Discussion Notes on the Interval-valued Time Series

Yifan Li

July 16, 2020

This document is mainly written to be discussed with Prof. Kulperger and Prof. Yu. This is also highly related to our joint work with Prof. Defei Zhang in the time series in the G-expectation framework.

```
## Loading required package: ggplot2
## Loading required package: reshape2
## Warning: package 'Rfast' was built under R version 3.5.2
## Loading required package: Rcpp
## Loading required package: RcppZiggurat
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Contents

1 Introduction (Brief Version)

We intend to start from a simplified construction on the interval data which is highly related to the interval-valued time series data. Since it is a simplified one, this construction would be essentially connected with but different from our previous construction. It will eventually lead us to an more intuitive interpretation of the sequential independence in the G-expectation framework (in terms of the ambiguity in the interval direction).

Reference notes in the existing literature on interval-valued time series: (brief version to show our concerns)

1. Methods based on Symbolic Data Analysis (SDA), one typical example is to model the center and range process using standard ARMA model (ref: 2008-Luis), one concern here is the intervals here are all proper ones, so the range process will always be strictly non-negative, a standard ARMA (without any transformations) may not be appropriate.
2. Methods originated from the Theory of Random Sets, One key concept here is the Kernel distance between intervals (different choice of kernels will significantly affect the interpretation of the distinction of intervals.) One example in (ref:2013-Han) is a AR conditional Interval model, where they allows the improper intervals, but the parameters in the intervals are all point-valued. One of our focused points here is the parameter uncertainty which is characterized by the interval-valued parameter.
3. In short, we intend to extend the existing design of interval-valued time series to a study of considering the direction of the observed intervals and also allowing the parameter to be interval-valued.

2 Interval Operations

Mainly recall the left and right scaling.

3 Theoretical Construction

3.1 Interval-valued Time Series: Simplified Version

In practice, we can observe the interval-valued time series $\tilde{\boldsymbol{X}}_t, t = 0, 1, 2, \dots$ which are almost always proper ones. It can be generally treated as

$$\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t,$$

for one specified model $\boldsymbol{X}_t$ which may or may not involve improper intervals. As a reminder, $1\boldsymbol{X}_t$ is a left scaling of the interval: $1[a, b] \coloneqq [a \wedge b, a \vee b]$ and it is also the ensured proper transformation.

One straightforward candidate model is $\boldsymbol{X}_t \coloneqq \tilde{\boldsymbol{X}}_t$, that only considers proper intervals which are the observed ones.

To be specific,

$$\begin{aligned}\tilde{\boldsymbol{X}}_t &= 1\boldsymbol{X}_t = 1[\boldsymbol{X}_{tl}, \boldsymbol{X}_{tr}] \\ &= [\boldsymbol{X}_{tl} \wedge \boldsymbol{X}_{tr}, \boldsymbol{X}_{tl} \vee \boldsymbol{X}_{tr}].\end{aligned}$$

In terms of the center and range of the intervals:

$$\mathbf{C}(\tilde{\boldsymbol{X}}_t) = \mathbf{C}(\boldsymbol{X}_t) = \frac{1}{2}(\boldsymbol{X}_{tl} + \boldsymbol{X}_{tr}), \tag{3.1}$$

and

$$\mathbf{R}(\tilde{\boldsymbol{X}}_t) = |\mathbf{R}(\boldsymbol{X}_t)| = |\boldsymbol{X}_{tr} - \boldsymbol{X}_{tl}|. \tag{3.2}$$

The relation shown in Equation (3.1) tells us the center of the observed interval will be the same as the center of the underlying model no matter how the underlying model is specified (which is unknown from the side of data analysts).

For simplicity, in this context, for a specified model

$$\boldsymbol{X}_t = [\boldsymbol{X}_{tl}, \boldsymbol{X}_{tr}],$$

where each end can be treated as a stochastic process in the classical $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$. For each time t, let us define its interval-valued expectation as

$$\boldsymbol{\mathcal{E}}\boldsymbol{X}_t := [\mathbf{E}[\boldsymbol{X}_{tl}], \mathbf{E}[\boldsymbol{X}_{tr}]].$$

If $\boldsymbol{X}_t$ is degenerate (i.e. $\boldsymbol{X}_{tl} \equiv \boldsymbol{X}_{tr}$), the interval-valued expectation $\boldsymbol{\mathcal{E}}$ is degenerated into the linear expectation $\mathbf{E}$.

(Later on, when we consider the ambiguity in the specification of $\boldsymbol{X}_t$ or the interval direction, the the interval supreme of a class of $\boldsymbol{\mathcal{E}}$ will be become the $\hat{\boldsymbol{\mathcal{E}}}$ where the two ends become the sublinear expectation. This is especially useful in the study of the observed $\tilde{\boldsymbol{X}}_t$ where we are not sure of the underlying interval direction.)

4 A Series of Examples with Simulation

4.1 Mean Certainty and Variance Uncertainty

4.1.1 Data Problem Setup

Imagine a data situation as follows: we want to study the property of a time series X_t which is known to be a strong stationary one with $\mathbf{E}[|X_t|^3] < \infty$ (for simplicity of this problem). Let $\mu := \mathbf{E}[X_t]$ and $\sigma^2 := \mathbf{E}[(X_t - \mu)^2]$.

Suppose we ask a lab to help us do the measurement. The lab tells us they have two (secret) equipments,labeled **A** and **B**, as measurers and they have measured the X_t for $t = 1, 2, \cdots, n$ using both of them. At each time t, Equipment **A** will get the measurement X_{tl} and Equipment **B** gives X_{tr}.

We are also informed from the lab that:

1. Based on the design of these two equipments and measurement errors, at each time t, the true X_t must be between X_{tl} and X_{tr}.
2. Nonetheless, we are told that both of them are valid equipments, in the sense that X_{tl} and X_{tr} are precise in the mean part of X_t but may not be precise in the variance part of X_t.

However, due to some reasons (e.g. privacy problem), the lab refuses to give us the measurements $\boldsymbol{X}_t = (X_{tl}, X_{tr})$ labelled with **A** and **B** (it may contain some privacy information of the two equipments), the lab only provides us with the minimum and maximum measurements at each time (which is $\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t$).

In short, we only observe $\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t$. Based on the information we may assume the two unobserved objects $\boldsymbol{X}_t$ and X_t as follows:

1. For each t, $X_t \in 1\boldsymbol{X}_t = \tilde{\boldsymbol{X}}_t$,
2. $\boldsymbol{\mathcal{E}}\boldsymbol{X}_t = [\mu, \mu] = \mu := \mathbf{E}[X_t]$,

3. $\mathbf{E}[(X_t-\mu)^2] =: \sigma^2 \in 1\boldsymbol{\mathcal{E}}(\boldsymbol{X}_t-\mu)^2 =: [\underline{\sigma}^2, \overline{\sigma}^2]$ with $0 < \underline{\sigma} \le \overline{\sigma}$ (For simplicity, we assume the variance of the either ends of $\boldsymbol{X}_t$ only switch betwen $\underline{\sigma}^2$ and $\overline{\sigma}^2$.

We are interested in two specific problems:

1. (P1) Esimation of parameters of $\boldsymbol{X}_t$: based on data $\tilde{\boldsymbol{X}}_t$, we want to estimate the parameter $(\mu, \underline{\sigma}, \overline{\sigma})$ of $\boldsymbol{X}_t$ and also construct a reasonable confidence interval for μ .
2. (P2) Study of the normalized sum of X_t: if we are able to estimate $(\mu, \underline{\sigma}, \overline{\sigma})$, for $\varphi(x) = x^3$ (or other monotone functions), suppose we treat the estimated values as the true ones (if they are precise enough), let

$$Y_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \mu),$$

and

$$\mathbf{Y}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\boldsymbol{X}_t - \mu).$$

Since$Y_n$must between the two ends of $\mathbf{Y}_n$, we have $\mathbf{E}[\varphi(Y_n)]$ must be covered by

$$[\mathbf{E}[\varphi(\mathbf{Y}_{nl})], \mathbf{E}[\varphi(\mathbf{Y}_{nr})]].$$

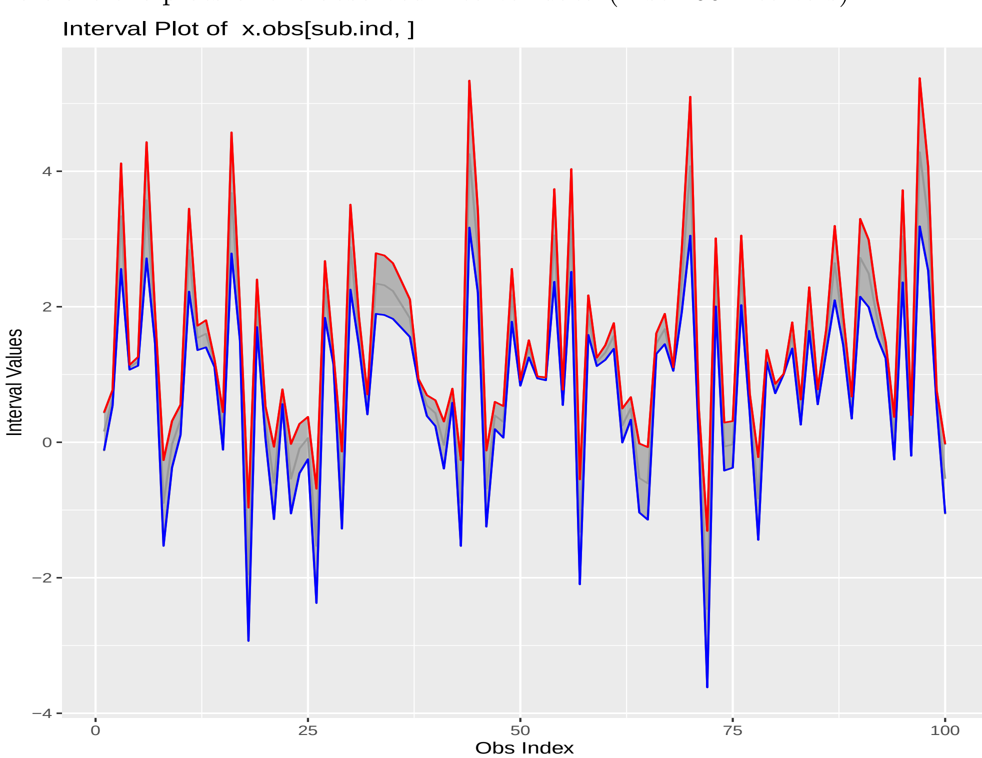
We want to study the possibilities of this envelope.

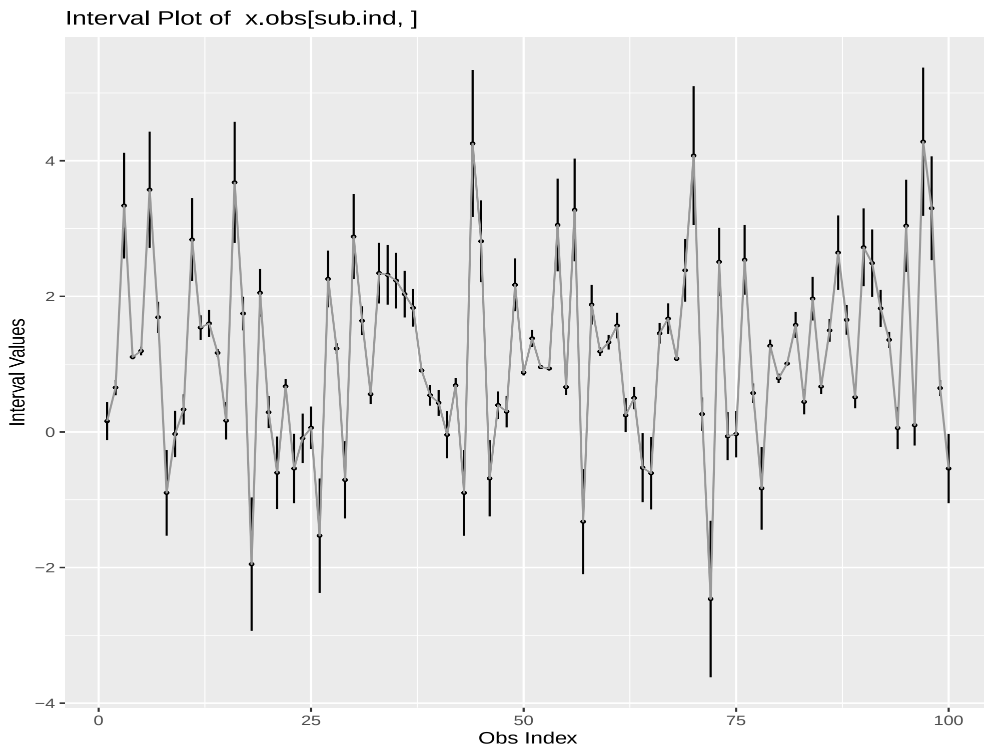
Let us forcus on (P1) first. One thing we can think about: does the lackness of information in interval direction cause a problem in the estimation of $(\mu, \underline{\sigma}, \overline{\sigma})$? (Intuitively, does the loss of interval direction mask information on the parameters of $\boldsymbol{X}_t$?)

4.1.2 Initial Analysis

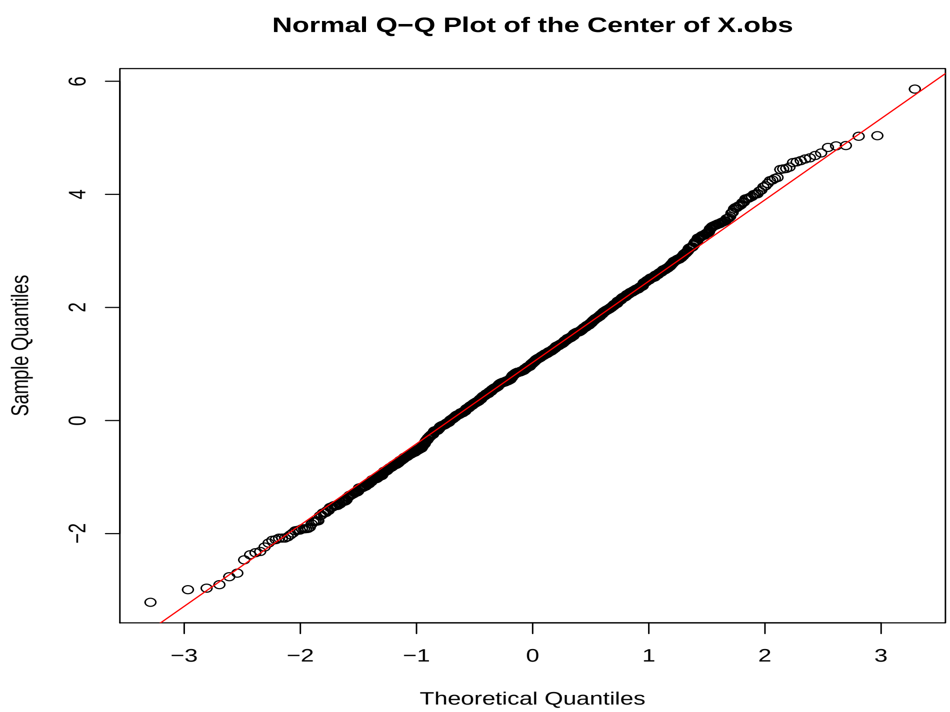
We first need to do some initial analysis on the observed interval time series.

Here are the plots of the observed interval data (first 100 intervals).

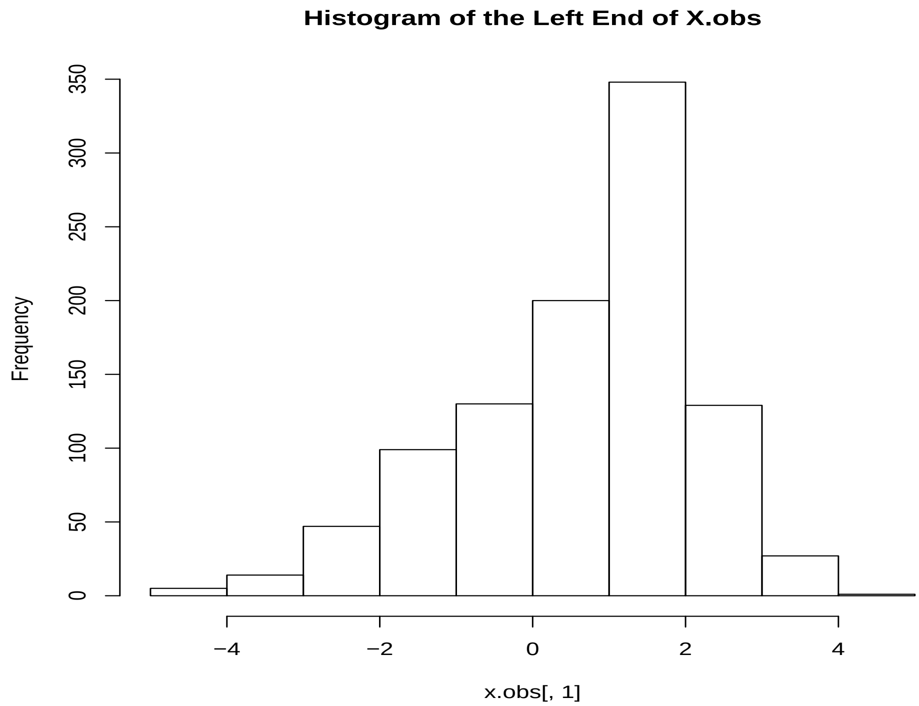


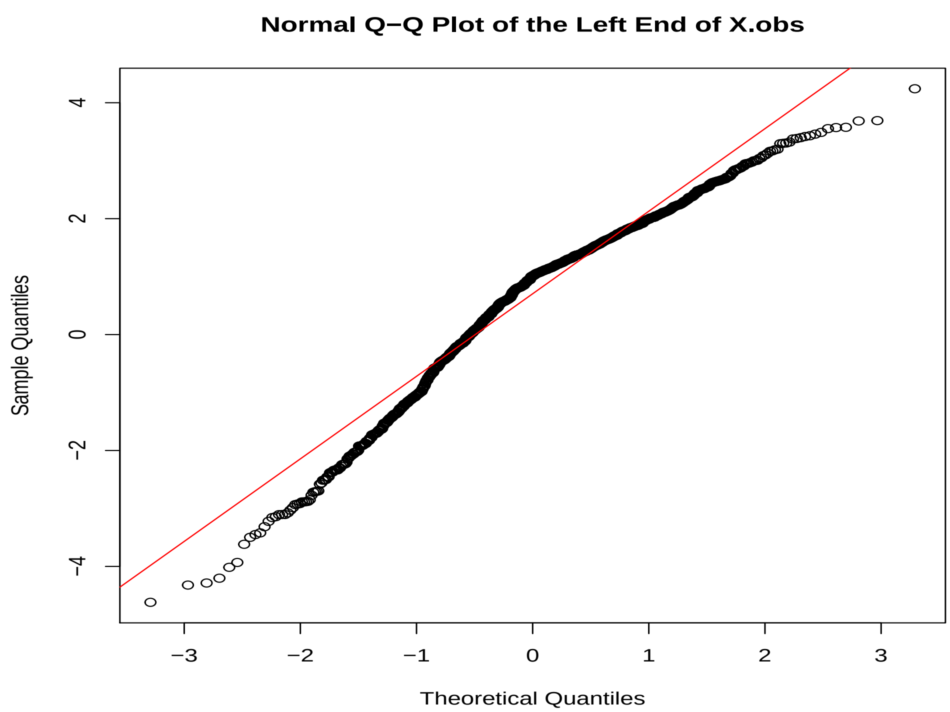


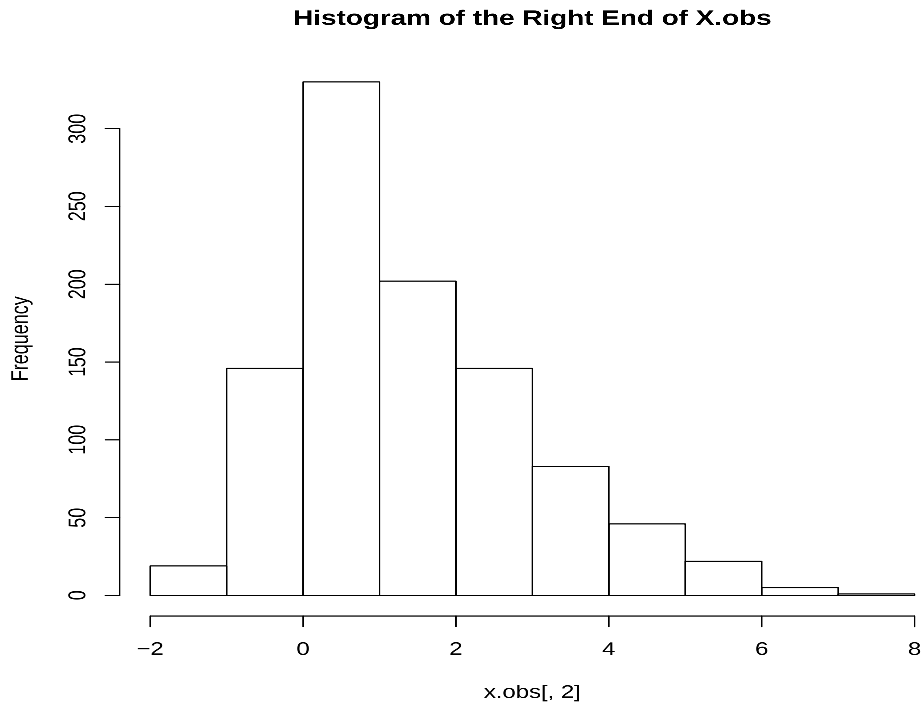
Equation (3.1) shows the observed interval has the same center as the true one, so we can use this property to at least check the distributional information.

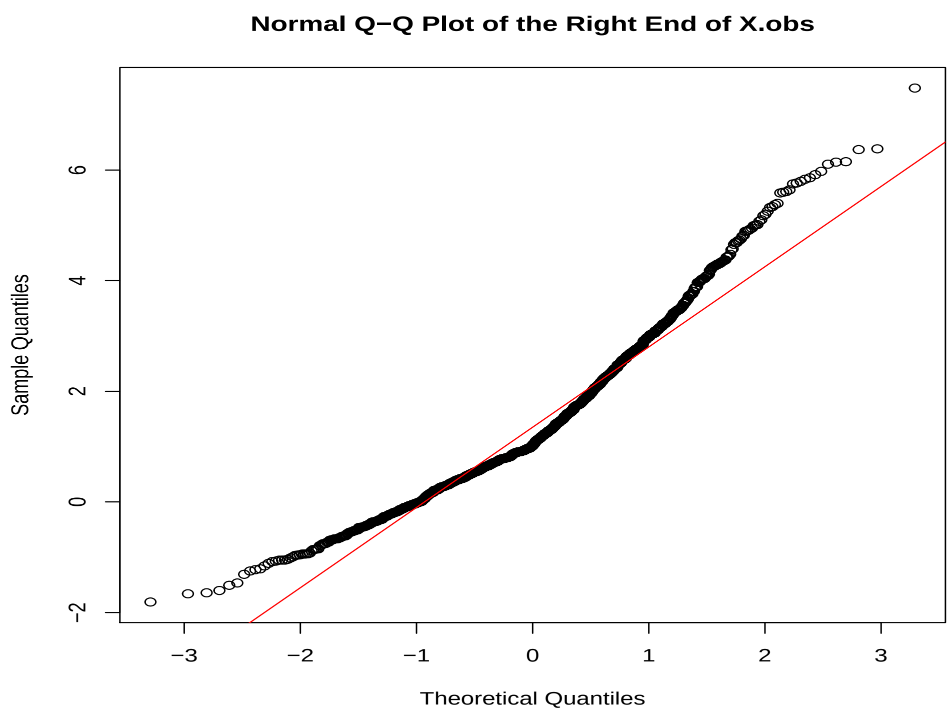


We can see that the center sequence of the intervals is approximately normally distributed. However, if we can check the distributions of the two ends.


Histogram of the Left End of X.obs
Frequency
0
50
100
150
200
250
300
350
−4
−2
0
2
4
x.obs[, 1]


Normal Q−Q Plot of the Left End of X.obs
Sample Quantiles
−4
−2
0
2
4
−3
−2
−1
0
1
2
3
Theoretical Quantiles





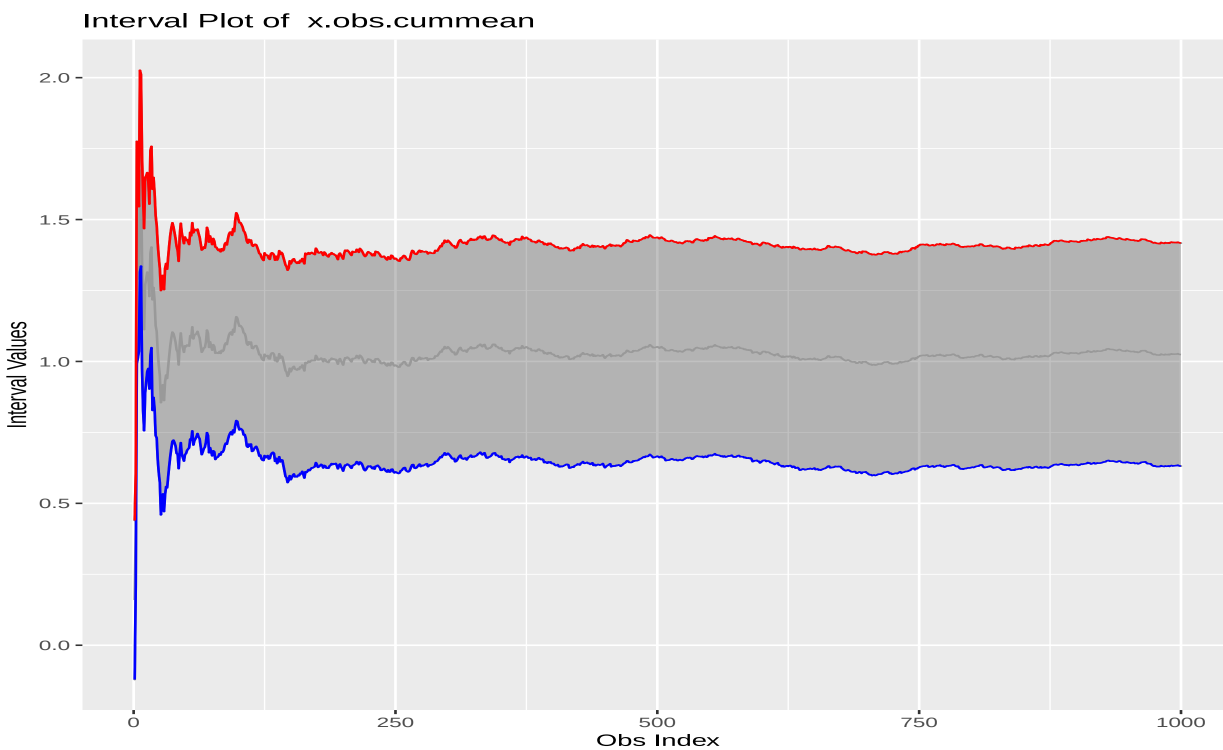
They seems not normally distributed at least from the skewness.

If we naively use the sample mean of the two ends of the observed intervals, it will lead to misleading results because we already know that the true mean should be a single value (or a degenerate interval). However, the sample mean of the two ends of the observed intervals will gives us a non-degenerate interval and this interval will never shrink to a single value as sample size increase.

This is also one of the distinctions of our work from part of the existing literature based on the LLN from Aumann expectation where the sample mean is directly applied to the observed intervals itself: it definitely a feasible method but it essentially treats the observed intervals as the true underlying process and the expectation will be the limit of the observed sample mean: it does not always the case, especially when we involves the direction of the intervals.

We also want readers to not overinterpret this non-degenerate interval too much (it may cover the true μ, but it has nothing to do with the confidence interval at any significance level.) We will see this from the study of the distribution of the two ends.

Our focus here is often the expectation (which should match the background of the dataset) which may or may not be the limit of the sample mean of the observed interval when the true underlying intervals are allowed to be improper ones. For instance, when we look at the interval-valued log return of a stock price, one may ask, what is the best expected return and the worst expected return we can get from the stock using some *realistic* trading strategy? One important note of the realistic trading strategy is based on the assumption that we cannot see into the future: it is unrealistic to ask one to always do the trading at the exact min and max point of the stock intraday price at each day accordingly to achieve the max or min daily return, but we can only do the trading based on some random rule and use the information up until the time point now to make the movement at the next time point (It is related to the predictability of the process). Therefore, if we simply use the sample mean of the upper and lower end of the observed log return, it will converge to an interval of expectations whose two ends are *impossible* to achieve based on some realistic trading strategies. Therefore, a more appropriate min and max expected return should forms a smaller interval, which is the focus of our work.



4.1.3 The First Candidate Model

One candidate model for the underlying true interval $\boldsymbol{X}_t$ of $\tilde{\boldsymbol{X}}_t$ is

$$\boldsymbol{X}_t^{(1)} = \mu + [\underline{\sigma}, \overline{\sigma}]\epsilon_t, \tag{4.1}$$

and then $\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t^{(1)}$. In the following context of this section, since we will mainly focus on $\boldsymbol{X}_t^{(1)}$, without causing any confusions, we will omit the superscript and write it as $\boldsymbol{X}_t$ but we want readers to know that it is only a candidate **model** for the true underlying interval process but not the true process itself.

In Equation (4.1), ϵ_t are classically i.i.d. $N(0,1)$ (or any other white noise process with $\mathbf{E}[\epsilon_t] = 0$ and $\mathbf{E}[\epsilon_t^2] = 1$). and the operation in $[\underline{\sigma}, \overline{\sigma}]\epsilon_t$ is the right scaling ($[x_l, x_r]c \coloneqq [x_l c, x_r c]$). We also require $0 < \underline{\sigma} \leq \overline{\sigma}$ for the identifiability of model (because we will consider $\tilde{\boldsymbol{X}}_t$ later on). When $\underline{\sigma} = \overline{\sigma} = \sigma$, $\boldsymbol{X}_t$ is degenerated into the shifted Gaussian white noise $\boldsymbol{X}_t \sim N(\mu, \sigma^2)$.

4.1.4 Properties of the First Model

Under the model specified by Equation (4.1) by assuming $\epsilon_t \sim N(0,1)$ in terms of the two ends, we can represent the true underlying process $\boldsymbol{X}_t$, which cannot be directly observed, as (we may call it the two-end representation)

$$\begin{cases} \boldsymbol{X}_{tl} = \mu + \underline{\sigma}\epsilon_t & \text{i.i.d.} \sim N(\mu, \underline{\sigma}^2) \\ \boldsymbol{X}_{tr} = \mu + \underline{\sigma}\epsilon_t & \text{i.i.d.} \sim N(\mu, \overline{\sigma}^2) \end{cases}.$$

The center and range of $\boldsymbol{X}_t$ are (we may call it the center-range representation),

$$\begin{cases} \mathbf{C}(\boldsymbol{X}_t) = \mu + \frac{\underline{\sigma}+\overline{\sigma}}{2}\epsilon_t & \text{i.i.d.} \sim N(\mu, \frac{(\underline{\sigma}+\overline{\sigma})^2}{4}) \\ \mathbf{R}(\boldsymbol{X}_t) = (\overline{\sigma} - \underline{\sigma})\epsilon_t & \text{i.i.d.} \sim N(0, (\overline{\sigma} - \underline{\sigma})^2) \end{cases}.$$

Under this model specification, we also have the properties that

$$\begin{aligned} \boldsymbol{\mathcal{E}}\boldsymbol{X}_t &= [\mu, \mu], \\ \boldsymbol{\mathcal{E}}\boldsymbol{X}_t^2 &= [\mu^2 + \underline{\sigma}^2, \mu^2 + \overline{\sigma}^2], \\ \boldsymbol{\mathcal{E}}\boldsymbol{X}_t^3 &= [0, 0], \\ \boldsymbol{\mathcal{E}}\mathbb{1}_{\{\boldsymbol{X}_t < c\}} &= [\Phi(\frac{c-\mu}{\underline{\sigma}}), \Phi(\frac{c-\mu}{\overline{\sigma}})]. \end{aligned}$$

and in terms of the sample mean

$$\bar{\boldsymbol{X}}_n = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t,$$

in terms of the two ends, we have

$$\bar{\boldsymbol{X}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\mathcal{E}}\boldsymbol{X}_t = \boldsymbol{\mu} = [\mu_l, \mu_r] = [\mu, \mu],$$

and

$$\begin{aligned} \sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\underline{\sigma}, \overline{\sigma}]\epsilon_t \\ &= [\underline{\sigma}, \overline{\sigma}]\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t \\ &\xrightarrow{\text{d}} [\underline{\sigma}, \overline{\sigma}]Z, \end{aligned}$$

where $Z \sim N(0,1)$.

If we have the data directly for $\boldsymbol{X}_t$, it is straightforward to use traditional statistical estimation methods (such as maximum likelihood estimation and method of moments) to make inference on $(\mu, \underline{\sigma}, \overline{\sigma})$. However, since we can only observe

$$\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t = [\boldsymbol{X}_{tl} \wedge \boldsymbol{X}_{tr}, \boldsymbol{X}_{tl} \vee \boldsymbol{X}_{tr}], \tag{4.2}$$

The problem here is whether there still exists a tractable method for us to still get reasonable estimation on $(\mu, \underline{\sigma}, \overline{\sigma})$ from the data $\tilde{\boldsymbol{X}}_t$? We will give a positive answer in this section.

However, we only have $\tilde{\boldsymbol{X}}_t$ observed and our goal is to the estimation for the parameters $(\mu, \underline{\sigma}, \overline{\sigma})$ in $\boldsymbol{X}_t$ which is are directly observed. Here are two possible directions:

1. Derive the distributions of the two ends of $\tilde{\boldsymbol{X}}_t$ based on the model of $\boldsymbol{X}_t$, then use traditional statistical methods (such as MLE, MM or estimating equations) to estimate the parameters;
2. Adjust the direction of the observed data $\tilde{\boldsymbol{X}}_t$ into $\boldsymbol{X}'_t$ to make it close to the model of $\boldsymbol{X}_t$, then treat $\boldsymbol{X}'_t$ as $\boldsymbol{X}_t$ to directly use the distributions of $\boldsymbol{X}_t$ to do the estimation.

4.1.5 Distributions of the observed data

Specifically, given the distributions of $\boldsymbol{X}_{tl}$ and $\boldsymbol{X}_{tr}$, from Equation (4.2), we can derive the distributions of the two ends of $\tilde{\boldsymbol{X}}_t$: with $x^+ := \max\{x, 0\}$ and $x^- := \min\{-x, 0\}$,

$$\tilde{\boldsymbol{X}}_{tl} = \begin{cases} \mu + \underline{\sigma}\epsilon_t & \text{if } \epsilon_t \geq 0 \\ \mu + \overline{\sigma}\epsilon_t & \text{if } \epsilon_t < 0 \end{cases} = \mu + (\underline{\sigma}\epsilon_t^+ - \overline{\sigma}\epsilon_t^-),$$

and similarly,

$$\tilde{\boldsymbol{X}}_{tr} = \mu + (\overline{\sigma}\epsilon_t^+ - \underline{\sigma}\epsilon_t^-).$$

We can first work on a simplified case: for $\epsilon \sim N(0,1)$ and $0 < \underline{\sigma} \leq \overline{\sigma}$, consider $Y_r = (\underline{\sigma}\epsilon) \vee (\overline{\sigma}\epsilon) = \overline{\sigma}\epsilon^+ - \underline{\sigma}\epsilon^-$ which will switch between $\overline{\sigma}\epsilon$ and $\underline{\sigma}\epsilon$ depending on the sign of ϵ.

