BOOTSTRAPPING

We discuss some bootstrapping methods for time series data, but we first introduce the bootstrap for the iid setting.

IID BOOTSTRAP

Let $X_1, \dots, X_n$ be indep. r.v.s with cdf F.

Let $T_n = t_n(X_1, \dots, X_n, F)$ be a function of $X_1, \dots, X_n$ as well as of the cdf F, so that it may involve some parameters which are functionals of F.

Suppose it is of interest to estimate either some quantities based on moments of the sampling distribution of T_n, say

$$\mathbb{E}\, T_n \quad \text{or} \quad \mathrm{Var}\, T_n$$

or the cdf of the sampling distribution $G_n(x) = P(T_n \le x)$.

<u>Example</u>: Consider $T_n = t_n(X_1, \dots, X_n, F) = \sqrt{n}(\bar{X}_n - \mu)/S_n$, where

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n}, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \quad \text{and} \quad \mu = \int x\, dF(x).$$

In order to build a confidence interval for μ, we would like to estimate the cdf G_n given by

$$G_n(x) = P\left(\sqrt{n}(\bar{X}_n - \mu)/S_n \le x\right).$$

Let $G_{n,\xi} = \inf\{x : G(x) \ge 1-\xi\}$ be the upper ξ-quantile of T_n. Then we have

$$P\left(G_{n,1-\alpha/2} < \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} < G_{n,\alpha/2}\right) = 1-\alpha,$$

which can be rearranged as

$$P\left(\bar{X}_n - G_{n,\alpha/2}\frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n - G_{n,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right) = 1-\alpha,$$

which shows us that a $(1-\alpha)*100\%$ C.I. for μ may be constructed as

$$\left(\bar{X}_n - G_{n,\alpha/2}\frac{S_n}{\sqrt{n}},\ \bar{X}_n - G_{n,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right).$$

But to construct this interval we must know G_n.

The idea of the bootstrap is to estimate the cdf F with some estimator $\hat{F}_n$ based on $X_1, \dots, X_n$ and then define

$$T_n^* = t_n(X_1^*, \dots, X_n^*, \hat{F}_n),$$

where $X_1^*, \dots, X_n^*$ are independent r.v.s with distribution $\hat{F}_n$, conditionally on $X_1, \dots, X_n$.

Let $\mathbb{E}_* T_n^* = \mathbb{E}[T_n^* | X_1, \dots, X_n]$ and $\text{Var}_* T_n^* = \text{Var}[T_n^* | X_1, \dots, X_n]$, so that $\mathbb{E}_*$ and Var_* are expectation and variance conditional on $X_1, \dots, X_n$. Then $\mathbb{E}_* T_n^*$ and $\text{Var}_* T_n^*$ are the bootstrap estimators of $\mathbb{E} T_n$ and $\text{Var}\, T_n$, respectively.

Let G_n^* be the cdf given by

$$G_n^*(x) = P_*(T_n^* \leq x) = P(T_n^* \leq x \,|\, X_1, \dots, X_n),$$

where P_* denotes probability conditional on $X_1, \dots, X_n$. Then G_n^* is the bootstrap estimator of G_n.

<u>Obtaining a Monte-Carlo approximation of G_n^*:</u>

In many bootstrap applications, we cannot get $\mathbb{E}_* T_n^*$, $\text{Var}_* T_n^*$, or G_n^* analytically. We therefore approximate them using a Monte Carlo procedure like the following:

Given $X_1, \dots, X_n$, compute $\hat{F}_n$. Choose a large integer B.

For $b = 1, \dots, B$ do:

Sample $X_1^{*,b}, \dots, X_n^{*,b}$ as indep r.v.s from $\hat{F}_n$.

Compute $T_n^{*,b} = t_n(X_1^{*,b}, \dots, X_n^{*,b}, \hat{F}_n)$.

Then set $\mathbb{E}_* T_n^* = \frac{1}{B} \sum_{b=1}^{B} T_n^{*,b}$ and $\text{Var}_* T_n^* = \frac{1}{B} \sum_{b=1}^{B} (T_n^{*,b} - \mathbb{E}_* T_n^*)^2$

as well as $G_n^*(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}(T_n^{*,b} \leq x)$ for all x.

To retrive quantiles of G_n^*, sort the values $T_n^{*,1}, \dots, T_n^{*,B}$.

