Dependent Random Weighting

Katherine Goode and Shan Yu

05/04/2018

Introduction

We were interested in learning about resampling methods for irregularly spaced time series data. This led us to read the paper

> *"The Dependent Random Weighting" (2015) by Srijan Sengupta, Xiaofeng Shao, and Yingchuan Wang.*

The paper:

- Introduces a method that assigns random weights to the irregular time series data
- Weights are created using a dependence structure that mimics that of the observed data

Irregular Time Series Data

Irregular time series data can occur in two ways.

1. **Missing Values**: Time series occurs at equally space intervals but not all data points are observed

2. **Unequal Intervals**: Times when the data are observed are generated from a 1-D point process

<<<<<<< HEAD ## Dependent Random Weighting (the process)

- A stationary time series $\{X_t\}_{t\in\mathbb{Z}}$. And the parameter of interest

Dependent Random Weighting

- The bootstrap sample is

$$\widehat{\theta}^*_{n,DRW} = T(F^*_n).$$

- **Example:** If we are interest in the marginal expectation of X_t, then we have

$$\bar{X}^*_{n,DRW} = \sum_{j=1}^{n} w(t_j) X_{t_j}.$$

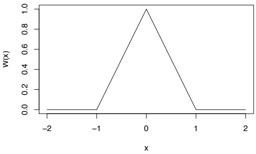
Dependent Random Weighting (theorems)

- Assume the random weights $\{w(t_i)\}_{i=1}^n$ take the form

$$w(t_i) = \frac{Z(t_i)}{\sum_{i=1}^n Z(t_i)}.$$

- $Z(t_i)$ are a realization from a **non-negative** and $l-$**dependent** process $Z(t),\ t \in R$.

- **Example:** $Z(t_i) = (Y(t_i) + c)^2$, where $\{Y(t_i)\}_{i=1}^n \sim N(0, \Sigma)$. Σ is a $n \times n$ matrix with $\Sigma(i,j) = W\left(\frac{t_i - t_j}{l}\right)$, where $W(\cdot)$ is a symmetric kernel function.



Dependent Random Weighting (why it is useful)

. . .

Dependent Random Weighting (theorems)

. . .

Our Simulations: Overview

We wanted to apply and compare DRW to methods learned in STAT 651. We decided to compare the following situations.

- **Methods**: DRW versus MBB
- **Data**: MA versus AR time series
- **Estimators**: mean versus median
- **Bandwidth**: block size versus l-dependence

Note on irregular data type:

- Paper used unequal time intervals (type 2)
- We used equal time intervals with missing values (type 1)

Our Simulations: The Procedure

We used the following procedure for our simulations.

1. **Generate irregular time series of size $n = 400$.**

   (i) Simulate Y_t for $t = 1, \ldots, n$ from
   - an MA(2) process with $\mu = 0$, $\theta_1 = -1$, and $\theta_2 = 0.7$ or
   - an AR(2) process with $\mu = 0$, $\phi_1 = -0.1$, or $\phi_2 = 0.6$.

   (ii) Assign a weight ω_t to Y_t where

   $$\omega_t = \sin\left(\frac{\pi \cdot t}{n}\right).$$

   (iii) Generate $X_t \sim \mathit{binomial}(\omega_t)$ for $t = 1, \ldots, n$.

   (iv) Let

   $$X_t = \begin{cases} Y_t & \text{if } X_t = 1 \\ \text{missing} & \text{if } X_t = 0 \end{cases}$$

   for $t = 1, \ldots, n$.

   (v) Re-index the non-missing X_t as X_i for i from 1 to n_j and use as the observed sample.

Our Simulations: The Procedure

2. **Let $\ell = 1$, and apply the resampling method to $K = 1000$ samples.**
   - MBB: Draw block bootstrap samples from $X_1, ..., X_{n_j}$ with blocks of size $b = \ell$. (ignores missing values)
   - DRW: Randomly assign weights to $X_1, ..., X_{n_j}$ using the method from the paper assuming m-dependence with $m = \ell$.
3. **Compute the mean and median from the K samples.**
4. **Use the distributions of means and medians to compute evaluative measures.**
   - Determine if the 95% confidence interval contains the true value. (True process medians were approximated using 100,000 Monte Carlo simulations.)
   - Compute the standard deviation of the distribution. (Denote this as $\sigma_{n_j}^{(j)}/\sqrt{n_j}$.)