Let Φ and ϕ be the cdf and pdf of $N(0,1)$, respectively. Then we can write the cdf of Y_r as

$$\begin{aligned} F_{Y_r}(y) &= \mathbf{P}(Y_r \leq y) \\ &= \mathbf{P}(Y_r \leq y, \epsilon < 0) + \mathbf{P}(Y_r \leq y, \epsilon > 0) \\ &= \mathbf{P}(\underline{\sigma}\epsilon \leq y, \epsilon < 0) + \mathbf{P}(\overline{\sigma}\epsilon \leq y, \epsilon > 0). \\ &= \begin{cases} 0 + \mathbf{P}(\epsilon \leq y/\underline{\sigma}, \epsilon < 0) = \mathbf{P}(\epsilon < y/\underline{\sigma}) & \text{if } y < 0 \\ \mathbf{P}(\epsilon < 0) + \mathbf{P}(0 < \epsilon \leq y/\overline{\sigma}) = \mathbf{P}(\epsilon \leq y/\overline{\sigma}) & \text{if } y \geq 0 \end{cases} \\ &= \begin{cases} \Phi(y/\underline{\sigma}) & \text{if } y < 0 \\ \Phi(y/\overline{\sigma}) & \text{if } y \geq 0 \end{cases}. \end{aligned}$$

For $Z_r = \mu + Y_r$, we have

$$F_{Z_r}(z) = \mathbf{P}(Z_r \leq z) = \mathbf{P}(Y_r \leq z - \mu) = F_{Y_r}(z - \mu).$$

Therefore,

$$F_{Z_r}(z) = \begin{cases} \Phi(\frac{z-\mu}{\underline{\sigma}}) & \text{if } z < \mu \\ \Phi(\frac{z-\mu}{\overline{\sigma}}) & \text{if } z \geq \mu \end{cases} = \begin{cases} \Phi(\frac{z-\mu}{\underline{\sigma}}) & \text{if } z < \mu \\ \frac{1}{2} & \text{if } z = \mu \\ \Phi(\frac{z-\mu}{\overline{\sigma}}) & \text{if } z > \mu \end{cases}.$$

Note that $F_{Z_r}(z)$ is not differentiable at $z = \mu$, but it has a continuous density function defined on $\mathbb{R}/\{\mu\}$,

$$f_{Z_r}(z) = \frac{1}{\underline{\sigma}}\phi(\frac{z-\mu}{\underline{\sigma}})\mathbb{1}_{\{z<\mu\}} + \frac{1}{\overline{\sigma}}\phi(\frac{z-\mu}{\overline{\sigma}})\mathbb{1}_{\{z>\mu\}}.$$

Similarly, the cdf and pdf $Z_l = \mu + Y_l = \mu + (\underline{\sigma}\epsilon^+ - \overline{\sigma}\epsilon^-)$ are, respectively,

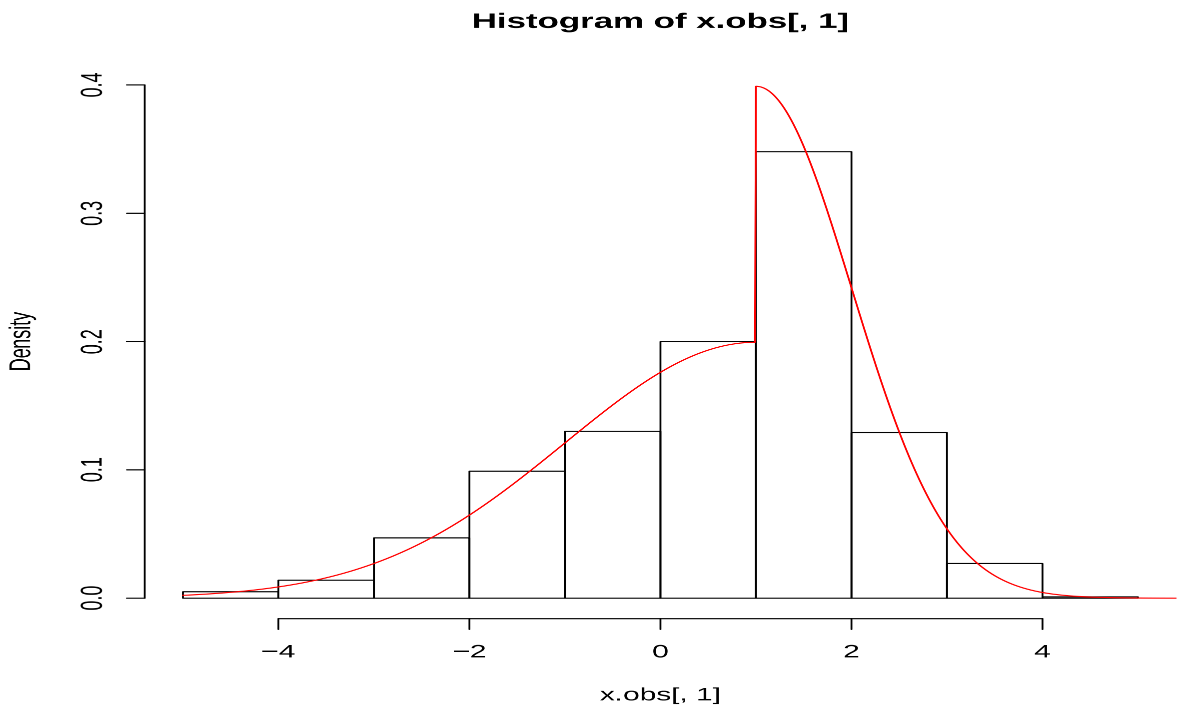
$$F_{Z_l}(z) = \mathbf{P}(Z_l \leq z) = \begin{cases} \Phi(\frac{z-\mu}{\overline{\sigma}}) & \text{if } z < \mu \\ \Phi(\frac{z-\mu}{\underline{\sigma}}) & \text{if } z \geq \mu \end{cases},$$

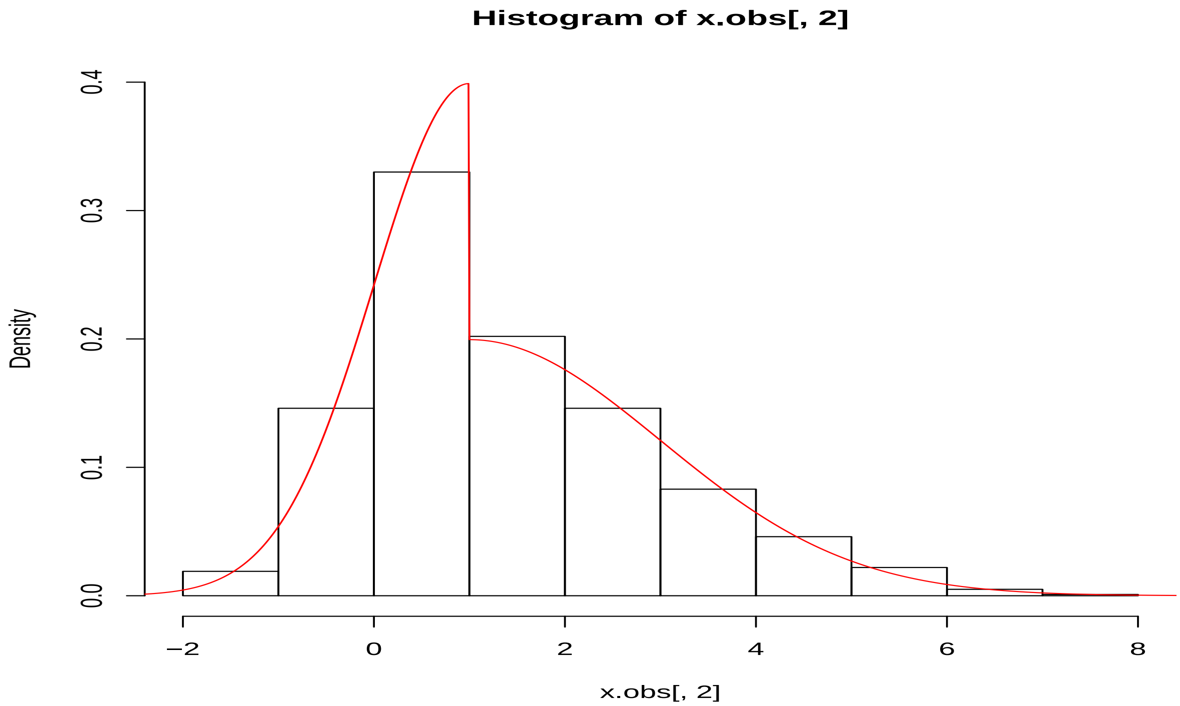
and

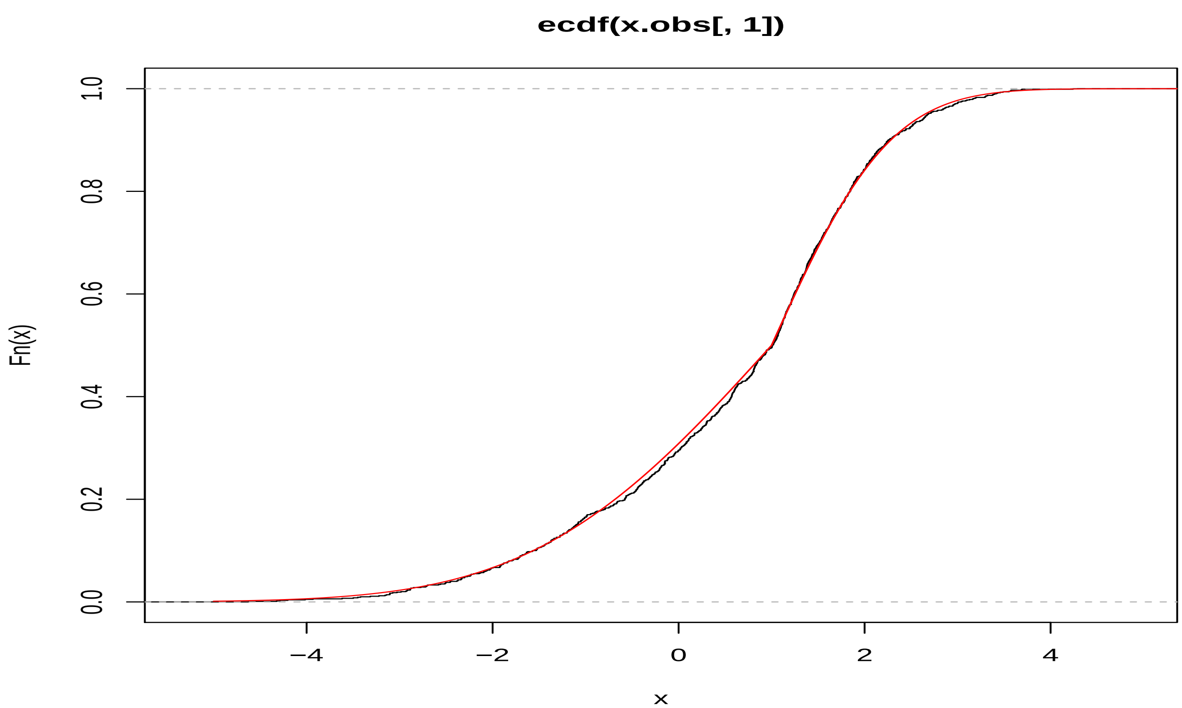
$$f_{Z_l}(z) = \frac{1}{\overline{\sigma}}\phi(\frac{z-\mu}{\overline{\sigma}})\mathbb{1}_{\{z<\mu\}} + \frac{1}{\underline{\sigma}}\phi(\frac{z-\mu}{\underline{\sigma}})\mathbb{1}_{\{z>\mu\}}.$$

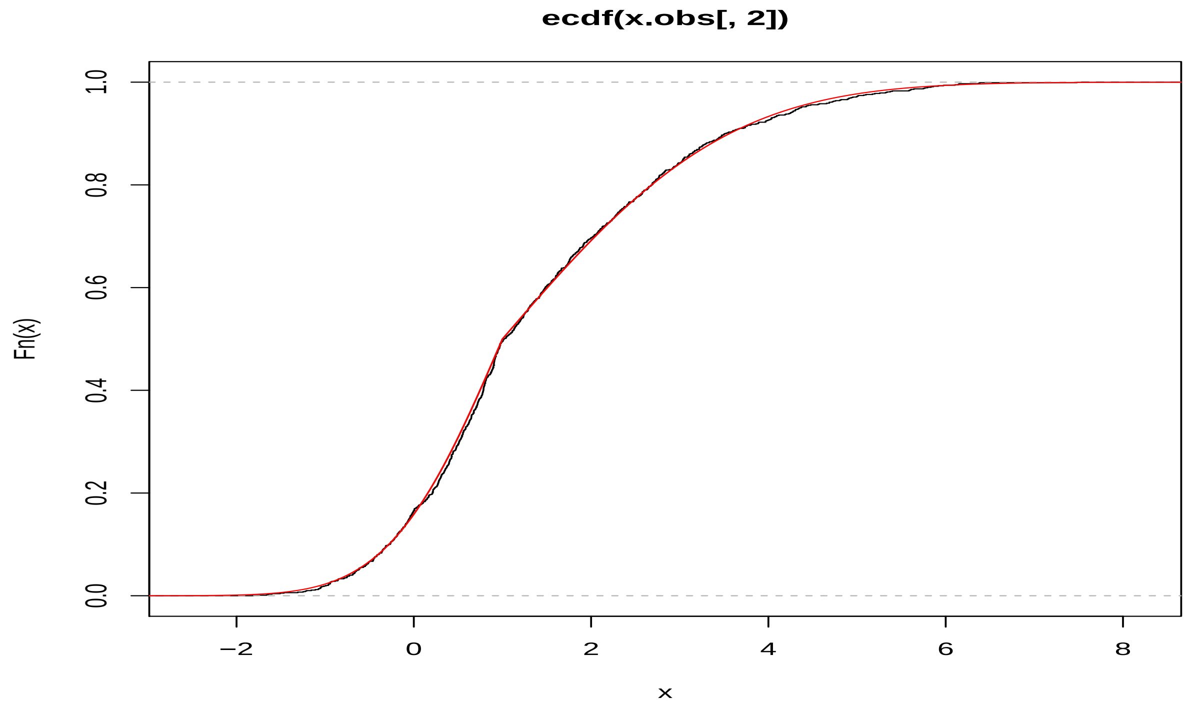
Note that $\tilde{\boldsymbol{X}}_{tr} \overset{\text{d}}{=} Z_r$ and $\tilde{\boldsymbol{X}}_{tl} \overset{\text{d}}{=} Z_l$ for each t.

Then we can validate our results by comparing with simulation.




Histogram of x.obs[, 2]
Density
0.0
0.1
0.2
0.3
0.4
−2
0
2
4
6
8
x.obs[, 2]


ecdf(x.obs[, 1])
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
x



We can also derive the distributions for the center and range of $\tilde{\boldsymbol{X}}_t$:

$$\mathbf{C}(\tilde{\boldsymbol{X}}_t) = \mathbf{C}(\tilde{\boldsymbol{X}}_t) = \mu + \frac{\underline{\sigma} + \overline{\sigma}}{2}\epsilon_t \text{i.i.d.} \sim N(\mu, \frac{(\underline{\sigma} + \overline{\sigma})^2}{4}),$$

and

$$\mathbf{R}(\tilde{\boldsymbol{X}}_t) = |\mathbf{R}(\boldsymbol{X}_t)| = |(\overline{\sigma} - \underline{\sigma})\epsilon_t|,$$

which follows the half normal distribution with scale parameter equal to $(\overline{\sigma} - \underline{\sigma})$.

For curiosity, we can compare the distributions of two ends of the observed data with the lower and upper cdf of the G-normal distributions.

4.1.6 Comparison with the G-normal cdf

Surprisingly, it seems there exists some similarity between the G-normal cdfs (upper and lower) with the two-ends of the observed intervals (left and right). Although so far it is not clear whether there exists any theoretical explanations on this similarity, it must be interesting if we find one.

The upper cdf of G-normal is (ref: G-VaR)

$$\begin{aligned}
\overline{F}_{\mathcal{N}}(x) &\coloneqq \mathbf{V}(X \leq x) \\
&= \int_{-\infty}^{x} \frac{\sqrt{2}}{(\overline{\sigma} + \underline{\sigma})\sqrt{\pi}} \left[e^{-x^2/(2\overline{\sigma}^2)} \mathbb{1}_{\{x \leq 0\}} + e^{-x^2/(2\underline{\sigma}^2)} \mathbb{1}_{\{x > 0\}} \right] \mathrm{d}x \\
&= \begin{cases} \frac{2\overline{\sigma}}{(\overline{\sigma}+\underline{\sigma})} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\overline{\sigma}} e^{-x^2/(2\overline{\sigma}^2)} \, \mathrm{d}x = \frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}} \Phi(\frac{x}{\overline{\sigma}}) & \text{if } x \leq 0 \\ \frac{2\overline{\sigma}}{(\overline{\sigma}+\underline{\sigma})} \Phi(0) + \frac{2\underline{\sigma}}{(\overline{\sigma}+\underline{\sigma})} \int_{0}^{x} \frac{1}{\sqrt{2\pi}\underline{\sigma}} e^{-x^2/(2\underline{\sigma}^2)} & \text{if } x > 0 \end{cases} \\
&= \frac{2\overline{\sigma}}{\overline{\sigma} + \underline{\sigma}} \Phi(\frac{x}{\overline{\sigma}}) \mathbb{1}_{\{x \leq 0\}} + (\frac{2\underline{\sigma}}{\overline{\sigma} + \underline{\sigma}} \Phi(\frac{x}{\underline{\sigma}}) + \frac{\overline{\sigma} - \underline{\sigma}}{\overline{\sigma} + \underline{\sigma}}) \mathbb{1}_{\{x > 0\}}.
\end{aligned}$$

Since

$$\begin{aligned}
\frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(\frac{x}{\underline{\sigma}}) + \frac{\overline{\sigma}-\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}} &= \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(\frac{x}{\underline{\sigma}}) + 1 - \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}} \\
&= 1 - \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}(1-\Phi(\frac{x}{\underline{\sigma}})) \\
&= 1 - \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(-\frac{x}{\underline{\sigma}}),
\end{aligned}$$

We further simplify the upper cdf as

$$\overline{F}_{\mathcal{N}}(x) = \frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(\frac{x}{\overline{\sigma}})\mathbb{1}_{\{x\le 0\}} + (1 - \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(-\frac{x}{\underline{\sigma}}))\mathbb{1}_{\{x>0\}},$$

which simply retrieves the result in the G-VaR paper. Accordingly, The upper pdf is

$$\begin{aligned}
\overline{f}_{\mathcal{N}}(x) &\coloneqq \frac{\mathrm{d}}{\mathrm{d}x}\overline{F}_{\mathcal{N}}(x) \\
&= \frac{\sqrt{2}}{(\overline{\sigma}+\underline{\sigma})\sqrt{\pi}}\big[e^{-x^2/(2\overline{\sigma}^2)}\mathbb{1}_{\{x\le 0\}} + e^{-x^2/(2\underline{\sigma}^2)}\mathbb{1}_{\{x>0\}}\big] \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{(\underline{\sigma}+\overline{\sigma})/2}\big[e^{-x^2/(2\overline{\sigma}^2)}\mathbb{1}_{\{x\le 0\}} + e^{-x^2/(2\underline{\sigma}^2)}\mathbb{1}_{\{x>0\}}\big].
\end{aligned}$$

It seems the G-normal pdf is one way (perhaps also the only way) to bind two half-parts of the normal densities together also maintain the form a continuous density.

As comparison, the pdf of the left end of the observed interval is (with $\mu = 0$): without changing its Lebesgue integral, we let $f_{Z_l}(0) \coloneqq f_{Z_l}(0-)$,

$$\begin{aligned}
f_{Z_l}(x) &= \frac{1}{\overline{\sigma}}\phi(\frac{x}{\overline{\sigma}})\mathbb{1}_{\{x\le 0\}} + \frac{1}{\underline{\sigma}}\phi(\frac{x}{\underline{\sigma}})\mathbb{1}_{\{x>0\}}. \\
&= \frac{1}{\sqrt{2\pi}}\big[\frac{1}{\overline{\sigma}}e^{-x^2/(2\overline{\sigma}^2)}\mathbb{1}_{\{x\le 0\}} + \frac{1}{\underline{\sigma}}e^{-x^2/(2\underline{\sigma}^2)}\mathbb{1}_{\{x>0\}}\big].
\end{aligned}$$

We can see that they are similar in terms of the likelihood ratio (by treating $\overline{f}_{\mathcal{N}}(x)$ as a classical pdf,)

$$\frac{\overline{f}_{\mathcal{N}}(x)}{f_{Z_l}(x)} = \begin{cases} \frac{2\overline{\sigma}}{\underline{\sigma}+\overline{\sigma}} & \text{if } x \le 0 \\ \frac{2\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}} & \text{if } x > 0 \end{cases} = \frac{2}{\underline{\sigma}+\overline{\sigma}}(\overline{\sigma}\mathbb{1}_{\{x\le 0\}} + \underline{\sigma}\mathbb{1}_{\{x>0\}}).$$

It indicates that we can do a straightforward change of measure if we simply want to transform the density from one to the other (by treating the G-normal cdf as a classical cdf). We can also do the rejection sampling to reject a proportion of samples from $f_{Z_l}(x)$ to get a sample from $\overline{f}_{\mathcal{N}}(x)$.

However, one need to be careful in terms of interpretation here. When we say an iid sample drawn from $\overline{f}_{\mathcal{N}}(x)$ and $\overline{F}_{\mathcal{N}}(x)$, we actually treat these two functions as classical pdf or cdf: imagine a random variable Y in classical probability space with $f_Y = \overline{f}_{\mathcal{N}}$ and $F_Y = \overline{F}_{\mathcal{N}}$ and we draw i.i.d. sample $Y_1, Y_2, \ldots, Y_n$ from Y.

Nonetheless, these sample is not "an classically i.i.d. sample from the G-normal distribution", which is not a well-defined statement, because G-normal distribution stays in a sublinear expectation space where the concept of independence is different.

The classically i.i.d. sample do not have much connection with the G-normal because

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i) \neq \hat{\mathbf{E}}[\varphi(X_i)],$$

except for those $\varphi(x) = \mathbb{1}_{\{x \le a\}}$, not even for $\varphi(x) = \mathbb{1}_{\{x > a\}}$ nor $\varphi(x) = \mathbb{1}_{\{a < x \le b\}}$, not to mention $\varphi(x) = x^n$.

When $\underline{\sigma} = \overline{\sigma} = \sigma$, they will both retrieve the classical normal $N(0, \sigma^2)$.

Similarly, we can compare the right end with the lower cdf of the G-normal distribution. We have shown before that, for $X \sim \mathcal{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$, since $X \overset{\mathrm{d}}{=} -X$, we can show that,

$$\begin{aligned}
\hat{\mathbf{E}}[\mathbb{1}_{\{X \le -x\}}] &= \hat{\mathbf{E}}[\mathbb{1}_{\{X \le -x\}}] \\
&= \hat{\mathbf{E}}[\mathbb{1}_{\{-X > x\}}] \\
&= \hat{\mathbf{E}}[1 - \mathbb{1}_{\{-X \le x\}}] \\
&= 1 + \hat{\mathbf{E}}[-\mathbb{1}_{\{-X \le x\}}] \\
&= 1 + \hat{\mathbf{E}}[-\mathbb{1}_{\{X \le x\}}] \\
&= 1 - \mathbf{v}[X \le x].
\end{aligned}$$

Let $\underline{F}_{\mathcal{N}}(x) \coloneqq \mathbf{v}[X \le x]$, then

$$\underline{F}_{\mathcal{N}}(x) + \overline{F}_{\mathcal{N}}(-x) = 1.$$

Hence, we have the lower cdf

$$\begin{aligned}
\underline{F}_{\mathcal{N}}(x) &= 1 - \overline{F}_{\mathcal{N}}(-x) \\
&= \mathbb{1}_{\{x \ge 0\}} + \mathbb{1}_{\{x < 0\}} - \frac{2\overline{\sigma}}{\overline{\sigma} + \underline{\sigma}} \Phi(\frac{-x}{\overline{\sigma}}) \mathbb{1}_{\{x \ge 0\}} - (1 - \frac{2\underline{\sigma}}{\overline{\sigma} + \underline{\sigma}} \Phi(\frac{x}{\underline{\sigma}})) \mathbb{1}_{\{x < 0\}} \\
&= \frac{2\underline{\sigma}}{\overline{\sigma} + \underline{\sigma}} \Phi(\frac{x}{\underline{\sigma}}) \mathbb{1}_{\{x < 0\}} + (1 - \frac{2\overline{\sigma}}{\overline{\sigma} + \underline{\sigma}} \Phi(\frac{-x}{\overline{\sigma}})) \mathbb{1}_{\{x \ge 0\}}.
\end{aligned}$$

and the lower pdf,

$$\begin{aligned}
\underline{f}_{\mathcal{N}}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} \underline{F}_{\mathcal{N}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} (1 - \overline{F}_{\mathcal{N}}(-x)) \\
&= \overline{f}_{\mathcal{N}}(-x) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{(\underline{\sigma} + \overline{\sigma})/2} \big[e^{-x^2/(2\overline{\sigma}^2)} \mathbb{1}_{\{x \ge 0\}} + e^{-x^2/(2\underline{\sigma}^2)} \mathbb{1}_{\{x < 0\}} \big].
\end{aligned}$$

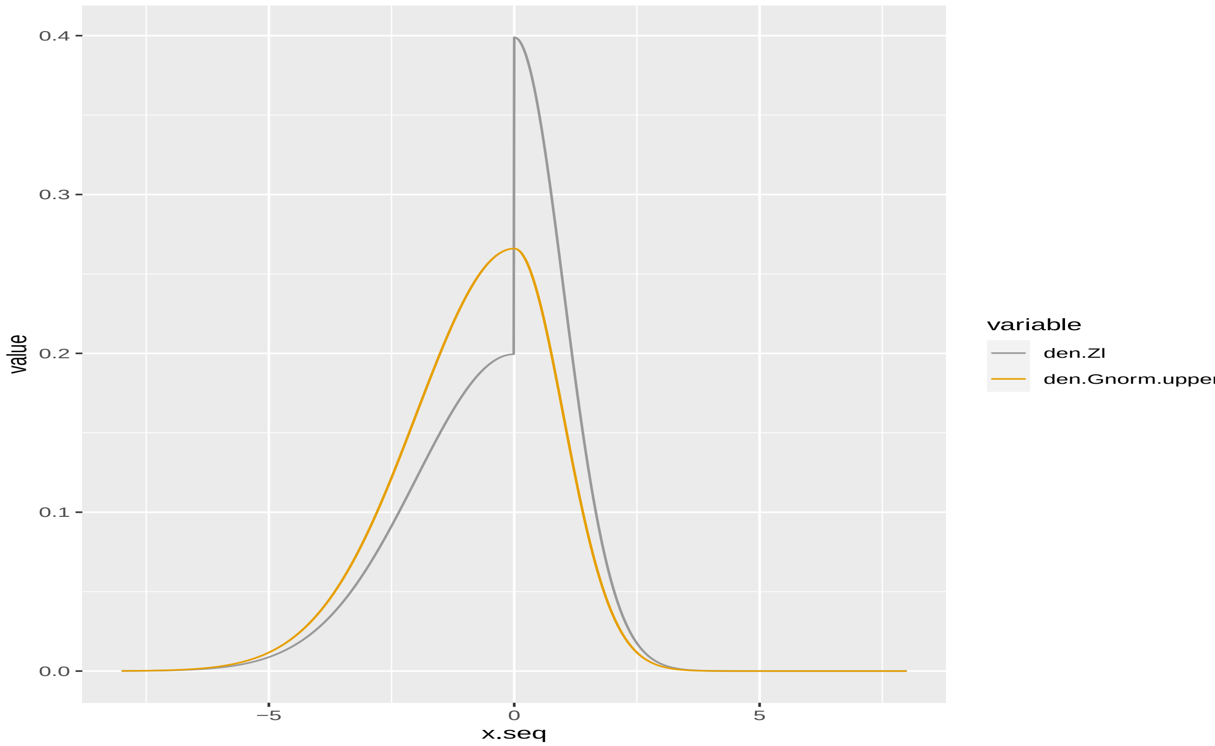
Meanwhile,

$$\begin{aligned}
f_{Z_r}(x) &= \frac{1}{\overline{\sigma}} \phi(\frac{x}{\overline{\sigma}}) \mathbb{1}_{\{x \ge 0\}} + \frac{1}{\underline{\sigma}} \phi(\frac{x}{\underline{\sigma}}) \mathbb{1}_{\{x < 0\}}. \\
&= \frac{1}{\sqrt{2\pi}} \big[\frac{1}{\overline{\sigma}} e^{-x^2/(2\overline{\sigma}^2)} \mathbb{1}_{\{x \ge 0\}} + \frac{1}{\underline{\sigma}} e^{-x^2/(2\underline{\sigma}^2)} \mathbb{1}_{\{x < 0\}} \big].
\end{aligned}$$

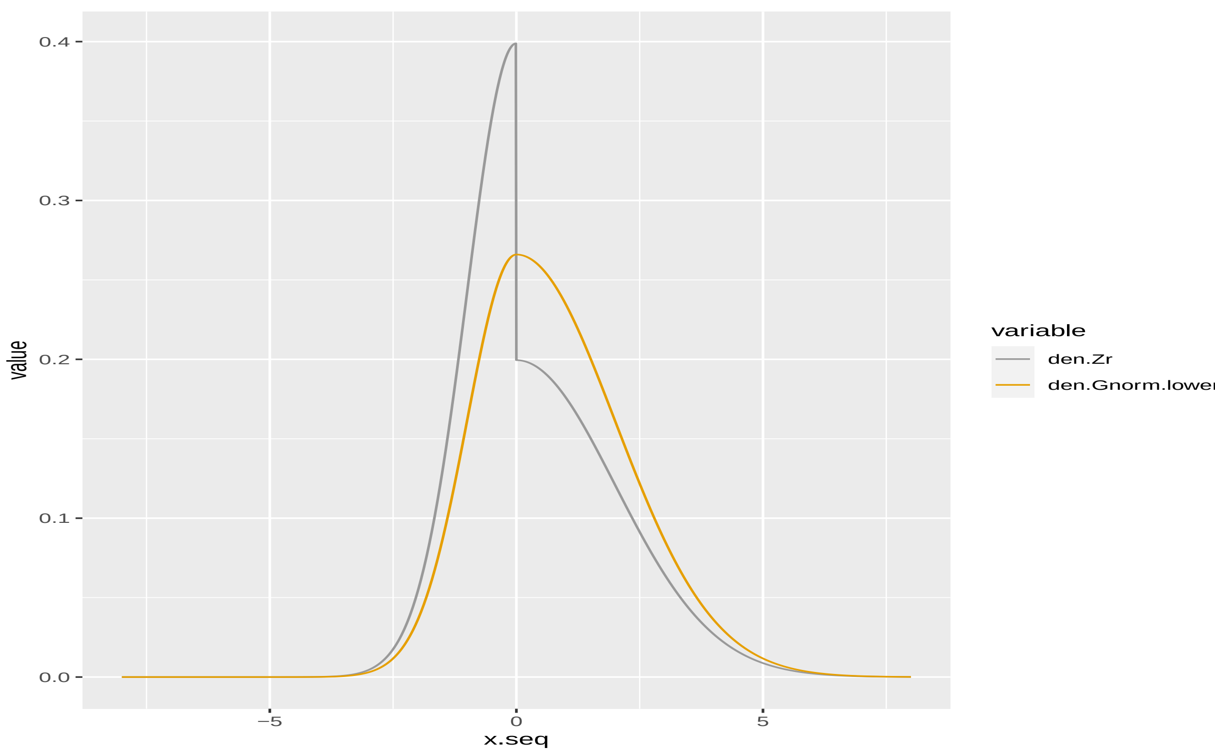
Therefore,

$$\frac{\underline{f}_{\mathcal{N}}(x)}{f_{Z_r}(x)} = \begin{cases} \frac{2\overline{\sigma}}{\underline{\sigma} + \overline{\sigma}} & \text{if } x \ge 0 \\ \frac{2\underline{\sigma}}{\underline{\sigma} + \overline{\sigma}} & \text{if } x < 0 \end{cases} = \frac{2}{\underline{\sigma} + \overline{\sigma}} (\overline{\sigma} \mathbb{1}_{\{x \ge 0\}} + \underline{\sigma} \mathbb{1}_{\{x < 0\}}).$$

Here is a comparison between the $\overline{f}_{\mathcal{N}}(x)$ and $f_{Z_l}(x)$.



We also have the comparison between the $\underline{f}_{\mathcal{N}}(x)$ and $f_{Z_r}(x)$.



4.1.7 Sampling from G-normal cdf

One question we may ask, since it is easy to sample from Z_l, how can we modify the sample from $f_{Z_l}(x)$, under this likelihood ratio, so that we can get a sample from $\overline{f}_{\mathcal{N}}(x)$? (So far I do not think there is a simple way to do the sampling directly from transforming the samples from $f_{Z_l}(x)$, except we use inverse cdf, as shown in the ref:sim-G-BM or rejection sampling techniques.)

If we use rejection sampling, let $g(x) = f_{Z_l}(x)$ and $f(x) = \overline{f}_{\mathcal{N}}(x)$, our goal is to sample from $f(x)$ based on the sampling from $g(x)$. (Try to see whether there is any simple forms coming out.) Since we have the nice result that,

$$\frac{f(x)}{g(x)} = \frac{2}{\underline{\sigma} + \overline{\sigma}}(\overline{\sigma}\mathbb{1}_{\{x\leq 0\}} + \underline{\sigma}\mathbb{1}_{\{x>0\}}) \leq \frac{2\overline{\sigma}}{\underline{\sigma} + \overline{\sigma}},$$

we can choose

$$M := \frac{2\overline{\sigma}}{\underline{\sigma} + \overline{\sigma}}.$$

Then we can perform the rejection sampling:

1. Sample X from $g(x)$,
2. Generate U from Unif(0,1), (independent from X)
3. (R1) If
$$U \leq \frac{f(X)}{Mg(X)} = \frac{1}{\overline{\sigma}}(\underline{\sigma}\mathbb{1}_{\{X>0\}} + \overline{\sigma}\mathbb{1}_{\{X\leq 0\}}) = \frac{\underline{\sigma}}{\overline{\sigma}}\mathbb{1}_{\{X>0\}} - \mathbb{1}_{\{X\leq 0\}},$$
set $X^* = X$ ("accept"); otherwise, go back to step 1 ("reject").
4. (or R2) If $X \leq 0$, accept. If $X > 0$, generate B from Bern($\underline{\sigma}/\overline{\sigma}$), accept X if and only if $B = 1$.
5. (or R3) If $X \leq 0$, accept. Accept the around first $\underline{\sigma}/\overline{\sigma}$ fraction of X with $X > 0$.

One note here, as long as the generated $X > 0$ it will always accept. The acceptance probability is

$$\frac{1}{M} = \frac{\underline{\sigma} + \overline{\sigma}}{2\overline{\sigma}}.$$

Accordingly, we can use the rejection sampling to sample from $f(x) = \underline{f}_{\mathcal{N}}(x)$ based on the sample from $g(x) = f_{Z_r}(x)$. (Try to see whether there is any simple forms coming out.) Since we have the nice result that,

$$\frac{f(x)}{g(x)} = \frac{2}{\underline{\sigma} + \overline{\sigma}}(\overline{\sigma}\mathbb{1}_{\{x\geq 0\}} + \underline{\sigma}\mathbb{1}_{\{x<0\}}) \leq \frac{2\overline{\sigma}}{\underline{\sigma} + \overline{\sigma}},$$

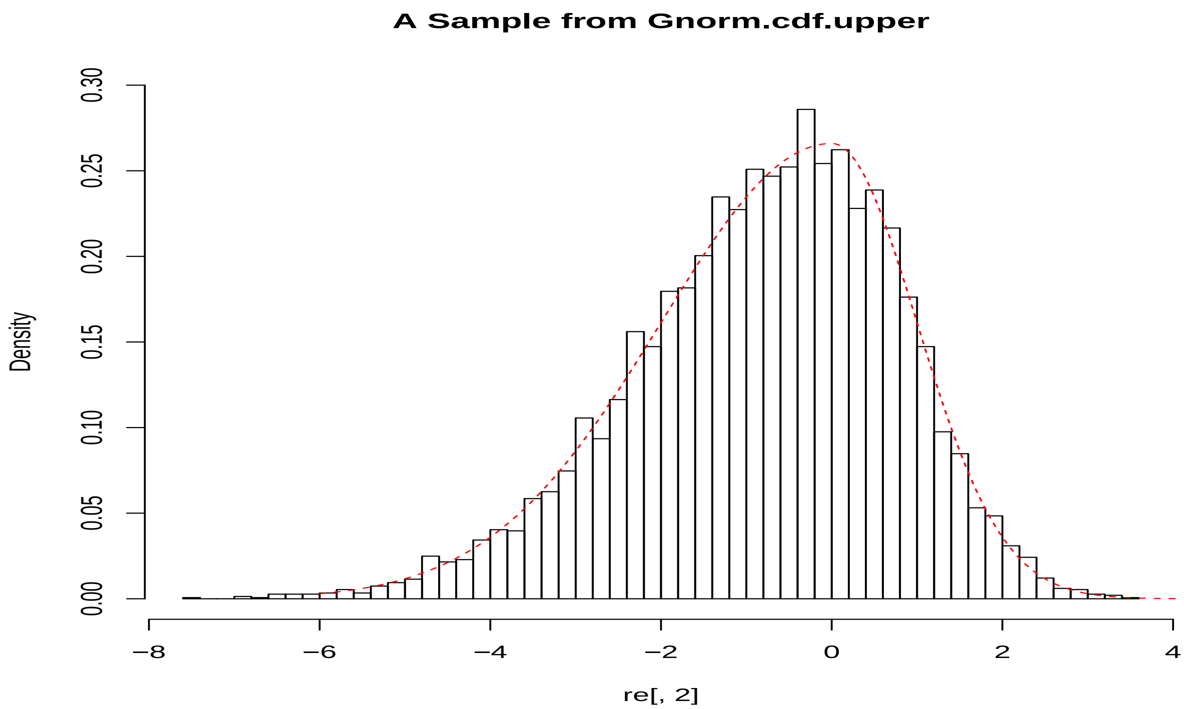
we can choose

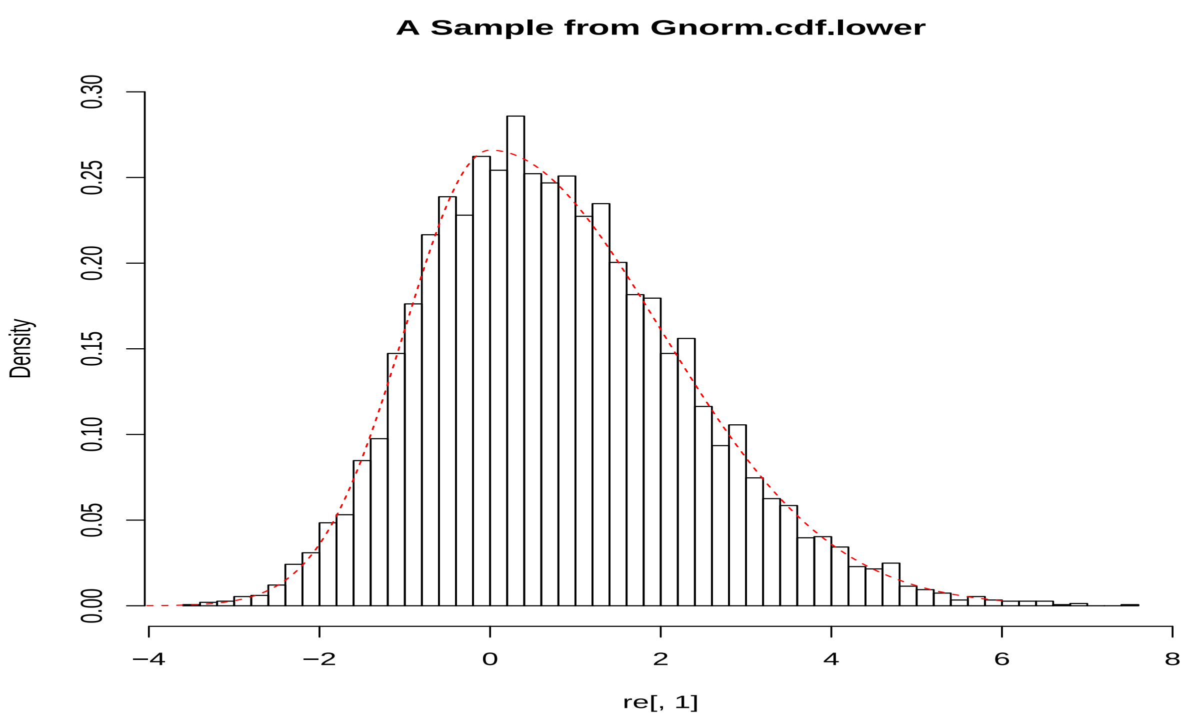
$$M := \frac{2\overline{\sigma}}{\underline{\sigma} + \overline{\sigma}}.$$

Then we can perform the rejection sampling:

1. Sample X from $g(x)$,
2. Generate U from Unif(0,1), (independent from X)
3. If
$$U \leq \frac{f(X)}{Mg(X)} = \frac{1}{\overline{\sigma}}(\overline{\sigma}\mathbb{1}_{\{X\geq 0\}} + \underline{\sigma}\mathbb{1}_{\{X<0\}}) = \frac{\underline{\sigma}}{\overline{\sigma}}\mathbb{1}_{\{X<0\}} + \mathbb{1}_{\{X\geq 0\}},$$
set $X^* = X$ ("accept"); otherwise, go back to step 1 ("reject").
