A unified approach to self-normalized block sampling

Murad S. Taqqu

joint work with Shuyang Bai and Ting Zhang

Boston University

August 14, 2016

Overview:

1. **Goal:** To perform inference in such a way that we do not need to estimate unknown parameters in a time series.

2. **Features of the method:** Combines the advantages of self-normalization so as to avoid having to know or to estimate the scale parameters, with block sampling so that one can use the sampling distribution instead of the asymptotic one.

3. **Advantage:** To derive a unified approach for short and long-range dependence, and heavy tailed distributions.

4. **Application:** Apply the method to inference about the mean.

5. **Scope:** We consider two basic cases:
   (a) the data is subordinated to the Gaussian;
   (b) the data is strong mixing.

Outline of the talk

1. Introduction
2. The suggested procedure
3. The asymptotic theory in the Gaussian subordinated case
4. Examples
5. The asymptotic theory in the strong mixing case
6. Examples

Confidence interval for the mean: review of the i.i.d. case

Samples $\{X_i, i = 1, \ldots, n\}$, i.i.d. with finite variance σ^2.
Sample size n reasonably large.

Confidence interval for the mean $\mu = \mathbb{E}X_i$ involves:

- Sample mean: $\bar{X}_n$;
- Sample variance: $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$;
- Normal $(1-\alpha/2)$-quantile: $q_{1-\alpha/2}$;

Two-sided $(1-\alpha)$-level confidence interval:

$$I_n = [\bar{X}_n - n^{-1/2}\hat{\sigma}_n q_{1-\alpha/2}\ ,\quad \bar{X}_n + n^{-1/2}\hat{\sigma}_n q_{1-\alpha/2}].$$

$$\text{Then} \quad P\left(\text{Random interval } I_n \text{ covers } \mu\right) \approx 1-\alpha.$$

Theoretical basis for the confidence interval I_n involves:

Central Limit Theorem: $n^{-1/2}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} N(0, \sigma^2)$.
Law of Large Numbers: $\hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2$.

If $\{X_i\}$ is short-range dependent, then σ^2 is replaced by $\sum_{k=-\infty}^{\infty}\mathrm{Cov}[X_k, X_0] =: \sum_k \gamma(k)$. How to deal with this first challenge?

Challenge 1: Short-range dependence

$\{X_i\}$ stationary weakly dependent (short-range dependent) with covariance

$$\gamma(k) = \mathrm{Cov}[X(k), X(0)], \quad k \in \mathbb{Z}, \quad \text{satisfying } \sum_k |\gamma(k)| < \infty.$$

Central Limit Theorem:

$$n^{-1/2} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{d} N(0, \sigma^2),$$

where now σ^2 is the so-called *long-run variance*

$$\sigma^2 = \sum_{k=-\infty}^{\infty} \gamma(k).$$

To construct a confidence interval, we need a consistent estimator for σ^2.

We have the sample covariance: $\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} (X_i - \bar{X}_n)(X_{i+k} - \bar{X}_n)$. But σ^2 cannot be estimated simply by $\sum_k \hat{\gamma}(k)$ (too few summands for large k).

Challenge 2: Heavy tails

$\{X_i\}$ i.i.d. heavy-tailed

$$P(X_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \quad P(X_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha} \quad x \to +\infty, \tag{1}$$

where constant $A > 0$, parameters $\beta \in [-1, 1]$, $\alpha \in (1, 2)$.

$\mathbb{E}|X_1| < \infty$ but $\mathbb{E}|X_1|^2 = \infty$.

Heavy tail Central Limit Theorem:

$$n^{-1/\alpha} \sum_{i=1}^{n} (X_i - \mu) \to S_\alpha(\sigma, \beta, 0)$$

where $S_\alpha(\sigma, \beta, 0)$ is the α-stable random variable with location parameter 0, scale parameter σ (depending on A and α) and skewness parameter β.

How about the unknown α, β, A?

Even more complicated situation: a slowly varying function replaces the constant A in (3).

Even more complicated: $\{X_i\}$ are weakly dependent (σ then depends on dependence structure).

Challenge 3: Long-range dependence

$\{X_i\}$ is strongly dependent (long-range dependent), with covariance function

$$\gamma(k) \sim c_\gamma k^{2H-2}, \quad H \in (1/2, 1). \tag{2}$$

Some models of $\{X_i\}$, e.g., nonlinear transform of a long-range dependent Gaussian process, give rise to limit theorem (Dobrushin Major (1979), Taqqu 1979):

$$\frac{1}{n^H}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} cZ_{m,H},$$

where c depends on c_γ and H, and m is a positive integer (the so-called Hermite rank), and

$$Z_{m,H} = v_{m,H}\int'_{\mathbb{R}^m}\int_0^1 \prod_{j=1}^{m}(s - x_j)_+^{(H-1)/m-1/2} ds\ B(dx_1)\dots B(dx_m), \quad B(\cdot)\text{: Brownian motion}$$

is a standardized random variable expressed by a multiple Wiener-Itô integral which is non-Gaussian if $m \geq 2$.

Need to estimate c_γ, H, m (no available method for m).
More complicated if c_γ in (4) is replaced by a slowly varying function.

Self-normalization under short-range dependence

Goal: design a way to avoid estimation of the nuisance parameter $\sigma^2 = \sum_k \gamma(k)$.

