SAMOS: an R package for unconstrained or constrained sensitivity analysis in matched observational studies via iterative optimization

Pengyun Wang*, Yanxin Shen*, Ping Huang, and Siyu Heng†

University of Oxford
Nankai University
New York University

10 Oct 2023

Paper DOI: Awaiting
Software Repository: Awaiting
Software Archive: Awaiting

Summary

Matching is one of the most common methods for conducting causal inference in observational (i.e., non-experimental) studies. In matched observational studies, treated subjects are matched with control subjects for measured confounders to move the corresponding overt bias effectively. However, like many other study designs in observational studies, matching can not remove hidden bias due to unmeasured confounders. Therefore, in matched observational studies, a sensitivity analysis is typically needed to examine how sensitive a causal conclusion (e.g., the p-value for testing the null effect or the confidence interval for effect) based on the no unmeasured confounding assumption is to potential unmeasured confounding. SAMOS (**S**ensitivity **A**nalysis in **M**atched **O**bservational **S**tudies) is an open-source R package for sensitivity analysis in matched observational studies based on an iterative optimization method. It works for a wide range of matching designs and test statistics and also allows researchers to conduct constrained sensitivity analysis, in which side information about unmeasured confounders can be incorporated to make sensitivity analysis more informative and less conservative.

Statement of need

In causal inference, conducting a randomized experiment is a gold standard but is often infeasible due to ethical or practical issues, and conducting an observational study is a common alternative. When conducting causal inference in observational studies, a central challenge is to adjust for bias due to confounders (i.e., the variables that are associated with both the treatment and the outcome variable). Among various study designs for adjusting for confounders in observational studies, matching is one of the most widely used ones. In matched observational studies, treated subjects are matched with control subjects with the same or similar measured confounders, so the overt bias due to measured confounders can be effectively removed or reduced. However, like other common study designs in observational studies, matching can not control hidden bias due to unmeasured confounders. Therefore, in matched observational studies, it is usually necessary to conduct a sensitivity analysis to assess the sensitivity of causal inference results to potential unmeasured confounding.

Among various sensitivity analysis frameworks in matched observational studies, the Rosenbaum bounds sensitivity analysis is arguably the most commonly used one. In the Rosenbaum bounds sensitivity analysis, researchers first consider a sensitivity parameter $\Gamma \geq 1$ that bounds the ratio of the odds of receiving the

treatment between subjects with the same or similar measured confounders. That is, the sensitivity parameter $\Gamma = 1$ corresponds to no unmeasured confounding, and larger Γ corresponds to a larger magnitude of potential unmeasured confounding. Then, under various prespecified sensitivity parameters, researchers report the worst-case (i.e., maximal) p-value for testing the null effect and the worst-case (i.e., longest) confidence interval for the treatment effect under all possible allocations of unmeasured confounders. A causal inference result (e.g., the p-value is statistically significant under the prespecified significance level α) assuming no unmeasured confounding (i.e., assuming $\Gamma = 1$) is regarded as insensitive to unmeasured confounding if it can still be retained under some large magnitude of potential unmeasured confounding (i.e., large Γ); otherwise, it is considered as sensitive to unmeasured confounding.

Several existing open-source packages implement the Rosenbaum bounds sensitivity analysis for various matching designs and various test statistics, such as R packages rbounds, sensitivitymw, and sensitivitymv. However, there is still a lack of open-source software that can universally work for various matching designs (e.g., pair matching, matching with multiple or variable controls, and full matching) and various common test statistics (those for binary outcomes, ordinal outcomes, or continuous outcomes). Moreover, existing software packages are built for unconstrained sensitivity analysis, which cannot directly implement additional constraints on unmeasured confounders (e.g., box constraints, order constraints, and moment constraints), and recent studies have shown that ignoring such useful side information can render a sensitivity analysis much less informative and much more conservative.

To fill the aforementioned gap, in this work, we use an iterative optimization approach to develop an open-source R package SAMOS (**S**ensitivity **A**nalysis in **M**atched **O**bservational **S**tudies), which can implement the widely used Rosenbaum bounds sensitivity analysis under most of the commonly used matching designs and test statistics (more precisely, the general class of sum statistics). In addition to the traditional unconstrained sensitivity analysis, our R package also allows researchers to conduct constrained sensitivity analyses by adding additional constraints on unmeasured confounding (e.g., box and order constraints based on background information or expertise knowledge) when needed to conduct more informative sensitivity analysis.