4. (or) If $X \geq 0$, accept. If $X < 0$, generate B from Bern($\underline{\sigma}/\overline{\sigma}$), accept X if and only if $B = 1$.

Then we can implement the rejection sampling.





4.1.8 Estimation using the distributions of the observed data

One may ask, in this simple setup, since we can explicitly derive the (marginal) distributions of $\tilde{\boldsymbol{X}}_{tl}$ and $\tilde{\boldsymbol{X}}_{tr}$ based on the property of $\boldsymbol{X}_t$, can we directly use traditional estimation approachs based on the observed $\tilde{\boldsymbol{X}}_{tl}$ and $\tilde{\boldsymbol{X}}_{tr}$?

To clarify this concern, we will explore this direction here.

If we directly work on the two-end or the center-range quantities of $\tilde{\boldsymbol{X}}_t$, based on their distributions, we have two directions to do the estimation of $(\mu, \underline{\sigma}, \overline{\sigma})$ (the estimation of $(\underline{\sigma}, \overline{\sigma})$ is needed to construct confidence interval for μ):

1. Use MM (method of moments) or MLE based on the distributions of $[\tilde{\boldsymbol{X}}_{tl}, \tilde{\boldsymbol{X}}_{tr}]$ to estimate $(\mu, \underline{\sigma}, \overline{\sigma})$;
2. Use estimate μ and $(\underline{\sigma} + \overline{\sigma})$ from $\mathbf{C}(\tilde{\boldsymbol{X}}_t)$ and then estimate $(\overline{\sigma} - \underline{\sigma})$ from on $\mathbf{R}(\tilde{\boldsymbol{X}}_t)$ which follows a half normal distribution.

Since the first one involve the moments (which is related to the folded normal distribution) which is relatively complex in the form (maybe save it for next week), here we first persue the second direction (using MM first). Based on the first moment of the half normal distribution,

$$\mathbf{E}[\mathbf{R}(\tilde{\boldsymbol{X}}_t)] = \sqrt{\frac{2}{\pi}}(\overline{\sigma} - \underline{\sigma}),$$

we have

$$\begin{aligned}
\hat{\mu}_{\text{cr}} =& \hat{\mu}_{\text{cen}} := \frac{1}{n}\sum_{t=1}^{n} \mathbf{C}(\tilde{\boldsymbol{X}}_t), \\
\hat{\overline{\sigma}}_{\text{cr}} &= (\hat{a} + \hat{b})/2, \\
\hat{\underline{\sigma}}_{\text{cr}} &= (\hat{a} - \hat{b})/2,
\end{aligned}$$

where

$$\begin{aligned}
\hat{a} &= 2\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(\mathbf{C}(\tilde{\boldsymbol{X}}_t) - \hat{\mu})^2}, \\
\hat{b} &= \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{t=1}^{n}\mathbf{R}(\tilde{\boldsymbol{X}}_t).
\end{aligned}$$

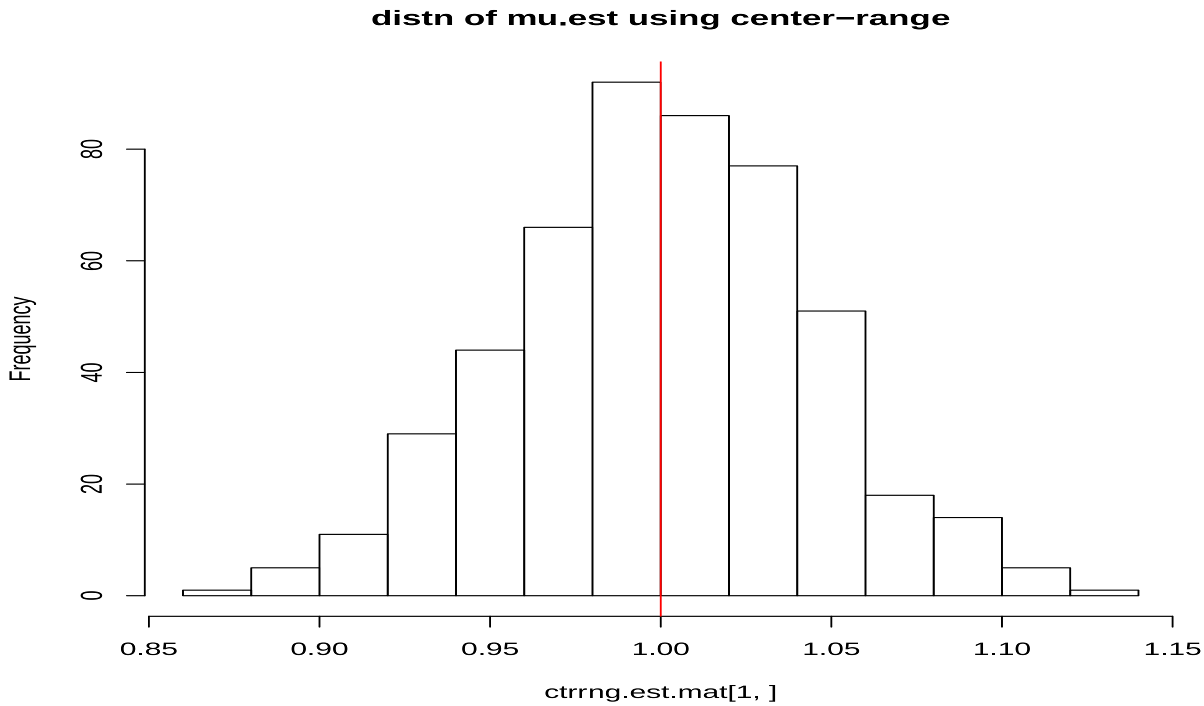
The $1 - \alpha$ confidence interval for μ is quite standard,

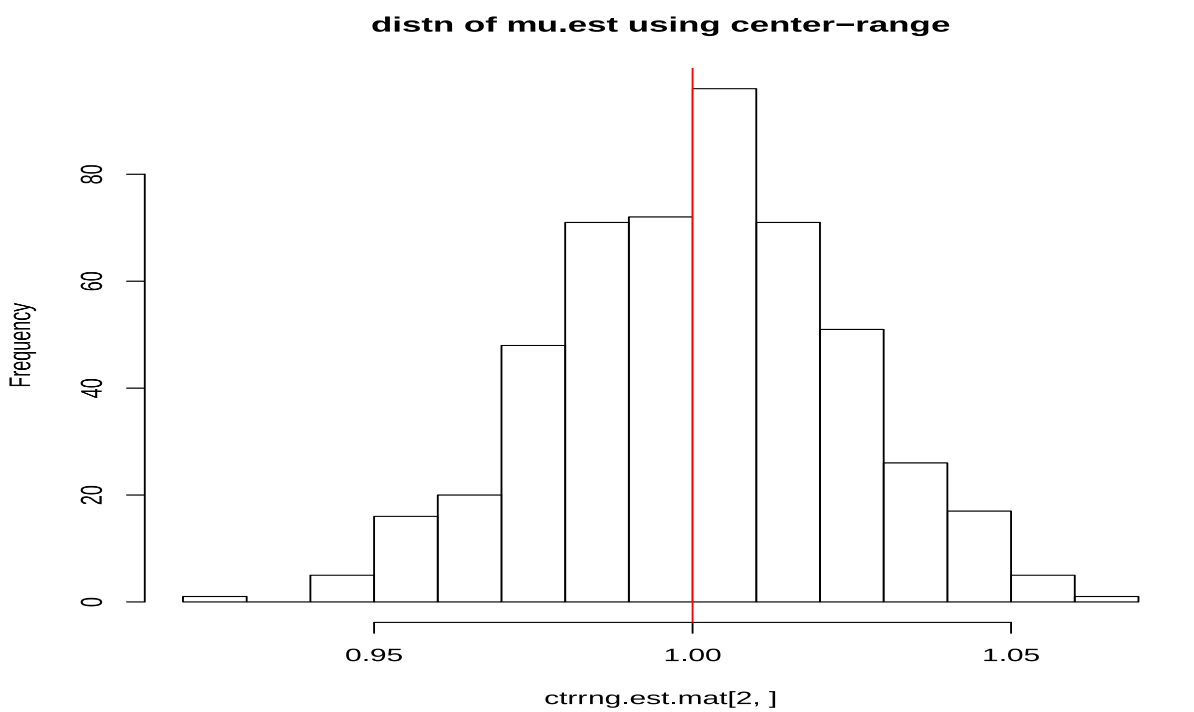
$$\hat{\mu}_{\text{cr}} \pm t_{\alpha/2, n-1}\sqrt{\frac{\hat{a}}{2n}}.$$

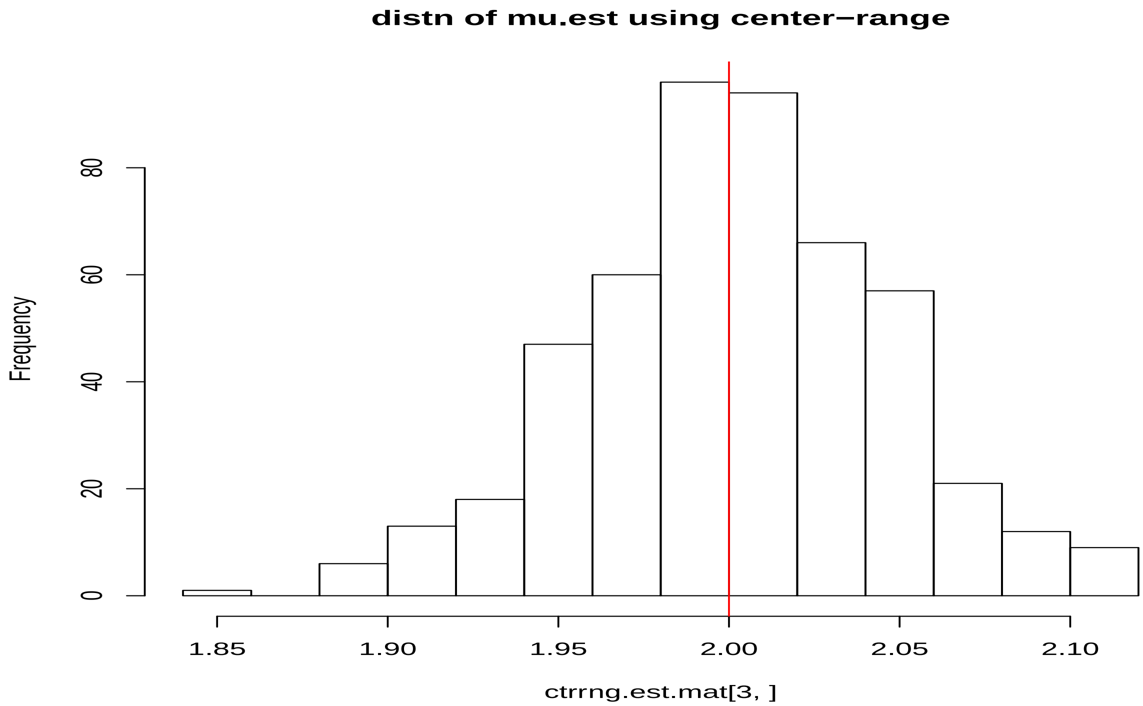
We can implement this method under parameter setup: $(\mu, \underline{\sigma}, \overline{\sigma}) = (1, 1, 2)$.

```
## $par.est
##    mu.est   sdl.est   sdr.est
## 1.0241918 0.9953161 1.9797688
##
## $confint.est
## [1] 0.9485069 1.0998767
```

Then we can check the sampling distribution.


distn of mu.est using center−range
Frequency
0
20
40
60
80
0.85
0.90
0.95
1.00
1.05
1.10
1.15
ctrrng.est.mat[1,]


distn of mu.est using center−range
Frequency
0
20
40
60
80
0.95
1.00
1.05
ctrrng.est.mat[2,]



We can also compute the bias and standard error for each estimator by MC.

```
##                 Bias         SE
## mu.est  8.429365e-05 0.04391049
## sdl.est 6.981626e-04 0.02263543
## sdr.est 1.707093e-03 0.04431292
```

4.1.9 Estimations based on the interval adjustment

Another useful technique we propose here is the so-called **interval adjustment**: under the model setup, we want to adjust $\tilde{\boldsymbol{X}}_t$ into one appropriate version $\boldsymbol{X}_t'$ that can approximately match the model specification of $\boldsymbol{X}_t$. In other words, adjust the observed dataset to retrieve the underlying interval process based our current model. The key idea here is the observed $\tilde{\boldsymbol{X}}_t$ has the same **center** as $\boldsymbol{X}_t$ which shows the sign of ϵ_t.

Let

$$\tilde{C}_t := \mathbf{C}(\tilde{\boldsymbol{X}}_t),$$

Consider

$$\hat{\mu}_{\text{cen}} := \frac{1}{n}\sum_{t=1}^{n}\tilde{C}_t.$$

Since we have seen that

$$\tilde{C}_t - \mu = \frac{\underline{\sigma} + \overline{\sigma}}{2}\epsilon_t.$$

We can first use ths sign of

$$e_t = \tilde{C}_t - \hat{\mu}_{\text{cen}}$$

to approximately show the sign of the underlying ϵ_i. Based on the relationship:

$$\begin{aligned}\mathbf{R}(\boldsymbol{X}_t) &= (\overline{\sigma} - \underline{\sigma})\epsilon_t \\ &= |(\overline{\sigma} - \underline{\sigma})\epsilon_t|\text{sign}(\epsilon_t) \\ &= \mathbf{R}(\tilde{\boldsymbol{X}}_t)\text{sign}(\epsilon_t).\end{aligned}$$

We have

$$\boldsymbol{X}_t = \tilde{\boldsymbol{X}}_t \mathbb{1}_{\{\epsilon_t \geq 0\}} + \overline{\tilde{\boldsymbol{X}}}_t \mathbb{1}_{\{\epsilon_t < 0\}}.$$

Since the sign ϵ_t is unknown, we can use the sign of e_t to approximate, to get the adjusted interval

$$\boldsymbol{X}'_t \coloneqq \tilde{\boldsymbol{X}}_t \mathbb{1}_{\{e_t \geq 0\}} + \overline{\tilde{\boldsymbol{X}}}_t \mathbb{1}_{\{e_t < 0\}},$$

as an approximation of $\boldsymbol{X}_t$ whose two ends follow normal distributions.

Then we can use the estimators,

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{\text{adj}} &= [\hat{\mu}_l, \hat{\mu}_r] = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}'_t, \\ \hat{\underline{\sigma}}_{\text{adj}} &= \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(\boldsymbol{X}'_{tl} - \hat{\mu}_l)^2}, \\ \hat{\overline{\sigma}}_{\text{adj}} &= \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(\boldsymbol{X}'_{tr} - \hat{\mu}_r)^2}.\end{aligned}$$

To derive the confidence interval, we need to consider the error in the adjustment step which can be described as, for each fixed $t = 1, 2, \ldots, n$, let $Y_t = \tilde{C}_t - \mu$, then $Y_t \sim N(0, (\underline{\sigma} + \overline{\sigma})^2/4)$, and then $e_t = Y_t - \bar{Y}_n$,

$$\begin{aligned}\mathbf{P}(\boldsymbol{X}'_t = \boldsymbol{X}_t) &= \mathbf{P}(Y'_t = Y_t) \\ &= \mathbf{P}(e_t\epsilon_t \geq 0) \\ &= \mathbf{P}((Y_t - \bar{Y}_n)Y_t \geq 0).\end{aligned}$$

Let $t = 1$, since

$$\bar{Y}_n \xrightarrow{\mathbf{P}} 0,$$

then we can apply the sluskty theorem

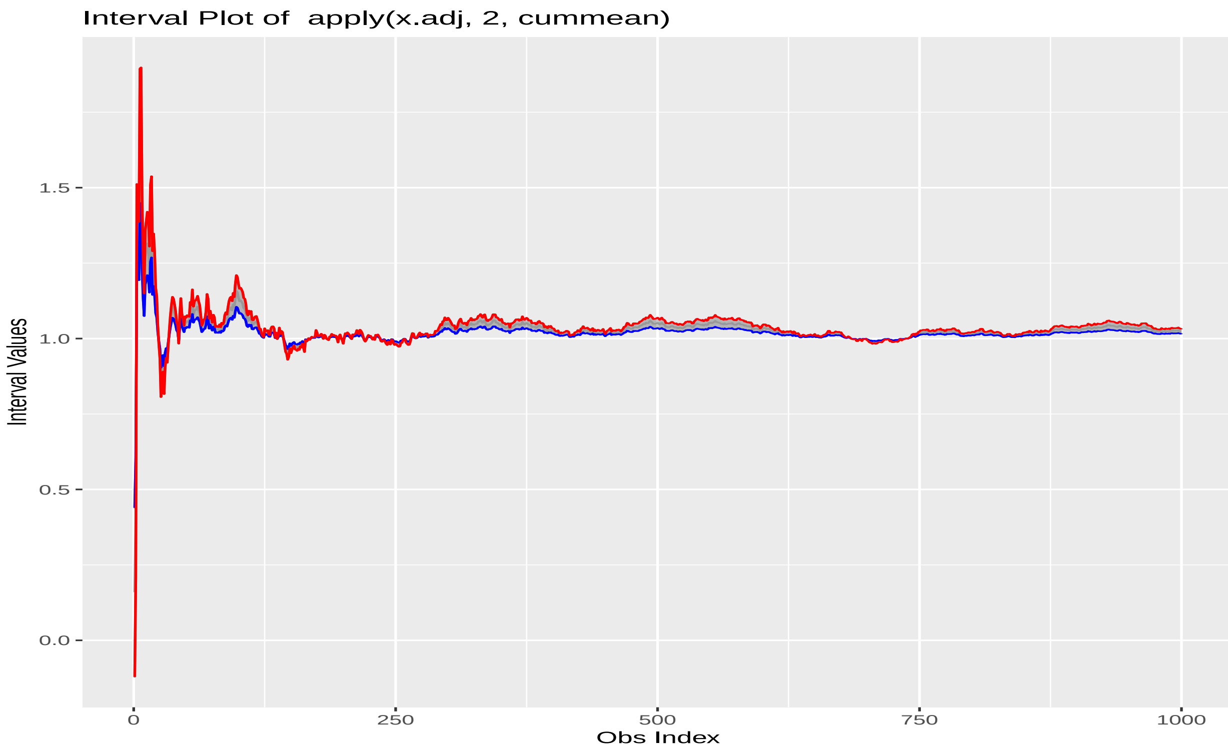
$$Y_1^2 - \bar{Y}_n Y_1 \xrightarrow{\text{d}} Y_1^2.$$

Therefore,

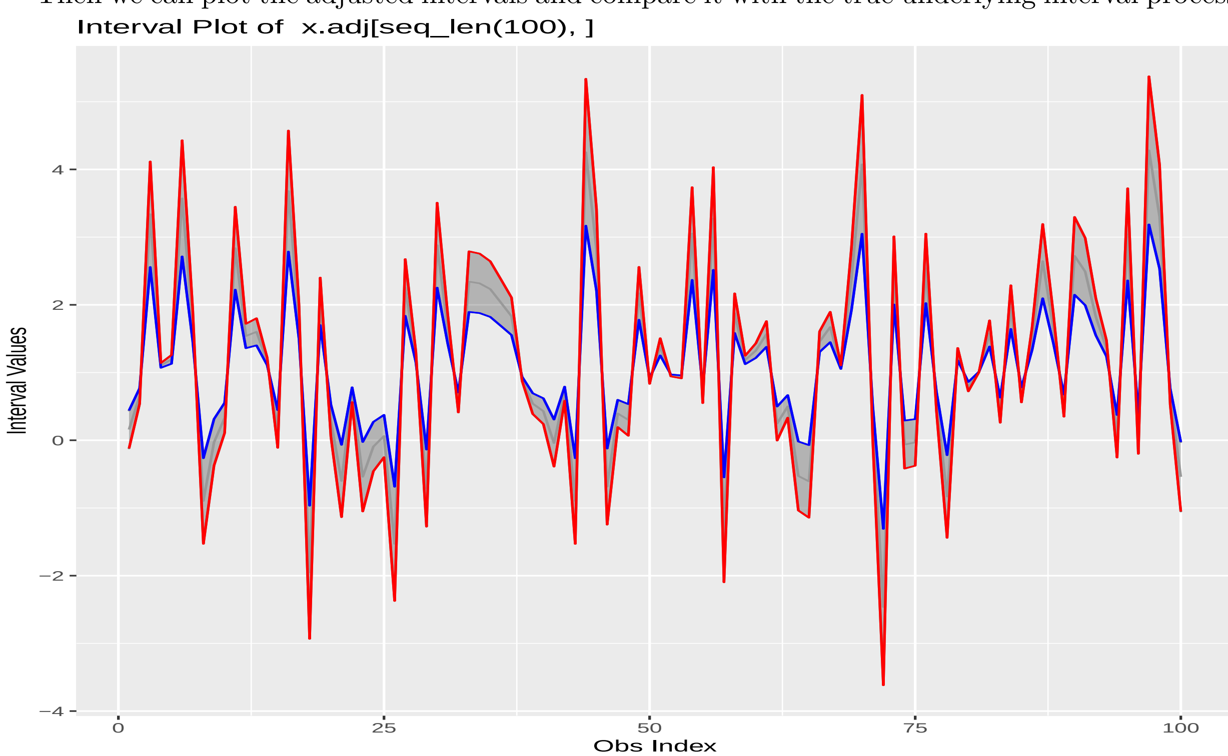
$$\mathbf{P}(\boldsymbol{X}'_t = \boldsymbol{X}_t) \to \mathbf{P}(Y_1^2 \geq 0) = 1.$$

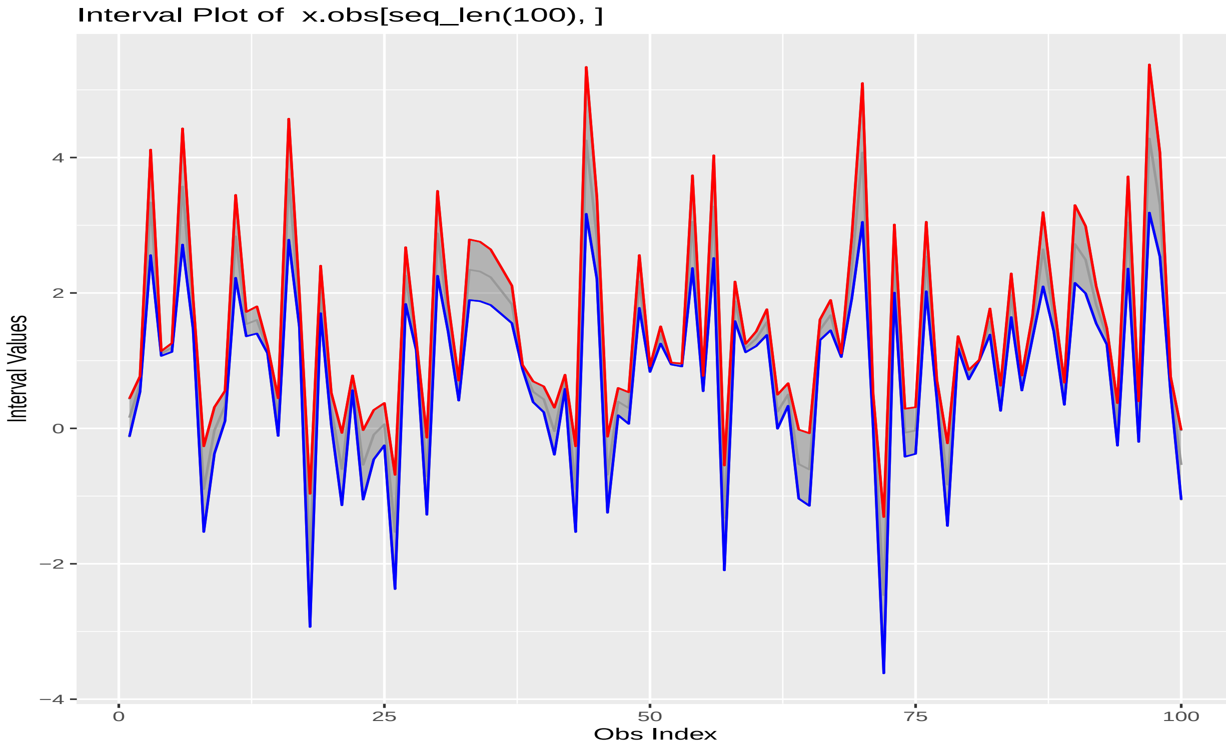
We still need to approximate the convergence rate in order to study the confidence interval rigorously.

```
##     meanl.adj mean.cen.adj    meanr.adj      sdl.est      sdr.est
##     1.0161593    1.0241918    1.0322243    0.9916948    1.9833903
```

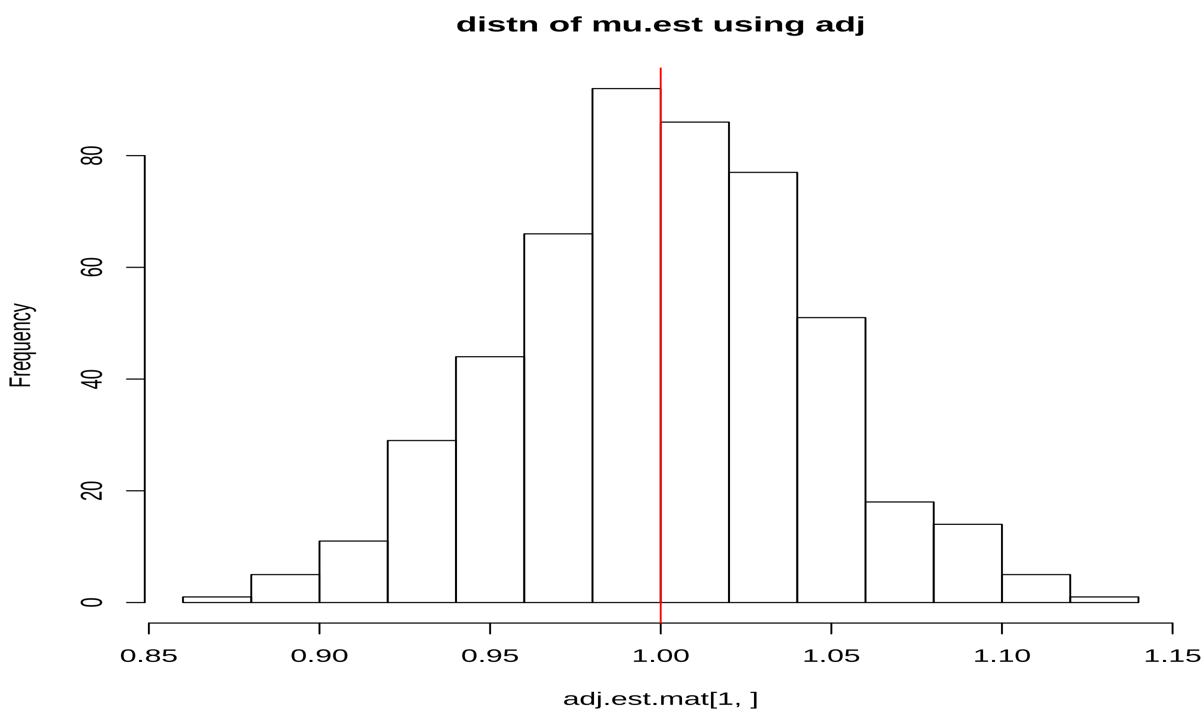


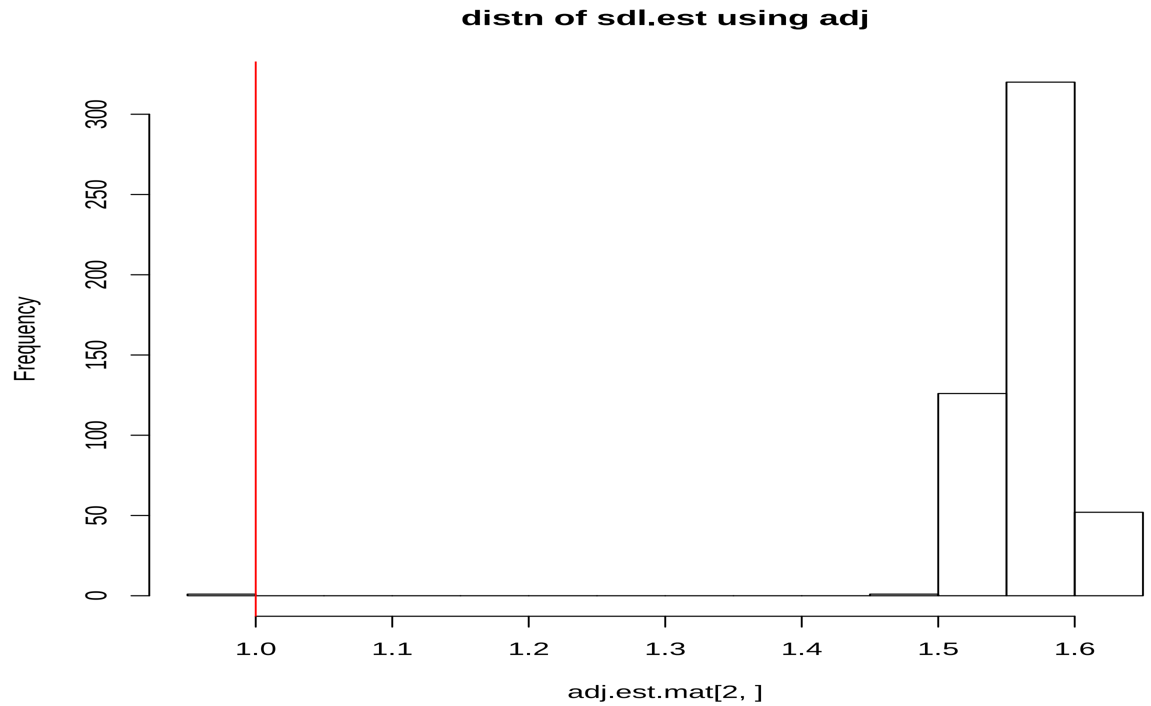
Then we can plot the adjusted intervals and compare it with the true underlying interval process.

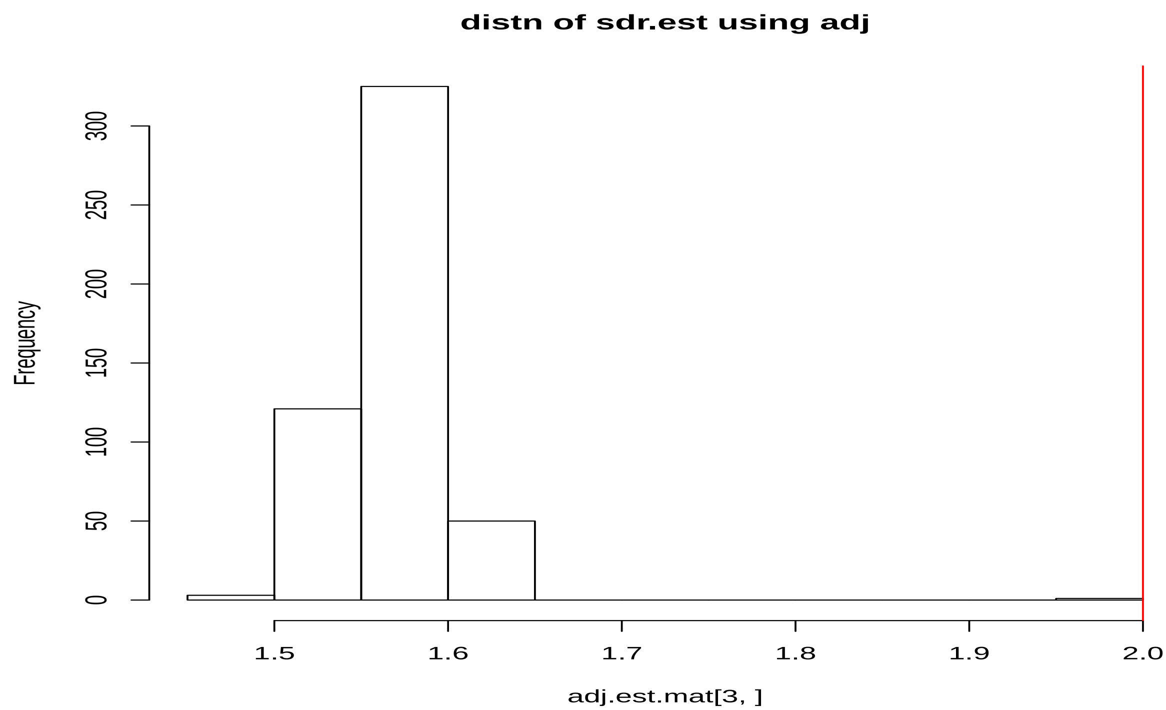




Check the sampling distribution:







Check the bias and standard error:

```
##                       Bias         SE
## mean.cen.adj  8.429365e-05 0.04391049
## sdl.est       5.666339e-01 0.03633990
## sdr.est      -4.314179e-01 0.03170808
```

4.1.10 Comparison of different estimations methods

Assessments on this method:

1. Estimation method based on center and range is straightforward to implement and so far it performs better than the adjustment method.
2. The key benefit of the interval adjustment method is that it can roughly retrieve the original intervals based on our specified model. It actually shows the fact that in this context, once we specify a model, we actually assume a dynamic for the unobserved $\boldsymbol{X}_t$.

More questions to explore:

1. How to further improve the center-range estimation method?
2. What kinds of things may be missed if we only consider the first candidate models?

(One thing is, the performance of one equipment may not be the same as time goes.)

4.2 One key concern: ambiguity in the interval direction

The first candidate model is simply a start. As we mentioned before, the key concern here is we do not know the true $\boldsymbol{X}_t$, therefore any fixed model is actually imposing one possible setup on the dynamic of $[\boldsymbol{X}_{tl}, \boldsymbol{X}_{tr}]$, that is, the probabalistic structure of each end.

For (P1), if we simply use the traditional methods to do estimation (which actually does not depend on the direction of the true $\boldsymbol{X}_t$), then we do not need to worry about the interval directions. It actually shows we are still able to learn the parameters under the ambiguity in the interval directions.

However, for (P1), if one wants to use the interval adjustment to check the possible dynamic of $\boldsymbol{X}_t$ then use the two ends of adjusted intervals to construct confidence intervals on μ. Then the interval directions does matter because we need to consider the asymptotic behaviour of

$$\boldsymbol{Y}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\boldsymbol{X}_t - \mu),$$

and put the two ends under the $\mathbf{E}[g(\cdot)]$ with $g(x) = \mathbb{1}_{\{a \le x \le b\}}$. This is related to (P2): to study the interval expectation of $[\boldsymbol{Y}_{nl}, \boldsymbol{Y}_{nr}]$ (which are not directly observed):

$$[\mathbf{E}[\varphi(\boldsymbol{Y}_{nl}), \mathbf{E}[\varphi(\boldsymbol{Y}_{nr}]].$$

When it comes to (P2), the story has just begun because the dynamic of $[\boldsymbol{X}_{tl}, \boldsymbol{X}_{tr}]$ does matter. Previously, we have specified the first candidate model:

$$\begin{aligned} \boldsymbol{X}_t^{(1)} &= \mu + [\underline{\sigma}, \overline{\sigma}]\epsilon_t, \\ \tilde{\boldsymbol{X}}_t &= 1\boldsymbol{X}_t^{(1)}. \end{aligned}$$

Then we have solved (P1) by using two ways to estimate $(\mu, \underline{\sigma}, \overline{\sigma})$. However, it turns out the underlying process $\boldsymbol{X}_t^{(1)}$ is only one possiblity of the true $\boldsymbol{X}_t$ such that

$$\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t. \tag{4.3}$$

If we consider $\boldsymbol{X}_t^{(2)} = \overline{\boldsymbol{X}_t^{(1)}}$, it also statisifies Equation (4.3) but has different dynamic from $\boldsymbol{X}_t$. In other words, if we specify the model as

$$\boldsymbol{X}_t^{(2)} = \mu + [\overline{\sigma}, \underline{\sigma}]\epsilon_t,$$

and if we do the similar estimation procedure (as shown before) based on the same dataset, we will end up with the same estimation $(\hat{\mu}, \hat{\underline{\sigma}}, \hat{\overline{\sigma}})$ (by rearranging the role of the estimators accordingly), because if we compare the range in the center-range method, $|(\underline{\sigma} - \overline{\sigma})\epsilon_t|$ and $|(\overline{\sigma} - \underline{\sigma})\epsilon_t|$ are equal (so they have the same half-normal distribution).

One may argue that $\boldsymbol{X}_t^{(2)}$ do not have any practical difference from $\boldsymbol{X}_t^{(1)}$ because it simply has opposite interval direction timewise compared with the other - it provides the same information on the true underlying process $\boldsymbol{X}_t$ if we treat equipment A and B symmetrically. This statement is only true if the true $\boldsymbol{X}_t$ only belongs to one of these two possible candidate models.

On the one hand, note that so far $\boldsymbol{X}_t^{(i)}, i = 1, 2$ both assume homoscedasticity for the two ends of $\boldsymbol{X}_t$ (they have constant variance across time). However, this assumption cannot be validated based on the information provided by the lab. Actually, in practice, it is highly likely that the equipment has unstable measurement performance in accuracy across time (due to system deterioration or other changing environmental factors).

On the other hand, in our context, since $\boldsymbol{X}_t^{(i)}, i = 1, 2$ are both candidate models statisfying Equation (4.3), if we consider a interval time series switches between $\boldsymbol{X}_t^{(1)}$ and $\boldsymbol{X}_t^{(2)}$, its proper version will still be $\tilde{\boldsymbol{X}}_t$.

Therefore, we are supposed to at least consider a larger family of candidate models. In general, if we introduce a discrete-time binary-valued process $s_t : \Omega \to \{0, 1\}$ in $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$,

$$\boldsymbol{X}_t^s := \boldsymbol{X}_t^{(1)} s_t + \boldsymbol{X}_t^{(2)}(1 - s_t),$$

then at each time t, $\boldsymbol{X}_t^s$ would be either $\boldsymbol{X}_t^{(1)}$ or $\boldsymbol{X}_t^{(2)}$, so we must have

$$\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t^s,$$

holds timewise. Meanwhile, we can notice that under this formulation,

$$\begin{aligned}
\boldsymbol{X}_t^s &= \boldsymbol{X}_t^{(1)} s_t + \boldsymbol{X}_t^{(2)}(1 - s_t) \\
&= \mu + [(\underline{\sigma} s_t + \overline{\sigma}(1 - s_t)), (\overline{\sigma} s_t + \underline{\sigma}(1 - s_t))]\epsilon_t \\
&=: \mu + [\sigma_{tl}, \sigma_{tr}]\epsilon_t.