An idea which works for short-range dependence: self-normalization (Lobato (2001) and Shao (2010)). Consider:

$$D_n = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[\sum_{i=1}^{k} X_i - \frac{k}{n}\sum_{i=1}^{n} X_i\right]^2} = \sqrt{\int_0^1 \left[\sum_{i=1}^{[ns]} X_i - \frac{[ns]}{n}\sum_{i=1}^{n} X_i\right]^2 ds}$$

Note: D_n involves $\sum_i X_i$ and not $\sum_i X_i^2$.
Functional Central Limit Theorem (weak convergence in $D[0,1]$):

$$\frac{1}{n^{1/2}}\sum_{i=1}^{[nt]}(X_i - \mu) \Rightarrow \sigma B(t), \;\; B(t): \text{ Brownian motion.}$$

By the continuous mapping theorem:

$$\frac{n^{-1/2}\sum_{i=1}^{n}(X_i - \mu)}{n^{-1/2} D_n} \xrightarrow{d} \frac{\sigma B(1)}{\sigma\sqrt{\int_0^1 [B(s) - sB(1)]^2\, ds}} =: T.$$

No nuisance parameter! Use the distribution of T to construct confidence interval.

However, there are additional nuisance parameters in the heavy tail and long-range dependence case.

Self-normalization under heavy tail or long-range dependence

In general, suppose we have the Functional Central Limit Theorem

$$\frac{1}{n^H \ell(n)} \sum_{i=1}^{[nt]} (X_i - \mu) \Rightarrow cY(t),$$

where $H \in (0,1)$, $\ell(n)$ slowly varying.
One gets the self-normalized statistic

$$\frac{\sum_{i=1}^n (X_i - \mu)}{D_n} = \frac{n^{-H}\ell(n)^{-1} \sum_{i=1}^n (X_i - \mu)}{n^{-H}\ell(n)^{-1} D_n} \xrightarrow{d} T = \frac{Y(1)}{\sqrt{\int_0^1 [Y(s) - sY(1)]^2 ds}}.$$

Heavy tail: $Y(t)$ is the α-stable Lévy process $L_{\alpha,\beta}(t)$.

Long-range dependence: $Y(t)$ is the Hermite process $Z_{m,H}(t)$.

Caveat: self-normalization only frees one from the normalization (including the scale parameter), but not from other parameters (e.g. α*,* m*,* H*).* How to deal with that?

Our goal:

To get a confidence interval for μ using

$$T_n(\mathbf{X}_1^n;\mu) = \frac{\sum_{i=1}^n X_i - n\mu}{D_n(\mathbf{X}_1^n)}, \quad D_n(\mathbf{X}_1^n) = \sqrt{\frac{1}{n}\sum_{k=1}^n\left[\sum_{i=1}^k X_i - \frac{k}{n}\sum_{i=1}^n X_i\right]^2}.$$

where $\mathbf{X}_1^n = (X_1,\ldots,X_n)$.

Outline:

- Brief description of the procedure
- Conditions under which it is justified
- Applications

Brief description of the procedure

Recall $\mathbf{X}_1^n = (X_1, \ldots, X_n)$ and

$$T_n(\mathbf{X}_1^n; \mu) = \frac{\sum_{i=1}^n X_i - n\mu}{D_n(\mathbf{X}_1^n)}, \quad D_n(\mathbf{X}_1^n) = \sqrt{\frac{1}{n}\sum_{k=1}^n \left[\sum_{i=1}^k X_i - \frac{k}{n}\sum_{i=1}^n X_i\right]^2}.$$

We want to get a confidence interval for μ using $T_n(\mathbf{X}_1^n; \mu)$.
But we don't know the distribution of $T_n(\mathbf{X}_1^n; \mu)$.
We shall approximate it by the empirical distribution

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

of $T_b(\mathbf{X}_i^{b+i-1}; \bar{X}_n)$, $i = 1, \ldots, n-b+1$.

Note: Using (a) successive i; (b) overall sample mean

How to justify the use of $\widehat{F}_{n,b}(x)$?

The idea is to combine self-normalization with block sampling

We assume:

$$T_n(\mathbf{X}_1^n;\mu) = \frac{\sum_{i=1}^n X_i - n\mu}{D_n(\mathbf{X})} \xrightarrow{d} T.$$

When the block size b is large, we expect (setting $\mu = \mathbb{E}X_i$):

$$T_n(\mathbf{X}_1^n;\mu) \underset{\textbf{self-normalization}}{\overset{d}{\approx}} T_b(\mathbf{X}_1^b;\mu) \overset{d}{\approx} T_b(\mathbf{X}_1^b;\bar{X}_n) \underset{\textbf{block sampling}}{\overset{d}{\approx}} \hat{F}_{n,b_n}(x).$$

1st $\overset{d}{\approx}$: because self-normalization equalizes the scales of $T_n(\mathbf{X}_1^n;\mu)$ and $T_b(\mathbf{X}_1^b;\mu)$, and does not require knowing them;

2nd $\overset{d}{\approx}$: because $\bar{X}_n$ is close to unknown μ when $n \gg b$;

3rd $\overset{d}{\approx}$: because $\widehat{F}_{n,b}(x) = \frac{1}{n-b+1}\sum_{i=1}^{n-b+1} 1\{T_b(\mathbf{X}_i^{i+b-1};\bar{X}_n) \le x\}$ is the empirical distribution of $T_b(\mathbf{X}_1^b;\bar{X}_n)$.

Assumptions under which this procedure is shown to work:

1. The Gaussian subordination framework
2. The strong mixing framework

The Gaussian subordination case: theoretical assumptions

$\{X_i\}$: the stationary process we observe.
$\{Z_i\}$: hidden Gaussian stationary process with covariance $\gamma(k) = \mathrm{Cov}[Z_k, Z_0]$.