Review of the sensitivity analysis framework in matched observational studies

In matched observational studies, sensitivity analysis scrutinizes the robustness of study results against possible latent biases. Given I matched sets with n_i subjects in each, we can represent each subject ij by treatment indicator Z_{ij} and outcome $Y_{ij} = Z_{ij}Y_{ij}(1) + (1 - Z_{ij})Y_{ij}(0)$.

Using test statistics like $T = \sum_{i=1}^{I} \sum_{j=1}^{n_i} Z_{ij} q_{ij}$ based on Fisher's null hypothesis $H_0 : Y_{ij}(1) = Y_{ij}(0)$, alternative versions of Fisher's hypothesis emerge, for instance, $H_\beta : Y_{ij}(1) = Y_{ij}(0) + \beta$. Rosenbaum's foundational model for sensitivity analysis, as presented in [5], addresses assignment bias from a possibly omitted confounder u_{ij} via a logit model:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \theta(\mathbf{x}_{ij}) + \gamma u_{ij}, \quad \text{with } u_{ij} \in [0,1]$$

and sensitivity parameter $\gamma \geq 0$. When $\Gamma = \exp(\gamma) \geq 1$, this model equivalently yields the Rosenbaum bounds:

$$\frac{1}{\Gamma} \leq \frac{\pi_{ij}/(1-\pi_{ij})}{\pi_{ij'}/(1-\pi_{ij'})} \leq \Gamma.$$

Given H_0 and the above model, the permutation distribution of the test statistic becomes:

$$P(T \geq t \mid \mathcal{F}, \mathcal{Z}) = \sum_{\mathbf{z} \in \mathcal{Z}} \mathbf{1}(T(\mathbf{z}, \mathbf{R}) \geq t) \cdot \prod_{i=1}^{I} \frac{\exp(\gamma \sum_{j=1}^{n_i} z_{ij} u_{ij})}{\sum_{j=1}^{n_i} \exp(\gamma u_{ij})}.$$

By adjusting Γ, sensitivity analysis yields a spectrum of potential p-values and confidence intervals. For each $\Gamma = \exp(\gamma)$, the worst-case p-value derived from T for observed t is:

$$\max_{0 \leq u_{ij} \leq 1} P(T \geq t \mid \mathcal{F})$$

In application, researchers increase Γ, calculate the associated worst-case p-value, and denote the maximal Γ at which the p-value crosses a threshold α. This maximal Γ denotes the sensitivity value, revealing the degree of latent bias that could influence initial causal conclusions.

Core functionality

The SAMOS package computes asymptotically sharp worst-case two-sided p-values and confidence intervals under general matching, considering constraints of u_{ij}.

Main algorithm

Given $I = I_1 + I_2$, we re-index each stratum such that for $i = 1, \ldots, I_1$ we have $\sum_{j=1}^{n_i} Z_{ij} = 1$, and for $i = I_1 + 1, \ldots, I_1 + I_2$ we have $\sum_{j=1}^{n_i} Z_{ij} = n_i - 1$. Define $\widetilde{\mathcal{Z}}$ as the collection of treatment assignment indicators satisfying these summation conditions for all i.

The probability p_{ij} represents the chance that $Z_{ij} = 1$ given constraints $\mathcal{F}$ and $\widetilde{\mathcal{Z}}$, whereas $\widetilde{p}_{ij} = 1 - p_{ij}$ represents the converse. Furthermore, $T_i = \sum_{j=1}^{n_i} Z_{ij} q_{ij}$ for each i and j, and our test statistic of interest is $T = \sum_{i=1}^{I} T_i$.

