Simultaneous Confidence Interval by Calibration

Abstract: We have many methods to construct a confidence interval. But we have fewer tools when it comes to multiple testing or simultaneous confidence interval. Many articles come up with specific solutions to particular questions. Relatively speaking, manipulating the size of the confidence interval is a more general method. If those tests are independent of each other, we have a very loose bound for the size α by using Bonferroni correction. But there is not much we can do if there exists a dependent structure beneath. The calibrated CI [1] [2] use calibration to find the right alpha by doing bootstraps. It is not sensitive to the relationship between those tests. Thus it could solve this simultaneous CI problem. In this report, we will focus on a medical experiment and get some useful results from the simulation.

Keywords: Simultaneous CI; Calibrated CI.

1 Introduction

New assessment technologies, such as ecological momentary assessment (EMA), provide powerful tools to collect intensive longitudinal data over numerous time points. It allows the study of dynamic mechanisms of behavior change and captures temporal changes, thereby allowing the estimation of time-varying effects. Numerous examples of these time-varying effect models can be found in the smoking literature. We are going to come up with the calibrated confidence interval of the mediation between covariate and outcome. Mediation is a causal chain. An intervention has an effect on a mediator, which in turn affects the outcome. Mediation analysis helps us understand the underlying mechanisms of underlying behavior change and facilitates the development of more efficacious treatments.

We organize the paper as follows. In Section 2, we introduce the notations and models. In Section 3, we revise how calibrated CI works. Simulations are given in Section 4, where we record the coverage rate and running time.

2 Preliminary

Now, suppose that receipt of pharmacological treatment, such as varenicline, has a strong reduction in the risk of relapse shortly after quitting, but the effect diminishes as time since quitting increases. Suppose we hypothesize that this effect is mediated by craving and that

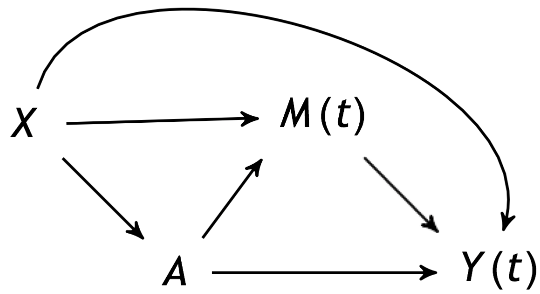


Figure 1: Time-varying mediation model

this mediated effect varies as a function of time since quitting. Figure 1 illustrates a time-varying mediation model where $Y \equiv$ outcome, $A \equiv$ treatment, $X \equiv$ pretreatment covariates, and $M \equiv$ mediator variable.

To test the treatment effect, we fit two linear models:

$$\begin{cases} Y_t = \beta_{0,t} + \beta_{A,t}A_t + \beta_{M,t}M_t + \epsilon_t \\ M_t = \gamma_{0,t} + \gamma_{A,t}A_t + \epsilon'_t \end{cases} \tag{1}$$

We combine these two equations together:

$$\begin{aligned} Y_t &= \beta_{0,t} + \beta_{A,t}A_t + \beta_{M,t}(\gamma_{0,t} + \gamma_{A,t}A_t + \epsilon'_t) + \epsilon_t & (2) \\ &= \beta_{0,t} + \beta_{A,t}A_t + \beta_{M,t}\gamma_{0,t} + \beta_{M,t}\gamma_{A,t}A_t + \beta_{M,t}\epsilon'_t + \epsilon_t & (3) \end{aligned}$$

Our goal is to give a simultaneous confidence interval for the indirect effect $\widehat{\beta_{M,t}}\widehat{\gamma_{A,t}}$ for every t by using calibrated methods. We are going to compare the following three methods:

1. Basic Percentile CI with Bonferroni correction
2. Bootstrap Calibrated-SD CI (Calibrate the CI length)
3. Bootstrap Calibrated-Percentile CI (Calibrate the size α)