Assumptions (one-dimensional simplified version):

A1. Subordination: $X_i = G(Z_i, \dots, Z_{i-l})$ with mean $\mu = \mathbb{E}X_i$, where l is a fixed non-negative integer;

A2. Weak convergence in $D[0,1]$: with a suitable Skorohod topology:

$$\left\{ \frac{1}{n^H \ell(n)} (S_{\lfloor nt \rfloor} - n\mu),\ 0 \le t \le 1 \right\} \Rightarrow \{Y(t),\ 0 \le t \le 1\},$$

for some process $Y(t)$, where $0 < H < 1$ and $\ell(\cdot)$: slowly varying;

A3. Weak canonical correlation: As $n \to \infty$, the block size $b_n \to \infty$, $b_n = o(n)$, and satisfies

$$\sum_{k=0}^{n} \rho_{k, l+b_n} = o(n),$$

where $\rho_{k,m}$ is the *between-block canonical correlation:*

$$\rho_{k,m} = \sup_{\mathbf{x},\mathbf{y} \in \mathbb{R}^m} \mathrm{Corr}\left[\langle \mathbf{x}, \mathbf{Z}_1^m \rangle \langle \mathbf{y}, \mathbf{Z}_{k+1}^{k+m} \rangle \right].$$

where $\mathbf{Z}_1^m = (Z_1, \cdots, Z_m)$, $\mathbf{Z}_{k+1}^{k+m} = (Z_{k+1}, \cdots, Z_{k+m})$

Note: $\rho_{k,m}$ involves the underlying Gaussian Z_i and not the nonlinear X_i.
Remark: *The assumptions can be extended to vector valued Z_i, which allows the inclusion of some nonlinear time series models.*

Why did the nonlinearity disappear?

The block canonical correlation

$$\rho_{k,m} = \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^m} \mathrm{Corr}\left[\langle \mathbf{x}, \mathbf{Z}_1^m\rangle \langle \mathbf{y}, \mathbf{Z}_{k+1}^{k+m}\rangle\right].$$

involves the underlying Gaussian Z_i and not the nonlinear X_i. This because in the proof we use a key result due to Kolmogorov and Rozanov (1960):

$$\sup_{F,G\in L^2(\mathbf{Z}_1^m)} \left|\mathrm{Corr}\big(F(\mathbf{Z}_1^m), G(\mathbf{Z}_{k+1}^{k+m})\big)\right| = \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^m} \left|\mathrm{Corr}\Big(\langle \mathbf{x}, \mathbf{Z}_1^m\rangle, \langle \mathbf{y}, \mathbf{Z}_{k+1}^{k+m}\rangle\Big)\right| = \rho_{k,m}$$

Main result: consistency of the self-normalized block sampling

Theorem

Under Assumptions A1–A3, as $n \to \infty$,

$$\sup_{x \in \mathbb{R}} \left| \widehat{F}_{n,b_n}(x) - P(T_n(\mathbf{X}_1^n; \mu) \le x) \right| \to 0 \quad \textit{in probability},$$

if T has a continuous distribution (otherwise the convergence holds without "sup", for x at continuity points).

So, we can use the empirical distribution

$$\widehat{F}_{n,b_n}(x) = \frac{1}{n - b_n + 1} \sum_{i=1}^{n-b_n+1} 1\{T_{b_n}(\mathbf{X}_i^{i+b_n-1}; \bar{X}_n) \le x\},$$

which is obtained from the block sampling, to approximate the unknown distribution of

$$T_n(\mathbf{X}_1^n; \mu) = \frac{\sum_{i=1}^n X_i - n\mu}{D_n(\mathbf{X}_1^n)}, \text{ where } D_n(\mathbf{X}_1^n) = \sqrt{\frac{1}{n} \sum_{k=1}^n \left[\sum_{i=1}^k X_i - \frac{k}{n} \sum_{i=1}^n X_i \right]^2}.$$

Basic steps in the proof

- Assumption (A2) implies that $T_n(\mathbf{X}_1^n; \mu) \xrightarrow{d} T$ (continuous mapping);
- Bias-variance decomposition ($\hat{F}^*$ is $\hat{F}$ with $\bar{X}_n$ replaced by μ):

$$\mathbb{E}\left[\hat{F}^*_{n,b_n}(x) - P(T \le x)\right]^2 = \left[P\big(T_{b_n}(\mathbf{X}_1^{b_n}; \mu) \le x\big) - P(T \le x)\right]^2 + \mathrm{Var}[T_{b_n}(X_1^{b_n}; \mu)]$$

- The first bias term goes to zero at continuity points of $P(T \le x)$ by A2.
- $\mathrm{Var}[T_{b_n}(X_1^{b_n}; \mu)] \to 0$ (follows from A1 and A3).
- $\hat{F}^*_{n,b_n}(x) \to P(T \le x)$ at continuity points (the centering is by μ).
- $\hat{F}_{n,b_n}(x) \to P(T \le x)$ at continuity points (the centering is by $\bar{X}_n$).
- $\sup_x |\hat{F}_{n,b_n}(x) - P(T \le x)| \to 0$ if $P(T \le x)$ is continuous.

When does Assumption A3 hold?

• Long memory case: Suppose that the spectral density of the underlying Gaussian $\{Z_i\}$ is given by

$$f(\lambda) = f_H(\lambda) f_0(\lambda),$$

where $f_H(\lambda) = |1 - e^{i\lambda}|^{-2H+1}$, $1/2 < H < 1$, and $f_0(\lambda)$ is a short-range dependent spectral density bounded away from zero. Then $b_n = o(n)$ implies Assumption A3.

Examples:
• FARIMA(p, d, q)
• fractional Gaussian noise with $H > 1/2$.

