate the distribution of $T_n(\mathbf{X}_1^n; \theta)$.

Does **NOT** work under *long-memory Gaussian subordination model*. Lahiri (1993).

Idea for remedy: keep the order (no artificial pasting) $\Rightarrow$ reduce sample size.

- **Subsampling** (Politis Romano Wolf 1999) or called *block sampling*, *sampling window*.

Subsampling

- **General procedure**:
  1. Estimate θ by $\widehat{\theta}_n = \widehat{\theta}_n(\mathbf{X}_1^n)$.
  2. Choose a block size b and form blocks $\mathbf{X}_1^b, \mathbf{X}_2^{b+1}, \ldots, \mathbf{X}_{n-b+1}^n$.
  3. Compute $T_b(\mathbf{X}_1^b; \widehat{\theta}_n)$, $T_b(\mathbf{X}_2^{b+1}; \widehat{\theta}_n)$, $\ldots$, $T_b(\mathbf{X}_{n-b+1}^b; \widehat{\theta}_n)$.
  4. Use the empirical distribution $\widehat{F}_{n,b}(x)$ of $\{T_b(\mathbf{X}_i^{i+b-1}; \widehat{\theta}_n)\}$ to approximate the distribution of $T_n(\mathbf{X}_1^n; \theta)$.

- **Example**: Inference of $\theta = \mathbb{E}X_i = \mu$.

$$T_n(\mathbf{x}; \mu) = \frac{\frac{1}{n}\sum_{i=1}^n x_i - \mu}{D_n(\mathbf{x})}, \qquad D_n(\mathbf{x}) = \sqrt{\frac{1}{n^3}\sum_{k=1}^n \left[\sum_{i=1}^k x_i - \frac{k}{n}\sum_{i=1}^n x_i\right]^2}.$$

Procedure for constructing a two-sided $(1-\alpha)$-confidence interval for μ:

1. Estimate μ by $\bar{X}_n$.
2. Choose a block size b and form blocks $\mathbf{X}_1^b, \mathbf{X}_2^{b+1}, \ldots, \mathbf{X}_{n-b+1}^n$.
3. Obtain the empirical distribution $\widehat{F}_{n,b}(x)$ of $\{T_b(\mathbf{X}_i^{b+i-1}; \bar{X}_n),\ i = 1, \ldots, n-b+1\}$.
4. Obtain the lower and upper $\alpha/2$ quantiles $L_{\alpha/2}$ and $U_{\alpha/2}$ of $\widehat{F}_{n,b}(x)$.
5. A $(1-\alpha)$-level confidence interval for the mean is given by

$$[\bar{X}_n - U_{\alpha/2}D_n(\mathbf{X}_1^n)\ ,\ \bar{X}_n - L_{\alpha/2}D_n(\mathbf{X}_1^n)].$$

Asymptotic Validity of Subsampling

$$\widehat{F}_{n,b}(x) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1\{T_b(\mathbf{X}_i^{b+i-1}; \widehat{\theta}_n) \le x\}.$$

When the sample size n and the block size b are reasonably large,

$$T_n(\mathbf{X}_1^n;\theta) \overset{\mathcal{L}}{\approx} T \overset{\mathcal{L}}{\approx} T_b(\mathbf{X}_1^b;\theta) \overset{\mathcal{L}}{\approx} T_b(\mathbf{X}_1^b;\widehat{\theta}_n) \underset{\textbf{subsampling}}{\overset{\mathcal{L}}{\approx}} \widehat{F}_{n,b}(x).$$

Consistency Result:

A 1 Gaussian subordination model: $\{X_i = G(Z_i)\}$,
The long-memory Gaussian $\{Z_i\}$ satisfies some regularity conditions.

A 2 $T_n(\mathbf{X}_1^n;\theta) \xrightarrow{\mathcal{L}} T$.

A 3 $T_b(\,\cdot\,;\widehat{\theta}_n)$ is asymptotically replaceable by $T_b(\,\cdot\,;\theta)$ in $\widehat{F}_{n,b}(x)$.
(E.g., holds for the common form $T_n(\mathbf{X}_1^n;\theta) = \frac{\widehat{\theta}_n - \theta}{D_n}$).

Theorem (Consistency of subsampling, Betken & Wenlder (2017), Bai & Taqqu (2017))

Suppose the sample size $n \to \infty$, the block size $b = b_n \to \infty$ and $b_n = o(n)$. Then

$$\left| \widehat{F}_{n,b_n}(x) - P\Big(T_n(\mathbf{X}_1^n;\theta) \le x\Big) \right| \xrightarrow{p} 0$$

at any continuity point x of the cdf of T.

Simulation Example

Data:

$$X_i = H_m(Z_i) \qquad \theta = \mu = \mathbb{E}X_i = 0.$$

$\{Z_i\}$: standardized fractional Gaussian noise ($\mathrm{Cov}(Z_0, Z_k) \sim k^{-\beta_0}$).
$H_m(x)$: Hermite polynomials. $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$.
Dichotomy:

$$\frac{1}{n^H}\sum_{i=1}^{[nt]} X_i \Rightarrow Y(t) \quad \begin{cases} \text{If } 2d-1 = (2d_0-1)m < -1, \ Y(t) = \sigma B(t), & H = 1/2. \\ \text{If } \beta = \beta_0 m < 1, \ Y(t) = \nu Z_{\beta_0,m}(t), & H = 1-\beta/2. \end{cases}$$

$$T_n(\mathbf{x};\mu) = \frac{\frac{1}{n}\sum_{i=1}^n x_i - \mu}{D_n(\mathbf{x})}, \qquad D_n(\mathbf{x}) = \sqrt{\frac{1}{n^3}\sum_{k=1}^n \left[\sum_{i=1}^k x_i - \frac{k}{n}\sum_{i=1}^n x_i\right]^2}.$$

m \ β_0	0.6	0.4	0.2
1	*86 vs 82*	*83 vs 39*	*76 vs 25*
2	90 vs 84	*91 vs 71*	*86 vs 41*
3	86 vs 86	90 vs 83	*89 vs 58*

Monte-Carlo evaluation of coverage percentage.
Sample size=500.
Nominal Level=90%.

Subsampling vs Block Bootstrap.

Block size: $\lfloor\sqrt{500}\rfloor = 22$.

Red: $\beta_0 m < 1$ (*long memory* regime)

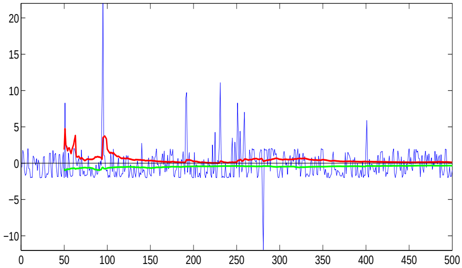


Figure: The running 90% confidence interval for a sample path of $\{X_i\}$. $\beta_0 = 0.2$, $m = 3$, $\beta = 0.6$.

Thank You!

Precise Statement of Main Result

Long-memory Gaussian (X_n), $\mathrm{Cov}[X_n, X_0] \sim n^{-\beta}$, $\beta \in (0,1)$. Assume the spectral density of (X_n) is given by

$$f(\lambda) = f_\beta(\lambda) f_0(\lambda),$$

where $f_\beta(\lambda)$ is the FARIMA(0,$d = \frac{1-\beta}{2}$,0) spectrum:

$$f_\beta(\lambda) = |1 - e^{i\lambda}|^{\beta - 1},$$

and $f_0(\lambda)$ satisfies short memory conditions ($\gamma_0(n)$ is the covariance of $f_0(\lambda)$):

(a) $\inf_\lambda f_0(\lambda) > 0$;   (b) $\gamma_0(n) = O(n^{-\alpha})$, $\alpha > 1$.

Then $\forall \lambda > 0$, $\exists\ 0 < c \leq C$

$$c \left(\frac{b}{k}\right)^\beta \leq \alpha_{k,b} \leq C \left(\frac{b}{k}\right)^\beta + \underbrace{\cancel{O(k^{-\alpha+1})}}_{\text{if } \alpha > 1+\beta}, \quad \text{for all } 1 \leq b \leq \lambda k.$$

- Time-domain interpretation: Let $d = \frac{1-\beta}{2}$, FARIMA model: $\Delta^d X_n = \epsilon_n$, (ϵ_n) has $f_0(\lambda)$.
- Examples: FARIMA(p,$d = \frac{1-\beta}{2}$,q), fractional Gaussian noise $H = 1 - \beta/2 > 1/2$.

Idea of Proof

Goal:

$$\left|\widehat{F}_{n,b_n}(x) - P\Big(T_n(\mathbf{X}_1^n;\theta) \le x\Big)\right| \xrightarrow{p} 0.$$

From Assumption 3, replace

$$\widehat{F}_{n,b}(x) = \frac{1}{n-b+1}\sum_{i=1}^{n-b+1} 1\{T_b(\mathbf{X}_i^{b+i-1};\widehat{\theta}_n) \le x\}.$$

by

$$\widehat{F}^*_{n,b}(x) = \frac{1}{n-b+1}\sum_{i=1}^{n-b+1} 1\{T_b(\mathbf{X}_i^{b+i-1};\theta) \le x\}.$$

Suffices to show

$$\left|\widehat{F}^*_{n,b_n}(x) - P\Big(T_n(\mathbf{X}_1^n;\theta) \le x\Big)\right| \xrightarrow{p} 0.$$

Bias-variance decomposition of mean-square error:

$$\mathbb{E}\left[\widehat{F}^*_{n,b}(x) - P\Big(T_n(\mathbf{X}_1^n;\theta) \le x\Big)\right]^2 = \underbrace{\left[P\Big(T_b(\mathbf{X}_1^b;\theta) \le x\Big) - P\Big(T_n(\mathbf{X}_1^n;\theta) \le x\Big)\right]^2}_{\text{Bias}} + \underbrace{\mathrm{Var}[\widehat{F}^*_{n,b}(x)]}_{\text{Variance}}$$

Bias $\to 0$ since by Assumption 2, both $T_n(\mathbf{X}_1^n;\theta)$ and $T_b(\mathbf{X}_1^b;\theta)) \xrightarrow{\mathcal{L}} T$ as $n, b \to \infty$.

How about the Variance term?

Control Variance Term

Recall $\widehat{F}^*_{n,b}(x) := \frac{1}{n-b+1}\sum_{i=1}^{n-b+1} 1\{T_b(\mathbf{X}_i^{i+b-1};\theta) \le x\}$. and want to show

$$\mathrm{Var}[\widehat{F}^*_{n,b}(x)] \to 0 \quad \text{as} \quad n \to \infty.$$

By a standard computation using stationarity of (X_n),

$$\begin{aligned}\mathrm{Var}[\widehat{F}^*_{n,b}(x)] &\le \frac{2}{n-b+1}\sum_{k=1}^{n-b+1}\left|\mathrm{Cov}\Big[\ 1\{T_b(\mathbf{X}_1^b;\theta) \le x\}\ ,\ 1\{T_b(\mathbf{X}_k^{k+b-1};\theta) \le x\}\ \Big]\right| \\ &\le \frac{2}{n-b+1}\sum_{k=1}^{n}\alpha_{k,b}, \qquad \text{(the reason of replacing } \widehat{\theta}_n \text{ by } \theta.)\end{aligned}$$

where $\alpha_{k,b}$ is the between-block mixing coefficient:

$$\alpha_{k,b} = \sup\{\ |\mathrm{Cov}[1_A, 1_B]|\ ,\ A \in \sigma(\mathbf{X}_1^b),\ B \in \sigma(\mathbf{X}_{k+1}^{k+b})\ \}.$$

Hence under $b_n = o(n)$,

$$\sum_{k=1}^{n}\alpha_{k,b_n} = o(n) \Rightarrow \mathrm{Var}[\widehat{F}^*_{n,b}(x)] \to 0.$$

which was mentioned before.