3 Calibrated Methods

3.1 Basic Percentile CI

First, let's clarify some notations. Let's say we have n samples, n_t time points $\{t_1, t_2, ..., t_{n_t}\}$. Thus the $Y_t \in \mathbb{R}^{n \times n_t}$. Similarly, $M_t, A_t \in \mathbb{R}^{n \times n_t}$. We fit two linear models mentioned above and get the estimated effect for each t, denoted as $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_{n_t})$. Then we do bootstrap C times. For the i-th bootstrap sample, we calculate the estimated effect, denoted as $\hat{\boldsymbol{\theta}}^*_{\boldsymbol{i}} = (\hat{\theta}^*_{i,1}, \hat{\theta}^*_{i,2}, ..., \hat{\theta}^*_{i,n_t})$. Then, the size-$\alpha$ basic percentile confidence interval at time point t is $(\hat{\theta}^*_{\alpha/2,t}, \hat{\theta}^*_{1-\alpha/2,t})$, where $\hat{\theta}^*_{\alpha,t}$ is the α-quantile of the $\{\hat{\theta}^*_{1,t}, \hat{\theta}^*_{2,t}, ...\hat{\theta}^*_{C,t}\}$. To get simultaneous converage confidence interval for all n_t time points, we consider the effects in each time point are independent. we use bonferroni correction and change α to α/n_t.

3.2 Calibrated Methods

Algorithm 1: z intervals with bootstrap estimate of SD

Input: Data matrix $\mathbf{X}$, calibrated α^*

Output: z interval

1 Given a set of data $\mathbf{X} = (X_1, ..., X_n)$, calculate $\hat{\theta}(t)$ for $t = 1, ..., T$;

2 Typical z interval is $J_\alpha(t) = \hat{\theta}(t) \pm z_{\alpha/2} s(t)$, where z_α is upper-α normal quantile and $s(t)$ is standard deviation of $\hat{\theta}$;

3 Use bootstrap to find $s(t)$;

4 **for** b = *1,...,B* **do**

5   Draw bootstrap sample $\mathbf{X}_b^*$ from $\mathbf{X}$;

6   Let $\hat{\theta}_b^*(t)$ be the estimate of $\theta(t)$ based on $\mathbf{X}_b^*$;

7 **end**

8 Calculate $\bar{\hat{\theta}} = B^{-1} \sum_b \hat{\theta}_b^*(t)$;

9 Calculate $s^2(t) = (B-1)^{-1} \sum_b (\hat{\theta}_b^*(t) - \bar{\hat{\theta}})^2$;

10 Calculate $J_\alpha(t) = \hat{\theta}(t) \pm z_{\alpha/2} s(t)$;

11 Given α, draw bootstrap samples from $\hat{F}$ to find α^* such that

$$P_{\hat{F}}\{\hat{\theta}(t) \in J_\alpha^*(t) \text{ for all } t = 1, ..., T\} = 1 - \alpha$$

Note: $s(t)$ must be re-estimated by inner bootstrap loop (4) for each $J_\alpha(t)$;

12 Then by bootstrap theory, if $P_{\hat{F}} \to P_F$ as $\hat{F} \to F$, then

$$P_F\{\theta(t) \in J_\alpha^*(t) \text{ for all } t = 1, ..., T\} \to 1 - \alpha$$

The method above would come up with a symmetric CI. To make it more general, we have the following calibrated bootstrap percentile methods, which break this symmetry by using the quantile of the data.

Algorithm 2: Bootstrap percentile intervals

Input: Data matrix $\mathbf{X}$, calibrated α^*

Output: Calibrated bootstrap percentile CI

Given a set of data $\mathbf{X} = (X_1, ..., X_n)$, calculate $\hat{\theta}(t)$ for $t = 1, ..., T$;

for b = *1,...,B* **do**

  Draw bootstrap sample $\mathbf{X}_b^*$ from $\mathbf{X}$;

  Let $\hat{\theta}_b^*(t)$ be the estimate of $\theta(t)$ based on $\mathbf{X}_b^*$;

end

Bootstrap percentile interval is $J_\alpha(t) = [L_{\alpha/2}(t), U_{\alpha/2}(t)]$, where the endpoints are $\alpha/2$-quantiles of bootstrap histogram of $\hat{\theta}_1^*, ..., \hat{\theta}_t^*$;