• Short memory case Suppose that $\inf_\lambda f(\lambda) > 0$, and $|\mathrm{Cov}[Z_0, Z_n]| \leq d_n$, where d_n is non-increasing and summable (typically, $d_n = cn^{-\beta}$ for some constant $c > 0$ and $\beta > 1$). If $b_n = o(n)$, then Assumption A3 holds.

• Strong mixing case: $b_n = o(n)$ always implies Assumption A3.

Practical choice of the block size b

Method 1: Rule of thumb: $b = cn^{1/2}$, with typically $1/2 \leq c \leq 2$ (Hall et al. (1998)).

Method 2: Data-dependent choice (Jach et al (2012)): choose the b which minimizes the changes in the Kolmogorov distance of the empirical distribution $\hat{F}_{n,b}(x)$ with respect to b (optimum is the most stable point).

1. Choose an evenly-spaced block size sequence $b_1, \ldots, b_{p+1}$ (e.g. $b_1 = 5, b_2 = 5 + \delta, \ldots, b_{p+1} = 5 + p\delta$).
2. Compute the empirical distributions $\hat{F}_{n,b_i}$, $i = 1, \ldots, p+1$.
3. Choose $b_{\text{opt}} = b_i$ which minimizes $d_{\text{kol}}\left(\hat{F}_{n,b_i}, \hat{F}_{n,b_{i+1}}\right)$ in $i = 1, \ldots, p$.

We use the rule of thumb $b = n^{1/2}$ in the examples below.

Chi-squared: SRD and LRD

$$X_i = G(Z_i) = Z_i^2,$$

where $\{Z_i\}$ is fractional Gaussian noise with Hurst index H_0 ($H_0 = 0.5$: white noise; $H_0 > 0.5$: long-range dependent; $H_0 < 0.5$: anti-persistent). The mean is $\mu = \mathbb{E}X_i = 1$. Assumption A2 holds with the following dichotomy:

$$\begin{cases} H = 1/2,\ \ell(n) = 1,\ Y(t) = \sigma B(t) & \text{if } H_0 < 3/4; \\ H = 2H_0 - 1,\ \ell(n) = 1,\ Y(t) = c_H Z_{2,H}(t) & \text{if } H_0 > 3/4, \end{cases}$$

where $\sigma^2 = \sum_n \mathrm{Cov}[X(n), X(0)]$, $c_H > 0$, $B(t)$ is the standard Brownian motion and $Z_{2,H}(t)$ is the standard Rosenblatt process (second-order Hermite process).

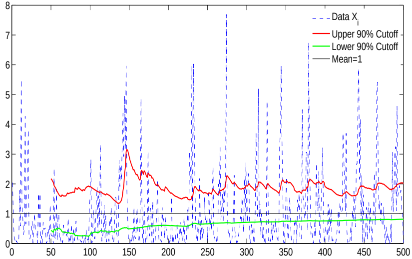


Figure: The running confidence cutoff for a sample path. $H_0 = 0.9$. The $\mathrm{Std}(\bar{X}_n) \sim n^{H-1} = n^{2H_0-2} = n^{-0.2}$.

H_0	0.5	0.7	0.9
	(86,95)	(88,94)	(92,83)

Table: Monte Carlo evaluation of coverage percentage (lower 90%, upper 90%). Sample size=500.

Example: t-transform

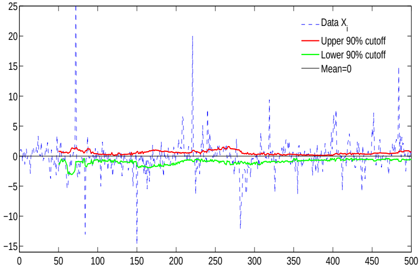
Data:

$$X_i = G(Z_i) = F_\alpha^{-1}(\Phi(Z_i)) \quad \Phi\text{: standard normal CDF}, \; F_\alpha\text{: } t_\alpha\text{-distribution CDF}, \alpha > 1,$$

where $\{Z_i\}$ is fractional Gaussian noise with Hurst index H_0. X_i mean 0 and has marginal t distribution with α degrees of freedom which is heavy-tailed: $P(|X_i| > x)$ behaves like $x^{-\alpha}$. $\mathrm{Var}[X_i] = \infty$ when $1 < \alpha < 2$ but $\mathbb{E}|X_i| < \infty$. $G(\cdot)$ has (generalized) Hermite rank 1. By Sly and Heyde (2008), Assumption A2 holds with the following dichotomy (for $0 < H_0 < 1$, $1 < \alpha < 2$):

$$\begin{cases} H = 1/\alpha, \; \ell(n) = 1, \; Y(t) = c_\alpha L_\alpha(t) & \text{if } H_0 < 1/\alpha; \\ H = H_0, \; \ell(n) = 1, \; Y(t) = c_H B_H(t) & \text{if } H_0 > 1/\alpha, \end{cases}$$

where $B_H(t)$ is the fractional Brownian motion and $L_\alpha(t)$ is the standard (scale parameter $\sigma = 1$) symmetric α-stable Lévy process. Since $\{Z_i\}$ is fractional Gaussian noise, Assumption A3 holds.



α \ H_0	0.25	0.5	0.75
1.5	(76,74)	(81,81)	(79,78)
2	(78,78)	(85,86)	(82,82)
5	(90,89)	(89,89)	(88,86)
10	(90,89)	(89,89)	(87,87)

Table: Monte Carlo evaluation of coverage percentage (lower 90%, upper 90%). Sample size=500.

Figure: The running confidence cutoff for a sample path. $\alpha = 1.5$, $H_0 = 0.75 > 1/\alpha = 2/3$.