\end{aligned}$$

It also indicates that the two ends of $\boldsymbol{X}_t^s$ could be heteroscedastic (at each t, σ_{tl} or σ_{tr} only switches between $\underline{\sigma}^2$ and $\overline{\sigma}^2$, which comes from the simplified setup of the problem by assuming the constant interval $[\underline{\sigma}, \overline{\sigma}]$; a future extension is to generalize it into $[\underline{\sigma}_t, \overline{\sigma}_t]$.)

Nonetheless, since we only have the observed data $\tilde{\boldsymbol{X}}_t$, it actually does not contain any information on the dynamic of s_t. For any two processes s and s' in a set of processes, $\boldsymbol{X}_t^s$ and $\boldsymbol{X}_t^{s'}$ may give different results in (P2) but they will result in the same observed interval data $\tilde{\boldsymbol{X}}_t$:

$$\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t^s = 1\boldsymbol{X}_t^{s'}.$$

Since s actually decides the interval direction of $\boldsymbol{X}_t$ (either the same as $[\underline{\sigma}, \overline{\sigma}]\epsilon_t$ or $[\overline{\sigma}, \underline{\sigma}]\epsilon_t$), this is called **the ambiguity in the interval direction**, which is a true ambiguity if we do not have more background information on the unobsevable interval $\boldsymbol{X}_t$, which does not really exist in the context of point-valued data because a point, as a degenerate interval, has no ambiguity in its direction. This leads to one of the essential differences from the traditional statistical methods in point-valued data.

(point-valued data may have distribution or model uncertainty in temporal sense. It is another kind of ambiguity, which we have studied before.)

(i.e. any parameters related to the dynamic of s_t are not identifiable based on the data $\tilde{\boldsymbol{X}}_t$.)

4.2.1 Family of Possible Models

When it comes to the study of the ambiguity (in the interval direction characterized by s), we can no longer only consider a single model for s but a set $\mathcal{S}$ of models, each of which will lead to an interval (linear) expectation concerned by (P2): for any $s \in \mathcal{S}$, we have the interval expectation

$$\boldsymbol{\mathcal{E}}[\boldsymbol{Y}_n^s] = [\mathbf{E}[\boldsymbol{Y}_{n\,l}^s], \mathbf{E}[\boldsymbol{Y}_{n\,r}^s]].$$

In order to construct an interval that is robust to the choice of s, as an precautious strategy, we can consider

$$\hat{\boldsymbol{\mathcal{E}}}_{\mathcal{S}}[\boldsymbol{Y}_n] \coloneqq \sup_{s\in\mathcal{S}}[\mathbf{E}[\boldsymbol{Y}_{n\,l}^s], \mathbf{E}[\boldsymbol{Y}_{n\,r}^s]],$$

where

$$\sup_{i\in I}[a_i, b_i] \coloneqq [\inf_{i\in I} a_i, \sup_{i\in I} b_i].$$

(In this context, the design of the set $\mathcal{S}$ is decided by the degree of leakage in the available information on $\boldsymbol{X}_t$, and user may also personally initialize it based on their attitude or belief and update it once having more information on $\boldsymbol{X}_t$.)

Actually, general picture for the possible specifications of candidate models: in the classical probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$, let $\epsilon_t \sim \mathrm{IID}(0,1)$, consider $s_t : \Omega \to \{0,1\}$ and s_t is (classically) independent from ϵ_t,

1. $\boldsymbol{X}_t^{(1)} = \mu + [\underline{\sigma}, \overline{\sigma}]\epsilon_t$,
2. $\boldsymbol{X}_t^{(2)} = \mu + [\overline{\sigma}, \underline{\sigma}]\epsilon_t = \overline{\boldsymbol{X}_t^{(1)}}$,
3. $\boldsymbol{X}_t^{(3)} = \boldsymbol{X}_t^{(1)} s_t + \boldsymbol{X}_t^{(2)}(1-s_t)$, where $s_t : \Omega \to \{0,1\}$ is any Markov process,
4. $\boldsymbol{X}_t^{(4)} = \boldsymbol{X}_t^{(1)} s_t + \boldsymbol{X}_t^{(2)}(1-s_t)$, where $s_t : \Omega \to \{0,1\}$ is any $\sigma(s_{t-k}, k\geq 1)$-measurable process,
5. $\boldsymbol{X}_t^{(5)} = \boldsymbol{X}_t^{(1)} s_t + \boldsymbol{X}_t^{(2)}(1-s_t)$, where $s_t : \Omega \to \{0,1\}$ is any $\sigma(s_{t-k}, \epsilon_{t-k}, k\geq 1)$-measurable process.

One example for the 3rd case, s_t has transition probability,

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix},$$

where $p_{ij} \coloneqq \mathbf{P}(s_t = j | s_t = i)$ with $i,j = 0,1$. Then in this case, $\boldsymbol{X}_t^{(3)}$ can be treated as an interval-valued hidden Markov model. For the 5th case,

$$\begin{aligned}\boldsymbol{X}_t^{(5)} &= \boldsymbol{X}_t^{(1)} s_t + \boldsymbol{X}_t^{(2)}(1-s_t) \\ &= \mu + [(\underline{\sigma} s_t + \overline{\sigma}(1-s_t)), (\overline{\sigma} s_t + \underline{\sigma}(1-s_t))]\epsilon_t \\ &\eqqcolon \mu + [\sigma_{tl}, \sigma_{tr}]\epsilon_t.\end{aligned}$$