Our Simulations: The Procedure

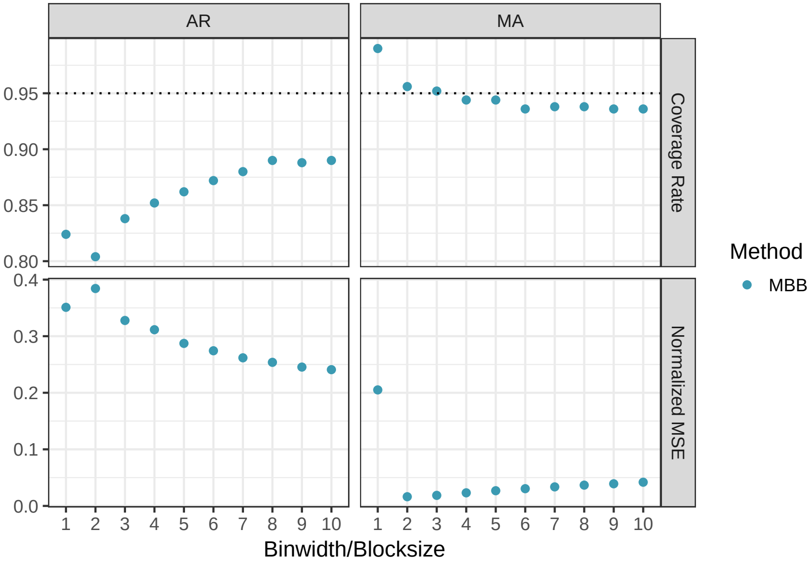
5. **Repeat steps 1 to 4 for $M = 500$ times.**
6. **Compute final evaluative measures.**
   - Coverage rate for both the mean and median
   - Normalized MSE:

$$\frac{1}{M}\sum_{j=1}^{M}\left(\frac{n_j\sigma_{n_j}^{(j)}}{n\sigma_n} - 1\right)^2$$

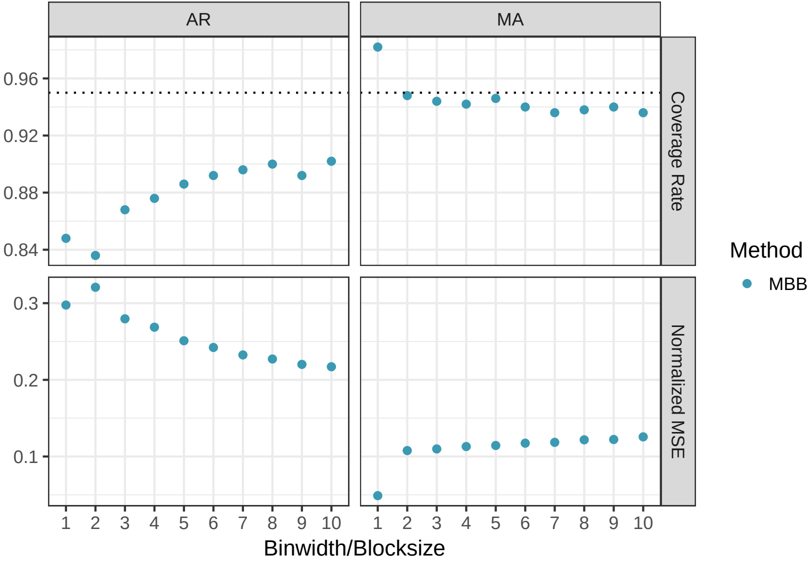
   where $\sigma_n = \sqrt{n}Var(\hat{\theta}_n)$ with $\hat{\theta}_n$ denoting the estimator of interest, was approximated using 100,000 Monte Carlo simulations for both the mean and median

7. **Repeat steps 1 to 6 for $\ell = 2, ..., 10$.**

Our Simulations: Results for Means



Our Simulations: Results for Medians



Our Simulations: Results for Computing Time

We wanted to compare computing times since the paper mentioned that DRW should be easier to implement.

- MBB simulations run on a personal computer
- DRW simulations run on the ISU Condo Cluster

We found that the process took much longer for the DRW than the MBB even when run on a more powerful computer.

	MBB (personal computer)	DRW (ISU Condo)
AR	24.9	
MA	24.6	

Table 1: Computing times (in minutes) for full simulation process within a category

Conclusions

Our simulations provided us with the following information.

- DRW results were...than MBB results
- DRW took more computation time than MBB

It would be interesting to run more simulations to consider:

- Would results change if different parameters were used to simulate AR and MA processes?
- How would different amounts or locations of missingness affect the results?
- Would different sample sizes affect the results?