Example: stochastic duration

$\{Z_i\}$: fractional Gaussian noise with Hurst index H_0. Data:

$$X_i = \exp(Z_i)\xi_i, \quad \xi_i \overset{i.i.d.}{\sim} F(1, 2\alpha), \quad \alpha > 1,$$

$F(1, 2\alpha)$: F-distribution with parameters 1 and 2α. X_i is positively skewed, dependent when $H_0 \in (1/2, 1)$, and heavy tailed: $P(X_i > x)$ behaves like $x^{-\alpha}$. $\mathbb{E}|X_i| < \infty$ but $\mathrm{Var}[X_i] = \infty$ when $1 < \alpha < 2$.
ξ_i can be rewritten as $G(Z_i')$ for suitable function $G(\cdot)$ and i.i.d. Gaussian $\{Z_i'\}$ which is independent of $\{Z_i\}$.
So X_i is subordinated to (Z_i, Z_i'). The mean is $\mu = \mathbb{E}X_i = \mathbb{E}\exp(Z_i)\mathbb{E}\xi_i = \exp(1/2)\alpha/(\alpha - 1)$.
By Beran et al. (2013), Assumption A2 holds with the following dichotomy:

$$\begin{cases} H = 1/\alpha,\ \ell(n) = 1,\ Y(t) = c_\alpha L_{\alpha,1,1}(t) & \text{if } H_0 < 1/\alpha; \\ H = H_0,\ \ell(n) = 1,\ Y(t) = c_H B_H(t) & \text{if } H_0 > 1/\alpha, \end{cases}$$

where $B_H(t)$ is the fractional Brownian motion and $L_{\alpha,1,1}(t)$ is standard $(\sigma = 1)$ α-stable Lévy process totally skewed to the right $(\beta = 1)$.

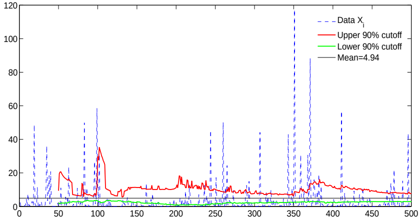


Figure: The running confidence cutoff for a sample path. $\alpha = 1.5, H_0 = 0.75 > 1/\alpha$.

α \ H_0	0.25	0.5	0.75
1.5	(86,91)	(86,92)	(84,93)
2	(85,94)	(84,95)	(82,94)
5	(84,95)	(84,96)	(80,93)
10	(83,95)	(83,96)	(80,93)

Table: Monte Carlo evaluation of coverage percentage (lower 90%, upper 90%). Sample size=500. Extended to $H_0 \leq 1/2$.

Weak dependence

Two types of weak dependence:

(1) $X_i = G(Z_i, \ldots, Z_{i-l})$ with Z_i LRD Gaussian but $\gamma(k) = \mathrm{Cov}[X(k), X(0)]$ is summable.

(2) X_i is strong mixing.

$$(1) \not\Rightarrow (2) \quad \text{and} \quad (2) \not\Rightarrow (1).$$

What happens if (2) replaces (1) in the assumptions?

The strong mixing case: theoretical assumptions

Gaussian subordination case:

A1. $X_i = G(Z_i, \ldots, Z_{i-l})$ with mean $\mu = \mathbb{E}X_i$, where l is a fixed non-negative integer;

A2. We have weak convergence in $D[0,1]$ with a suitable Skorohod topology:

$$\left\{ \frac{1}{n^H \ell(n)} (S_{\lfloor nt \rfloor} - n\mu),\ 0 \le t \le 1 \right\} \Rightarrow \{Y(t),\ 0 \le t \le 1\},$$

for some process $Y(t)$, where $0 < H < 1$ and $\ell(\cdot)$: slowly varying;

A3. As $n \to \infty$, the block size $b_n \to \infty$, $b_n = o(n)$, and satisfies $\sum_{k=0}^{n} \rho_{k,l+b_n} = o(n)$.

Strong mixing: $\alpha(k) = \sup \left\{ |P(A)P(B) - P(A \cap B)|,\ A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_k^{\infty} \right\} \to 0$ as $k \to \infty$.

B1. $\{X_i\}$ is a strong mixing stationary process with mean $\mu = \mathbb{E}X_i$.

B2. Same as A2.

B3. The block size $b_n \to \infty$ and $b_n = o(n)$ as $n \to \infty$.

Example: moving average

$$X_i = \epsilon_i + a\epsilon_{i-1}, \quad \epsilon_i \overset{i.i.d.}{\sim} t_\alpha,$$

where $a > 0$, t_α is the t-distribution with degrees of freedom α.
$\{X_i\}$ is 2-dependent and thus strong mixing.
By Avram and Taqqu (1992), Assumption B2 holds with (in the Skorohod M_2 topology) the following dichotomy:

$$\begin{cases} H = 1/2, \ \ell(n) = 1, \ Y(t) = \sigma B(t) & \text{if } \alpha > 2; \\ H = 1/\alpha, \ \ell(n) = 1, \ Y(t) = c_\alpha L_\alpha(t) & \text{if } 1 < \alpha < 2, \end{cases}$$

where $\sigma^2 = \sum_n \mathrm{Cov}[X(n), X(0)]$, $c_\alpha > 0$, $B(t)$ is the standard Brownian motion, $L_\alpha(t)$ is the symmetric α-stable Lévy motion.

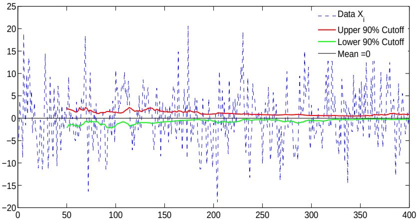


Figure: The running confidence cutoff for a sample path. $a = 5, \alpha = 5$.