If we let $\mathcal{F}_t \coloneqq \sigma(s_{t-k}, \epsilon_{t-k}, k\geq 0)$, we can see that

$$\sigma_{tl} : \Omega \to \{\underline{\sigma}, \overline{\sigma}\}$$

is a $\mathcal{F}_{t-1}$-measurable process, in other words, a predictable process; so is σ_{tr} because $\sigma_{tr} = (\underline{\sigma} + \overline{\sigma}) - \sigma_{tl}$.

Any more practical examples that we need to consider (P2):

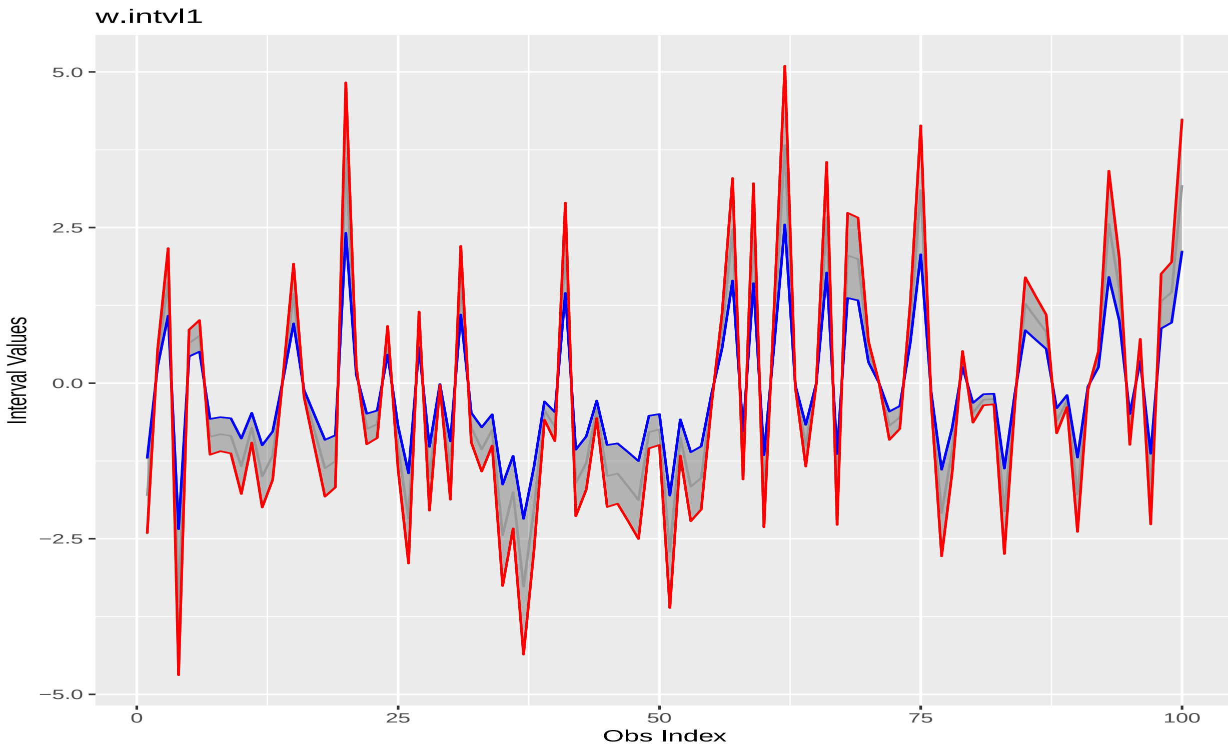
1. (in the context of interval-valued log return data)

4.2.2 Illustration of this ambiguity

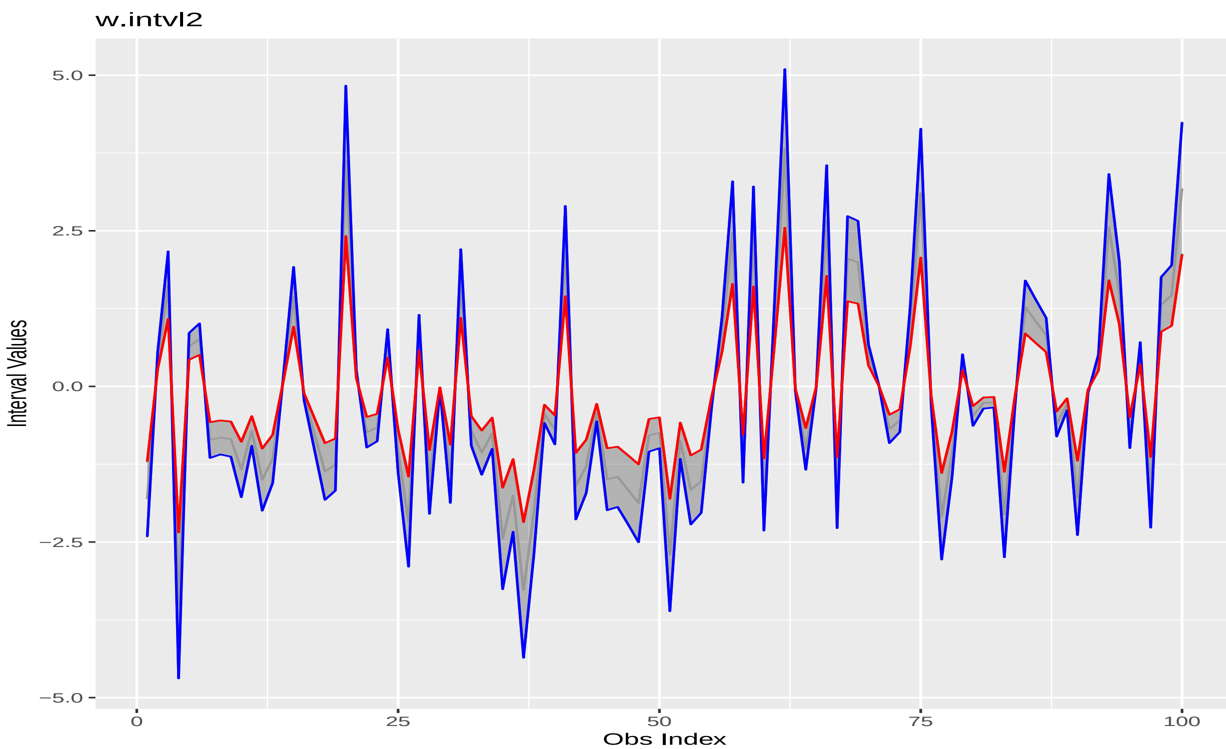
1. $W_t^{(1)} = [\underline{\sigma}, \overline{\sigma}]\epsilon_t$,
2. $W_t^{(2)} = [\overline{\sigma}, \underline{\sigma}]\epsilon_t$,
3. $W_t^{(3)} = W_t^{(1)} s_t + W_t^{(2)}(1 - s_t)$, $s_t \sim Bern(1/2)$. An independent Bernoulli Mixture of version 1 and 2.
4. $W_t^{(4)} = W_t^{(1)} s_t + W_t^{(2)}(1 - s_t)$, s_t has transition probability,

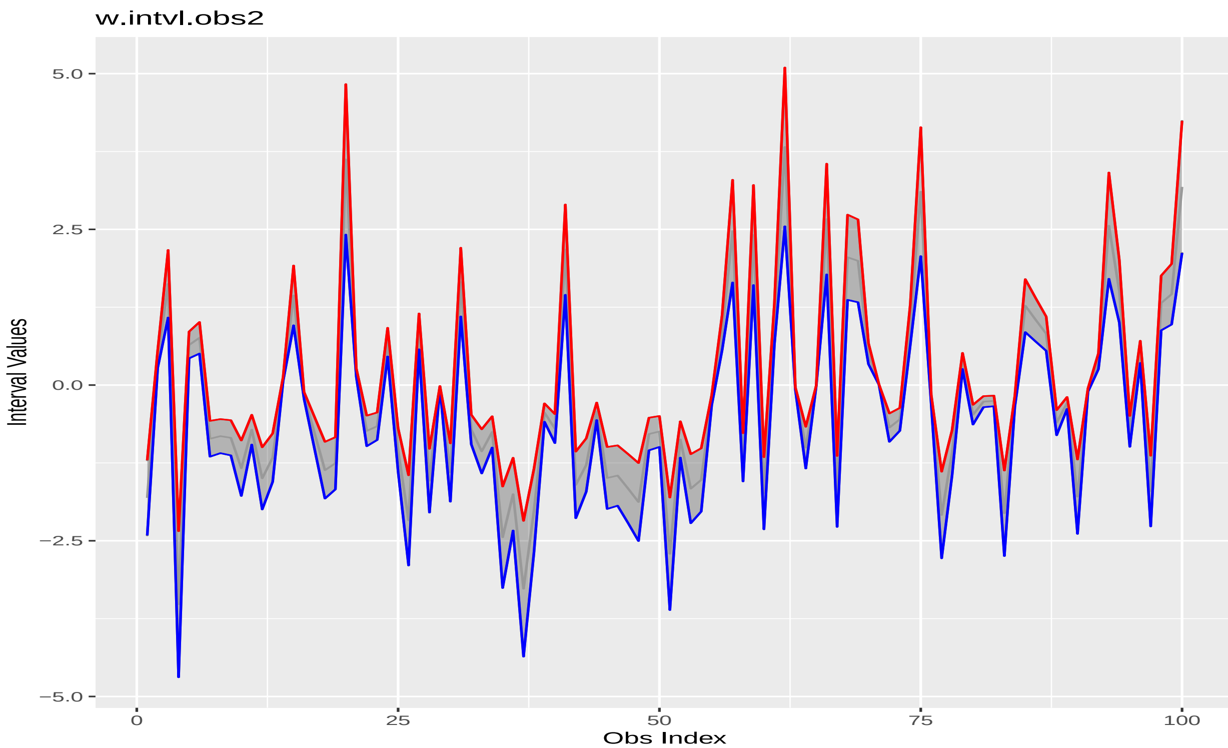
$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}.$$

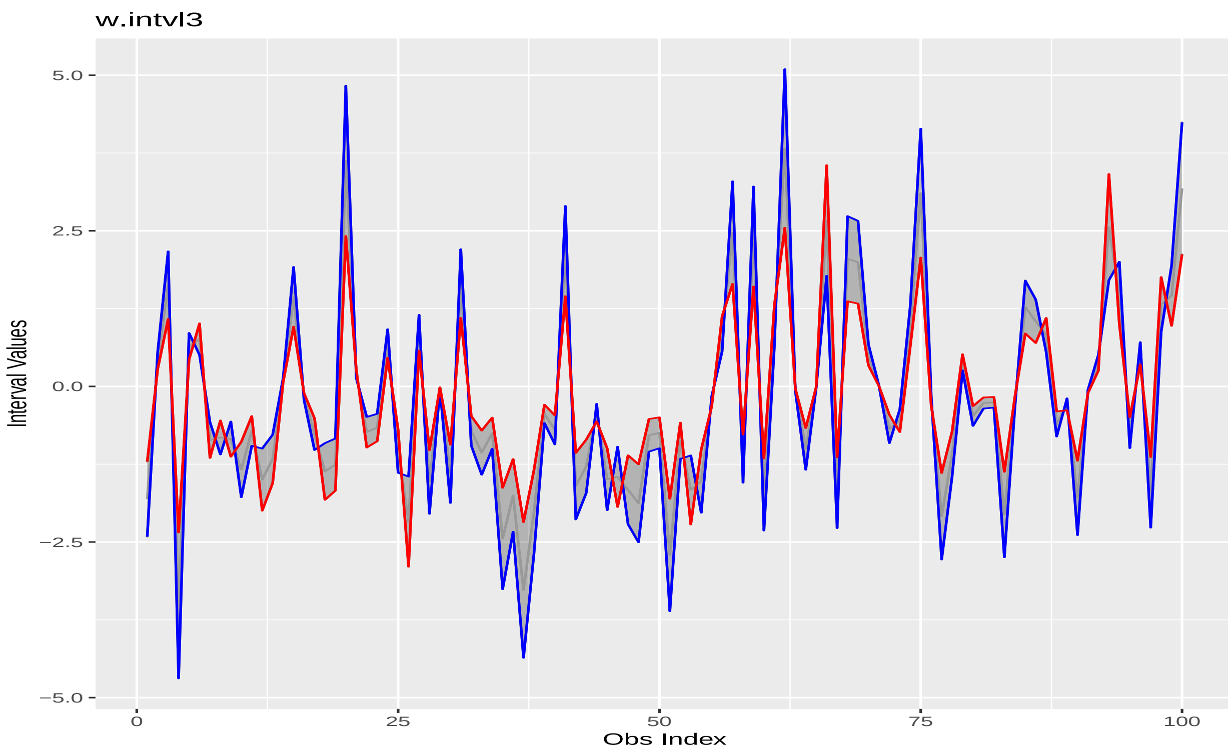
   A HMM-mixture of version 1 and 2.
5. $W_t^{(5)} = W_t^{(1)} s_t + W_t^{(2)}(1 - s_t)$, $s_t = \mathbb{1}_{\{\epsilon_{t-1} > 0\}}$.

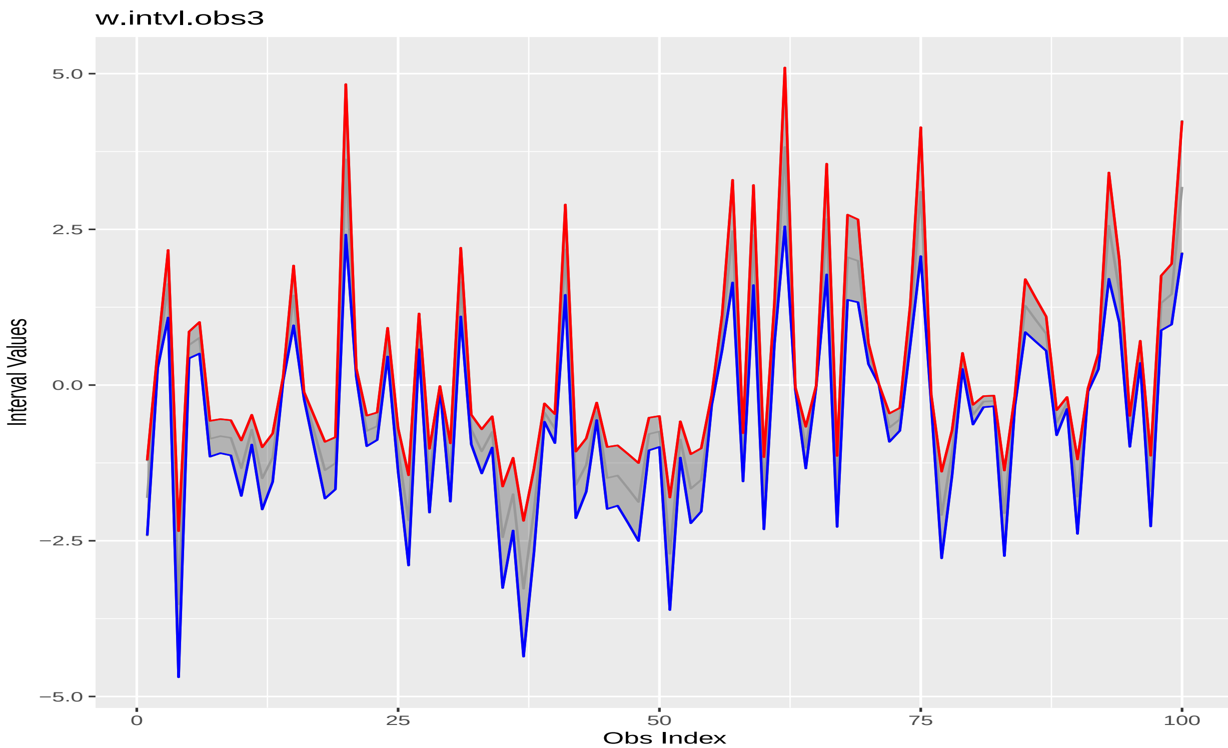


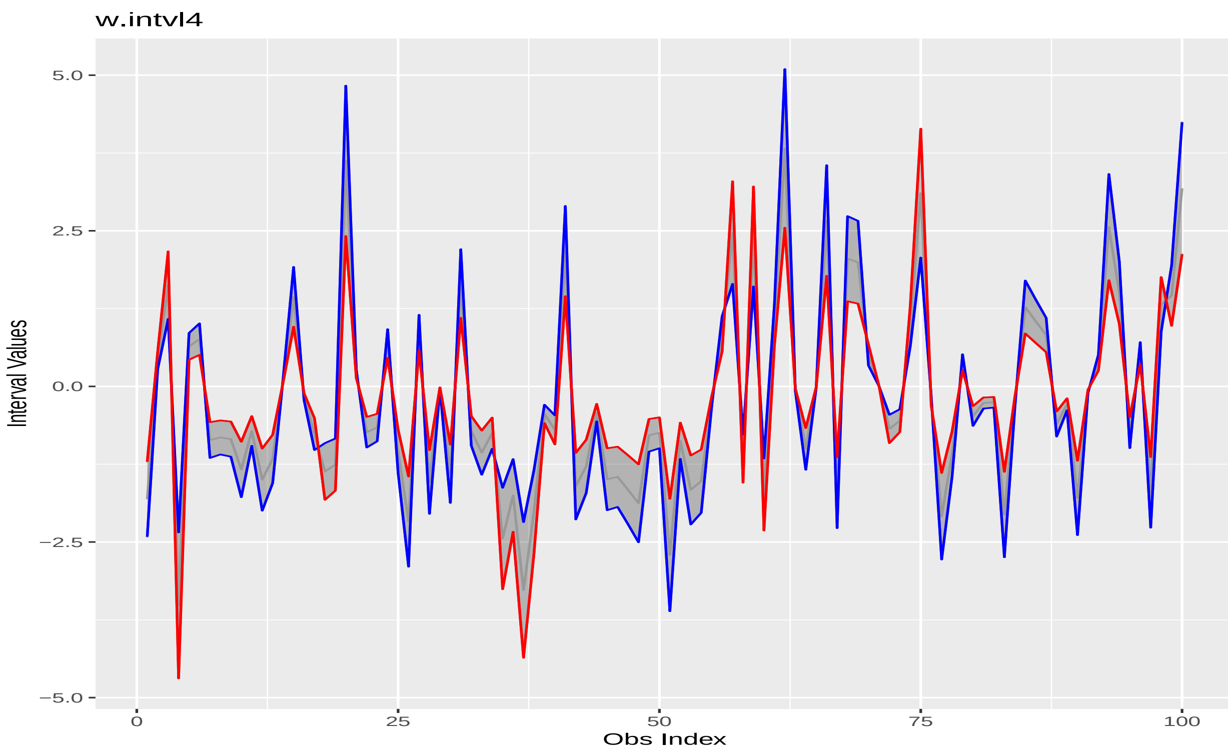

w.intvl.obs1
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

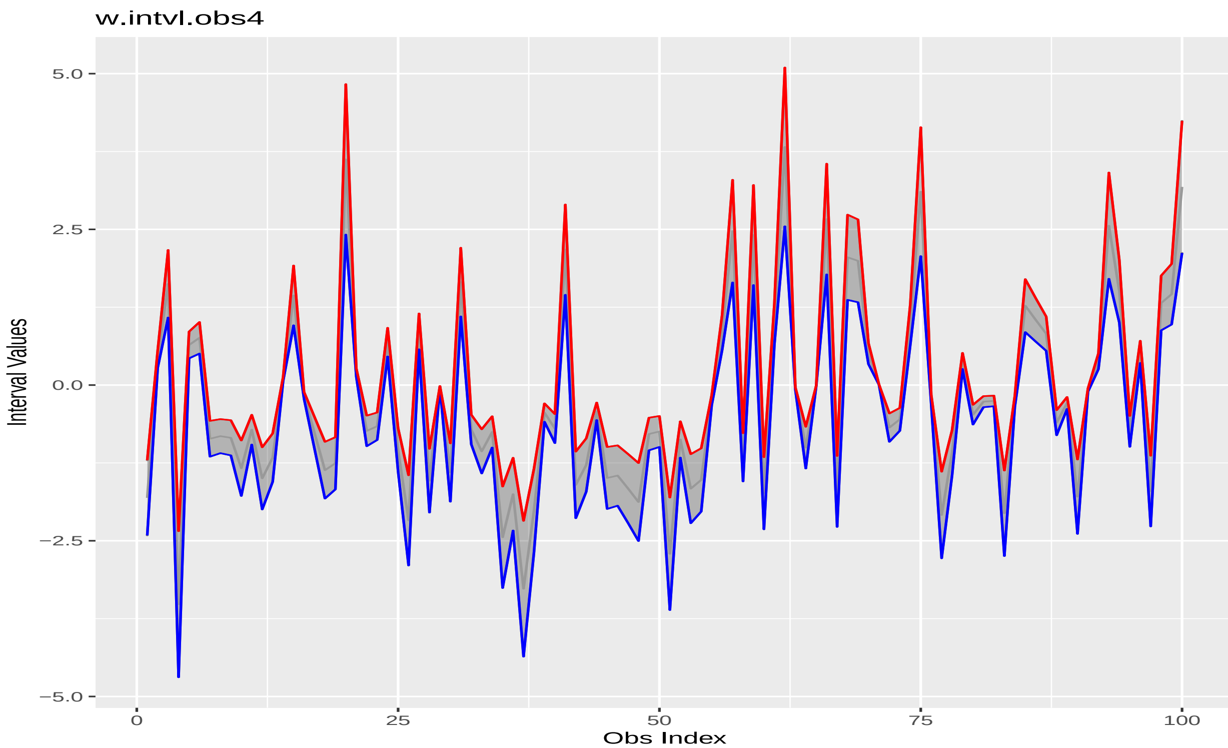

w.intvl2
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

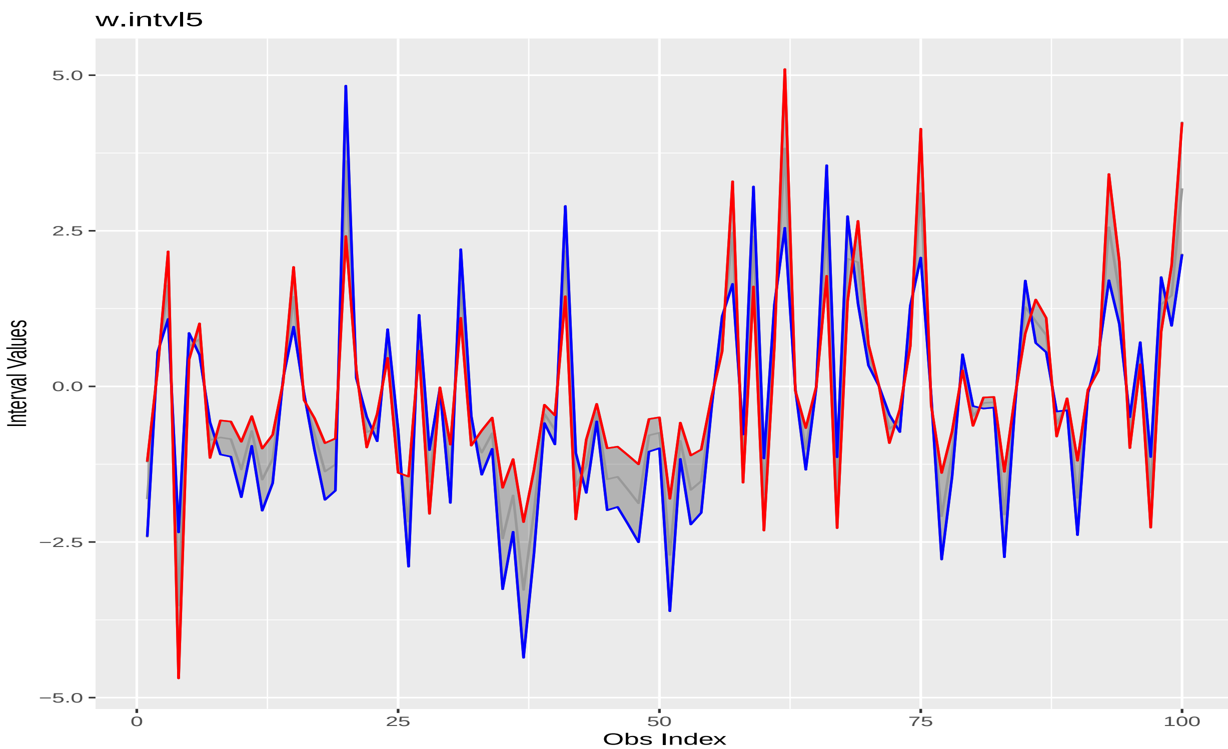

w.intvl.obs2
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

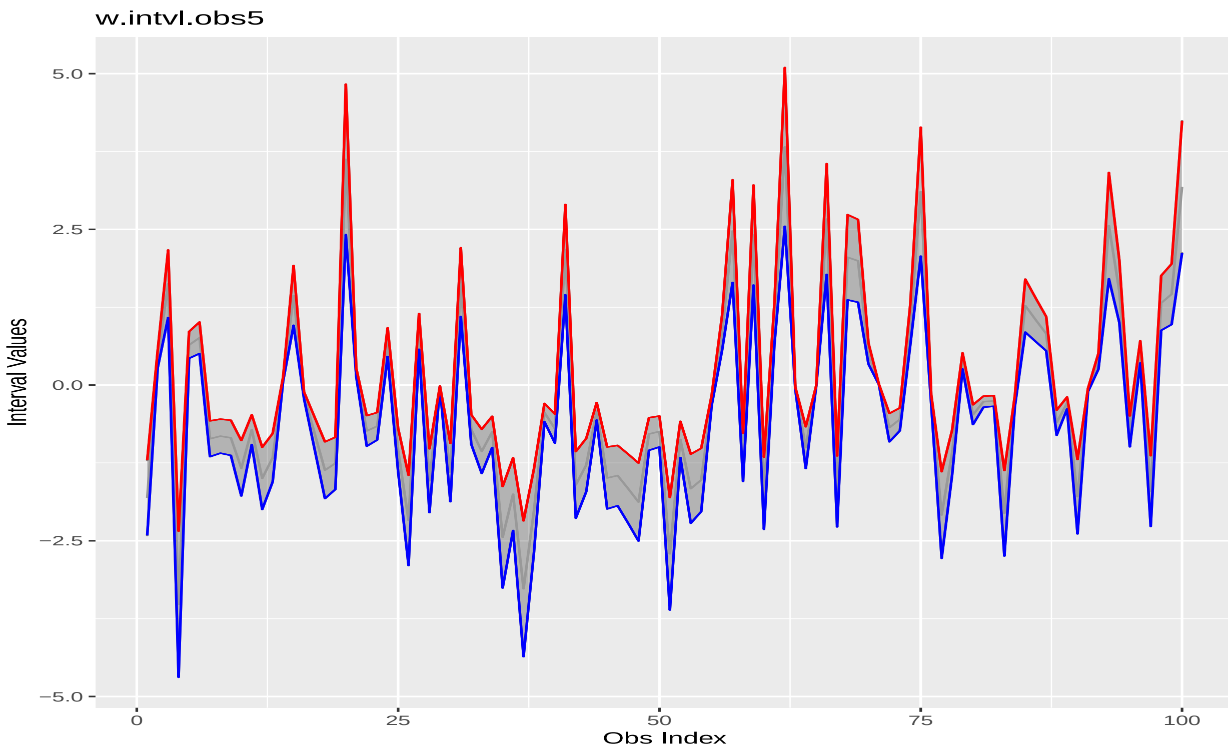

w.intvl3
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index


w.intvl.obs3
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

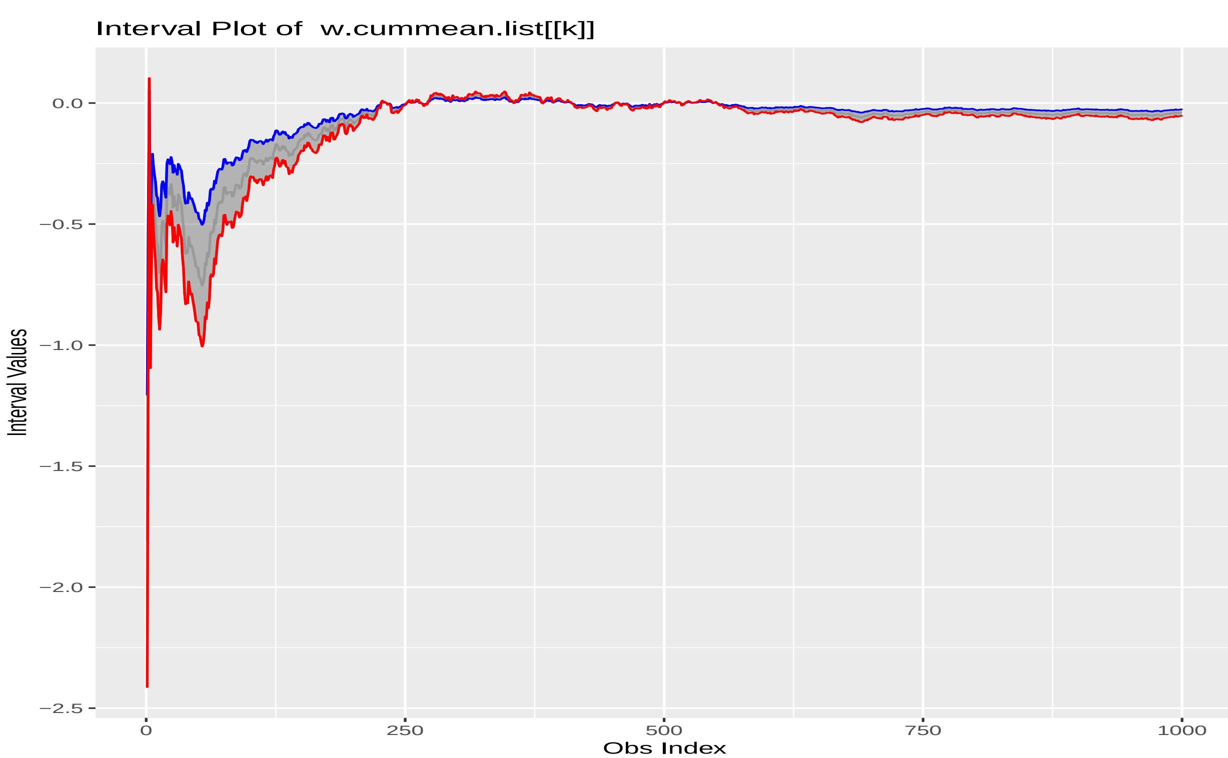

w.intvl4
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

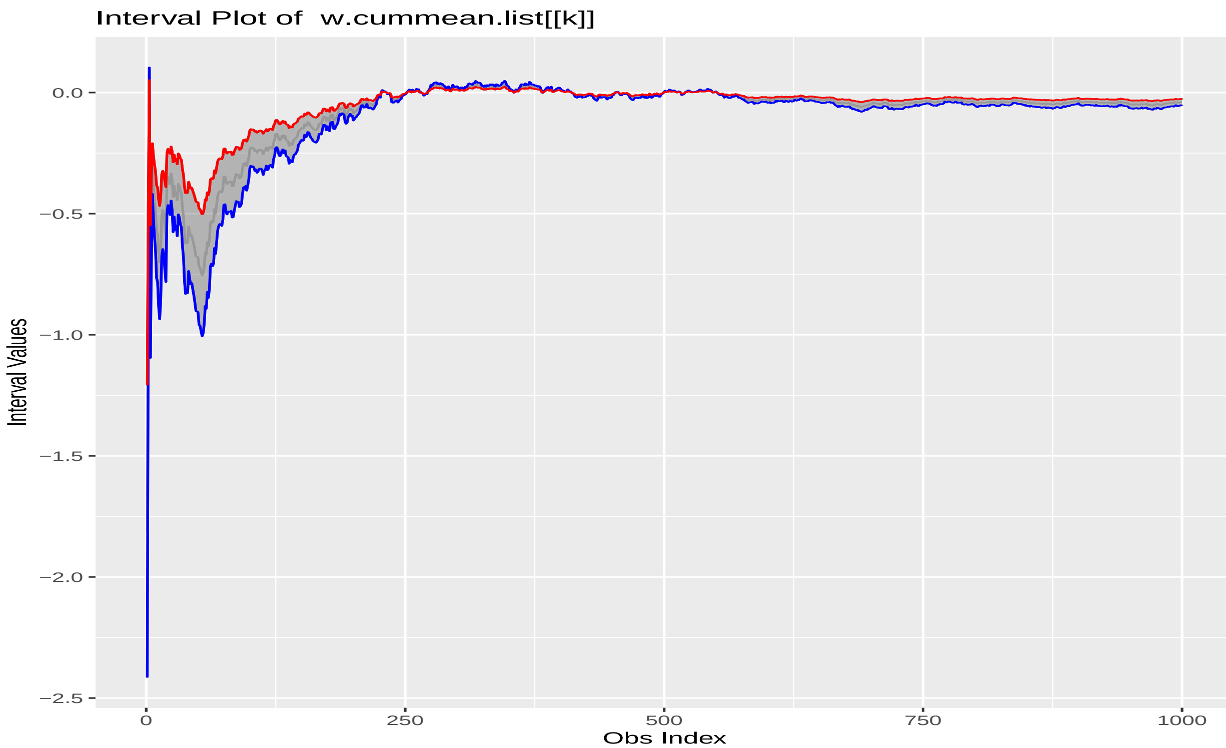

w.intvl.obs4
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

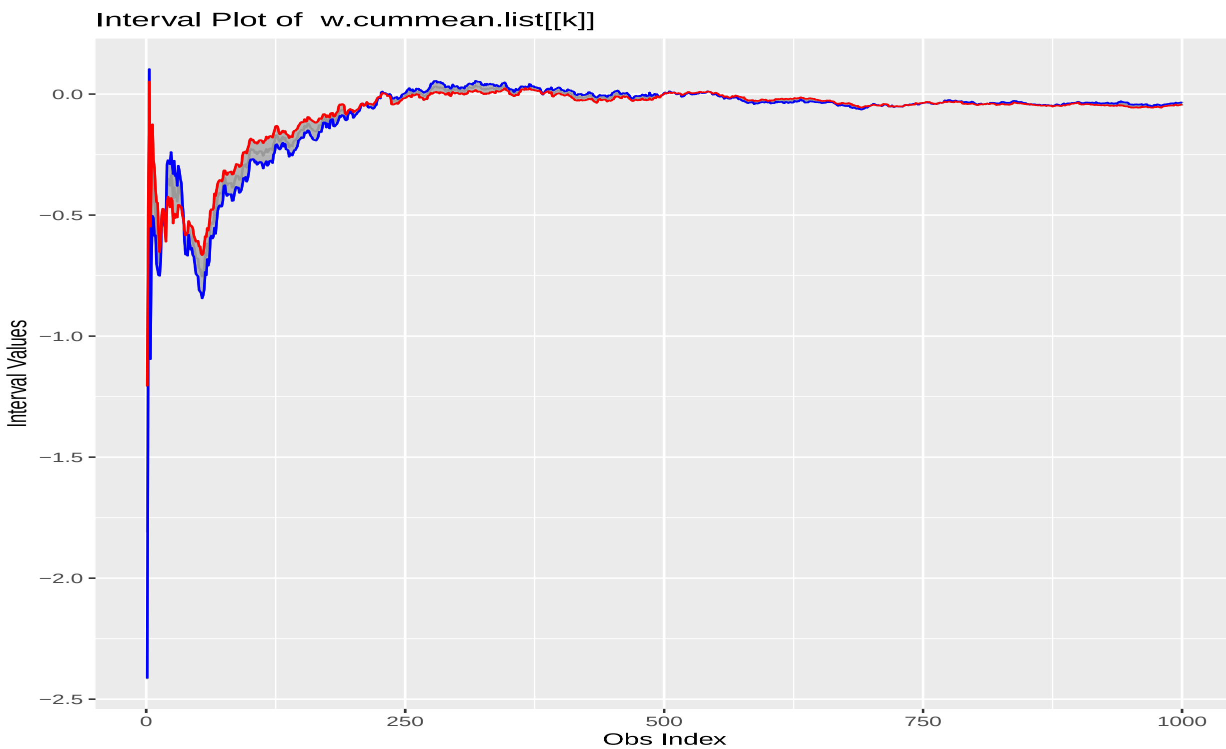

w.intvl5
Interval Values
5.0
2.5
0.0
−2.5
−5.0
0
25
50
75
100
Obs Index

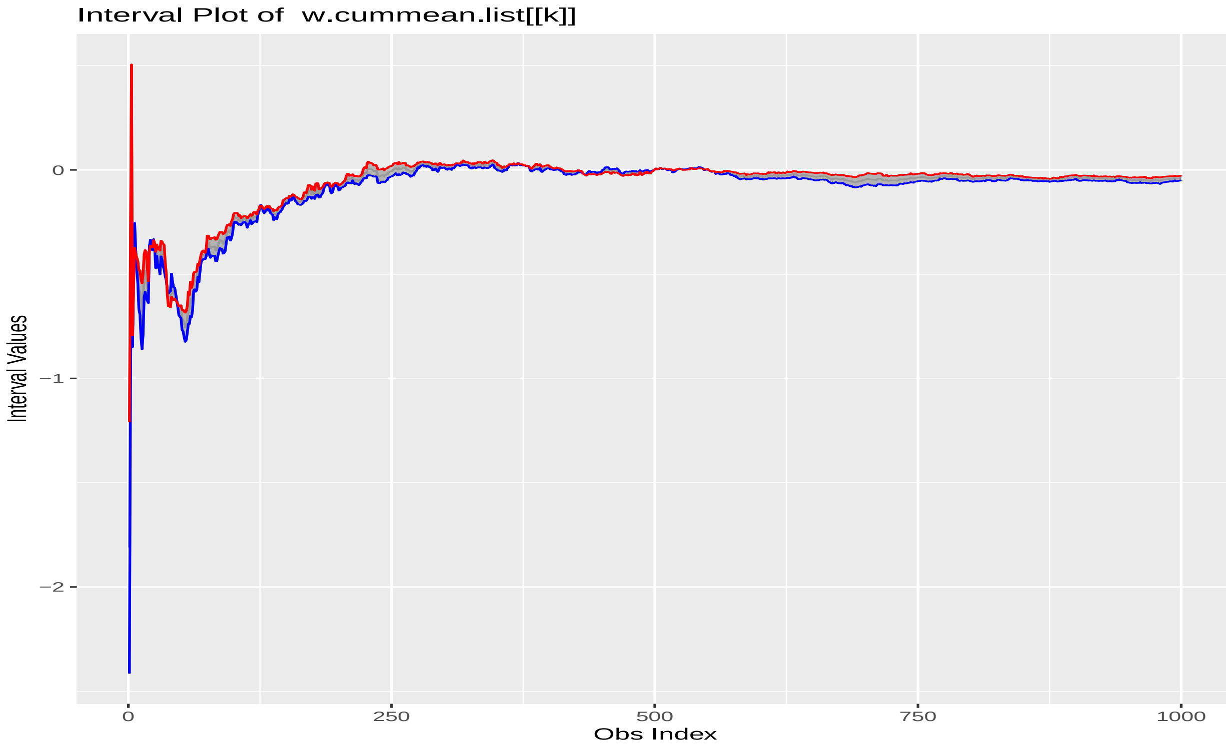



```
## [1] TRUE
## [1] TRUE
## [1] TRUE
## [1] TRUE
```

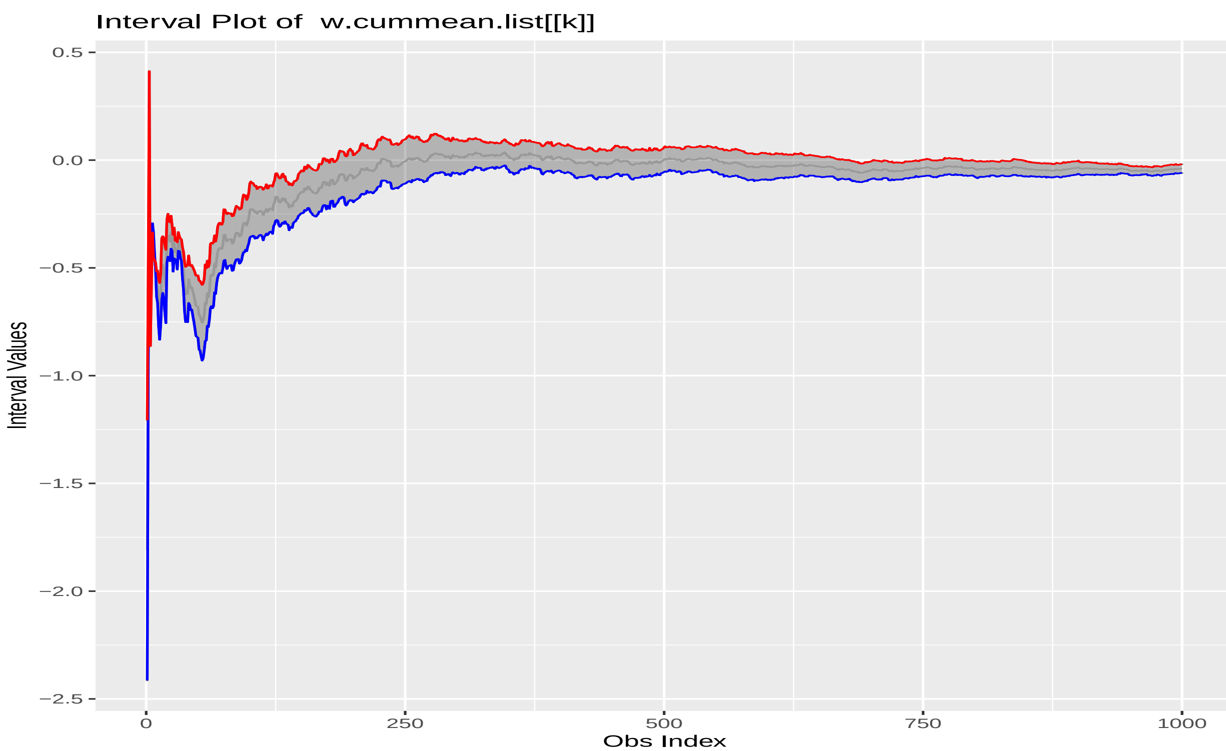



Interval Plot of w.cummean.list[[k]]
Interval Values
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0
250
500
750
1000
Obs Index

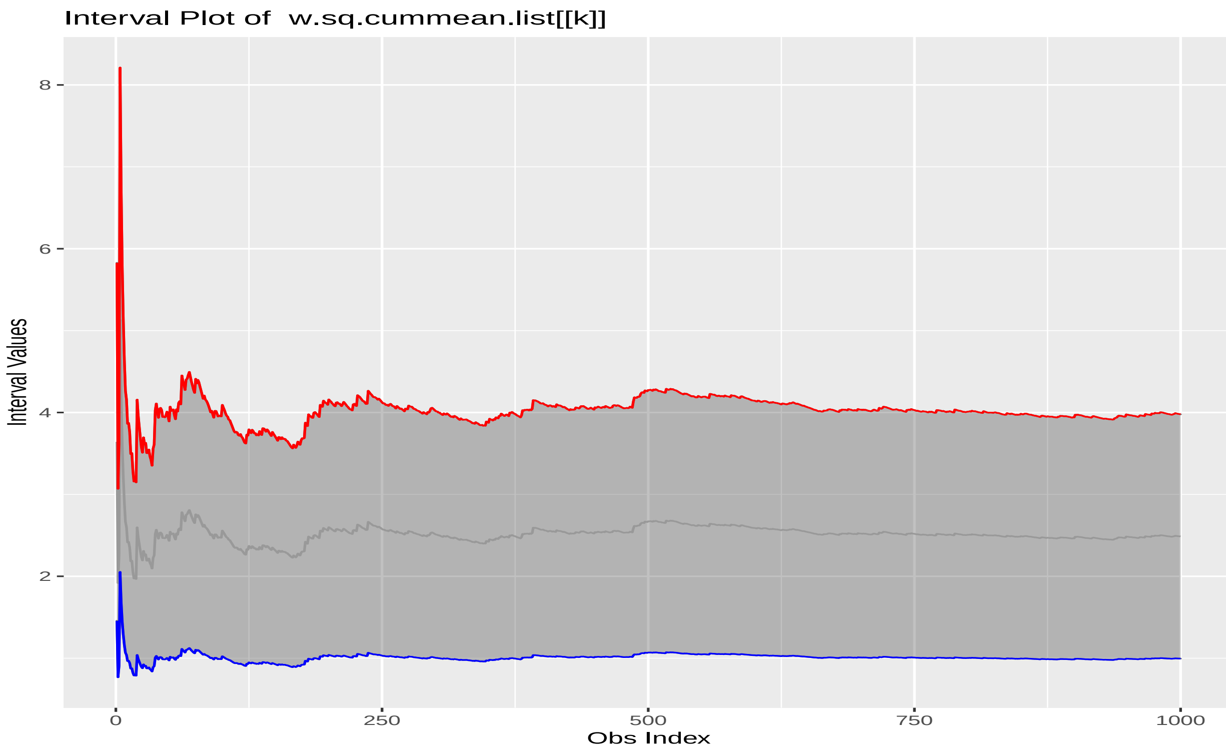

Interval Plot of w.cummean.list[[k]]
Interval Values
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0
250
500
750
1000
Obs Index

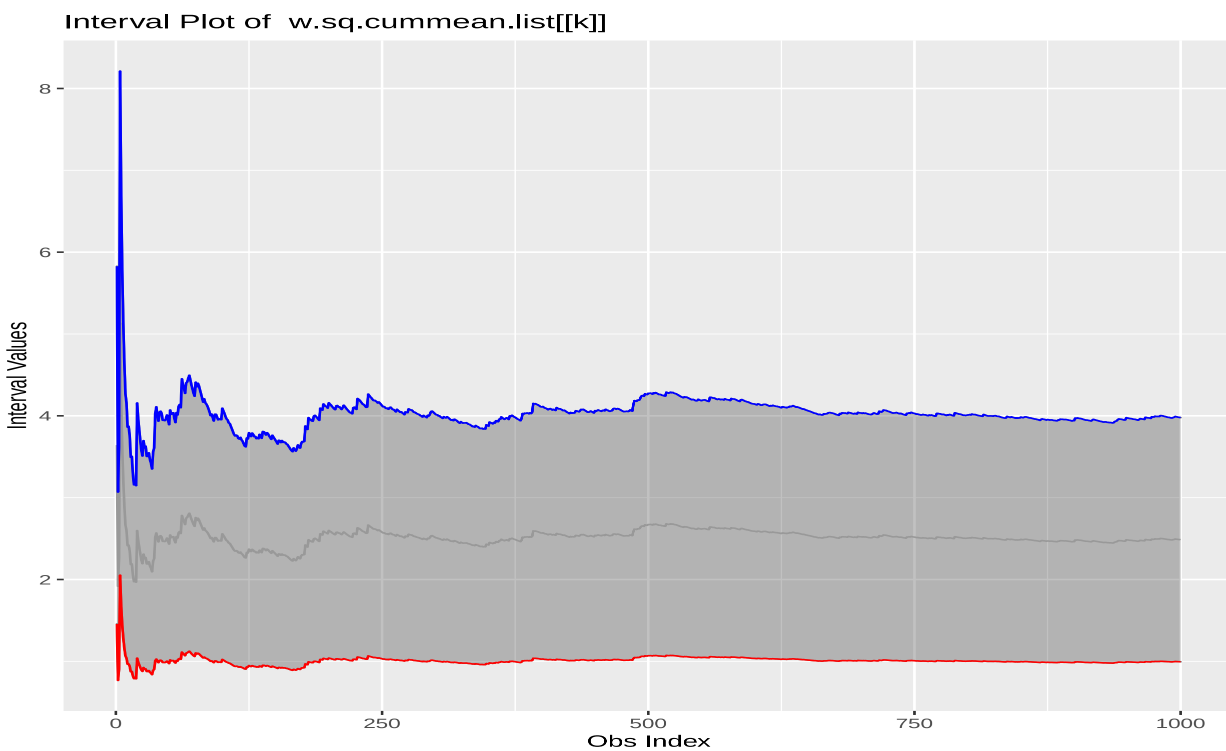


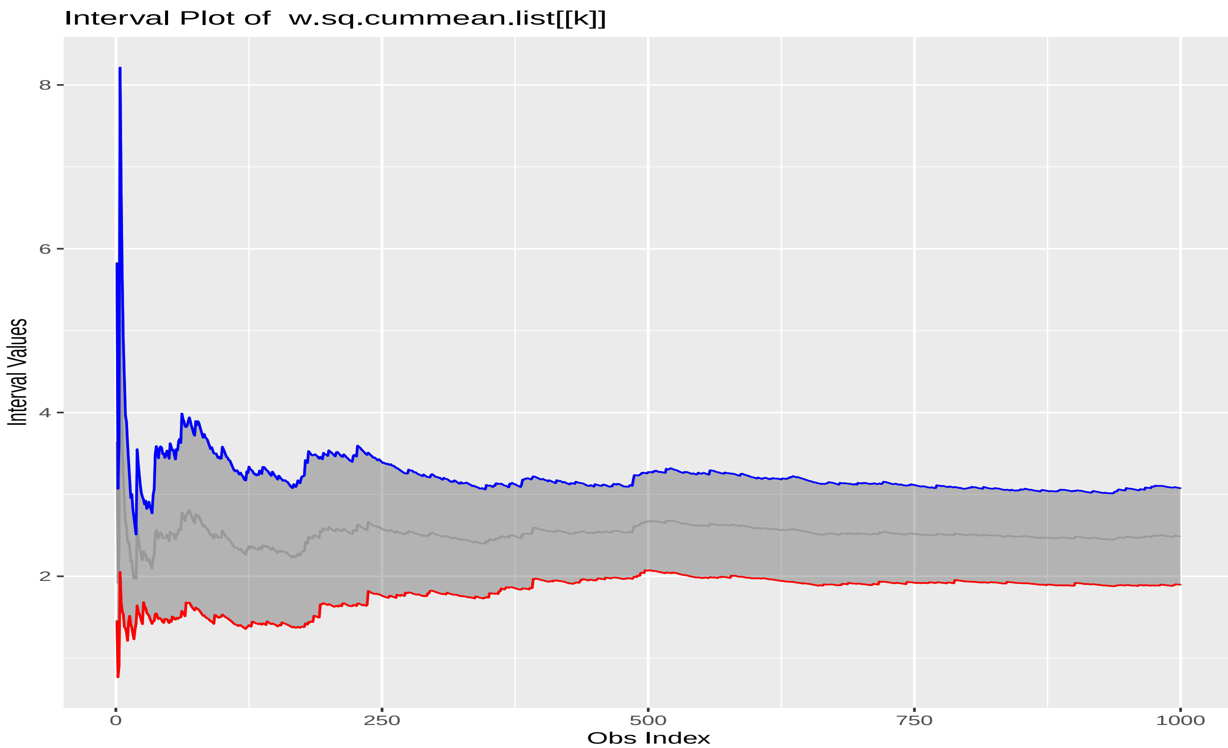
Interval Plot of w.cummean.list[[k]]
Interval Values
0
−1
−2
0
250
500
750
1000
Obs Index

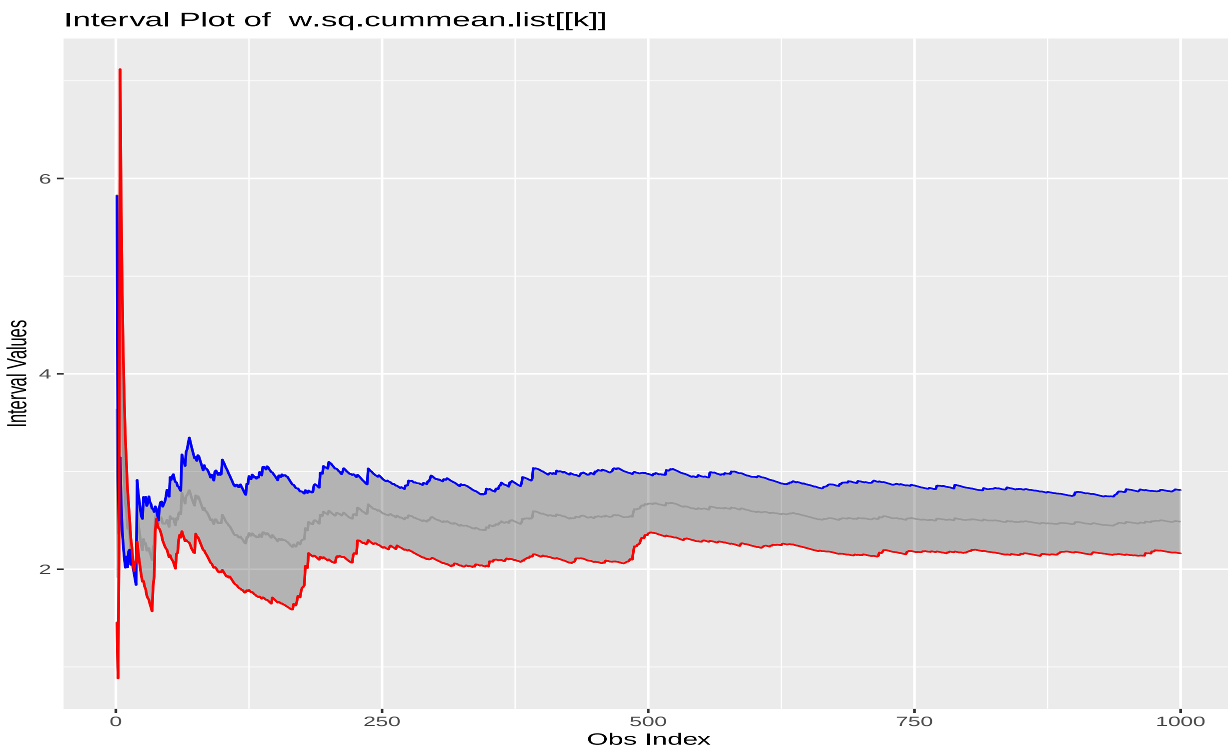


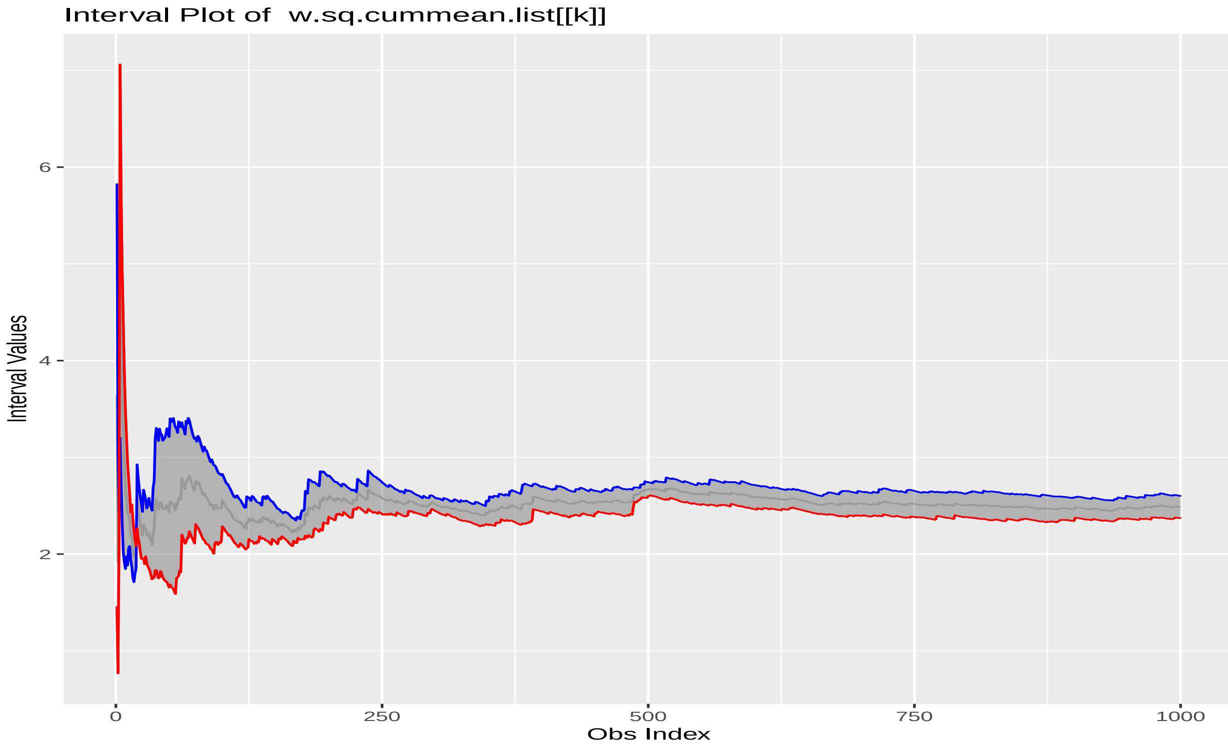
Interval Plot of w.cummean.list[[k]]
Interval Values
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
0
250
500
750
1000
Obs Index

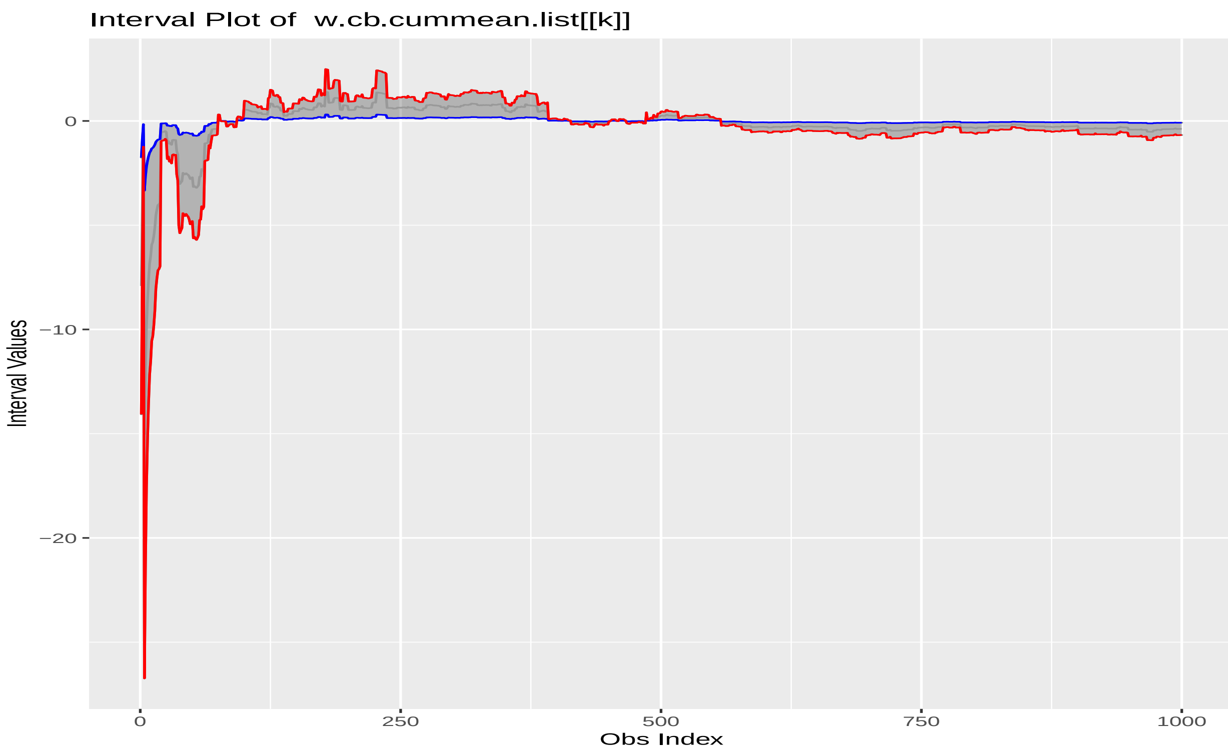

Interval Plot of w.sq.cummean.list[[k]]
Interval Values
8
6
4
2
0
250
500
750
1000
Obs Index

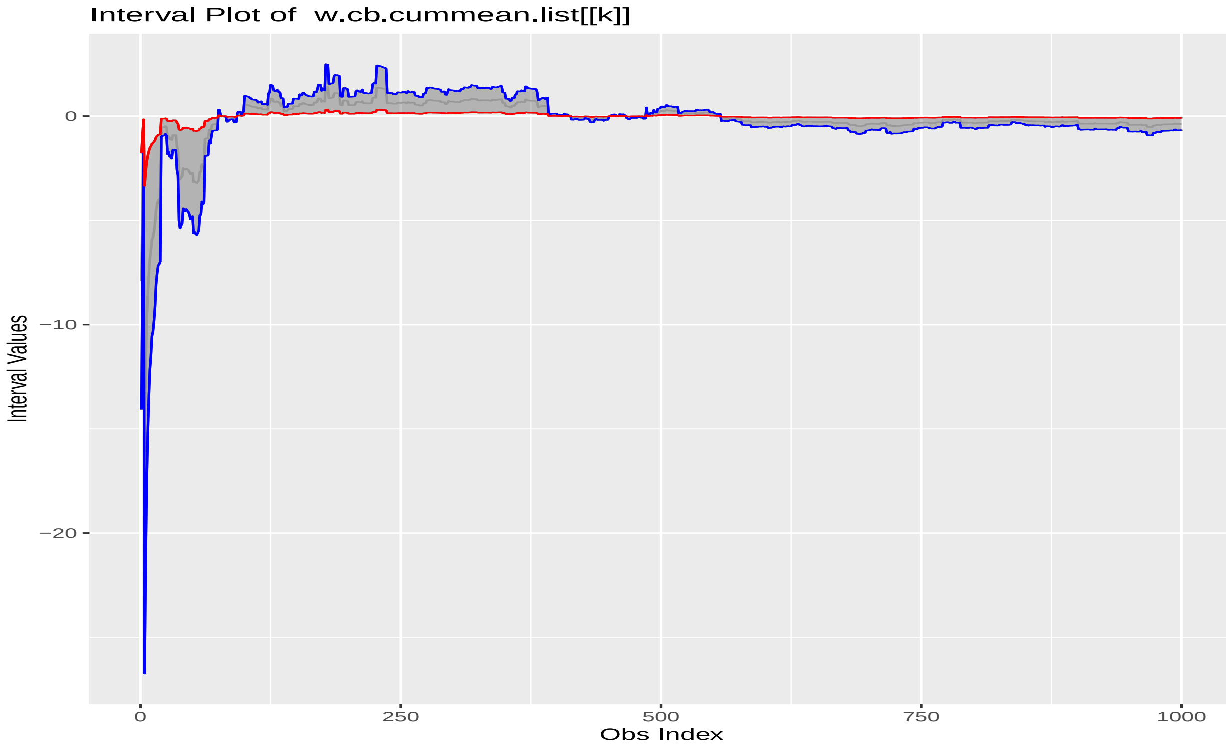

Interval Plot of w.sq.cummean.list[[k]]
Interval Values
8
6
4
2
0
250
500
750
1000
Obs Index

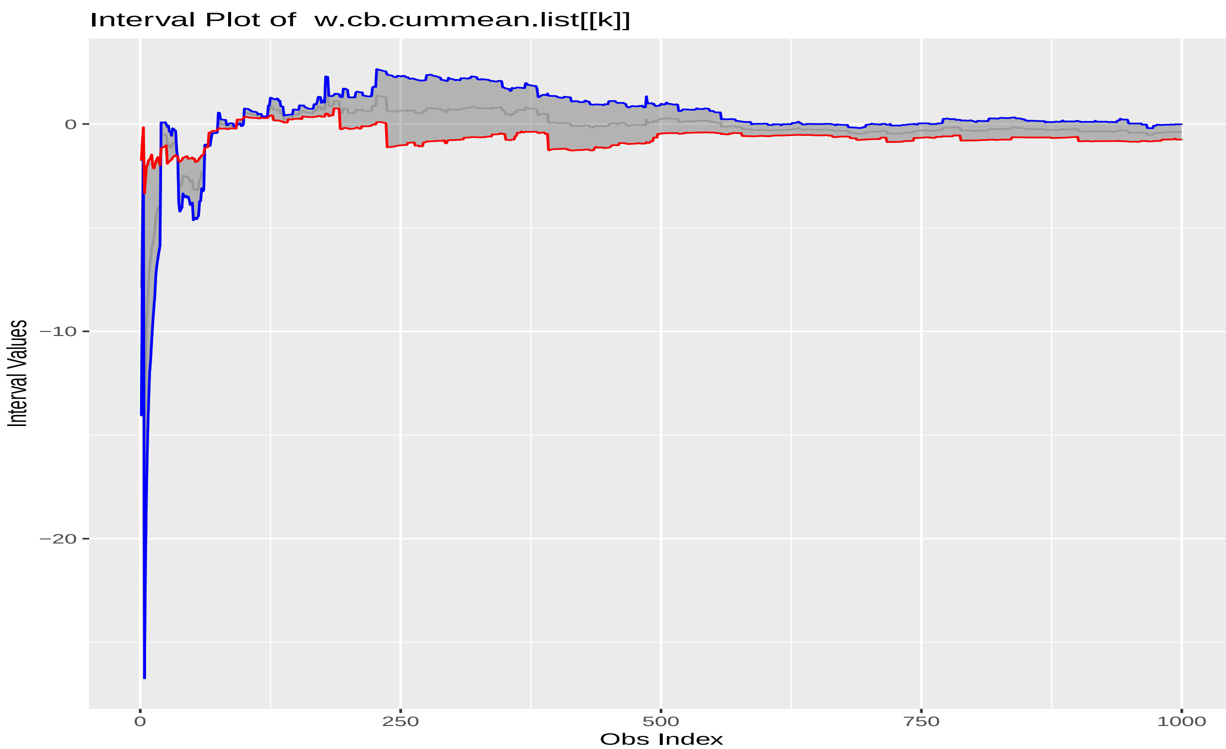

Interval Plot of w.sq.cummean.list[[k]]
Interval Values
Obs Index
0
250
500
750
1000
2
4
6
8

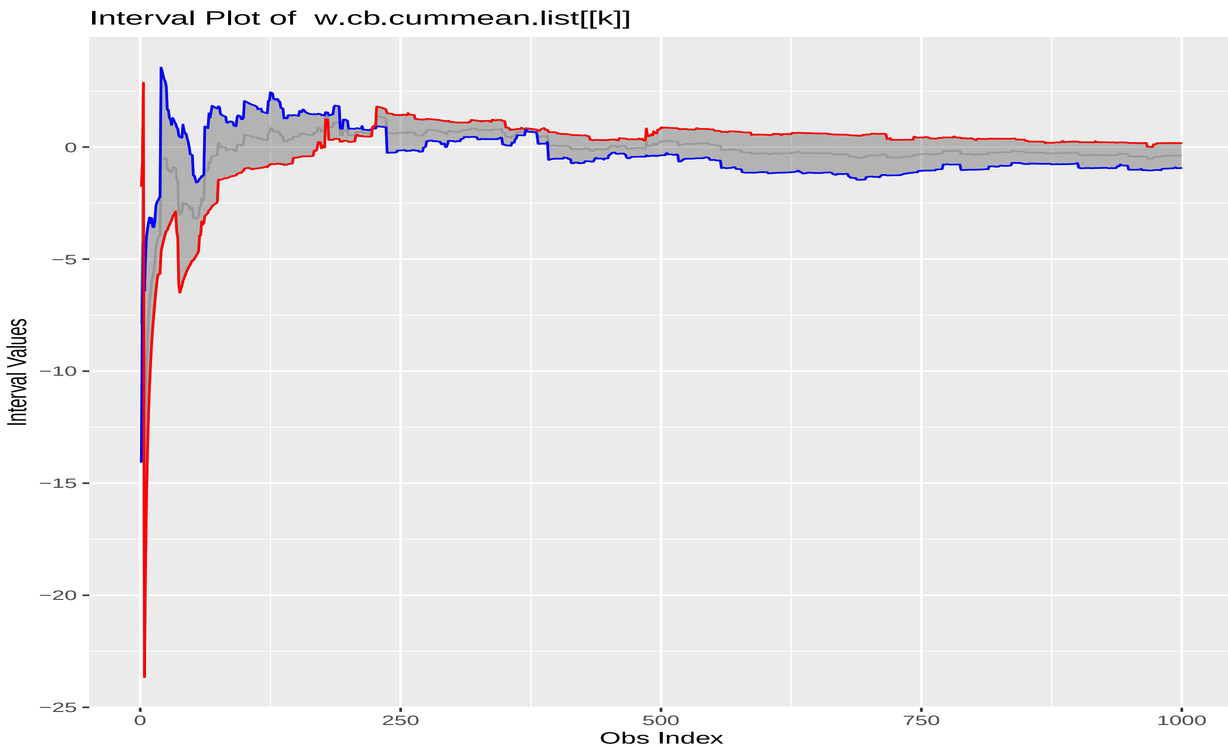

Interval Plot of w.sq.cummean.list[[k]]
Interval Values
Obs Index
0
250
500
750
1000
2
4
6

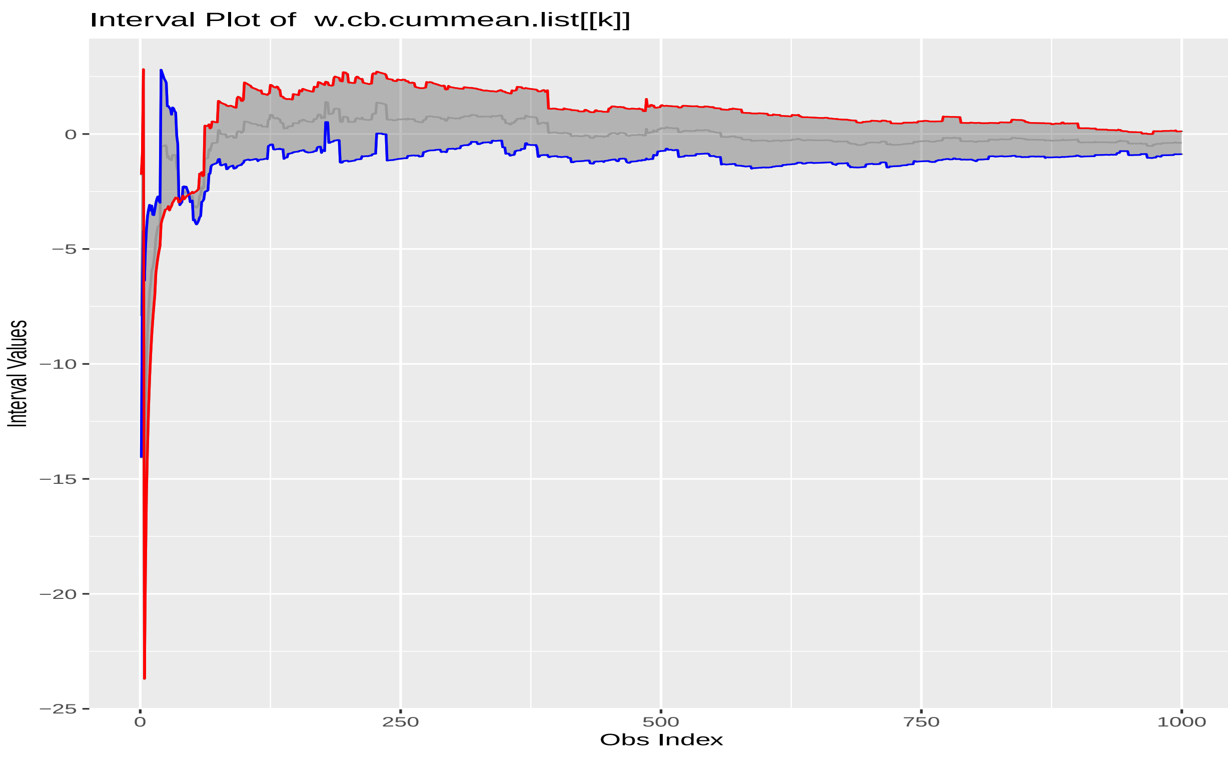

Interval Plot of w.sq.cummean.list[[k]]
Interval Values
Obs Index
0
250
500
750
1000
2
4
6


Interval Plot of w.cb.cummean.list[[k]]
Interval Values
Obs Index
0
250
500
750
1000
0
−10
−20


Interval Plot of w.cb.cummean.list[[k]]
Interval Values
0
−10
−20
0
250
500
750
1000
Obs Index


Interval Plot of w.cb.cummean.list[[k]]
Interval Values
0
−10
−20
0
250
500
750
1000
Obs Index





Notes: For a directed interval $[a, b]$ with $ab \geq 0$, the sign of the center can show the sign of the interval. The sign of the range can show the direction of the interval. In practice, we can usually observe the sign of the center but not the sign of the range: the sign of the range has not be recorded but only its magnitude. This is one of the reasons we want to point out the ambiguity in the interval direction.

