An Augmented Method for Empirical *p*-value Calibration in Observational Studies

Huiyuan Wang[1], Yuru Zhu[1], Schuemie Martijn[2,3], Yuqing Lei[1], Yong Chen[1]

[1] Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania

[2] Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey, USA

[3] Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, California, USA

Background

p-values are ubiquitously utilized to make statistical assertions, particularly in evaluating the effects of medical products. However, their validity in observational studies is often compromised due to unmeasured confounding or biased selection. Existing methods for calibrating p-values using multiple negative control outcomes (NCOs) fail to account for internal correlation structures, leading to potentially reduced power.

Method

In response to this issue, we propose a novel subsampling method that is robust to correlation structures and demonstrates greater efficiency. Additionally, we establish a rule-of-thumb for the number of NCOs needed, depending on the correlation structures and sample size. The reliability and validity of our proposed approach are evidenced by intensive simulations and real-world data analysis. For a dataset consisting of N samples with J NCOs, our proposed method mainly consists of three steps:

1. Choose a number $n = N^{\rho}$ for some $\rho \in (0,1)$ and subsample n observations without replacement from the whole sample for B times.
2. For the bth subsample ($b = 1, \dots, B$), apply the conventional causal pipeline to obtain estimators and their asymptotic variances for the causal effect of the treatment on the NCOs, which we denote by $(Y_{b,j}, \sigma_{b,j}^2)$.

3. Compute the average of NCO estimates derived from subsamples weighted by their variances for each NCO

$$\hat{Y}_j = \frac{\sum_b \frac{Y_{b,j}}{\sigma_{b,j}^2}}{\sum_b \frac{1}{\sigma_{b,j}^2}}, \hat{\sigma}_j^2 = \frac{1}{\sum_b \frac{1}{\sigma_{b,j}^2}}.$$

4. Apply the method proposed by Schuemie et al. (2014) to obtain the calibrated p-value:

$$p = \Phi\left(\frac{|Y_{J+1} - \hat{\mu}|}{\sqrt{\hat{\sigma}_{J+1}^2 + \hat{\tau}^2}}\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution,

$$\hat{\mu} = \frac{\sum_j \frac{\hat{Y}_j}{\hat{\sigma}_j^2 + \hat{\tau}^2}}{\sum_j \frac{1}{\hat{\sigma}_j^2 + \hat{\tau}^2}},$$

$$\hat{\tau}^2 = \max\{0, \frac{\sum_j \frac{\left(\hat{Y}_j - \mu_F\right)^2}{\hat{\sigma}_j^2} - (J-1)}{\sum_j \hat{\sigma}_j^{-2} - \sum_j \frac{\hat{\sigma}_j^{-4}}{\sum_j \hat{\sigma}_j^{-2}}}\},$$

and

$$\mu_F = \sum_j \frac{\frac{\hat{Y}_j}{\hat{\sigma}_j^2}}{\sum \frac{1}{\hat{\sigma}_j^2}}.$$

Results

We provide simulation results in this section. The simulation design follows the model specified below.

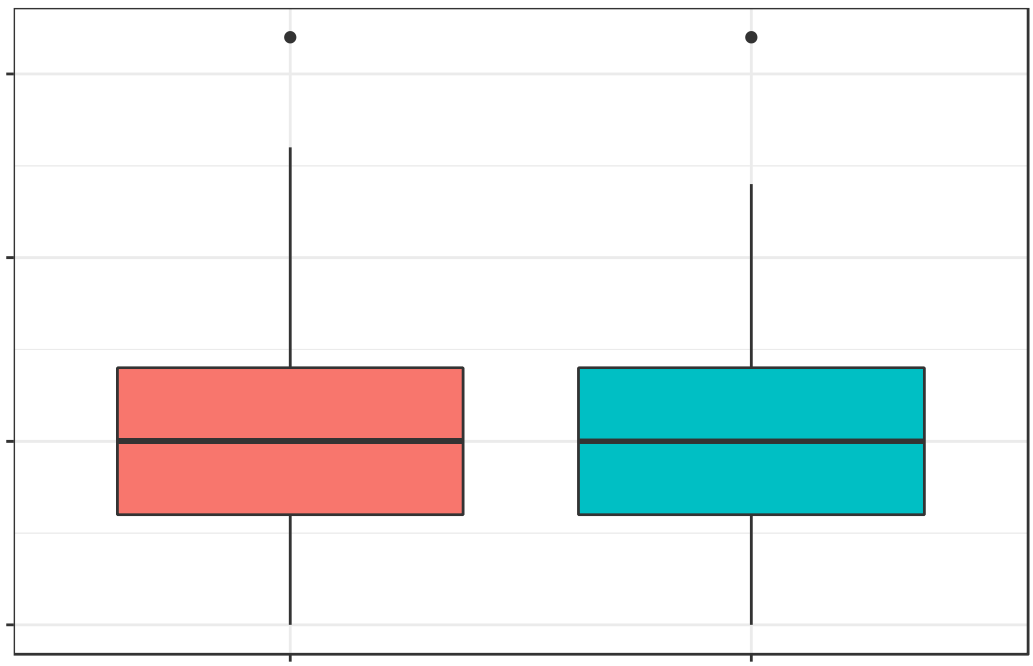
- The NCO is generated from a linear model:

$$W_j = U\beta_j + \varepsilon_j$$

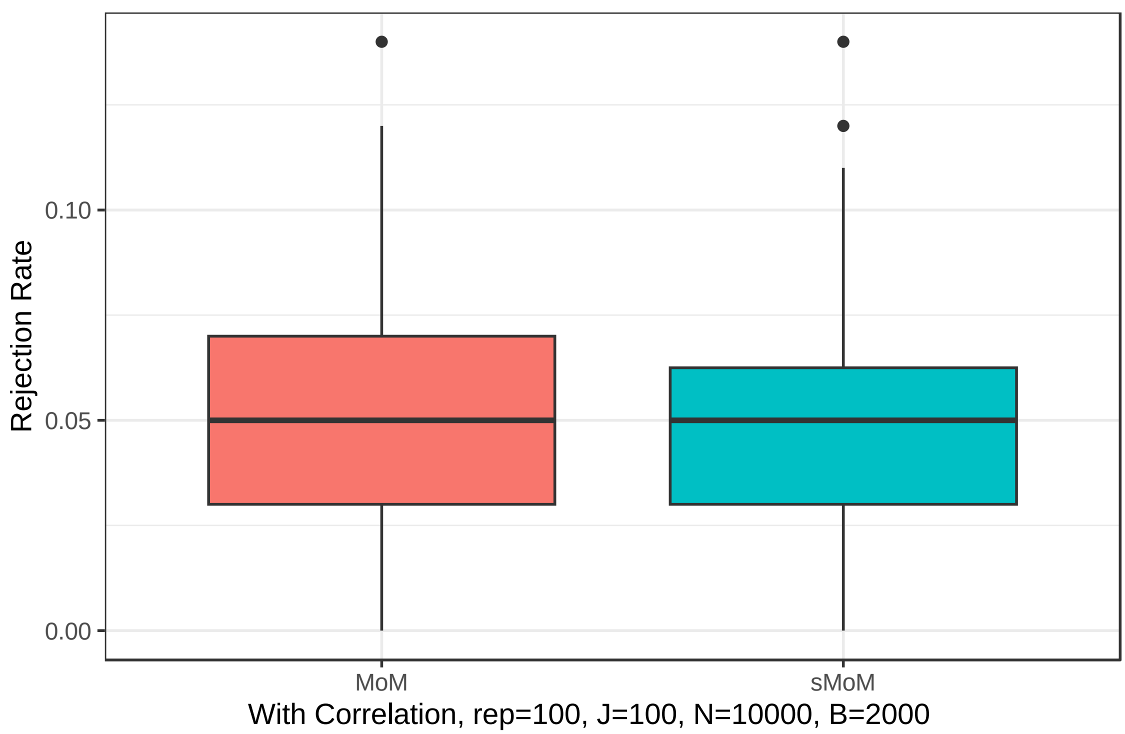
for $j = 1, \ldots, J$, where U denotes the unmeasured confounder, β_j denotes the regression coefficient, and ε_j denotes the additive random noise.

- The treatment assignment follows a logistic regression model:
$$A \mid U \sim \mathrm{Bernoulli}(1/(1+\mathrm{e}^{-U\gamma}))$$
where γ denotes the regression coefficient.
- The unmeasured confounder and regression coefficients are *independently* drawn from the standard normal distribution
$$U, \beta_j, \gamma \sim N(0,1)$$
- The additive noises are correlated
$$(\varepsilon_1, \ldots, \varepsilon_J) \sim N(0_J, \Sigma),$$
where the covariance matrix has the banding structure; that is, its (i, j)th entry is $r^{-|i-j|}$ for some positive constant $r \in (0,1)$.

Simulation results of 100 independent repeated experiments are present as follows.



out



Here, we use MoM to represent the method proposed by Schuemie et al. (2014), and use sMoM to denote our proposed method.

Conclusion

From a toy simulation study, we can find that

1. The method proposed by Schuemie et al. (2014) is robust to the internal correlation among NCOs but exhibit a slightly wide range of rejection rates over the 100 times repeated experiments.
2. Our proposed method attains the required rejection rate on average and meanwhile with a relatively small range of rejection rates, which shows that our proposed method attains a higher efficiency.

Reference

Schuemie, Martijn J., Patrick B. Ryan, William DuMouchel, Marc A. Suchard, and David Madigan. “Interpreting observational studies: why empirical calibration is needed to correct p-values.” *Statistics in Medicine* 33, no. 2 (2014): 209-218.