If $T_n^{*(1)} < \dots < T_n^{*(B)}$ represent the ordered values, then set

$$G_{n,\xi}^* = T_n^{*,(\lfloor B(1-\xi) \rfloor)}.$$

Example (continued): Consider $T_n = t_n(X_1,\dots,X_n, F) = \sqrt{n}(\bar{X}_n - \mu)/S_n$.

A natural choice of $\hat{F}_n$ is the empirical distribution $\mathbb{F}_n$, which has cdf given by

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \le x).$$

The boostrap version of T_n is

$$T_n^* = t_n(X_1^*,\dots,X_n^*, \mathbb{F}_n) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n)/S_n^*,$$

where

$$\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^n X_i^*, \quad S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^n (X_i^* - \bar{X}_n^*)^2, \qquad \bar{X}_n = \int x \, d\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^n X_i.$$

Then G_n^* is the cdf given by

$$G_n^*(x) = P_*\left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n)/S_n^* \le x\right).$$

So the $(1-\alpha)*100\%$ C.I. for μ based on the bootstrap estimate of G_n would be given by

$$\left(\bar{X}_n - G^*_{n,\alpha/2}\frac{S_n}{\sqrt{n}},\ \bar{X}_n - G^*_{n,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right),$$

but we need to get Monte-Carlo approximations to $G^*_{n,\alpha/2}$ and $G^*_{n,1-\alpha/2}$.

To obtain a Monte-Carlo approximation to G_n^*, choose a large integer B, and for $b = 1,\dots,B$ do:

Draw $X_1^{*(b)},\dots,X_n^{*(b)}$ as indep. r.v.s from dist with cdf $\mathbb{F}_n$, which can be done by sampling with replacement from $X_1,\dots,X_n$.

Compute $T_n^{*,b} = \sqrt{n}(\bar{X}_n^{*,b} - \bar{X}_n)/S_n^{*,b}$.

We use as the upper ξ-quantile of G_n^* the value $T_n^{*,(\lfloor B(1-\xi)\rfloor)}$, so the bootstrap $(1-\alpha)*100\%$ C.I. for μ would be given by

$$\left(\bar{X}_n - T_n^{*(\lfloor B(1-\xi)\rfloor)}\frac{S_n}{\sqrt{n}},\ \bar{X}_n - T_n^{*(\lfloor B\xi\rfloor)}\frac{S_n}{\sqrt{n}}\right)$$

First results for the IID bootstrap for the sample mean:

Suppose $X_1, \dots, X_n$ are independent r.v.s with cdf F, mean μ, and variance $\sigma^2 < \infty$.

Let

$$T_n = t_n(X_1, \dots, X_n, F) = \sqrt{n}(\bar{X}_n - \mu),$$

and suppose we are interested in estimating

$$\mathbb{E}\, T_n = \mathbb{E}\sqrt{n}(\bar{X}_n - \mu) = 0$$

$$\operatorname{Var} T_n = \operatorname{Var}\left(\sqrt{n}(\bar{X}_n - \mu)\right) = \sigma^2$$

using the IID bootstrap.

Let $\hat{F}_n$ be the empirical distribution function based on $X_1, \dots, X_n$. Then

$$T_n^* = t_n(X_1^*, \dots, X_n^*, \hat{F}_n) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n),$$

since $\quad \bar{X}_n = \int x \, d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n X_i$.

We have

$$\begin{aligned}
\mathbb{E}_* \sqrt{n}(\bar{X}_n^* - \bar{X}_n) &= \sqrt{n}\, \mathbb{E}_* \frac{1}{n}\sum_{i=1}^n (X_i^* - \bar{X}_n) \\
&= \sqrt{n}\, \mathbb{E}_* (X_1^* - \bar{X}_n) \\
&= \sqrt{n}\, \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n) \\
&= 0,
\end{aligned}$$

so the bootstrap estimate of $\mathbb{E}\, T_n$ is exactly equal to $\mathbb{E}\, T_n = 0$.