The model specifies:

$$p_{ij} = \frac{\exp(\gamma u_{ij})}{\sum_{j'=1}^{n_i} \exp(\gamma u_{ij'})} \text{ for } i = 1, \ldots, I_1,$$

and

$$\widetilde{p}_{ij} = \frac{\exp(-\gamma u_{ij})}{\sum_{j'=1}^{n_i} \exp(-\gamma u_{ij'})} \text{ for } i = I_1 + 1, \ldots, I_1 + I_2.$$

From the above, the expected value of T under the constraints is given by:

$$\widetilde{\mu}_{\mathbf{u}} = \sum_{i=1}^{I_1} \sum_{j=1}^{n_i} p_{ij} q_{ij} + \sum_{i=I_1+1}^{I_1+I_2} \sum_{j=1}^{n_i} q_{ij} - \sum_{i=I_1+1}^{I_1+I_2} \sum_{j=1}^{n_i} \widetilde{p}_{ij} q_{ij},$$

and its variance as:

$$\widetilde{\sigma}^2_{\mathbf{u}} = \sum_{i=1}^{I_1} \sum_{j=1}^{n_i} p_{ij} q_{ij}^2 - \left(\sum_{i=1}^{I_1} \sum_{j=1}^{n_i} p_{ij} q_{ij} \right)^2 + \sum_{i=I_1+1}^{I_1+I_2} \sum_{j=1}^{n_i} \widetilde{p}_{ij} q_{ij}^2 - \left(\sum_{i=I_1+1}^{I_1+I_2} \sum_{j=1}^{n_i} \widetilde{p}_{ij} q_{ij} \right)^2.$$

The identification of the worst-case two-sided p-value and confidence interval aligns with the resolution of the following minimal squared deviate problem

$$\min_{\mathbf{u} \in \mathcal{U}} \quad \frac{(t - \widetilde{\mu}_{\mathbf{u}})^2}{\widetilde{\sigma}^2_{\mathbf{u}}} \qquad \text{(P0)}$$

Following [2] and [4], we then set

$$\begin{cases} s_i = 1/\sum_{j'=1}^{n_i} \exp(\gamma u_{ij'}) & \text{for } i = 1, \ldots, I_1 \\ \widetilde{s}_i = 1/\sum_{j'=1}^{n_i} \exp(-\gamma u_{ij'}) & \text{for } \ i = I_1 + 1, \ldots, I_1 + I_2. \end{cases}$$

and define the following linear constrained quadratic programming:

$$
\begin{aligned}
\underset{p_{ij},\widetilde{p}_{ij},s_i,\widetilde{s}_i}{\text{minimize}} \quad & (t-\widetilde{\mu}_{\mathbf{u}})^2 - c\cdot\widetilde{\sigma}^2_{\mathbf{u}} \qquad (\mathrm{H}_c)\\
\text{subject to} \quad & \sum_{j=1}^{n_i} p_{ij} = 1 \quad \forall i = 1,\ldots,I_1,\\
& \sum_{j=1}^{n_i} \widetilde{p}_{ij} = 1 \quad \forall i = I_1+1,\ldots,I_1+I_2,\\
& s_i \le p_{ij} \le \Gamma s_i \quad \forall i = 1,\ldots,I_1, j = 1,\ldots,n_i,\\
& \Gamma^{-1}\widetilde{s}_i \le \widetilde{p}_{ij} \le \widetilde{s}_i \quad \forall i = I_1+1,\ldots,I_1+I_2, j = 1,\ldots,n_i,\\
& p_{ij} \ge 0 \quad \forall i = 1,\ldots,I_1, j = 1,\ldots,n_i,\\
& \widetilde{p}_{ij} \ge 0 \quad \forall i = I_1+1,\ldots,I_1+I_2, j = 1,\ldots,n_i.
\end{aligned}
$$

Our quadratic programming implemented via algorithm listed below, using gurobi, for nonlinear optimization.

Input: Sensitivity parameter Γ; treatment assignment indicator vector $\mathbf{Z} = (Z_{11},\ldots,Z_{In_I})^T$; the score vector $\mathbf{q} = (q_{11},\ldots,q_{In_I})^T$ associated with $\mathbf{Y} = (Y_{11},\ldots,Y_{In_I})^T$
Step 1: Start with an initial value $c^{(0)}$.