Given α, draw bootstrap samples from $\hat{F}$ to find α^* such that

$$P_{\hat{F}}\{\hat{\theta}(t) \in J_\alpha^*(t) \text{ for all } t = 1, ..., T\} = 1 - \alpha$$

Note: $L_{\alpha/2}(t), U_{\alpha/2}(t)$ must be re-computed by inner bootstrap loop (4) for each $J_\alpha(t)$;

Then by bootstrap theory, if $P_{\hat{F}} \to P_F$ as $\hat{F} \to F$, then

$$P_F\{\theta(t) \in J_\alpha^*(t) \text{ for all } t = 1, ..., T\} \to 1 - \alpha$$

We will use the following calibrated method to find the α^* for the previous two methods.

Algorithm 3: Using iterated bootstrap to find α^*

Input: Data matrix $\mathbf{X}$

Output: Calibrated α^*

Given a set of data $\mathbf{X} = (X_1, ..., X_n)$, calculate $\hat{\theta}(t)$ for $t = 1, ..., T$;

for b = *1,...,B* **do**

  Draw bootstrap sample $\mathbf{X}_b^*$ from $\mathbf{X}$;

  Let $\hat{\theta}_b^*(t)$ be the estimate of $\theta(t)$ based on $\mathbf{X}_b^*$;

  for c = *1,...,C* **do**

    Draw bootstrap sample $\mathbf{X}_{bc}^{**}$ from $\mathbf{X}_b^*$;

    Let $\hat{\theta}_{bc}^{**}(t)$ be the estimate of $\theta(t)$ based on $\mathbf{X}_{bc}^{**}$;

  end

  Let $s_b^*(t)$ be the SD and $[L_{\alpha/2}^*(b,t), U_{\alpha/2}^*(b,t)]$ the lower and upper $\alpha/2$-quantiles of bootstrap histogram $\hat{\theta}_{b1}^{**}(t), ..., \hat{\theta}_{bC}^{**}(t)$;

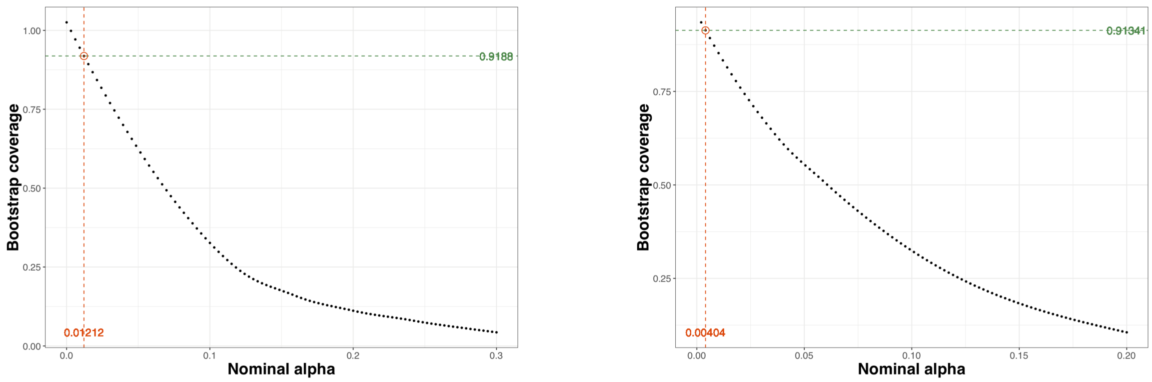
  z-interval is $J_\alpha(b,t) = \hat{\theta}_b^*(t) \pm z_{\alpha/2} s^*(t)$;

  Percentile inveral is $J_\alpha(b,t) = [L_{\alpha/2}^*(b,t), U_{\alpha/2}^*(b,t)]$;

end

Let $\text{Cover}(\alpha) = B^{-1} \sum_{b=1}^{B} I\{\hat{\theta}(t) \in J_\alpha(b,t) \text{ for all } t\}$, $\alpha = 0.05, 0.04, ...$;

Linearly interpolate to find α^* such that $\text{Cover}(\alpha^*) = 1 - \alpha$



(a) α^* for Percentile CI

(b) α^* for z-interval

Figure 2: α^* selection

4 Simulation

4.1 Settings

Remember the linear models are:

$$\begin{aligned} M(t) &= \alpha_0(t) + \alpha_1(t)A + E_M(t) \\ Y(t) &= \beta_0(t) + \beta_1(t)A + \beta_2(t)M + E_Y(t) \end{aligned}$$

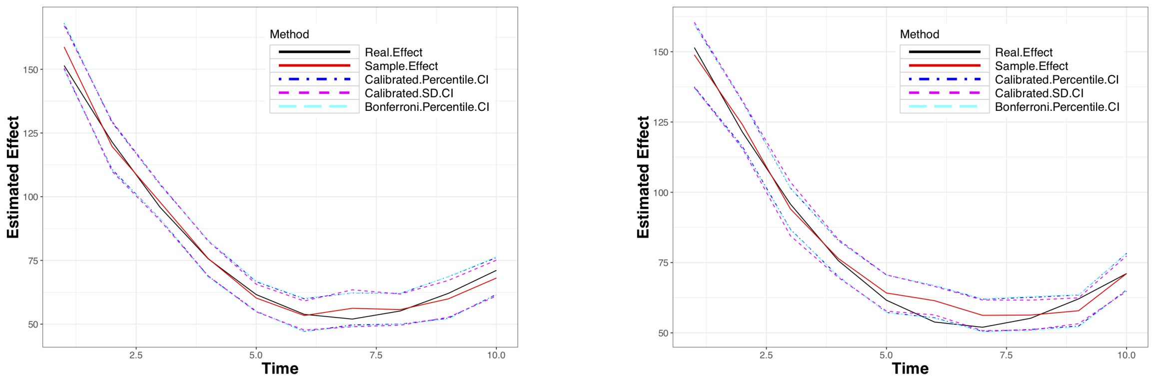
I use the following settings:

$$\begin{aligned} \alpha_0(t) &= \beta_0(t) = 0 \\ \alpha_1(t) &= 15 + 8.7\sin(0.5\pi t) \\ \beta_1(t) &= 4 - 17(t - 1/2)^2 \\ \beta_2(t) &= 1 + 2t^2 + 11.3(1-t)^3 \\ E_M(t) &= E_Y(t) \sim N(0, 2^2) \\ \alpha &= 0.1 \end{aligned}$$

$A \sim$ Bernoulli(0.5)
Time of observation are 10 equally-spaced points on [0,1]
Sample size = 100
Number of Bootstrap(B) = 2371
Number of inner iteration(C) = 1000

4.2 Results

Figure 2 shows the selection of α^* for z-interval and percentile confidence interval in one bootstrap sample.



(a) Simultaneous coverage

(b) Not simultaneous coverage

Figure 3: Confidence intervals

Figure 3 shows two sets of confidence intervals. One of them achieve simultaneous coverage while the other one not.

Table 1 shows the coverage of the three methods. The estimated standard deviations are in the parentheses.

Table 1: Summary table(SEs in parentheses)

Method	Ave. CI length	Nominal α	Coverage rate	Ave. Runtime (mins)
Calibrated-SD CI	13.31(0.02)	0.01348(4e-5)	0.894(0.007)	12.3
Calibated-percentile CI	13.30(0.02)	0.01361(9e-5)	0.900(0.006)	241.2
Bonferroni-percentile CI	12.48(0.01)	0.02	0.872(0.006)	0.3

From the result we can see that Calibrated bootstrap percentile CI has the best performance, although the runtime is also the largest. The reason it outperforms other methods is that it takes the advantage of the information in each iteration while SD method only keep SD and Bonferroni method does not keep any information. When there are some dependence structures beneath (we set the M and Y to be a function of time t), we find that Bonferroni correction lose its power and perform worse than those two calibrated methods.

References

[1] Wei-Yin Loh. Calibrating confidence coefficients. *Journal of the American Statistical Association*, 82(397):155–162, 1987.

[2] Wei-Yin Loh. Bootstrap calibration for confidence interval construction and selection. *Statistica Sinica*, 1(2):477–491, 1991.