We also have

$$\begin{aligned}
\operatorname{Var}_* T_n^* &= \operatorname{Var}_*\left[\sqrt{n}(\bar{X}_n^* - \bar{X}_n)\right] \\
&= n\, \mathbb{E}_* (\bar{X}_n^* - \bar{X}_n)^2
\end{aligned}$$

$$= n\,\mathbb{E}_*\left(\frac{1}{n}\sum_{i=1}^{n}(X_i^* - \bar{X}_n)\right)^2$$

$$= \frac{1}{n}\,\mathbb{E}_*\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i^* - \bar{X}_n)(X_j^* - \bar{X}_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_*(X_i^* - \bar{X}_n)^2$$

$$= \mathbb{E}_*(X_1^* - \bar{X}_n)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$\xrightarrow{p} \sigma^2 \quad \text{as} \quad n \to \infty.$$

We see that the boostrap estimator of $\operatorname{Var} T_n = \sigma^2$ is consistent.

Note that in this situation, we could find $\mathbb{E}_* T_n$ and $\operatorname{Var}_* T_n$ without doing any Monte-Carlo simulations!

The bootstrap does not always entail resampling from the data!

* R example *

TIME SERIES BOOTSTRAP APPROACHES

Suppose $\{X_t, t \in \mathbb{Z}\}$ is a stationary time series such that $X_1, \dots, X_n$ have joint distribution $\mathbb{F}_n$. We will use $\mathbb{F}_n$ to represent the joint distribution as well as the joint distribution function.

If we were to follow the same steps as in the IID case, given some quantity

$$T_n = t_n(X_1, \dots, X_n, \mathbb{F}_n),$$

we would define the bootstrap version T_n^* of T_n as

$$T_n^* = t_n(X_1^*, \dots, X_n^*, \hat{\mathbb{F}}_n),$$

where $\hat{\mathbb{F}}_n$ is an estimator of the joint distribution $\mathbb{F}_n$ and $X_1^*, \dots, X_n^*$ are random variables with joint distribution $\hat{\mathbb{F}}_n$.

The problem with this is that it is difficult to get a good estimator $\hat{\mathbb{F}}_n$ of $\mathbb{F}_n$.

There are many bootstrapping schemes for time series data. We first consider block-bootstrap methods.

BLOCK-BASED METHODS

Block-bootstrap methods are based on the assumption that the joint distribution $\mathbb{F}_n$ of $X_1, \dots, X_n$ can be approximated by the product of joint distributions of blocks of random variables among $X_1, \dots, X_n$. That is, if we break $X_1, \dots, X_n$ into blocks of size ℓ, assuming for the moment that n/ℓ is integer-valued for the sake of simplicity, then $\mathbb{F}_n$ may be written as

$$\begin{aligned} \mathbb{F}_n(x_1, \dots, x_n) &= P(X_1 \le x_1, \dots, X_n \le x_n) \\ &\approx \prod_{j=1}^{n/\ell} P\left(X_{(j-1)\ell+1} \le x_{(j-1)\ell+1}, \dots, X_{j\ell} \le x_{j\ell}\right) \\ &= \prod_{j=1}^{n/\ell} \mathbb{F}_\ell(x_{(j-1)\ell+1}, \dots, x_{j\ell}), \end{aligned}$$

where $\mathbb{F}_\ell(x_1, \dots, x_\ell) = P(X_1 \le x_1, \dots, X_\ell \le x_\ell)$. Such an approximation to $\mathbb{F}_n$ can be made under strict stationarity and if the dependence between r.v.s decreases the further apart they are in time.

The blocks are constructed according to this diagram:

$$\underbrace{X_1 \cdots X_\ell}_{\text{Block } 1} \ \underbrace{X_{\ell+1} \cdots X_{2\ell}}_{\text{Block } 2} \ \cdots \ \underbrace{X_{(j-1)\ell+1} \cdots X_{j\ell}}_{\text{Block } j} \ \cdots \ \underbrace{X_{n-\ell+1} \cdots X_n}_{\text{Block } n/\ell}$$

A block-bootstrap version of $T_n = t_n(X_1, \dots, X_n, \mathbb{F}_n)$ may be defined as

$$T_n^* = t_n(X_1^*, \dots, X_n^*, \hat{\mathbb{F}}_n^{\text{block}}),$$

where $\hat{\mathbb{F}}_n^{\text{block}}$ is the distribution with cdf given by

$$\hat{\mathbb{F}}_n^{\text{block}}(x_1, \dots, x_n) = \prod_{j=1}^{n/\ell} \hat{\mathbb{F}}_\ell(x_{(j-1)\ell+1}, \dots, x_{j\ell}),$$

where $\hat{\mathbb{F}}_\ell$ is an estimator of $\mathbb{F}_\ell$.

We consider two ways of estimating $\mathbb{F}_\ell$:

Non-overlapping-block bootstrap (NBB):

Denote by $\hat{\mathbb{F}}_n^{\text{NBB}}$ the NBB estimator of $\mathbb{F}_n$. It has cdf given by

$$\hat{\mathbb{F}}_n^{\text{NBB}}(x_1, \dots, x_n) = \prod_{j=1}^{n/\ell} \hat{\mathbb{F}}_\ell^{\text{NBB}}(x_{(j-1)\ell+1}, \dots, x_{j\ell}),$$

where $\hat{\mathbb{F}}_\ell^{\text{NBB}}$ is the distribution placing mass $1/(n/\ell)$ on each of the blocks

$$(X_{(j-1)\ell+1}, \dots, X_{j\ell}), \quad j = 1, \dots, n/\ell.$$

To generate a Monte-Carlo draw $X_1^*, \dots, X_n^*$ from $\hat{\mathbb{F}}_n^{\text{NBB}}$, sample with replacement n/ℓ times from the blocks

$$(X_{(j-1)\ell+1}, \dots, X_{j\ell}), \quad j = 1, \dots, n/\ell$$

and concatenate the sampled blocks to obtain $X_1^*, \dots, X_n^*$.

Application of NBB to the mean:

Let $T_n = t_n(X_1, \dots, X_n, \mathbb{F}_n) = \sqrt{n}(\bar{X}_n - \mathbb{E}\bar{X}_n)$, and suppose we wish to estimate $\operatorname{Var} T_n$ using the NBB.

The NBB version of T_n is

$$T_n^* = t_n(X_1^*, \dots, X_n^*, \hat{\mathbb{F}}_n^{\text{NBB}}) = \sqrt{n}(\bar{X}_n^* - \mathbb{E}_* \bar{X}_n^*),$$

where $\mathbb{E}_*$ is expection conditional on $X_1, \dots, X_n$ and based on the distribution $\hat{\mathbb{F}}_n^{\text{NBB}}$.

The NBB estimator of $\operatorname{Var} T_n$ is $\operatorname{Var}_* T_n^*$, where Var_* operates according to $\hat{\mathbb{F}}_n^{NBB}$.

To find $\operatorname{Var}_* T_n^*$, we first find $\mathbb{E}_* \bar{X}_n^*$. We have

$$\begin{aligned}
\mathbb{E}_* \bar{X}_n^* &= \mathbb{E}_* \frac{1}{n} \sum_{i=1}^{n} X_i^* \\
&= \frac{1}{n} \mathbb{E}_* \sum_{j=1}^{n/\ell} \left(X^*_{(j-1)\ell+1} + \dots + X^*_{j\ell} \right) \\
&= \frac{1}{\ell} \mathbb{E}_* \left(X_1^* + \dots + X_\ell^* \right) \\
&= \frac{1}{\ell} \frac{1}{n/\ell} \sum_{j=1}^{n/\ell} \left(X_{(j-1)\ell+1} + \dots + X_{j\ell} \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} X_j \\
&= \bar{X}_n .
\end{aligned}$$

Now $\operatorname{Var}_* T_n^*$ is given by

$$\begin{aligned}
\operatorname{Var}_* \left[\sqrt{n} \left(\bar{X}_n^* - \bar{X}_n \right) \right] &= n \, \mathbb{E}_* \left(\bar{X}_n^* - \bar{X}_n \right)^2 \\
&= n \, \mathbb{E}_* \left[\frac{1}{n} \sum_{i=1}^{n} \left(X_i^* - \bar{X}_n \right) \right]^2 \\
&= \frac{1}{n} \mathbb{E}_* \left(\sum_{j=1}^{n/\ell} \left[\left(X^*_{(j-1)\ell+1} - \bar{X}_n \right) + \dots + \left(X^*_{j\ell} - \bar{X}_n \right) \right] \right)^2 \\
&= \frac{1}{n} \sum_{j=1}^{n/\ell} \mathbb{E}_* \left[\left(X^*_{(j-1)\ell+1} - \bar{X}_n \right) + \dots + \left(X^*_{j\ell} - \bar{X}_n \right) \right]^2 \\
&= \frac{1}{\ell} \mathbb{E}_* \left[\left(X_1^* - \bar{X}_n \right) + \dots + \left(X_\ell^* - \bar{X}_n \right) \right]^2 \\
&= \frac{1}{\ell} \frac{1}{n/\ell} \sum_{j=1}^{n/\ell} \left[\left(X_{(j-1)\ell+1} - \bar{X}_n \right) + \dots + \left(X_{j\ell} - \bar{X}_n \right) \right]^2
\end{aligned}$$