Step 2: In each iteration $m \ge 1$, solve $(\mathrm{H}_{c^{(m-1)}})$, obtain the optimal solution $(p^{*(m)}_{ij}, \widetilde{p}^{*(m)}_{ij}, s^{*(m)}_i, \widetilde{s}^{*(m)}_i)$, and set $c^{(m)} = (t-\widetilde{\mu}^{*(m)}_{\mathbf{u}})^2/\widetilde{\sigma}^{2*(m)}_{\mathbf{u}}$, where

$$
\begin{aligned}
\widetilde{\mu}^{*(m)}_{\mathbf{u}} &= \sum_{i=1}^{I_1}\sum_{j=1}^{n_i} p^{*(m)}_{ij} q_{ij} + \sum_{i=I_1+1}^{I_1+I_2}\sum_{j=1}^{n_i} q_{ij} - \sum_{i=I_1+1}^{I_1+I_2}\sum_{j=1}^{n_i} \widetilde{p}^{*(m)}_{ij} q_{ij},\\
\widetilde{\sigma}^{2*(m)}_{\mathbf{u}} &= \sum_{i=1}^{I_1}\sum_{j=1}^{n_i} p^{*(m)}_{ij} q^2_{ij} - \sum_{i=1}^{I_1}(\sum_{j=1}^{n_i} p^{*(m)}_{ij} q_{ij})^2 + \sum_{i=I_1+1}^{I_1+I_2}\sum_{j=1}^{n_i} \widetilde{p}^{*(m)}_{ij} q^2_{ij} - \sum_{i=I_1+1}^{I_1+I_2}(\sum_{j=1}^{n_i} \widetilde{p}^{*(m)}_{ij} q_{ij})^2.
\end{aligned}
$$

Step 3: Repeat iterations until $|c^{(m)} - c^{(m-1)}| < \epsilon$. Then approximate c^* with $c^{(m)}$.

Output: Report the worst-case two-sided p-value p^* such that $\chi^2_{1,1-p^*} = c^*$.

To calculate the confidence interval, we use the same objective and constraints, with no iterative computation for c^*. We change the initial value $c^{(0)}$ to $\chi^2_{1,1-\alpha}$, utilizing different treatment effects β . Subsequently, one simply checks if $(t-\widetilde{\mu}_{\mathbf{u}})^2 - c\cdot\widetilde{\sigma}^2$ is less than zero. If true, β is inside the confidence interval; if not, it's outside.

The user can input other constraints for each u_{ij}, then incorporate them into the above quadratic programming framework. Below, we provide an example of a linear constraint and a quadratic constraint.

- Subset constraints: The subset of u_{ij} represents scenarios where specific values are indeterminate but within known ranges, as exemplified below:

$$
\begin{cases}
u_{ij} \in [a_{ij}, b_{ij}] \iff p_{ij} \in [\gamma^{-1} s_i \log a_{ij}, \gamma^{-1} s_i \log b_{ij}] & \text{for } i = 1,\ldots,I_1, j = 1,\ldots,n_i,\\
u_{ij} \in [a_{ij}, b_{ij}] \iff p_{ij} \in [\gamma^{-1} s_i \log b_{ij}, \gamma^{-1} s_i \log a_{ij}] & \text{for } i = I_1+1,\ldots,I_1+I_2, j = 1,\ldots,n_i.
\end{cases}
$$

- Rank constraints: The values of u_{ij} remain unknown, the relative ranking of u_{ij} is discernible. For, $p_{ij}/s_i = \exp(\gamma u_{ij})$, $\widetilde{p}_{ij}/\widetilde{s}_i = \exp(-\gamma u_{ij})$. Thus, rank constraints denote the relative magnitude of u_{ij}, as exemplified below:

$$
\begin{cases}
u_{ij} > u_{mn} \iff p_{ij}/s_i > p_{mn}/s_m & \text{for } i,m = 1,\ldots,I_1, j,n = 1,\ldots,n_i,\\
u_{qw} > u_{er} \iff \widetilde{p}_{qw}/\widetilde{s}_q > \widetilde{p}_{er}/\widetilde{s}_e & \text{for } q,e = I_1+1,\ldots,I_1+I_2, w,r = 1,\ldots,n_i.
\end{cases} \tag{1}
$$

Example

Users need to input the following:

- α**:** the significance level
- β**:** the treatment effect
- Γ**:** the sensitivity parameter
- **threshold:** the ϵ, controls the convergence precision of iterative optimization
- **max_iteration:** maximum number of iterations.
- **index:** vector of identifiers for matched sets after conducting matching.
- **q_vec:** score vector.
- **Z:** binary vector identifying controls (0s) and treated (1s) within matched sets.
- **linearmatrix (optional):** user-desired additional linear constraint matrix.
- **quadraticlist (optional):** user-desired additional quadratic constraint list.
- **LB:** potential lower confidence interval bound estimated by user
- **UB:** potential upper confidence interval bound estimated by user

Illustratively, we explore a causal link between teenage pregnancy and child stunting using the 2003 Kenya Demographic and Health Surveys (DHS), informed by prior work [1, 7]. Children of mothers ≤ 18 years form the treatment group, and those with mothers ≥ 19 years are controls, using height-for-age z-scores as outcomes (q_vec). We matched pairs on seven covariates, implemented optimal matching with a rank-based Mahalanobis distance within a propensity score caliper [3], and assessed 397 treated individuals, each paired with two controls. Standardized differences approached zero, indicating balance [6].