4.2.3 Use G-expectaiton to deal with this Ambiguity

In (P2), suppose $\mu = 0$, and consider the normalized sum of the true interval process $\boldsymbol{X}_t$ (with unknown interval direction),

$$\begin{aligned}\boldsymbol{Y}_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{X}_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[(\underline{\sigma}s_t + \overline{\sigma}(1-s_t)), (\overline{\sigma}s_t + \underline{\sigma}(1-s_t))]\epsilon_t \\ &= [\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tl}\epsilon_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tr}\epsilon_t].\end{aligned}$$

We are interested in

$$\begin{aligned}\boldsymbol{\mathcal{E}}[\mathbf{Y}_n\varphi] &= \boldsymbol{\mathcal{E}}[\varphi(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tl}\epsilon_t), \varphi(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tr}\epsilon_t)] \\ &= [\mathbf{E}[\varphi(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tl}\epsilon_t)], \mathbf{E}[\varphi(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tr}\epsilon_t)]].\end{aligned}$$

In many cases, these two quantities are related,

$$\begin{aligned}\boldsymbol{\mathcal{E}}[\varphi(\boldsymbol{Y}_n)] &= \boldsymbol{\mathcal{E}}[\varphi\boldsymbol{Y}_n] \\ &= \boldsymbol{\mathcal{E}}[1(\boldsymbol{Y}_n\varphi)] \\ &= \boldsymbol{\mathcal{E}}[(\boldsymbol{Y}_n\varphi) \sqcup (\overline{\boldsymbol{Y}_n\varphi})].\end{aligned}$$

The key problem here is, the interval expectation we are interested in does depend on the dynamic of s_t which we cannot learn from the observed data $\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t$. However, we are able to learn $(\underline{\sigma}, \overline{\sigma})$ from the data $\tilde{\boldsymbol{X}}_t$. Before we have more information on the true underlying interval process $\boldsymbol{X}_t$, we can consider all possible reasonable dynamic for s_t and create an convex hull as the envelope to have a control on the best and worst case scenario for a given φ. Recall the notations: for two directed intervals $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{I}(\mathbb{R})$

$$\begin{aligned}\sup\{\boldsymbol{x}, \boldsymbol{y}\} &\coloneqq \boldsymbol{x} \sqcup \boldsymbol{y} \\ &= [\inf\{x_l, y_l\}, \sup\{x_r, y_r\}].\end{aligned}$$

Consider all possible dynamic for s_t: (double check the terminology) let $\mathcal{S}$ represent the set of all predictable processes valuing in $\{0, 1\}$.

$$\mathcal{S} = \{s_t : \Omega \to \{0, 1\}, \mathcal{F}_{t-1}\text{-measurable}, t = 0, 1, 2, \ldots, n\}$$

Then

$$\begin{aligned}\hat{\boldsymbol{\mathcal{E}}}_{\mathcal{S}}[\boldsymbol{Y}_n\varphi] &\coloneqq \sup_{s\in\mathcal{S}}\boldsymbol{\mathcal{E}}[\boldsymbol{Y}_n^s\varphi] = \sqcup_{s\in\mathcal{S}}\boldsymbol{\mathcal{E}}[\boldsymbol{Y}_n^s\varphi] \\ &= [\inf_{s\in\mathcal{S}}\mathbf{E}[\varphi(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tl}^s\epsilon_t)], \sup_{s\in\mathcal{S}}\mathbf{E}[\varphi(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_{tl}^s\epsilon_t)]] \\ &\to [-\hat{\mathbf{E}}[-\varphi(W)], \hat{\mathbf{E}}[\varphi(W)]] = [-\mathcal{N}_G[-\varphi], \mathcal{N}_G[\varphi]].\end{aligned}$$

where $W \sim \mathcal{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$.

This part will be especially useful in the study of interval-valued log return dataset. For instance, similar to the spirit in G-VaR, if we consider $\varphi(x) = \mathbb{1}_{\{X \le a\}}$, then the results above can be rewritten as the statement that

$$
\begin{aligned}
&[\inf_{s\in\mathcal{S}} \mathbf{P}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \sigma_{tl}^{s}\epsilon_t \le a), \sup_{s\in\mathcal{S}} \mathbf{P}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \sigma_{tl}^{s}\epsilon_t \le a)] \\
&\qquad\qquad \to [\mathbf{v}(W \le a), \mathbf{V}(W \le a)] \\
&\qquad\qquad = [\underline{F}_{\mathcal{N}}(a), \overline{F}_{\mathcal{N}}(a)].
\end{aligned}
$$

4.3 Mean Uncertainty and Variance Uncertainty

The first candidate model is, with $0 < \underline{\sigma} < \overline{\sigma}$, and $\underline{\mu} < \overline{\mu}$,

$$
\boldsymbol{X}_t = \boldsymbol{X}_t^{(1)} = [\underline{\mu}, \overline{\mu}] + [\underline{\sigma}, \overline{\sigma}]\epsilon_t,
$$

as the underlying process and $\tilde{X}_t = 1\boldsymbol{X}_t$. Suppose $\epsilon_t \sim N(0,1)$, then we have the two-end representation of $\boldsymbol{X}_t$,

$$
\begin{cases}
\boldsymbol{X}_{tl} = \underline{\mu} + \underline{\sigma}\epsilon_t & \text{i.i.d.} \sim N(\underline{\mu}, \underline{\sigma}^2) \\
\boldsymbol{X}_{tr} = \overline{\mu} + \overline{\sigma}\epsilon_t & \text{i.i.d.} \sim N(\overline{\mu}, \overline{\sigma}^2)
\end{cases}.
$$

The center-range representation of $\boldsymbol{X}_t$ are,

$$
\begin{cases}
\mathbf{C}(\boldsymbol{X}_t) = \frac{\underline{\mu}+\overline{\mu}}{2} + \frac{\underline{\sigma}+\overline{\sigma}}{2}\epsilon_t & \text{i.i.d.} \sim N(\frac{\underline{\mu}+\overline{\mu}}{2}, \frac{(\underline{\sigma}+\overline{\sigma})^2}{4}) \\
\mathbf{R}(\boldsymbol{X}_t) = \overline{\mu} - \underline{\mu} + (\overline{\sigma} - \underline{\sigma})\epsilon_t & \text{i.i.d.} \sim N(\overline{\mu} - \underline{\mu}, (\overline{\sigma} - \underline{\sigma})^2)
\end{cases}.
$$

Accordingly, we can write the two-end representation of $\tilde{\boldsymbol{X}}_t$,

$$
\tilde{\boldsymbol{X}}_{tl} = (\underline{\mu} + \underline{\sigma}\epsilon_t) \wedge (\overline{\mu} + \overline{\sigma}\epsilon_t),
$$

and similarly,

Consider the center-range representation

$$
\mathbf{C}(\tilde{\boldsymbol{X}}_t) = \mathbf{C}(\boldsymbol{X}_t) = \frac{\underline{\mu}+\overline{\mu}}{2} + \frac{\underline{\sigma}+\overline{\sigma}}{2}\epsilon_t \sim N(\frac{\underline{\mu}+\overline{\mu}}{2}, \frac{(\underline{\sigma}+\overline{\sigma})^2}{4}),
$$

and

$$
\mathbf{R}(\tilde{\boldsymbol{X}}_t) = |\mathbf{R}(\boldsymbol{X}_t)| = |(\overline{\mu} - \underline{\mu}) + (\overline{\sigma} - \underline{\sigma})\epsilon_t|.
$$

$$
\begin{aligned}
\mathbf{r}(\tilde{\boldsymbol{X}}_t) = \frac{1}{2}\mathbf{R}(\tilde{\boldsymbol{X}}_t) &= |\frac{\overline{\mu} - \underline{\mu}}{2} + \frac{\overline{\sigma} - \underline{\sigma}}{2}\epsilon_t| \\
&\sim N_f(\frac{\overline{\mu} - \underline{\mu}}{2}, \frac{(\overline{\sigma} - \underline{\sigma})^2}{4}).
\end{aligned}
$$

Folded normal distribution: $Y \sim N_f(\mu, \sigma^2)$ if $Y = |X|$ with $X \sim N(\mu, \sigma^2)$, note that we assume $\mu > 0$ here to ensure the identifiability of the parameter (otherwise, (μ, σ^2) and $(-\mu, \sigma^2)$ will result in the same density.)

Let

$$\mu_C := \frac{\underline{\mu} + \overline{\mu}}{2},$$
$$\sigma_C := \frac{|\underline{\sigma} + \overline{\sigma}|}{2},$$

and

$$\mu_r := \frac{|\overline{\mu} - \underline{\mu}|}{2} = \frac{\overline{\mu} - \underline{\mu}}{2},$$
$$\sigma_r := \frac{|\overline{\sigma} - \underline{\sigma}|}{2}.$$

Under the assumption $\underline{\sigma} \leq \overline{\sigma}$ and $|\underline{\sigma}| \leq |\overline{\sigma}|$, if $\overline{\sigma} < 0$, then

$$\underline{\sigma} \leq \overline{\sigma} < 0,$$

it is contradicts to $|\underline{\sigma}| \leq |\overline{\sigma}|$. Then $\overline{\sigma} \geq 0$. Hence

$$\overline{\sigma} = |\overline{\sigma}| \geq |\underline{\sigma}| \geq -\underline{\sigma},$$

namely,

$$\overline{\sigma} + \underline{\sigma} \geq 0.$$

Meanwhile, we obviously have $\overline{\sigma} - \overline{\sigma} \geq 0$. Therefore, we must have

$$\sigma_C = \frac{\underline{\sigma} + \overline{\sigma}}{2},$$

and

$$\sigma_r = \frac{\overline{\sigma} - \underline{\sigma}}{2}.$$

There is a one-to-one correspondence between $(\underline{\mu}, \overline{\mu}, \underline{\sigma}, \overline{\sigma})$ and $(\mu_C, \mu_r, \sigma_C, \sigma_r)$:

$$\begin{cases} \underline{\mu} & = \mu_C - \mu_r \\ \overline{\mu} & = \mu_C + \mu_r \\ \underline{\sigma} & = \sigma_C - \sigma_r \\ \overline{\sigma} & = \sigma_C + \sigma_r \end{cases}.$$

In matrix form,

$$\mathbf{A} = \mathbf{K}\tilde{\mathbf{A}},$$

where

$$\mathbf{K} := \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\mu} & \boldsymbol{\sigma} \end{pmatrix} := \begin{pmatrix} \underline{\mu} & \underline{\sigma} \\ \overline{\mu} & \overline{\sigma} \end{pmatrix},$$

and

$$\tilde{\mathbf{A}} = \begin{pmatrix} \tilde{\boldsymbol{\mu}} & \tilde{\boldsymbol{\sigma}} \end{pmatrix} := \begin{pmatrix} \mu_C & \sigma_C \\ \mu_r & \sigma_r \end{pmatrix}.$$

Then we can use estimation equations or directly use MLE to estimate $\tilde{\mathbf{A}}$ first, then use $\tilde{\mathbf{A}}$ to estimate $\mathbf{A}$.

4.3.1 Ambiguity in the interval direction

We have $\tilde{\boldsymbol{X}}_t = 1\boldsymbol{X}_t^{(i)}$, $i = 1, 2, 3, 4$ where

1. $\boldsymbol{X}_t^{(1)} = [\underline{\mu}, \overline{\mu}] + [\underline{\sigma}, \overline{\sigma}]\epsilon_t$, with $\underline{\mu} \leq \overline{\mu}$, $\underline{\sigma} \leq \overline{\sigma}$ and $|\underline{\sigma}| \leq |\overline{\sigma}|$,
2. $\boldsymbol{X}_t^{(2)} := \overline{\boldsymbol{X}_t^{(1)}} = [\overline{\mu}, \underline{\mu}] + [\overline{\sigma}, \underline{\sigma}]\epsilon_t$,
3. $\boldsymbol{X}_t^{(3)} = [\underline{\mu}, \overline{\mu}] + [\overline{\sigma}', \underline{\sigma}']\epsilon_t$, with $\underline{\mu} \leq \overline{\mu}$, $\underline{\sigma}' \leq \overline{\sigma}'$ and $|\underline{\sigma}'| \leq |\overline{\sigma}'|$,
4. $\boldsymbol{X}_t^{(4)} := \overline{\boldsymbol{X}_t^{(3)}} = [\overline{\mu}, \underline{\mu}] + [\underline{\sigma}', \overline{\sigma}']\epsilon_t$,

(Draft version) Consider $s_t : \Omega \to \{1, 2, 3, 4\}$ (independent from ϵ_t)

$$\boldsymbol{X}_t^s = \sum_{i=1}^{4} \boldsymbol{X}_t^{(i)} \mathbb{1}_{\{s_t = i\}}.$$

s_t could belong to

1. $\mathcal{M}$: the set of all Markov processes (s_t is independent from ϵ_t), this belongs to HMM-mixture model,
2. $\mathcal{S}_0$: the set of all $\sigma(s_{t-k}, k \geq 0)$-measurable process,
3. $\mathcal{S}$: the set of all $\sigma(s_{t-k}, k \geq 0, \epsilon_{t-l}, l \geq 1)$-measurable process.

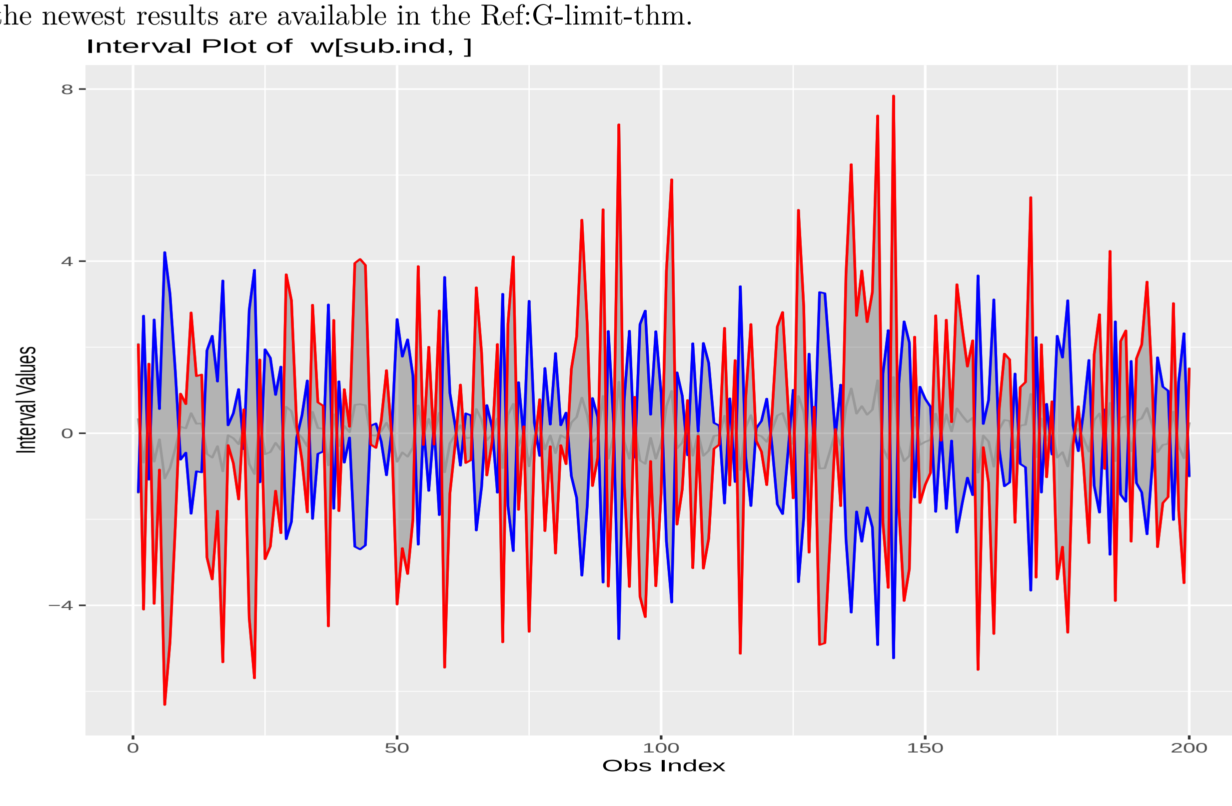
We have

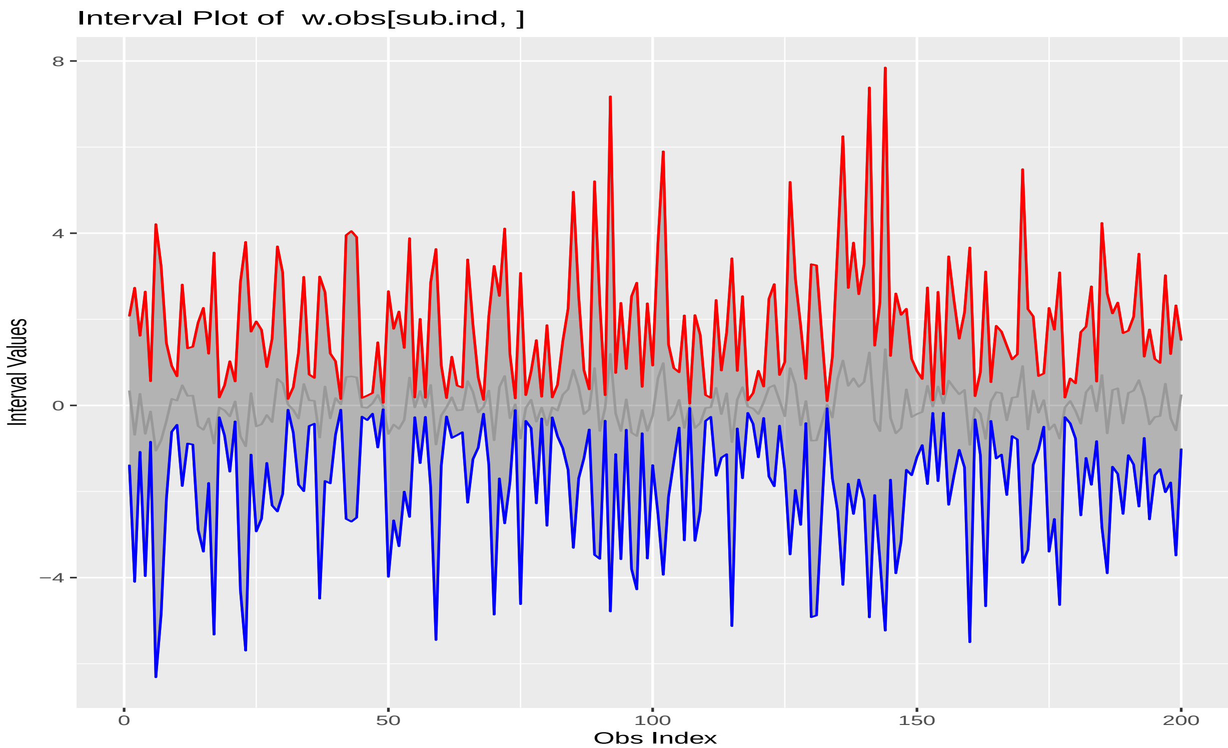
$$\mathcal{M} \subset \mathcal{S}_0 \subset \mathcal{S}.$$

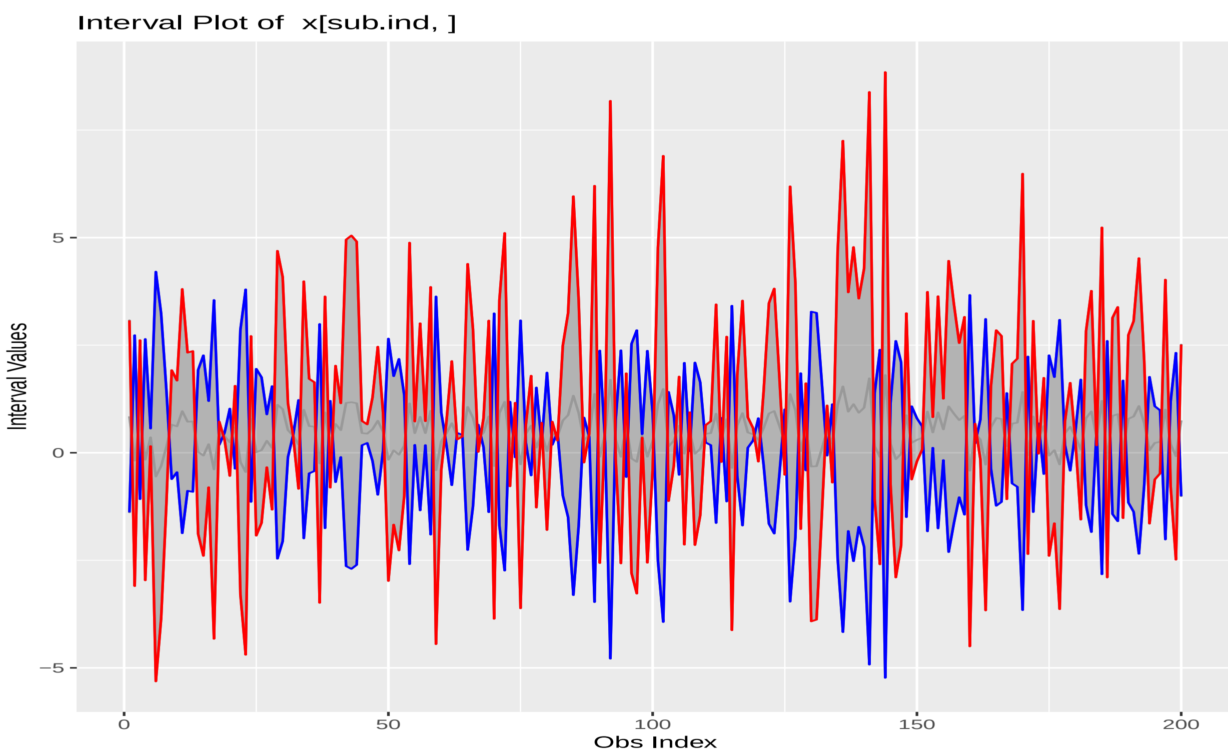
For the last one, if we let $\mathcal{F}_t := \sigma(s_{t-k}, \epsilon_{t-k}, k \geq 0)$, then we have s_t is $\mathcal{F}_t$-measurable (also note that s_t is independent from ϵ_t).

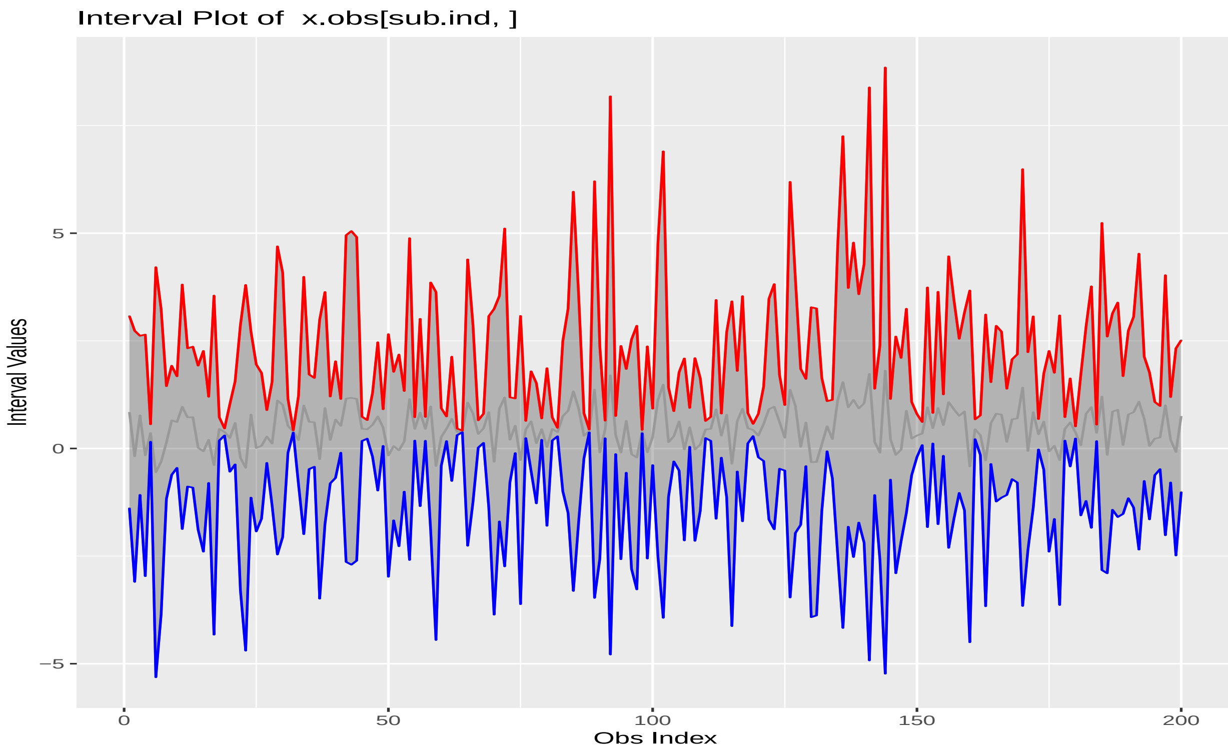
any Markov process $\subset$ $\sigma(s_{t-k}, k \geq 0)$-measurable $\subset$ $\sigma(s_{t-k}, k \geq 0, \epsilon_{t-l}, l \geq 1)$-measurable) process.

The limiting theorems to talk about ambiguity in this situation is still under discussions. Part of the newest results are available in the Ref:G-limit-thm.

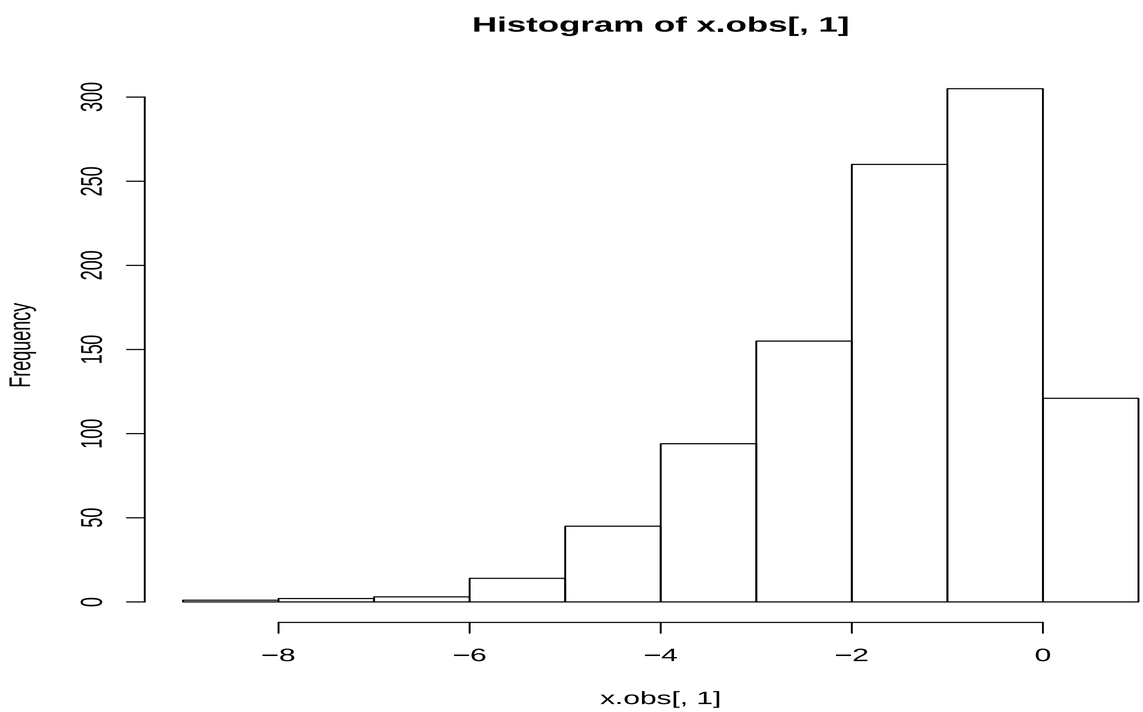



Interval Plot of w.obs[sub.ind,]
Interval Values
Obs Index
8
4
0
−4
0
50
100
150
200

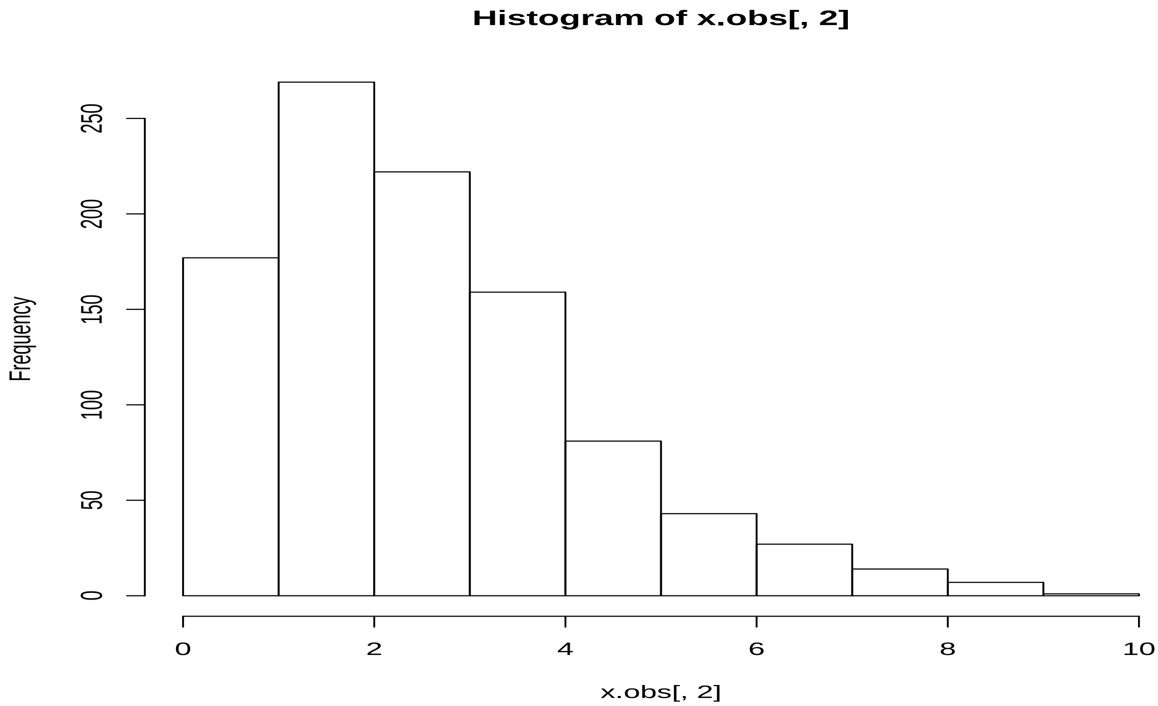

Interval Plot of x[sub.ind,]
Interval Values
Obs Index
5
0
−5
0
50
100
150
200

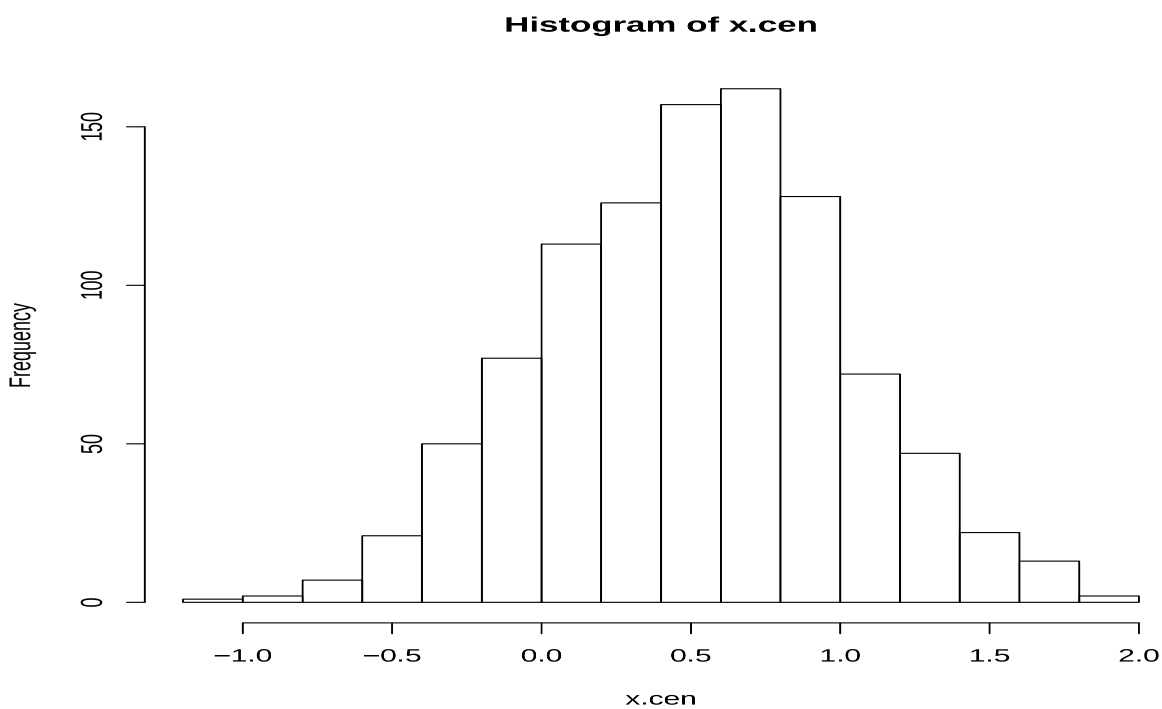


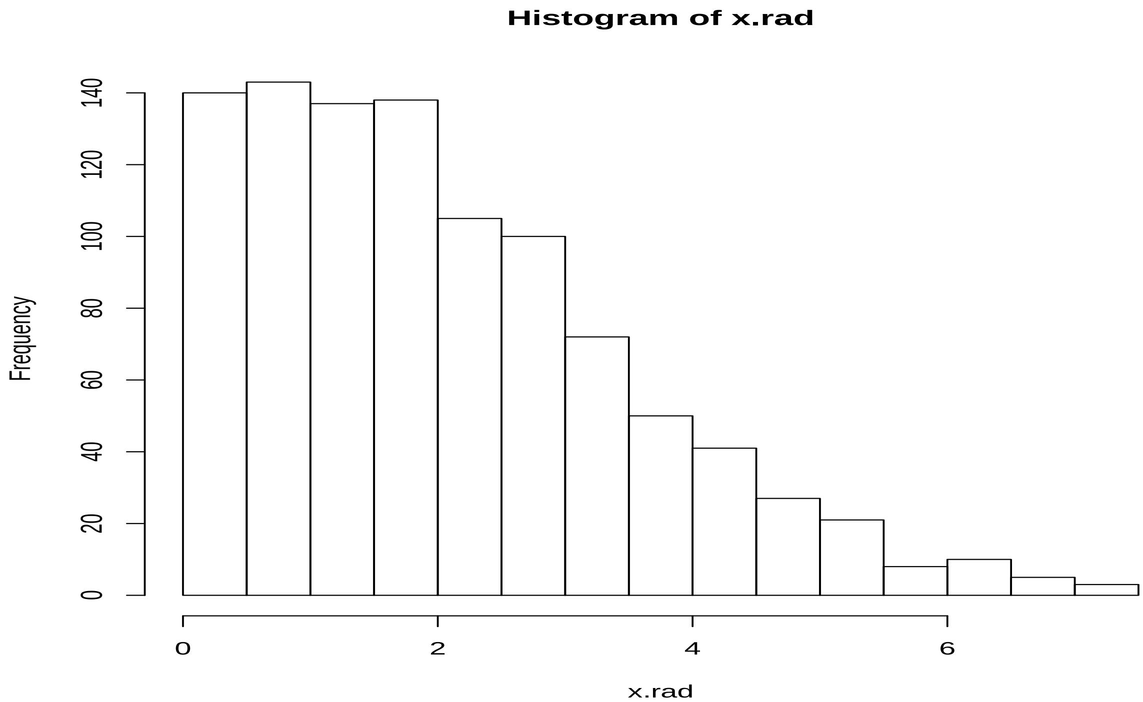
Interval Plot of x.obs[sub.ind,]
Interval Values
Obs Index
5
0
−5
0
50
100
150
200

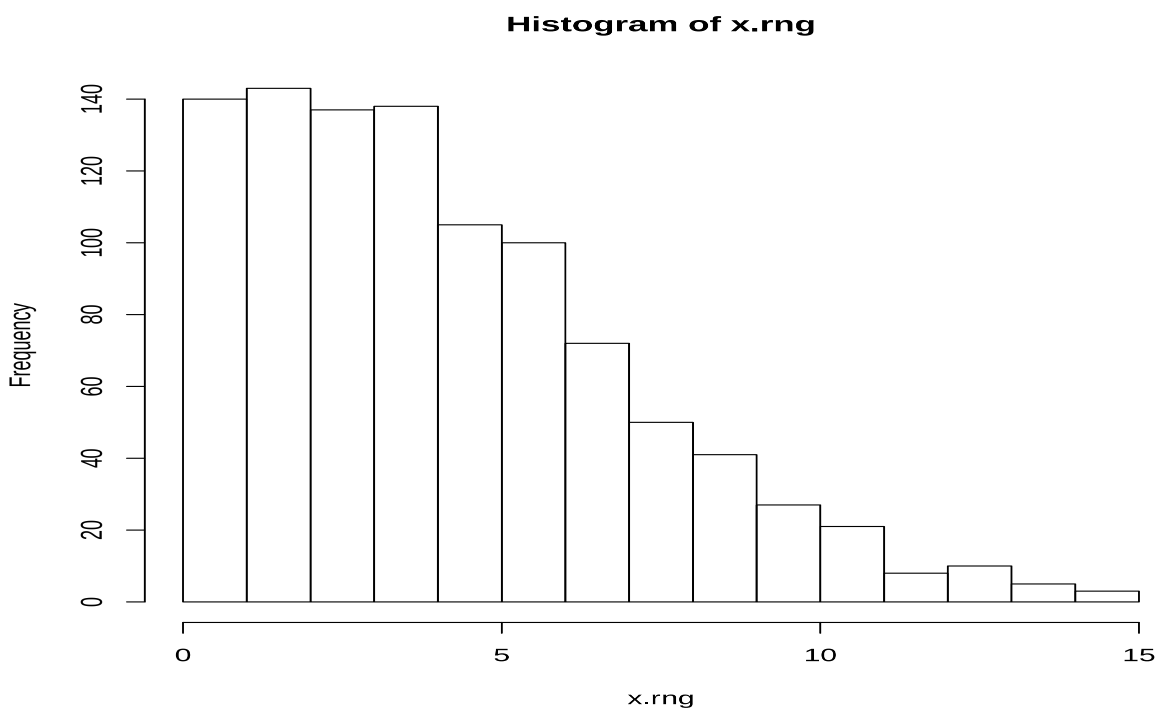


Histogram of x.obs[, 1]
Frequency
x.obs[, 1]
0
50
100
150
200
250
300
−8
−6
−4
−2
0


Histogram of x.obs[, 2]
Frequency
0
50
100
150
200
250
0
2
4
6
8
10
x.obs[, 2]


Histogram of x.cen
Frequency
0
50
100
150
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x.cen





5 Real Data Examples

5.1 The Big Picture

Consider a stochastic process X_t with $t \in \mathcal{T}$ and $\mathcal{T}$ is an arbitrary index set.