by indep. of blocks and by $\frac{1}{\ell} \mathbb{E} \left(X_1^* + \dots + X_\ell^* \right) = \bar{X}_n$

$$= \frac{1}{n/\ell} \sum_{j=1}^{n/\ell} \frac{1}{\ell} \sum_{r=1}^{\ell} \sum_{s=1}^{\ell} (X_{(j-1)\ell+r} - \bar{X}_n)(X_{(j-1)\ell+s} - \bar{X}_n)$$

$$= \frac{1}{n/\ell} \sum_{j=1}^{n/\ell} \sum_{|h| \le \ell} \underbrace{\frac{1}{\ell} \sum_{t=(j-1)\ell+1}^{j\ell-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)}_{\text{like sample acvf of block } j}$$

$$= \sum_{|h|<\ell} \hat{\gamma}_{n,\ell}^{NBB}(h),$$

where

$$\hat{\gamma}_{n,\ell}^{NBB}(h) = \begin{cases} \frac{1}{n/\ell} \sum_{j=1}^{n/\ell} \frac{1}{\ell} \sum_{t=(j-1)\ell+1}^{j\ell-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n), & h = -(\ell-1), \dots, \ell-1 \\ 0, & \text{otherwise}. \end{cases}$$

The function $\hat{\gamma}_{n,\ell}^{NBB}(h)$ is basically the mean of sample autocovariance functions computed on the blocks

$$(X_{(j-1)\ell+1}, \dots, X_{j\ell}), \quad j = 1, \dots, n/\ell.$$

Under some mild conditions on the dependence structure of the time series, we have

$$\text{Var}\, T_n \to \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{as} \quad n \to \infty,$$

where $\gamma(\cdot)$ is the acvf of $\{X_t, t \in \mathbb{Z}\}$, as well as

$$\sum_{|h|<\ell} \hat{\gamma}_{n,\ell}^{NBB}(h) \xrightarrow{p} \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{as} \quad n \to \infty$$

provided $\ell \to \infty$ and $\ell/n \to 0$ as $n \to \infty$.

Moving-block bootstrap (MBB):

Denote by $\hat{\mathbb{F}}_n^{MBB}$ the MBB estimator of $\mathbb{F}_n$. It has cdf given by

$$\hat{\mathbb{F}}_n^{MBB}(x_1, \dots, x_n) = \prod_{j=1}^{n/\ell} \hat{\mathbb{F}}_\ell^{MBB}(x_{(j-1)\ell+1}, \dots, x_{j\ell}),$$

where $\hat{\mathbb{F}}_\ell^{MBB}$ is the distribution placing mass $1/(n-\ell+1)$ on each of the blocks

$$(X_j, \dots, X_{j+\ell-1}), \qquad j = 1, \dots, n-\ell+1.$$

To generate a Monte-Carlo draw $X_1^*, \dots, X_n^*$ from $\hat{\mathbb{F}}_n^{MBB}$, sample with replacement n/ℓ times from the blocks

$$(X_j, \dots, X_{j+\ell-1}), \qquad j = 1, \dots, n-\ell+1$$

and concatenate the sampled blocks to obtain $X_1^*, \dots, X_n^*$.

Application of the MBB to the mean:

Let $T_n = t_n(X_1, \dots, X_n, \mathbb{F}_n) = \sqrt{n}(\bar{X}_n - \mathbb{E}\bar{X}_n)$, and suppose we wish to estimate $\operatorname{Var} T_n$ using the MBB.