The SAMOS can be straightforwardly executed, as demonstrated below:

```r
# Install package required
install.packages("\textsf{SAMOS}")

# Load required package
library(gurobi)
library(slam)
library(Matrix)
library(Rcpp)

# Load Data
data <- load(KenyaBirth.RData)

# Input
alpha <- 0.05
beta <- 0.5
Gamma <- 1.1
threshold <- 1e-3
max_iterations <- 15

# Generate Z
Z <- rep(NA, length(matchedset.index))
Z[treated.subject.index] <- 1
matched.control.subject.index <- as.vector(matched.control.subject.index.mat)
Z[matched.control.subject.index] <- 0
Z <- Z[!is.na(Z)]

# Generate index
```

```
index <- rep(NA, length(matchedset.index))
groups <- ceiling(seq_along(matched.control.subject.index) / 2)
for(i in seq_along(groups)) {
  index[matched.control.subject.index[i]] <- groups[i]
}
for(i in seq_along(treated.subject.index)) {
  index[treated.subject.index[i]] <- i
}
index[treated.subject.index] <- seq_along(treated.subject.index)
index <- index[!is.na(index)]

# Generate q_vec
q_vec <- z_score[!is.na(index)]

# Sort the Z and qvec according to index
sorted_index <- order(index)
index <- index[sorted_index]
Z <- Z[sorted_index]
q_vec <- q_vec[sorted_index]
index
[1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5 ......
Z
[1] 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 ......
q_vec
[1] -0.65  1.27 -0.84 -0.59 -3.31 -1.05 -2.33 -0.63 -1.62 ......

# Calculate worst-case two-sided p-value
out <- \textsf{SAMOS}(data, alpha, beta, Gamma, threshold, max_iterations, index,
    q_vec, Z, linearmatrix=NULL, quadraticlist=NULL, pvalue=TRUE, CI=FALSE)
out
[1] beta Gamma p_value
    0.5 1.1 0.02222245

# Calculate confidence interval
out <- \textsf{SAMOS}(data, alpha, beta, Gamma, threshold, max_iterations, index,
    q_vec, Z, linearmatrix=NULL, quadraticlist=NULL, pvalue=FALSE, CI=TRUE, LB=0,
    UB=1)
out
[1] CI: 0.13 0.43
```

Acknowledgements

Siyu Heng was partially supported by a grant from the New York University Research Catalyst Prize and a New York University School of Global Public Health Research Support Grant. This work was done when Yanxin Shen and Pengyun Wang were doing a remote research internship under Siyu Heng's mentorship.

References

[1] E.K.M. Darteh, E. Acquah, and A. Kumi-Kyereme. "Correlates of stunting among children in Ghana". In: *BMC Public Health* 14.1 (2014), p. 504.

[2] Colin B. Fogarty and Dylan S. Small. "Sensitivity Analysis for Multiple Comparisons in Matched Observational Studies Through Quadratically Constrained Linear Programming". In: *Journal of the American Statistical Association* 111.516 (2016), pp. 1820–1830. DOI: 10.1080/01621459.2015.1120675.

[3] Ben B Hansen and Stephanie Olsen Klopfer. "Optimal full matching and related designs via network flows". In: *Journal of Computational and Graphical Statistics* 15.3 (2006), pp. 609–627.

[4] Siyu Heng et al. “Increasing power for observational studies of aberrant response: An adaptive approach”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.3 (2021), pp. 482–504.
[5] Paul R Rosenbaum. *Observational Studies*. Springer, 2002.
[6] Paul R. Rosenbaum. *Design of Observational Studies*. New York: Springer, 2010.
[7] E. Van de Poel et al. “Malnutrition and the disproportional burden on the poor: The case of Ghana”. In: *International Journal for Equity in Health* 6.1 (2007), p. 21.