1. If $\mathcal{T} = \mathbb{N} := \{0, 1, 2, \dots\}$, X_t is a discrete-time process;
2. If $\mathcal{T} = [0, +\infty]$, X_t is a continuous-time process.

How should we understand the possible Knightian uncertainty or Ambiguity characterized by $\mathcal{Q} := \{F_\theta\}_{\theta \in \Theta}$ for X_t?

1. (point-valued data) $X_t \sim F_{\theta_t} \in \mathcal{Q}$ and θ_t may switch in Θ as t increases (in an unknown way). If user already has a model as a filter, X_t may play a role as the (varying) parameter in the model or the filtered noise part.
2. (interval-valued data) $\mathcal{X}_t = \{X_t^\theta \sim F_{\theta_t}, \theta_t \in \Theta \text{ or } X_t^\theta \sim F_\theta, \theta \in \Theta\}$, but we can only observe $1\boldsymbol{X}_t = \mathcal{X}_t^h$. (Here we also consider the ambiguity in the interval direction.)

5.2 Objectives

How should we view the interval-valued daily log return dataset?

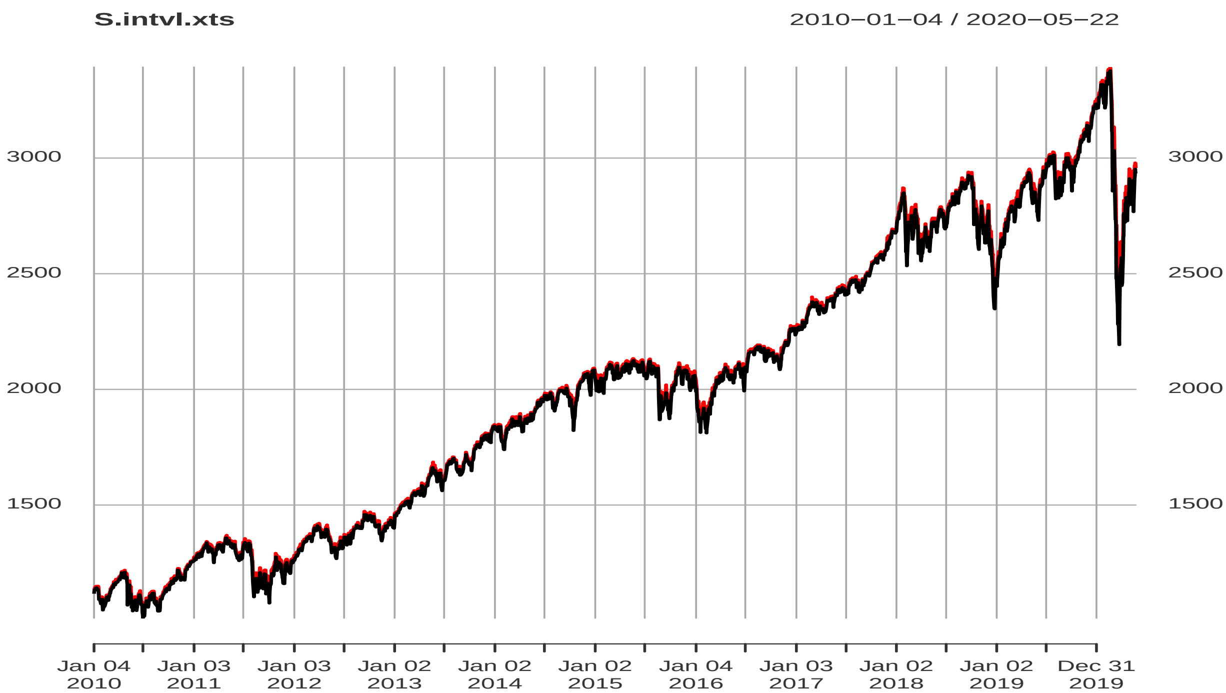
Two objectives:

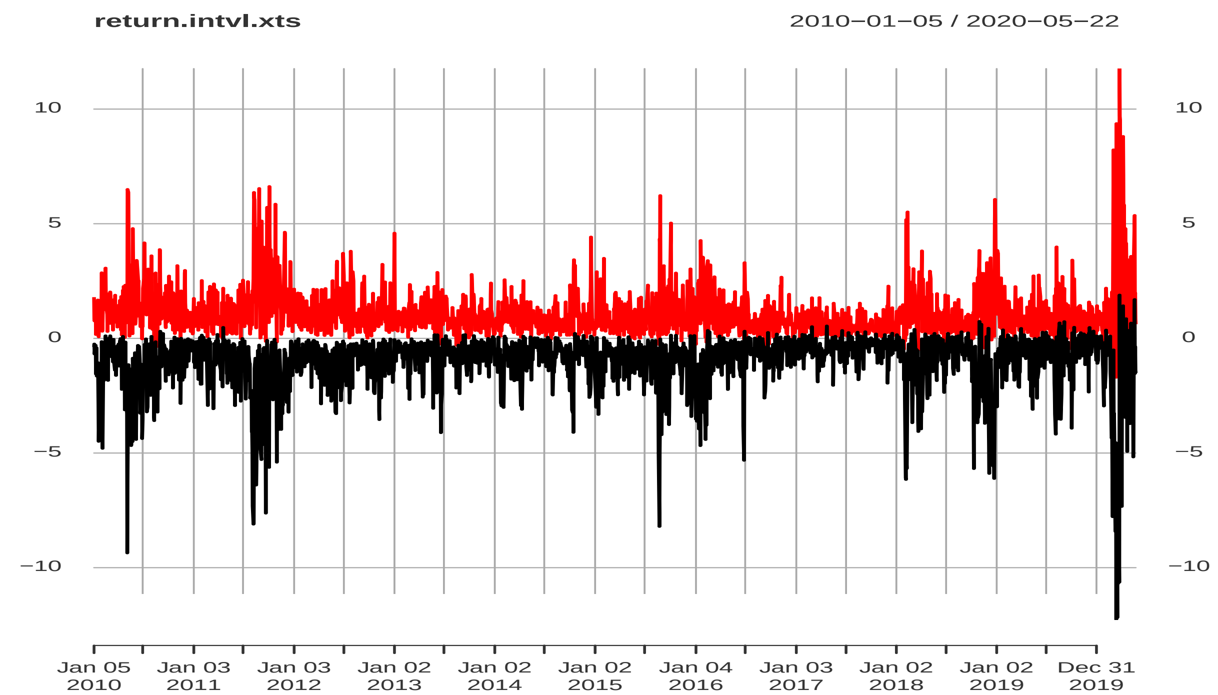
1. (O1) Find a model that can mimic the data pattern of the interval-valued log return (to do forecasting),
2. (O2) We want to study the expected daily return of an agent with some trading strategy, which is related to the ambiguity in the interval direction. (Our current focus)

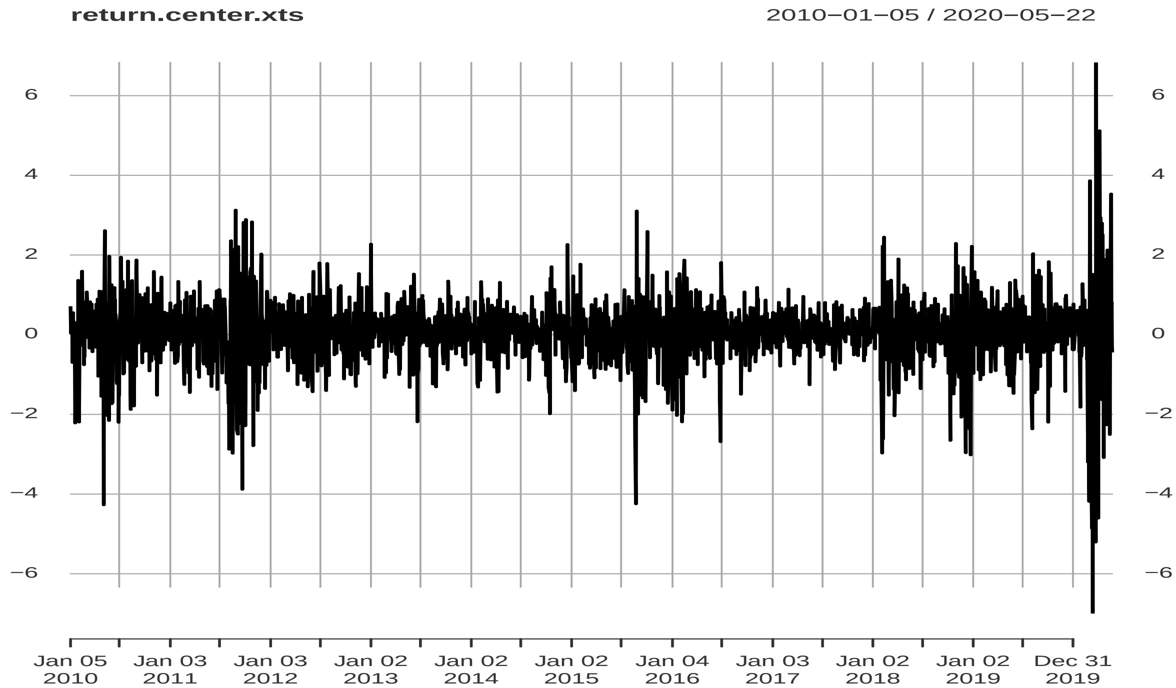
For (O1), we can try following methods.

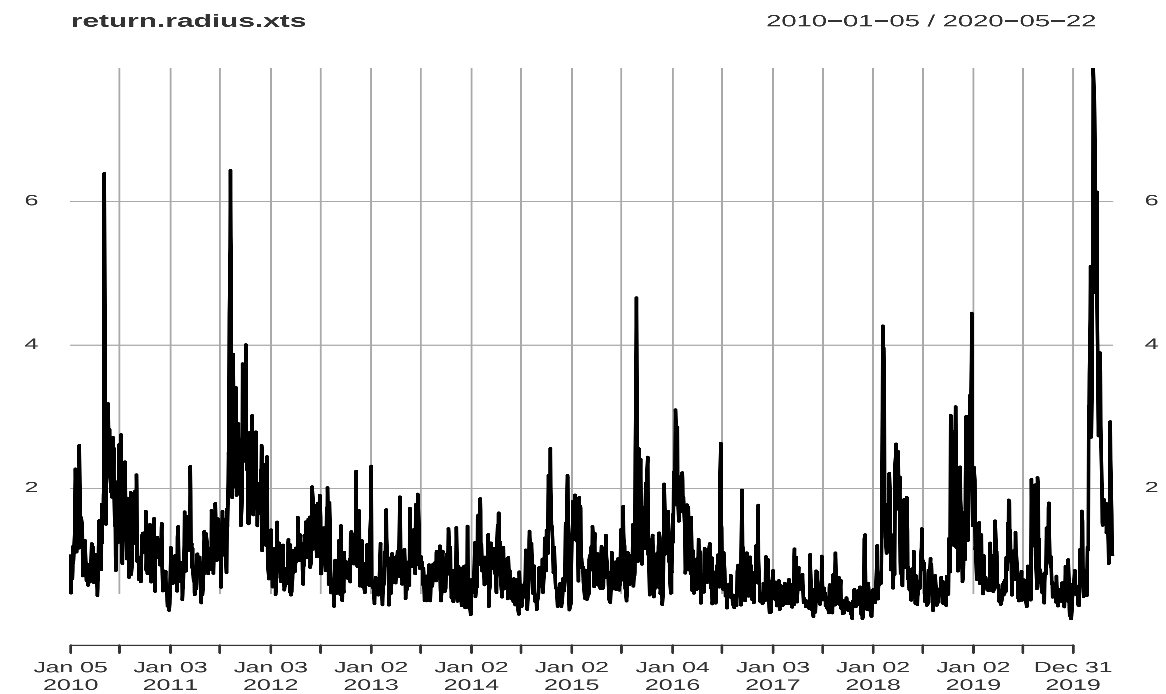
5.3 Interval-valued log return

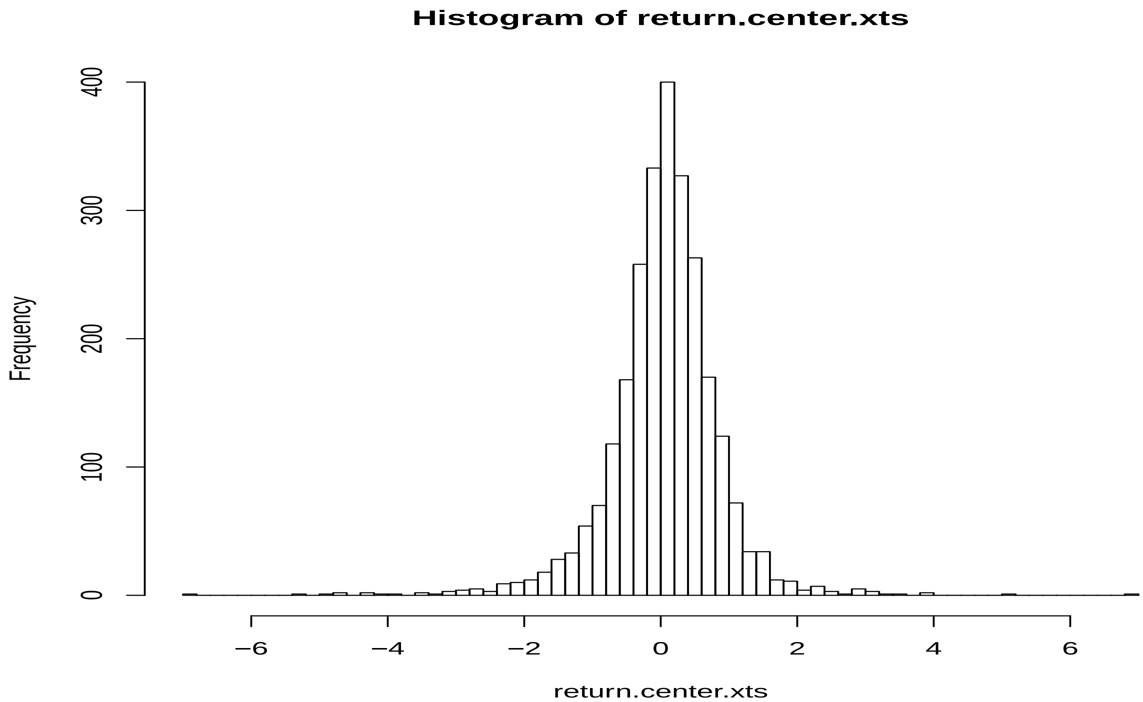
```
## Loading required package: xts
## Loading required package: TTR
## Version 0.4-0 included new data defaults. See ?getSymbols.
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.
##
## WARNING: There have been significant changes to Yahoo Finance data.
## Please see the Warning section of '?getSymbols.yahoo' for details.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.yahoo.warning"=FALSE).
## [1] "GSPC"
##            GSPC.Open GSPC.High GSPC.Low GSPC.Close GSPC.Volume GSPC.Adjusted
## 2000-01-03   1469.25   1478.00  1438.36    1455.22   931800000       1455.22
## 2000-01-04   1455.22   1455.22  1397.43    1399.42  1009000000       1399.42
## 2000-01-05   1399.42   1413.27  1377.68    1402.11  1085500000       1402.11
## 2000-01-06   1402.11   1411.90  1392.10    1403.45  1092300000       1403.45
## 2000-01-07   1403.45   1441.47  1400.73    1441.47  1225200000       1441.47
## 2000-01-10   1441.47   1464.36  1441.47    1457.60  1064800000       1457.60
```

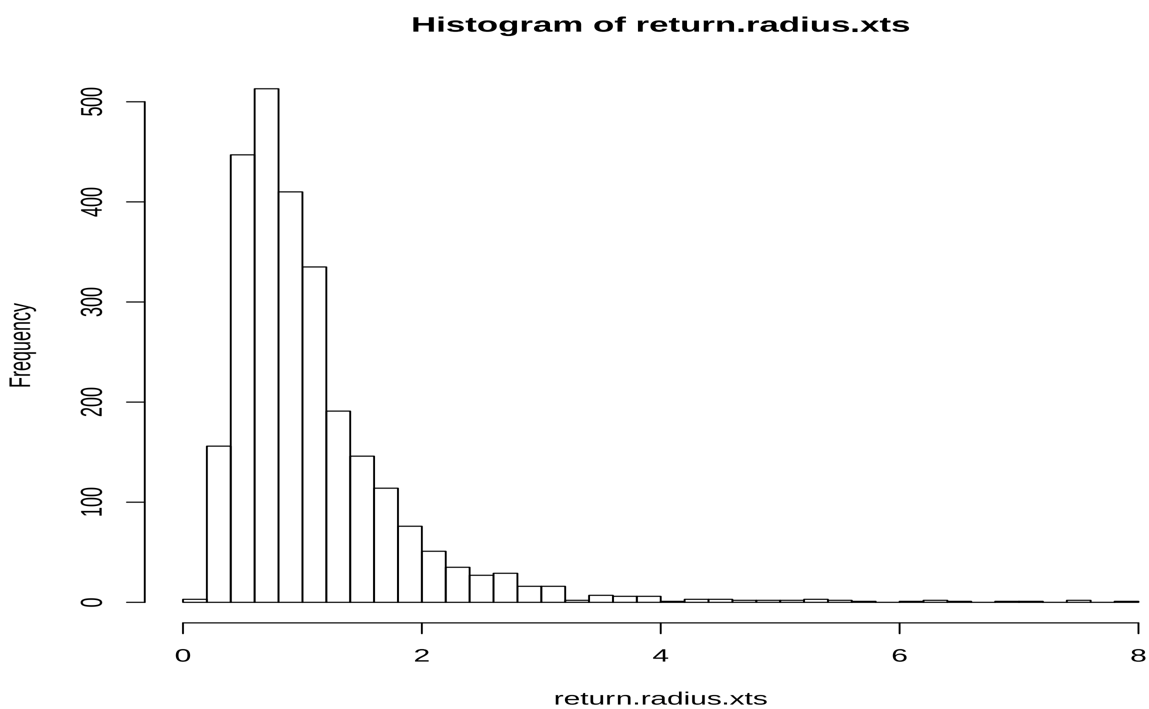

S.intvl.xts
2010−01−04 / 2020−05−22
3000
2500
2000
1500
Jan 04 2010
Jan 03 2011
Jan 03 2012
Jan 02 2013
Jan 02 2014
Jan 02 2015
Jan 04 2016
Jan 03 2017
Jan 02 2018
Jan 02 2019
Dec 31 2019


return.intvl.xts
2010−01−05 / 2020−05−22
10
5
0
−5
−10
Jan 05 2010
Jan 03 2011
Jan 03 2012
Jan 02 2013
Jan 02 2014
Jan 02 2015
Jan 04 2016
Jan 03 2017
Jan 02 2018
Jan 02 2019
Dec 31 2019

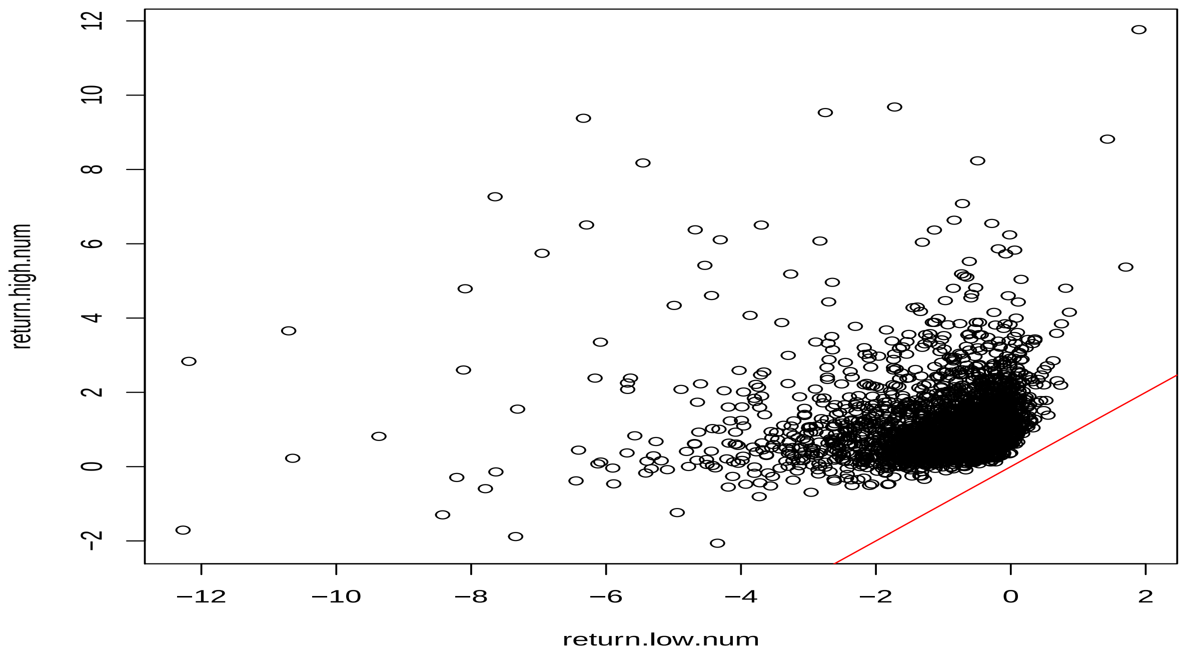

return.center.xts
2010−01−05 / 2020−05−22
6
4
2
0
−2
−4
−6
Jan 05 2010
Jan 03 2011
Jan 03 2012
Jan 02 2013
Jan 02 2014
Jan 02 2015
Jan 04 2016
Jan 03 2017
Jan 02 2018
Jan 02 2019
Dec 31 2019

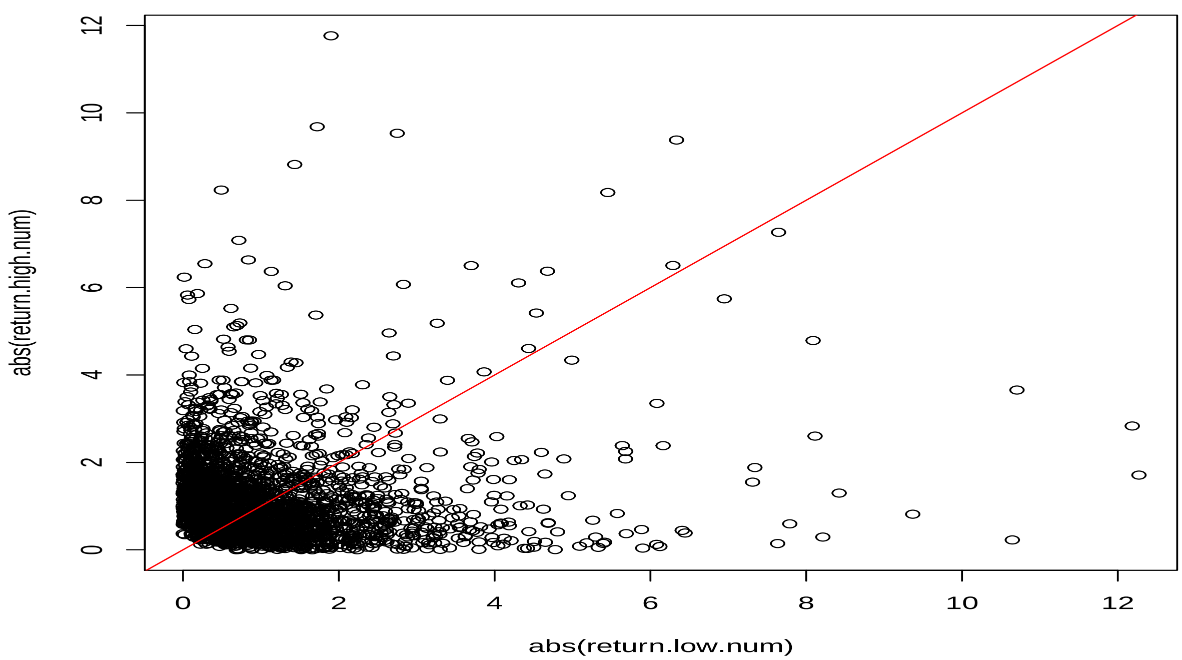

return.radius.xts
2010−01−05 / 2020−05−22
6
4
2
Jan 05 2010
Jan 03 2011
Jan 03 2012
Jan 02 2013
Jan 02 2014
Jan 02 2015
Jan 04 2016
Jan 03 2017
Jan 02 2018
Jan 02 2019
Dec 31 2019

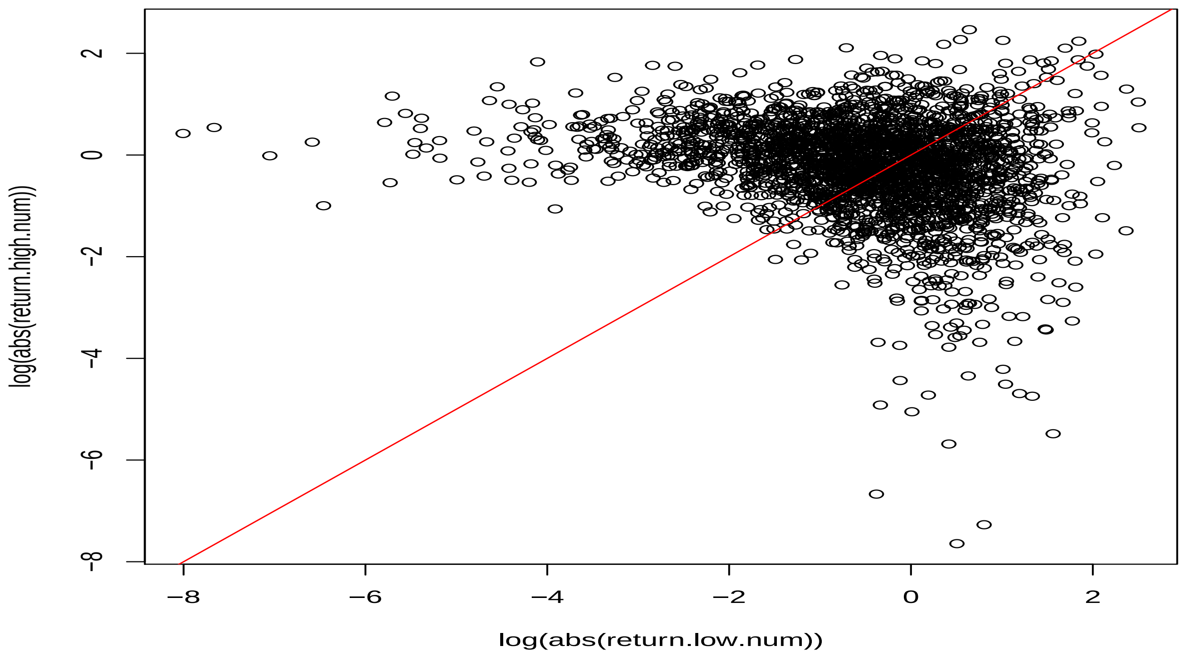




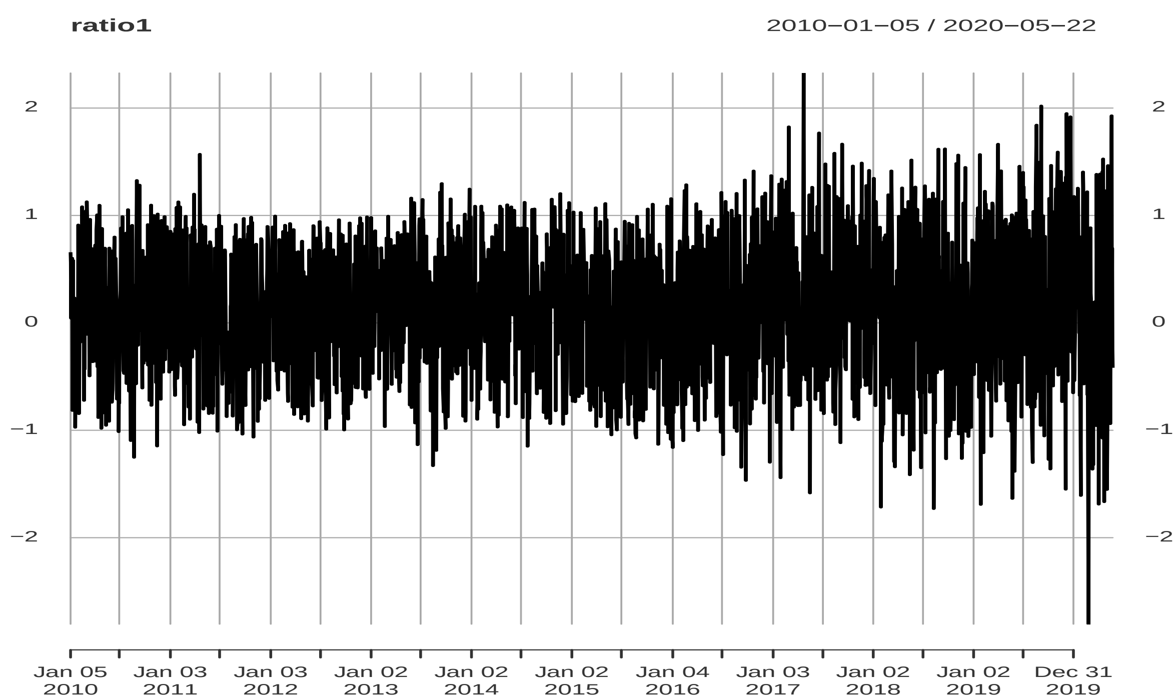
Check the data patterns.

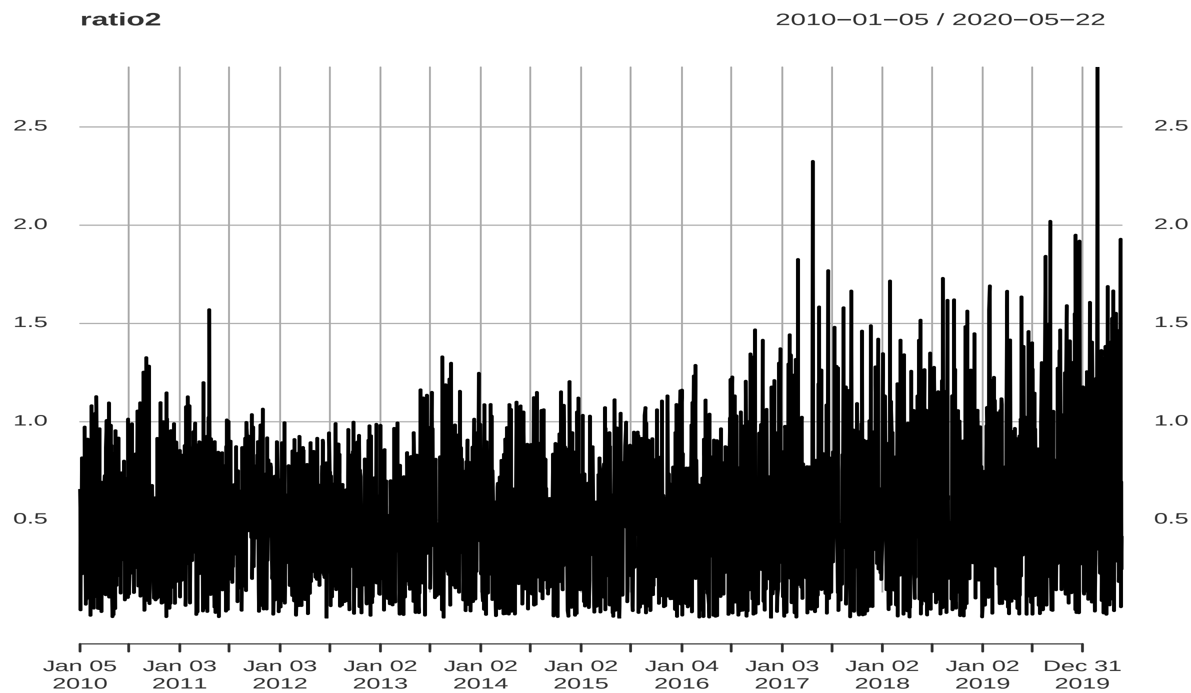

return.high.num
return.low.num
12
10
8
6
4
2
0
-2
−12
−10
−8
−6
−4
−2
0
2

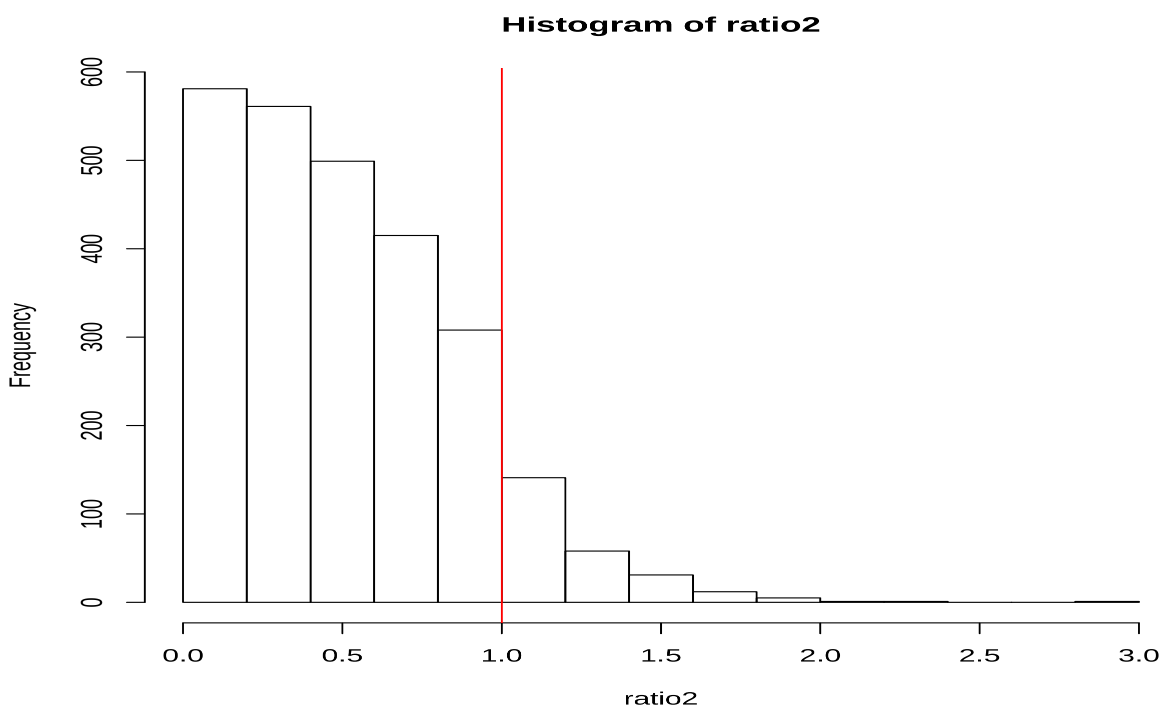

abs(return.high.num)
abs(return.low.num)
12
10
8
6
4
2
0
0
2
4
6
8
10
12

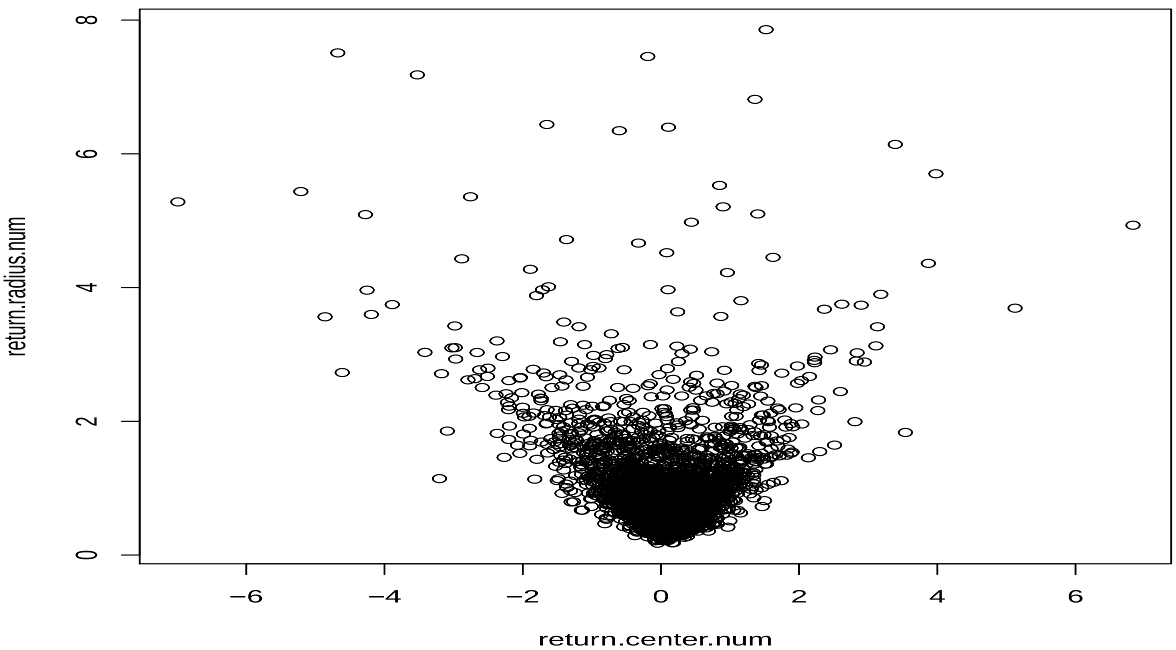


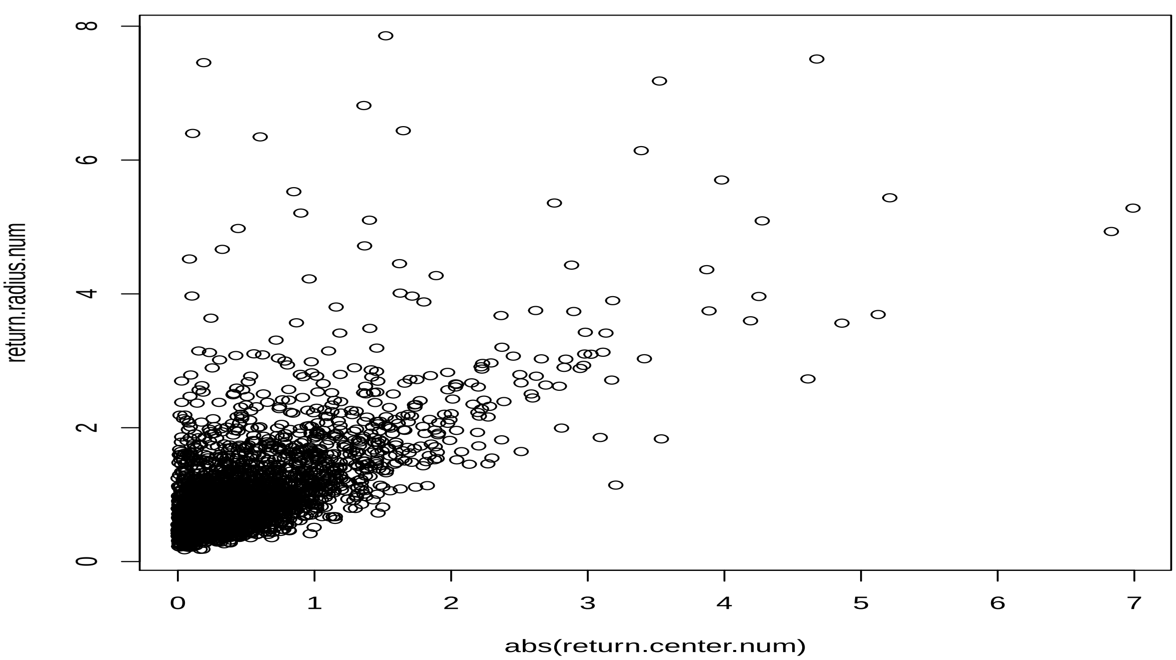
We can also study the ratio: abs(center)/radius.