The MBB version of T_n is

$$T_n^* = t_n(X_1^*, \dots, X_n^*, \hat{\mathbb{F}}_n^{MBB}) = \sqrt{n}(\bar{X}_n^* - \mathbb{E}_* \bar{X}_n^*),$$

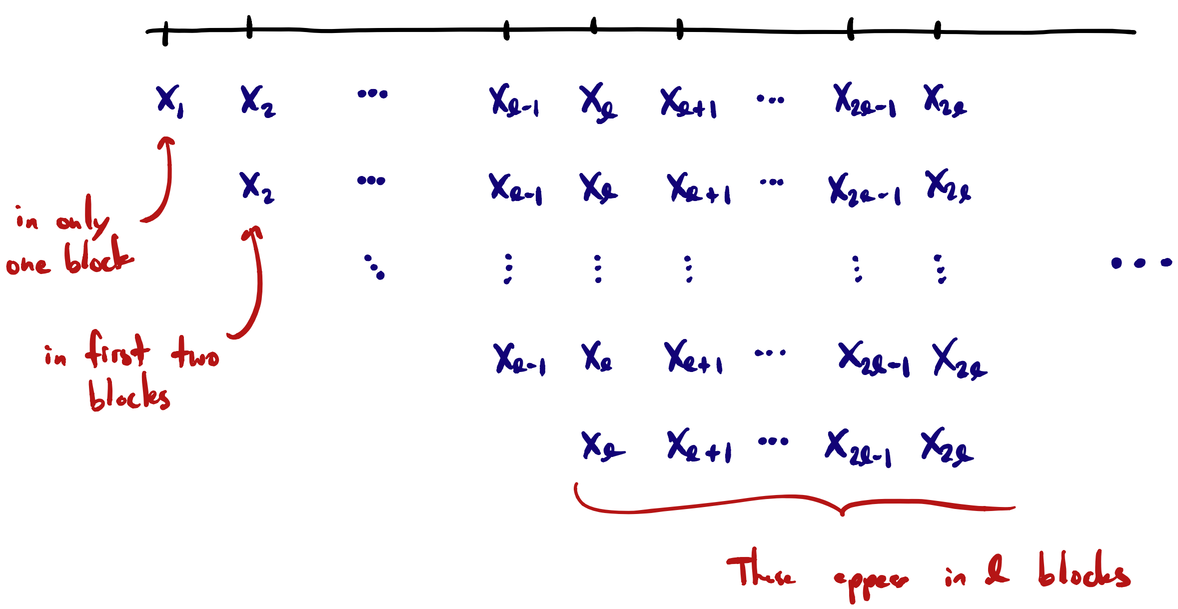
where $\mathbb{E}_*$ is expectation conditional on $X_1, \dots, X_n$ and based on the distribution $\hat{\mathbb{F}}_n^{MBB}$.

The MBB estimator of $\operatorname{Var} T_n$ is $\operatorname{Var}_* T_n^*$, where Var_* operates according to $\hat{\mathbb{F}}_n^{MBB}$.

To find $\mathrm{Var}_* T_n^*$, we first find $\mathbb{E}_* \bar{X}_n^*$. We have

$$\begin{aligned}
\mathbb{E}_* \bar{X}_n^* &= \mathbb{E}_* \frac{1}{n} \sum_{i=1}^{n} X_i^* \\
&= \frac{1}{n} \mathbb{E}_* \sum_{j=1}^{n/\ell} \left(X^*_{(j-1)\ell+1} + \cdots + X^*_{j\ell} \right) \\
&= \frac{1}{\ell} \mathbb{E}_* \left(X_1^* + \cdots + X_\ell^* \right) \\
&= \frac{1}{\ell} \frac{1}{n-\ell+1} \sum_{j=1}^{n-\ell+1} \left(X_j + \cdots + X_{j+\ell-1} \right) \\
&= \frac{1}{\ell} \frac{1}{n-\ell+1} \left[\sum_{j=1}^{\ell-1} j X_j + \sum_{j=\ell}^{n-\ell+1} \ell X_j + \sum_{j=n-\ell+2}^{n} (n-j+1) X_j \right] \\
&= \frac{1}{n-\ell+1} \left[\sum_{j=1}^{\ell-1} \left(\frac{j}{\ell} \right) X_j + \sum_{j=\ell}^{n-\ell+1} X_j + \sum_{j=n-\ell+2}^{n} \left(\frac{n-j+1}{\ell} \right) X_j \right] \\
&=: \bar{X}_{n,\ell}^{MBB},
\end{aligned}$$

noting that less weight is assigned by $\hat{F}_n^{MBB}$ to data points near the boundaries, as these appear in fewer blocks. See the diagram:



So $T_n^* = \sqrt{n}\left(\bar{X}_n^* - \bar{X}_{n,\ell}^{MBB} \right)$. The MBB estimator of $\mathrm{Var}\, T_n$ is

$$\mathrm{Var}_* T_n^* = \mathrm{Var}_* \left[\sqrt{n} \left(\bar{X}_n^* - \bar{X}_{n,\ell}^{MBB} \right) \right]$$

$$= n\,\mathbb{E}_*\left(\bar{X}_n^* - \bar{X}_{n,\ell}^{MBB}\right)^2$$

$$= n\,\mathbb{E}_*\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i^* - \bar{X}_{n,\ell}^{MBB}\right)\right]^2$$

$$= \frac{1}{n}\,\mathbb{E}_*\left[\sum_{j=1}^{n/\ell}\left(\left(X_{(j-1)\ell+1}^* - \bar{X}_{n,\ell}^{MBB}\right) + \dots + \left(X_{j\ell}^* - \bar{X}_{n,\ell}^{MBB}\right)\right)\right]^2$$

by indep. of blocks and by $\frac{1}{\ell}\mathbb{E}_*\left(X_1^* + \dots + X_\ell^*\right) = \bar{X}_{n,\ell}^{MBB}$

$$= \frac{1}{n}\sum_{j=1}^{n/\ell}\mathbb{E}_*\left[\left(X_{(j-1)\ell+1}^* - \bar{X}_{n,\ell}^{MBB}\right) + \dots + \left(X_{j\ell}^* - \bar{X}_{n,\ell}^{MBB}\right)\right]^2$$

$$= \frac{1}{\ell}\,\mathbb{E}_*\left[\left(X_1^* - \bar{X}_{n,\ell}^{MBB}\right) + \dots + \left(X_\ell^* - \bar{X}_{n,\ell}^{MBB}\right)\right]^2$$

$$= \frac{1}{\ell}\,\frac{1}{n-\ell+1}\sum_{j=1}^{n-\ell+1}\left[\left(X_j - \bar{X}_{n,\ell}^{MBB}\right) + \dots + \left(X_{j+\ell-1} - \bar{X}_{n,\ell}^{MBB}\right)\right]^2$$

$$= \frac{1}{n-\ell+1}\sum_{j=1}^{n-\ell+1}\frac{1}{\ell}\sum_{r=1}^{\ell}\sum_{s=1}^{\ell}\left(X_{j-1+r} - \bar{X}_{n,\ell}^{MBB}\right)\left(X_{j-1+s} - \bar{X}_{n,\ell}^{MBB}\right)$$

$$= \frac{1}{n-\ell+1}\sum_{j=1}^{n-\ell+1}\sum_{|h|\le\ell}\underbrace{\frac{1}{\ell}\sum_{t=j}^{j+\ell-1-|h|}\left(X_{t+|h|} - \bar{X}_{n,\ell}^{MBB}\right)\left(X_t - \bar{X}_{n,\ell}^{MBB}\right)}_{\text{like sample acvf of block } j}$$

$$= \sum_{|h|<\ell}\hat{\gamma}_{n,\ell}^{MBB}(h),$$

where

$$\hat{\gamma}_{n,\ell}^{MBB}(h) = \begin{cases} \frac{1}{n-\ell+1}\sum_{j=1}^{n-\ell+1}\frac{1}{\ell}\sum_{t=j}^{j+\ell-1-|h|}\left(X_{t+|h|} - \bar{X}_{n,\ell}^{MBB}\right)\left(X_t - \bar{X}_{n,\ell}^{MBB}\right), & h = -(\ell-1),\dots,\ell-1 \\ 0, & \text{otherwise.} \end{cases}$$

The function $\hat{\gamma}_{n,\ell}^{MBB}$ is essentially the mean of sample autocovariance functions computed on the blocks

$$(X_j, \dots, X_{j+\ell-1}), \qquad j = 1, \dots, n-\ell+1.$$

Under some mild conditions on the dependence structure of the time series, we have

$$\operatorname{Var} T_n \to \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{as} \quad n \to \infty,$$

where $\gamma(\cdot)$ is the acvf of $\{X_t, t \in \mathbb{Z}\}$, as well as

$$\sum_{|h|<\ell} \hat{\gamma}_{n,\ell}^{MBB}(h) \xrightarrow{P} \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{as} \quad n \to \infty$$

provided $\ell \to \infty$ and $\ell/n \to 0$ as $n \to \infty$.