α \ a	1	2	5
1.5	(82,82)	(82,82)	(81,81)
2	(88,86)	(86,88)	(86,86)
5	(90,91)	(91,90)	(90,90)
10	(90,91)	(90,90)	(90,90)

Table: Monte Carlo evaluation of coverage percentage (lower 90%, upper 90%). Sample size=500.

Example: GARCH(1,1)

ϵ_i i.i.d. standard Gaussian

Data:

$$X_i = \sigma_i \epsilon_i$$

$$\sigma_i^2 = c + aX_{i-1} + b\sigma_{i-1}^2, \quad a, b, c > 0, \; a + b < 1.$$

It is strong mixing with a geometric decay mixing coefficient and $\mathbb{E}|X_i|^{2+\delta} < \infty$ for $\delta > 0$ small enough (Lindner (2009)). Hence by Herrndorf (1984), Assumption B2 holds with

$$H = 1/2, \; \ell(n) = 1, \; Y(t) = \sigma^2 B(t),$$

where $\sigma^2 = \sum_n \mathrm{Cov}[X(n), X(0)]$ and $B(t)$ is the standard Brownian motion.

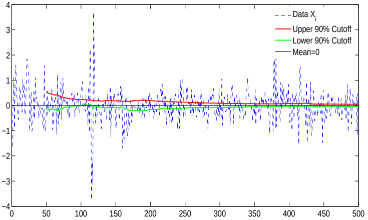


Figure: The running confidence cutoff for a sample path. $(a, b, c) = (0.2, 0.6, 0.1)$.

(a, b):	$(0.7, 0.1)$	$(0.5, 0.3)$	$(0.2, 0.6)$
	(89,89)	(88,88)	(87,86)

Table: Monte Carlo evaluation of coverage percentage (lower 90%, upper 90%). Sample size=500. $c = 0.1$

S. Bai, M.S. Taqqu, and T. Zhang. A unified approach to self-normalized block sampling. *arXiv Preprint arXiv:1512.00820*, to appear in *Stochastic Processes and Their Applications*, 2016.

Bai, S. and Taqqu, M. S. On the validity of resampling methods under long memory. *arXiv Preprint arXiv:1512.00819*, 2015.

Thank you!

Additional slides

Challenge 1: Short-range dependence

$\{X_i\}$ stationary weakly dependent (short-range dependent) with covariance

$$\gamma(k) = \mathrm{Cov}[X(k), X(0)], \quad k \in \mathbb{Z}, \quad \text{satisfying } \sum_k |\gamma(k)| < \infty.$$

Central Limit Theorem:

$$n^{-1/2} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{d} N(0, \sigma^2),$$

where now σ^2 is the so-called *long-run variance*

$$\sigma^2 = \sum_{k=-\infty}^{\infty} \gamma(k).$$

To construct a confidence interval, we need a consistent estimator for σ^2.

We have the sample covariance: $\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} (X_i - \bar{X}_n)(X_{i+k} - \bar{X}_n)$. But σ^2 cannot be estimated simply by $\sum_k \hat{\gamma}(k)$ (too few summands for large k).

Typical estimator is the lag window which regularizes an infinite-dimensional problem by exploiting the "sparsity" $\gamma(k) \approx 0$ for large k:

$$\hat{\sigma} = \sum_{|k| \le h} \hat{\gamma}(k) W(k/h),$$

where $W(k)$ is the lag-window function, $h \in \mathbb{Z}_+$ is the bandwidth.

Challenge 2: Heavy tails

$\{X_i\}$ i.i.d. heavy-tailed

$$P(X_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \quad P(X_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha} \quad x \to +\infty, \tag{3}$$

where constant $A > 0$, parameters $\beta \in [-1, 1]$, $\alpha \in (1, 2)$.

$\mathbb{E}|X_1| < \infty$ but $\mathbb{E}|X_1|^2 = \infty$.

Heavy tail Central Limit Theorem:

$$n^{-1/\alpha}\sum_{i=1}^{n}(X_i - \mu) \to S_\alpha(\sigma, \beta, 0)$$

where $S_\alpha(\sigma, \beta, 0)$ is the α-stable random variable with location parameter 0, scale parameter σ (depending on A and α) and skewness parameter β.

How about the unknown α, β, A?

Even more complicated situation: a slowly varying function replaces the constant A in (3).

Even more complicated: $\{X_i\}$ are weakly dependent (σ then depends on dependence structure).

Challenge 3: Long-range dependence

$\{X_i\}$ is strongly dependent (long-range dependent), with covariance function

$$\gamma(k) \sim c_\gamma k^{2H-2}, \quad H \in (1/2, 1). \tag{4}$$

Some models of $\{X_i\}$, e.g., nonlinear transform of a long-range dependent Gaussian process, give rise to limit theorem (Dobrushin Major (1979), Taqqu 1979):

$$\frac{1}{n^H} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{d} cZ_{m,H},$$

where c depends on c_γ and H, and m is a positive integer (the so-called Hermite rank), and

$$Z_{m,H} = v_{m,H} \int'_{\mathbb{R}^m} \int_0^1 \prod_{j=1}^{m} (s - x_j)_+^{(H-1)/m - 1/2} ds \; B(dx_1) \dots B(dx_m), \quad B(\cdot)\text{: Brownian motion}$$

is a standardized random variable expressed by a multiple Wiener-Itô integral which is non-Gaussian if $m \geq 2$.

Need to estimate c_γ, H, m (no available method for m).
More complicated if c_γ in (4) is replaced by a slowly varying function.

How about bootstrap?

If $\{X_i\}$ is short-range dependent, one can do the following block bootstrap. Let

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

1. Choose a block size b. Form $n - b + 1$ successive blocks (with overlap) $\mathbf{X}_1^b, \mathbf{X}_2^{b+1}, \ldots, \mathbf{X}_{n-b+1}^n$.
2. Sample randomly with replacement $[n/b]$ blocks. Paste them into a new time series $\mathbf{X}^*$ of length $b \times [n/b]$. Obtain the sample mean $\bar{X}^*$.
3. Repeat this N times, getting N bootstrapped sample mean $\bar{X}_1^*, \ldots, \bar{X}_N^*$.
4. Use the empirical distribution of $\{\bar{X}_i^*\}$ to construct confidence interval.

But this does NOT work for long-range dependent case. The strong dependence is destroyed by randomly sampling and pasting the blocks in Step 2.

Idea for remedy: replace pasting by re-scaling.

Block sampling (sampling window bootstrap)

Idea: No resampling. Include all blocks.
Form $n - b + 1$ successive blocks (overlapping) $\mathbf{X}_1^b, \mathbf{X}_2^{b+1}, \ldots, \mathbf{X}_{n-b+1}^n$, $b \ll n$.

For each block $\mathbf{X}_i^{b+i-1}$, obtain the block mean $\bar{X}_i^* = b^{-1} \sum_{j=i}^{i+b-1} X_j$.
Renormalize it (deterministicly) to get convergence to some limit T.

We cannot directly use the empirical distribution of $\{\bar{X}_i^*\}$, because the block means $\bar{X}_i^*$ fluctuate more than the overall sample mean $\bar{X}_n$ since $b \ll n$. To get the same level of fluctuation, rescale $\bar{X}_i^*$ by

$$r_{b,n} = \frac{\sqrt{\mathrm{Var}[\bar{X}_n]}}{\sqrt{\mathrm{Var}[\bar{X}_i^*]}},$$

and use the empirical distribution of $\{r_{b,n} \bar{X}_i^*\}$ as a surrogate to that of T.

Hall et al. (1998) and Zhang et al. (2013) estimate $r_{b,n}$ under long-range dependence. They use further block sampling and thus involve some tuning parameters in addition to b.

Caveat: block sampling frees one from knowing the asymptotic distribution, but one needs to estimate the normalization.

Proof of $\mathrm{Var}[\widehat{F}^*_{n,b_n}(x)] \to 0$

$$\mathrm{Var}[\widehat{F}^*_{n,b_n}(x)] = \mathrm{Var}\left[\frac{1}{n-b_n+1}\sum_{i=1}^{n-b_n+1} \mathrm{I}\{T^*_{i,b_n} \le x\}\right]$$
$$\le \frac{2}{n-b_n+1}\sum_{k=0}^{n}\left|\mathrm{Cov}\left[\mathrm{I}\{T^*_{1,b_n} \le x\}, \mathrm{I}\{T^*_{k+1,b_n} \le x\}\right]\right|.$$

Gaussian maximal correlation equality (Kolmogorov and Rozanov (1960)):

$$\sup_{F,G\in L^2(\mathbf{Z}_1^m)}\left|\mathrm{Corr}\big(F(\mathbf{Z}_1^m), G(\mathbf{Z}_{k+1}^{k+m})\big)\right| = \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^m}\left|\mathrm{Corr}\Big(\langle \mathbf{x}, \mathbf{Z}_1^m\rangle, \langle \mathbf{y}, \mathbf{Z}_{k+1}^{k+m}\rangle\Big)\right| =: \rho_{k,m}$$

One has

$$\left|\mathrm{Cov}[\mathrm{I}\{T^*_{1,b_n} \le x\}, \mathrm{I}\{T^*_{k+1,b_n} \le x\}]\right| \le \frac{1}{4}\left|\mathrm{Corr}[\mathrm{I}\{T^*_{1,b_n} \le x\}, \mathrm{I}\{T^*_{k+1,b_n} \le x\}]\right|$$
$$\le \frac{1}{4}\rho_{k,b_n+l}.$$

Bounding the correlation by 1 for $k < l + b_n$, we have

$$\mathrm{Var}[\widehat{F}^*_{n,b_n}(x)] \le \frac{1}{2(n-b_n+1)}\sum_{k=0}^{n}\rho_{k,b_n+l},$$

which converges to zero by Assumption A3.

Alternative sufficient condition for Assumption A3

Recall that $\gamma(k) = \mathrm{Cov}[Z_0, Z_k]$, $\{Z_i\}$ Gaussian, $X_i = G(Z_i, \dots, Z_{i-l})$.

$$M_\gamma(n) = \max_{k>n} |\gamma(k)|, \quad \lambda_m = \text{ minimum eigenvalue of } (\gamma(i-j))_{i,j=1,\dots,m}$$

Then

$$\sum_{k=0}^{n} \min\left\{ \frac{b_n}{\lambda_{b_n+l}} M_\gamma(k), 1 \right\} = o(n) \quad \Longrightarrow \quad \text{A3.} \tag{5}$$

If the spectral density has zeros, the minimum eigenvalue λ_m converges to zero with a rate which depends on the order of the zeros.

In the case of long memory, $\gamma(n) \sim cn^{2H-2}$, $H \in (1/2, 1)$ and the spectral density of $\{Z_i\}$ is bounded below away from zero. Then we have $M_\gamma(k) \sim ck^{2H-2}$ and so

$$b_n = o(n^{2-2H}), \quad 0 < 2 - 2H < 1. \quad \Longrightarrow \quad (5)$$

The proof thus avoids dealing with the complicated specific forms of $F(\cdot)$ and $G(\cdot)$.

Lemma

$$\rho_{k,m} := \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^m} \left|\mathrm{Corr}\Big(\langle \mathbf{x}, \mathbf{Z}_1^m\rangle, \langle \mathbf{y}, \mathbf{Z}_{k+1}^{k+m}\rangle\Big)\right| \le m\frac{M_\gamma(k-m)}{\lambda_m}.$$

Proof.

$$\text{2nd term} = \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^m} \frac{\mathbf{x}^T\Sigma_{k,m}\mathbf{y}}{\sqrt{\mathbf{x}^T\Sigma_m\mathbf{x}}\sqrt{\mathbf{y}^T\Sigma_m\mathbf{y}}} \le \sup_{\mathbf{x},\mathbf{y}\in\mathbb{R}^m} \frac{\left|\mathbf{x}^T\Sigma_{k,m}\mathbf{y}\right|}{\lambda_m\|\mathbf{x}\|\|\mathbf{y}\|} \le \frac{|\sigma_{k,m}|}{\lambda_m} \le m\frac{M_\gamma(k-m)}{\lambda_m}.$$

$$\Sigma_m = \text{covariance matrix } \left(\mathbb{E}Z_{i_1}Z_{i_2}\right)_{1\le i_1,i_2\le m},$$
$$\Sigma_{k,m} = \text{covariance matrix } \left(\mathbb{E}Z_{i_1}Z_{i_2+k}\right)_{1\le i_1,i_2\le m}.$$

$$\lambda_m = \text{smallest eigenvalue of } \Sigma_m,$$
$$\sigma_{k,m} = \text{largest singular value of } \Sigma_{k,m}.$$

$$\sigma_{k,m} \le \text{linear size } \times \text{ largest entry } \le m \max_{n>k-m} |\mathbb{E}Z_0Z_n| = mM_\gamma(k-m).$$

Note: $\Sigma_{k,m}$ is not a symmetric matrix. The square of its singular values are the eigenvalues of $\Sigma_{k,m}^T\Sigma_{k,m}$, which is symmetric and non-negative definite. □

Weak dependence

Two types of weak dependence:

(1) $X_i = G(Z_i, \dots, Z_{i-l})$ with Z_i LRD Gaussian but $\gamma(k) = \mathrm{Cov}[X(k), X(0)]$ is summable.

(2) X_i is strong mixing.

Proof (1) $\not\Rightarrow$ (2): If Z_i is LRD and $P(Z_i)$ is SRD, then $P(Z_i)$ may not be strong mixing. If it were, then there are cases where we may be able to find a polynomial Q such that $Q(P(Z_i))$ is strong mixing, obeying the CLT, but the at the same time $Q(P(Z_i))$ is LRD.

Proof (2) $\not\Rightarrow$ (1): Consider for example the trivial case $\{X_i\}$ i.i.d. Gaussian. There is no $\{X_i'\} \overset{f.d.d.}{=} \{X_i\}$ so that $X_i' = G(Z_i')$, where $\{Z_i'\}$ is LRD Gaussian, because the covariance $\mathrm{Cov}[X_i', X_0'] \neq 0$ for large i.

Self-normalization:

Ignacio N. Lobato (2001)
Testing that a dependent process is uncorrelated.
Journal of the American Statistical Association, 96(455):1066–1076.
*** Uses D_n.

X. Shao (2010)
A self-normalized approach to confidence interval construction in time series.
Journal of the Royal Statistical Society: Series B, 72:343–366.
*** Extended the use of D_n.

Block sampling:

P. Hall, B.-Y. Jing, and S. N. Lahiri (1998)
On the sampling window method for long-range dependent data.
Statistica Sinica, 8:1189–1204.
*** Block sampling under "LRD" with deterministic normalization.

A. Jach, T. McElroy, and D. N. Politis (2012).
Subsampling inference for the mean of heavy-tailed long-memory time series.
Journal of Time Series Analysis, 33(1):96–111.
*** Specific random normalization. Selection of the block size.

T. Zhang, H-C Ho, M. Wendler, and W.B. Wu (2013)
Block sampling under strong dependence.
Stochastic Processes and their Applications, 123:2323–2339.
*** Same as Hall et al. for finite variance transformation of LRD linear processes with deterministic normalization.

Bai, S. and Taqqu, M. S. (2015).
On the validity of resampling methods under long memory.
arXiv Preprint arXiv:1512.00819.

Betken, A. and Wendler, M. (2015).
Subsampling for general statistics under long range dependence.
arXiv preprint arXiv:1509.05720.

Functional limit theorems under strong mixing:

N. Herrndorf (1984)
A functional central limit theorem for weakly dependent sequences of random variables.
The Annals of Probability, pages 141–153.

Limit theorems under long-range dependence:

R. L. Dobrushin and P. Major (1979)
Non-central limit theorems for non-linear functional of gaussian fields.
Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 50:27–52.

M.S. Taqqu (1979)
Convergence of integrated processes of arbitrary Hermite rank.
Probability Theory and Related Fields, 50(1):53–83.

A. Sly and C. Heyde (2008)
Nonstandard limit theorem for infinite variance functionals.
The Annals of Probability, 36(2):796–805.

J. Beran, Y. Feng, S. Ghosh, and R. Kulik (2013).
Long-Memory Processes.
Springer.

Gaussian maximal correlation inequality:

A.N. Kolmogorov and Y.A. Rozanov (1960)
On strong mixing conditions for stationary Gaussian processes.
Theory of Probability & Its Applications, 5(2):204–208.

GARCH

[Lindner(2009)] Alexander M Lindner. Stationarity, mixing, distributional properties and moments of garch (p, q)–processes. In *Handbook of financial time series*, pages 43–69. Springer, 2009.