ratio2
2010−01−05 / 2020−05−22
2.5
2.0
1.5
1.0
0.5
Jan 05 2010
Jan 03 2011
Jan 03 2012
Jan 02 2013
Jan 02 2014
Jan 02 2015
Jan 04 2016
Jan 03 2017
Jan 02 2018
Jan 02 2019
Dec 31 2019


Histogram of ratio2
Frequency
0
100
200
300
400
500
600
0.0
0.5
1.0
1.5
2.0
2.5
3.0
ratio2






```
##
## Call:
## lm(formula = return.radius.num ~ abs(return.center.num))
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.0523 -0.3332 -0.1251  0.1734  6.6679
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.63594    0.01592   39.94   <2e-16 ***
## abs(return.center.num)  0.79848    0.01933   41.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5966 on 2612 degrees of freedom
## Multiple R-squared:  0.3951,Adjusted R-squared:  0.3949
## F-statistic:  1706 on 1 and 2612 DF,  p-value: < 2.2e-16
```

5.4 Construction from a Finer Time Grid: Point-valued Data

5.4.1 A Single Stochastic Volatility Model

(As suggested by Prof. Kulperger) Since the the interval form of the log return data itself comes from the lowest and highest value of the intraday stock price, we can consider a finer time grid and stochastic volatility: W_t is the classical Brownion motion in $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$. Let $\mathcal{F}_t \coloneqq \sigma(W_s, s \leq t)$. We can assume a reasonable model for σ_t (any $\mathcal{F}_t$-measurable process).

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t \, \mathrm{d}W_t,$$

namely,

$$\mathrm{d}\log S_t = \sigma_t \, \mathrm{d}W_t,$$

or in integral form from time 0 to time t,

$$\log S_t = \log S_0 + \int_0^t \sigma_s \, \mathrm{d}W_s,$$

or, the log return from t to $t + \Delta t$ (e.g. Δt is one day):

$$\log S_{t+\Delta t} - \log S_t = \int_t^{t+\Delta t} \sigma_s \, \mathrm{d}W_s.$$

In order to consider the interval-valued log return, we first need to have the lowest and highest stock price. Set up the larger time grid (for daily values) as: $t_i = t_0 + i\Delta t$.

$$\underline{lS}_i \coloneqq \min_{t_i \leq t < t_{i+1}} \log S_t = \log S_0 + \min_{t_i \leq t < t_{i+1}} \int_0^t \sigma_s \, \mathrm{d}W_s$$

and

$$\overline{lS}_i \coloneqq \max_{t_i \leq t < t_{i+1}} \log S_t.$$

We have the daily return (based on the closing price) as

$$R_i \coloneqq \log S_{t_i} - \log S_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \sigma_s \, \mathrm{d}W_s.$$

Then we also have the observed interval-valued log return $[\underline{R}_i, \overline{R}_i]$ as

$$\underline{R}_i = \underline{lS}_i - \overline{lS}_{i-1},$$

and

$$\overline{R}_i = \overline{lS}_i - \underline{lS}_{i-1}.$$

We must have $R_i \in [\underline{R}_i, \overline{R}_i]$. For a given function (as a test/loss/gain/utitility function), we can consider

$$[\mathbf{E}[\varphi(\underline{R}_i)], \mathbf{E}[\varphi(\overline{R}_i)].$$

We also have

$$(\mathbf{E}[\varphi(\log \frac{S_{t_i}}{S_{t_{i-1}}})] =)\mathbf{E}[\varphi(R_i)] = \mathbf{E}[\varphi(\int_{t_{i-1}}^{t_i} \sigma_s \, \mathrm{d}W_s)].$$

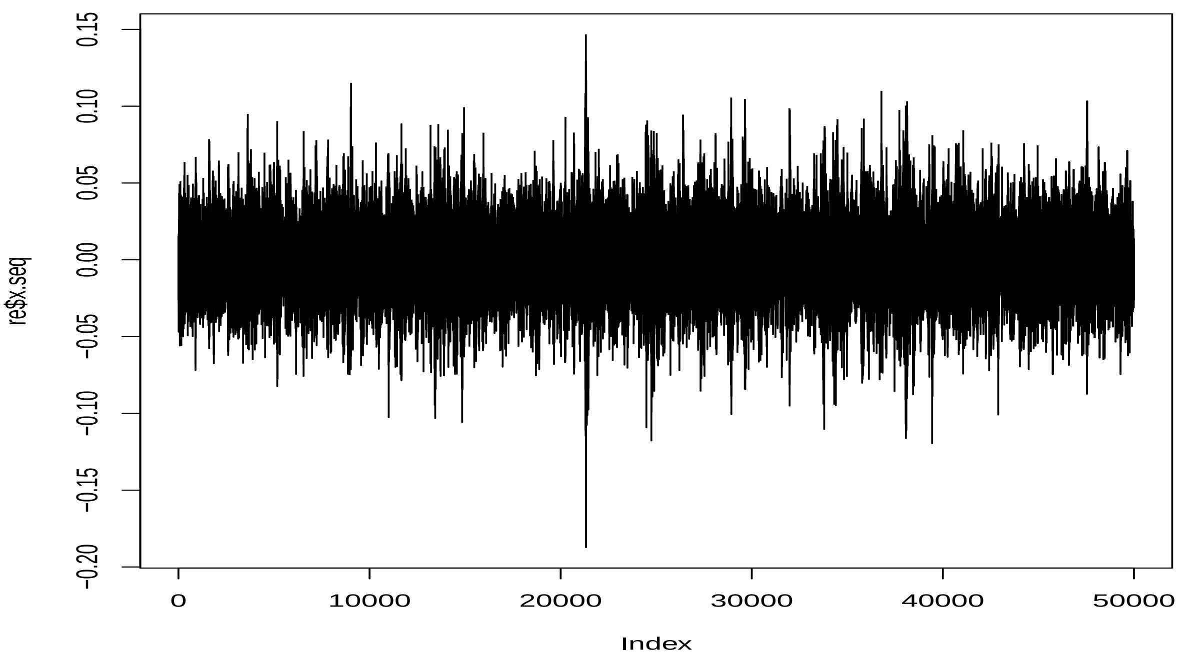
Next we can implement a simulation study by specifying σ_t as a GARCH(1,1) model. Let $r_t = \log R_t$. Assume

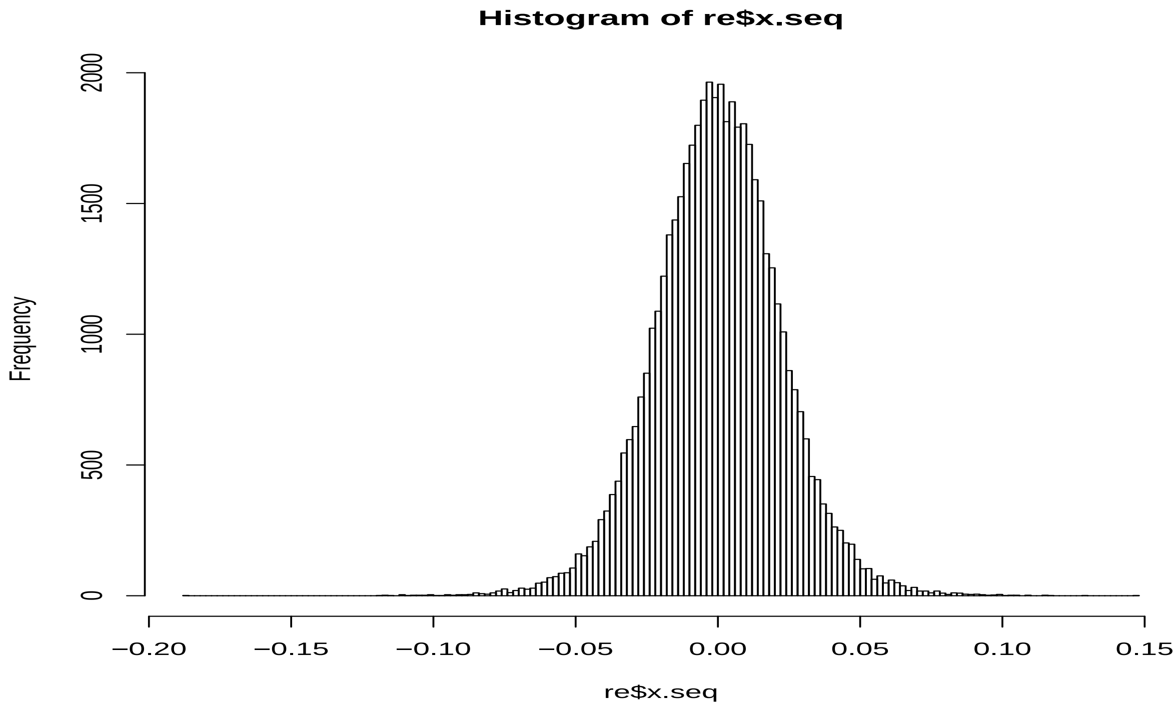
$$r_t = \sigma_t \epsilon_t,$$

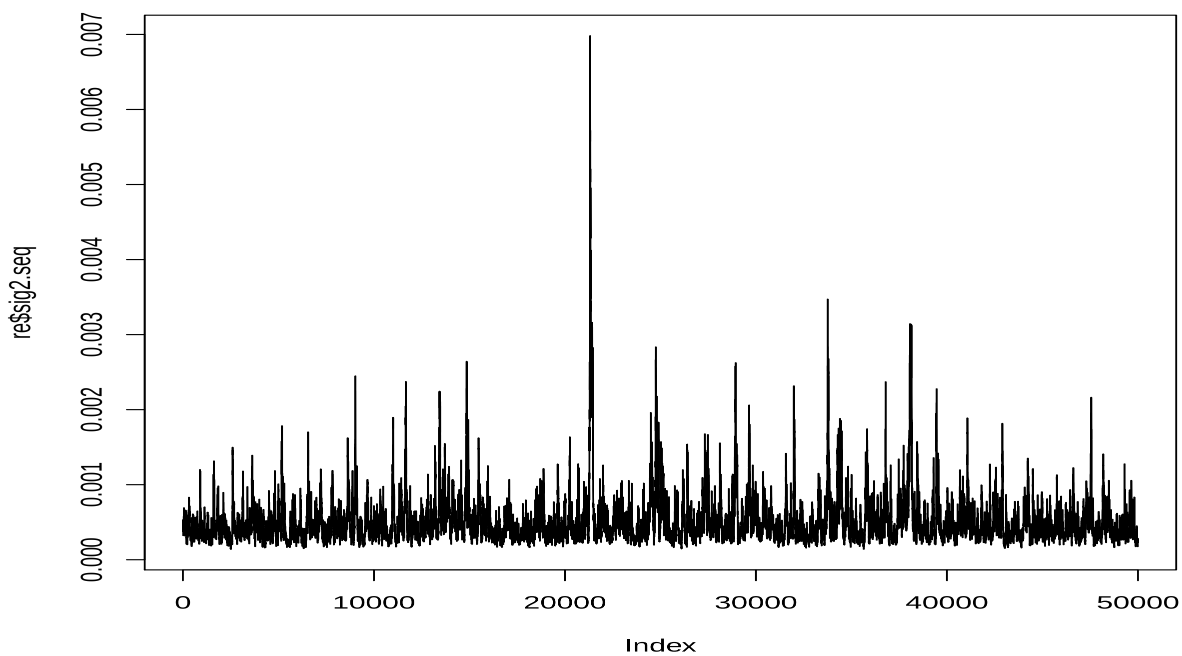
and

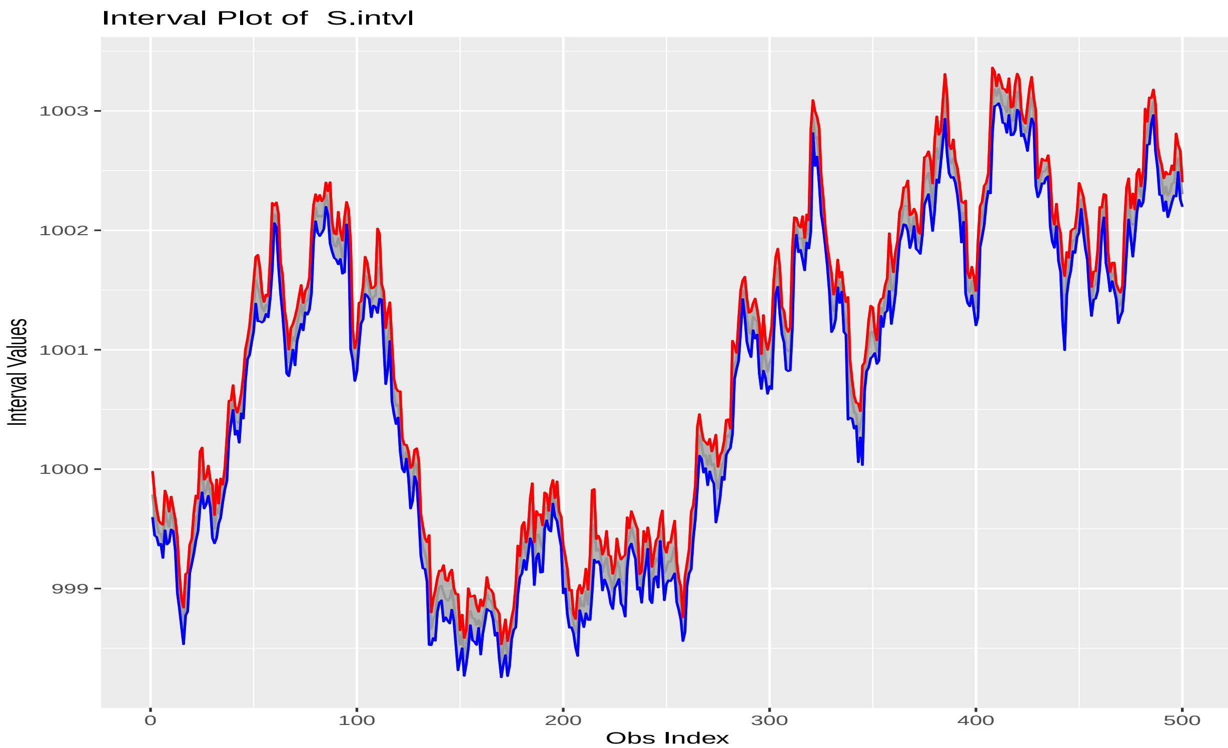
$$\sigma_t^2 = \alpha + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

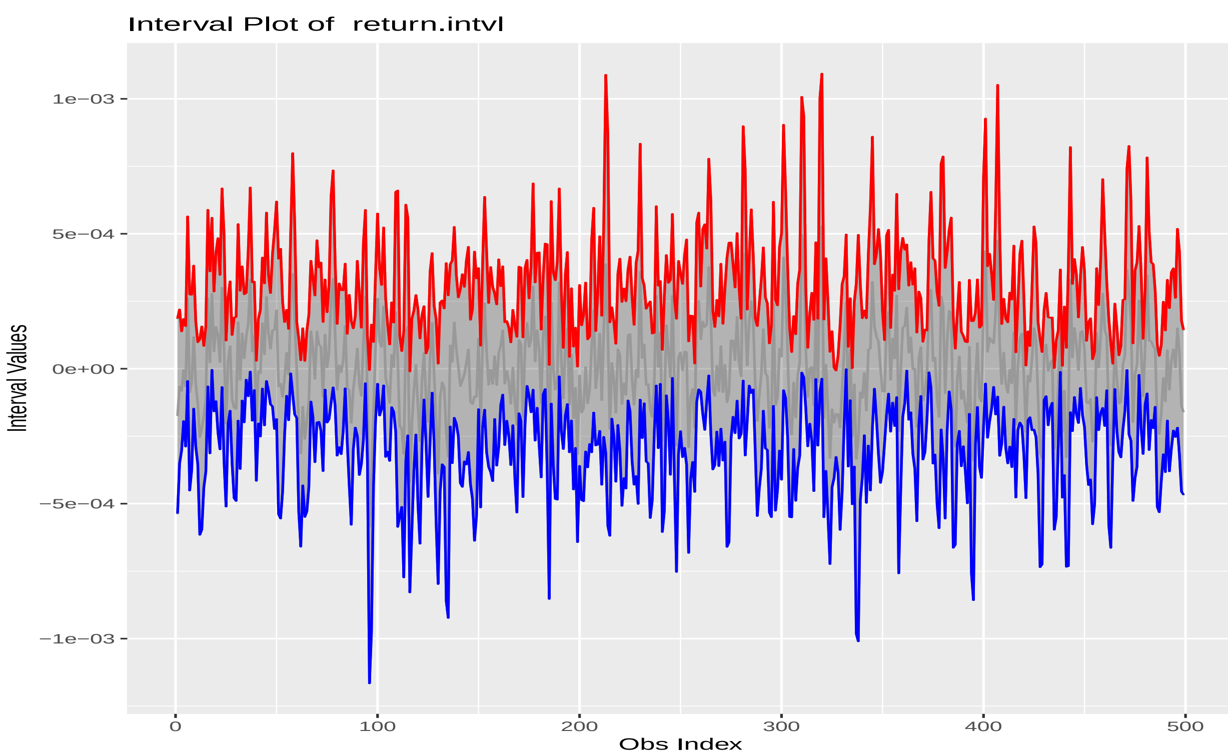
Here we set $(\alpha, \alpha_1, \beta_1) = (1e-5, 0.08, 0.9)$.

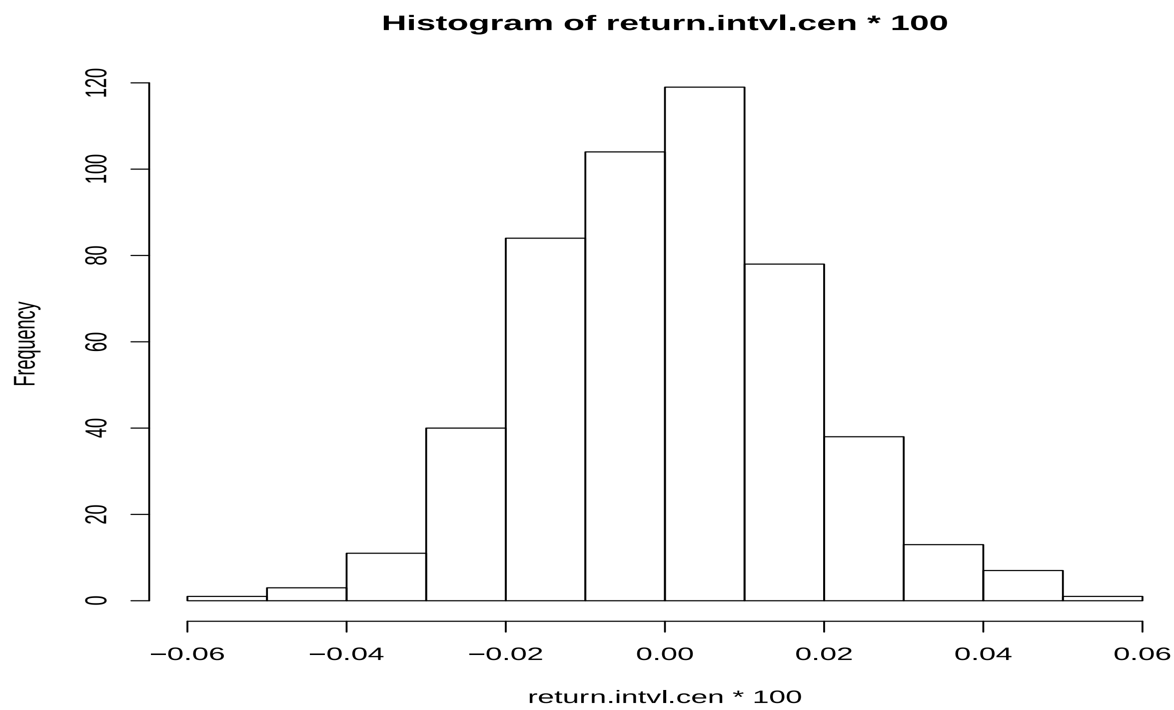



Histogram of re$x.seq
Frequency
0
500
1000
1500
2000
−0.20
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
re$x.seq

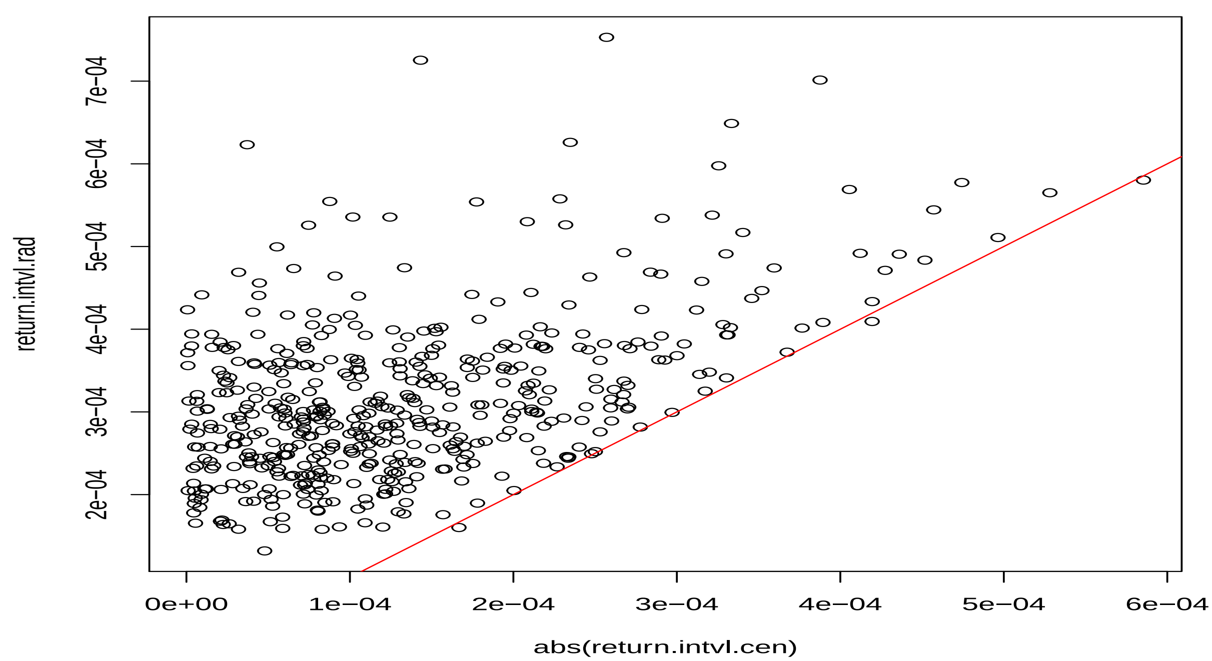

re$sig2.seq
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0
10000
20000
30000
40000
50000
Index

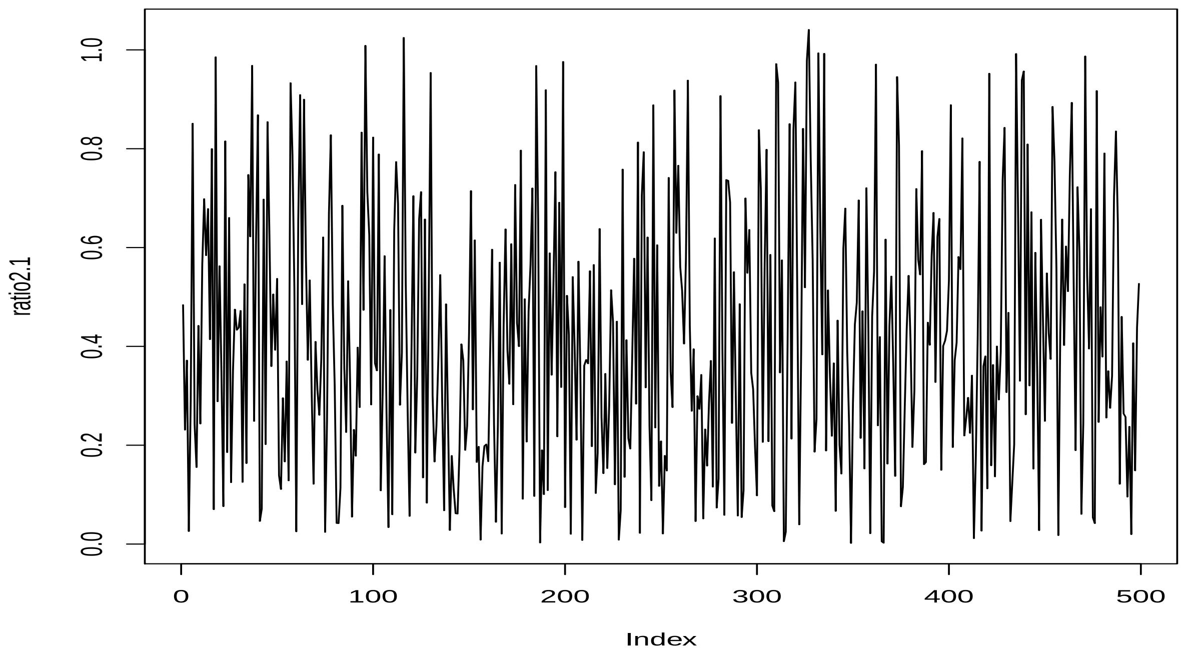

Interval Plot of S.intvl
Interval Values
1003
1002
1001
1000
999
0
100
200
300
400
500
Obs Index

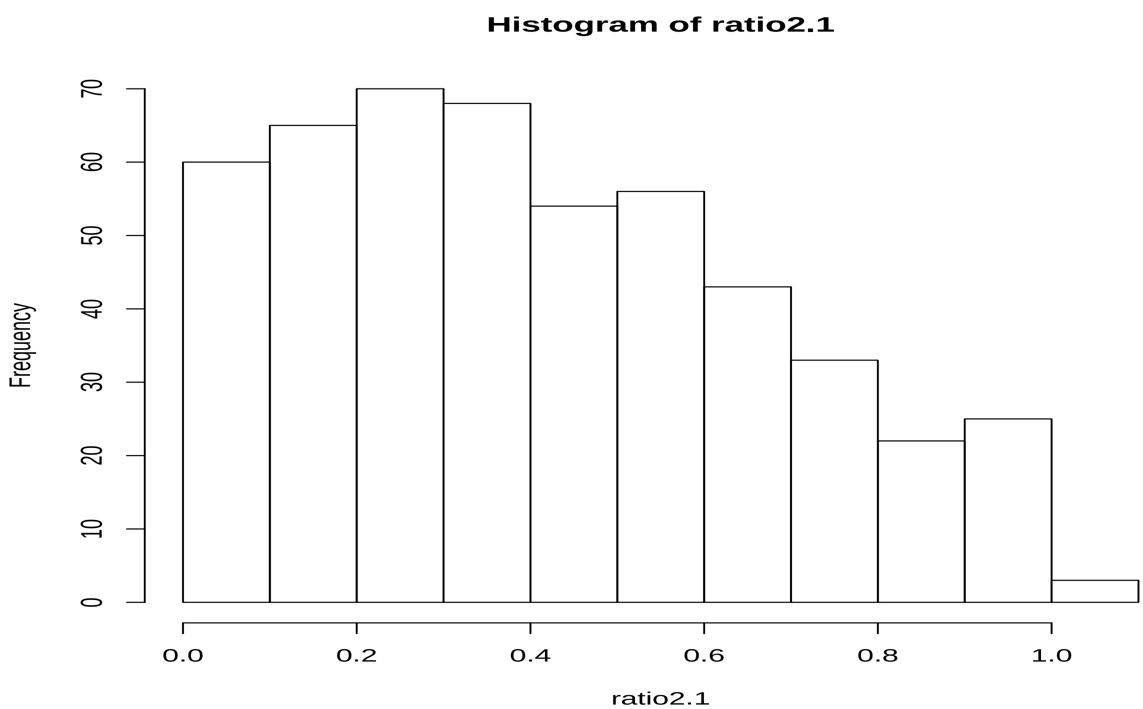

Interval Plot of return.intvl
Interval Values
1e−03
5e−04
0e+00
−5e−04
−1e−03
0
100
200
300
400
500
Obs Index


Histogram of return.intvl.cen * 100
Frequency
0
20
40
60
80
100
120
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
return.intvl.cen * 100

```
## [1] 3.23562
```


return.intvl.rad
2e−04
3e−04
4e−04
5e−04
6e−04
7e−04
0e+00
1e−04
2e−04
3e−04
4e−04
5e−04
6e−04
abs(return.intvl.cen)





Then the next question is how to do the estimation or calibration in this case. Before that, how should we study the property of $[\underline{lS}_i, \overline{lS}_i]$.

This could be one possible future direction for us to imagine and study. Since $\log S_t$ is a martingale, one general result in classical stochastic analysis so far we can apply here is the Martingale inequality, which is an extension of Kolmogorov's maximum inequality, let

$$X_t \coloneqq \log S_t.$$

Let

$$M_u = \max_{t \leq u} X_t,$$

and

$$m_u = \min_{t \leq u} X_t.$$

Then

$$\overline{lS}_i \coloneqq \max_{t_i \leq t < t_{i+1}} X_t = X_{t_i} + \max_{t_i \leq t < t_{i+1}} (X_t - X_{t_i}).$$

Since both $-X_t$ and X_t are submartingale, we can apply the martingale inequalities (Bernstein and Lévy). For any $r \geq 0$ and $u \in [0, \infty)$, since X_t is a submartingale, we have

$$r\mathbf{P}(M_u \geq r) \leq \mathbf{E}[X_u \mathbb{1}_{\{M_u \geq r\}}] \leq \mathbf{E}[X_u^+],$$
$$r\mathbf{P}(m_u \leq -r) \leq -\mathbf{E}[X_0] + \mathbf{E}[X_t \mathbb{1}_{\{m_u > -r\}}] \leq \hat{\mathbf{E}}[X_u^+] - \mathbf{E}[X_0].$$

We can also apply these inequalities to $-X_t$. If we have a parametric form for σ_t, we can derive the distribution of $[\underline{lS}_i, \overline{lS}_i]$.

5.4.2 Study of Uncertainty in the Volatility Model

The ideas here will come back to our previous discussions in the model uncertainty for point-valued data.

If we specify a certain parametric model for the dynamic of σ_t and we calibrate the model based on the current dataset, as time goes and new data come in, in practice, it is required to re-calibrate the model and update the parameters. This essentially indicates we cannot ignore the time-varying feature of the parameter for a given model specification. (It is usually hard to absorb the dynamic of this change into our model to achieve the persistency of the parameter: a classical regime-switching model will be one candidate but its transition probability matrix cannot be ensured that there is no need to be updated after some time. We simply do not know how the future dataset will behave. We are discussing a fundamental concern of the model specification here.)

Furthermore, in different periods, the underlying model structure (i.e. the parametric form) itself may also change: for example, the normal versus crisis period of a financial market. (This is also one of the motivations of the Bayesian pooling or other more sophisicated hierarchical models to impose different model structures on different periods and there is a regime-switching structure assumed.)

We want to start from the other direction to think about these concerns: what is the largest set of possibilities for σ_t (driven by only endogenous dynamic)?

(If we are uncertain about the dynamic of σ_t, then the first question in this point-valued data situation is, how uncertain we are?)

We are still in this log return data example:

$$\log S_t = \log S_0 + \int_0^t \sigma_t \, \mathrm{d}W_t.$$

If we consider the volatility ambiguity/uncertainty, that is, the ambiguity in the dynamic of σ_t, which can be characterized by

$$\mathcal{A} \coloneqq \{\sigma_t : \text{ any } \mathcal{F}_t\text{-measurable process valuing in } [\underline{\sigma}, \overline{\sigma}]\}.$$

This is indeed a really large set, we can update it once we have more information on the datasets (e.g. if the user want to match the stylized patterns of the dataset, we can shrink this set.) A smaller set:

$$\mathcal{A}_0 = \{\sigma_t : \text{any } \sigma(\sigma_s, s \le t)\text{-measurable process valuing in } [\underline{\sigma}, \overline{\sigma}]\}.$$

If we only consider the point-valued daily return (based on the closing price), for any possible dynamic for σ_t, we have the associated log return R_i^σ.

Then we can consider the extreme cases for $\mathbf{E}[\varphi(R_i)]$ (this is still for point-valued data),

$$\begin{aligned}\sup_{\sigma\in\mathcal{A}} \mathbf{E}[\varphi(R_i^\sigma)] &= \sup_{\sigma\in\mathcal{A}} \mathbf{E}[\varphi(\int_{t_{i-1}}^{t_i} \sigma_s \, \mathrm{d}W_s)] \\ &= \hat{\mathbf{E}}[\varphi(B_{t_i} - B_{t_{i-1}})] \\ &= \hat{\mathbf{E}}[\varphi(\sqrt{t_i - t_{i-1}} B_1)],\end{aligned}$$

where $B_1 \sim \mathcal{N}(0, [\underline{\sigma}^2, \overline{\sigma}^2])$. B_t is the G-Brownian motion in $(\Omega, \mathcal{H}, \hat{\mathbf{E}})$.

Furthermore, from (ref: Denis-Hu-Peng) we have

$$\begin{aligned}\sup_{\sigma\in\mathcal{A}} [\varphi(R_1, R_2, \ldots, R_n)] &= \sup_{\sigma\in\mathcal{A}} \mathbf{E}[\varphi(\int_{t_{i-1}}^{t_i} \sigma_s \, \mathrm{d}W_s, i = 1, 2, \ldots, n)] \\ &= \hat{\mathbf{E}}[\varphi(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})].\end{aligned}$$

(The relation with the interval data) Since one agent may get a daily return $R_i \in [\underline{R}_i, \overline{R}_i]$ depending her trading strategy which can achieve outcome that can be approximated by one σ. (it does make sense that, if we assume the expected value of log return is persistently zero, we can assume any trading strategy which could not see into the future is also expected to have zero return.)

For given W_t, we can assume that

$$[\underline{R}_i, \overline{R}_i] = \{\int_{t_{i-1}}^{t_i} \sigma_s \, \mathrm{d}W_s, \sigma \in \mathcal{A}\}.$$

(We can also consider any element σ, then we can see that $\mathbf{E}[\varphi(\sum_i R_i)]$ is sufficiently covered by the G-expectation of the G-BM, which is a strict subset of the expectation of $[\sum_i \underline{R}_i, \sum_i \overline{R}_i]$.

Readers may ask, when will we consider this kind of large set in theory or practice? From the knowledge of authors, here are several possibilities:

1. Consider a general statistical inference for the structure of σ_t:

$$H_0 : \sigma \in \mathcal{A}_0, \ \mathbf{vs} \ H_a : \sigma \in \mathcal{A}/\mathcal{A}_0.$$

2. Provide a scheme to check the capability of a general volatility model which have large flexibility in its structure (e.g. the neuron stochastic volatility model): for a given φ, we can use the nonlinear expectation to provide the theoretical maximum of

$$\hat{\mathbf{E}}[\varphi(\sum_i \sigma_i W_i)],$$

and also provide the argmax for this extreme case (the optimal σ_t), use it as a way to check the capability of a neuron network (to check whether it can learn the optimal σ_t to achieve the extreme case.)

If we are uncertain about the dynamic of σ_t, what is the envelope to cover the possible values of $\mathbf{E}[\varphi(R_i)]$ for a given φ?

5.5 Next development

1. We may also consider other distributions for ϵ_t or only consider it as $\mathrm{IID}(0, 1)$.
2. Mean uncertainty (or the mean certainty is unknown) and variance uncertainty: first candidate model
$$\boldsymbol{X}_t = [\underline{\mu}, \overline{\mu}] + [\underline{\sigma}, \overline{\sigma}]\epsilon_t.$$
3. A direct extension based on this example is the simple linear regression: