A high-breakdown point and high efficiency robust estimator for Mendelian randomization

Wang Yuquan

Fudan University

21110700112@m.fudan.edu.cn

April 1, 2024

Overview

1. Model
   - Model settings
   - Assumptions
2. Method
   - Review
   - MM-estimate for MR
3. Validation
4. Unfinished tasks

Two-sample MR model

$$\hat{\Gamma}_i \sim N(\Gamma_i = \beta_0\gamma_i + \alpha_i, \tau_0^2\sigma_{Y_i}^2),$$
$$\hat{\gamma}_i \sim N(\gamma_i, \tau_0^2\sigma_{X_i}^2).$$

- $\hat{\Gamma}_i, \hat{\gamma}_i$: the coefficient of ith IV to the outcome Y and the exposure X, respectively
- $\sigma_{Y_i}^2, \sigma_{X_i}^2$: the variance of the coefficients
- α_i: pleiotropic effects
- β_0: causal effect
- τ_0: dispersion parameter

Two-sample MR model

Assumptions:
(A1) (i) $\rho(0) = 0$; (ii) $\rho(-u) = \rho(u)$; (iii) ρ is nondecreasing on $[0, \infty)$; (iv) ρ is continuous; (v) $\sup \rho(u) = 1$;
(A2) (i) $\psi(0) = 0$; (ii) $\psi(-u) = -\psi(u)$; (iii) ψ is continuous; (iv) ψ is bounded; (v) $\psi(u) > 0$ on $[0, \infty)$;
(A3) $\|\gamma\|_2^2 = O(1), \|\gamma\|_1 = \Theta(\sqrt{p})$;
(A4) $\sigma x_i^2 = \Theta\left(\frac{1}{n}\right), \sigma x_i^2 = \Theta\left(\frac{1}{n}\right)$, for $1 \leqslant i \leqslant p$;
(A5) $p = \Theta(n)$;
(A6) $\#\left\{i : \gamma_i = \Theta\left(\frac{1}{\sqrt{p}}\right)\right\} \geqslant \frac{p}{K}$ for some $K \geqslant 1$.
(A1) and (A2) are general assumptions for robust loss functions and their derivatives. (A3) and (A4) follows from Zhao et al. (2020).
$\|\gamma\|_1 = \Theta(\sqrt{p})$, and (A6) assume γ_j tend to 0 uniformly, which are needed for identifiability of the parameters together with (A5).

Robust adjusted profile likelihood (RAPS)

The model settings of robMR is a little different from RAPS:

$$\hat{\Gamma}_i \sim N(\Gamma_i = \beta_0\gamma_i + \alpha_i, \sigma^2_{Y_i}),$$
$$\hat{\gamma}_i \sim N(\gamma_i, \sigma^2_{X_i}).$$

RAPS assumes the pleiotropic effects $\alpha_i \sim N(0, \tau^2)$, i.i.d. to explain the symmetric pleiotropic effect, and for the idiosyncratic pleiotropic IVs, which can not be explained, RAPS uses the robust loss function to reduce the influence of these IVs.

The normal assumption for the pleiotropic effect is too strong. Actually if one uses *mr.raps* to conduct MR for the example datasets given by the package, all the over dispersion model (i.e., symmetric pleiotropy) related functions will warn that τ^2 is too small.

The model settings of robMR

Compared to RAPS, robMR assumes the pleiotropic effects are fixed but the proportion of the invalid IVs should be less than 50% (equal when GWAS sample size n and the number of IVs p go to infinity). In this case, the estimator given by robMR is bounded. Furthermore, note that the variances of the coefficients, $\sigma_{Y_i}^2$ and $\sigma_{X_i}^2$, are also estimated from the data, it is possible that the normal assumptions of the coefficients could be violated. Therefore, the dispersion parameter τ_0 explains the model misspecification caused by the accuracy of these variances.

Review of robust statistics: M-estimator of scale

Consider observations x_i satisfying the multiplicative model

$$x_i = \sigma u_i$$

where the u_i are i.i.d with density f_0 and $\sigma > 0$ is the unknown parameter. The distributions of the x_i constitute a scale family, with density

$$\frac{1}{\sigma} f_0 \left(\frac{x}{\sigma} \right).$$

The MLE of σ is

$$\widehat{\sigma} = \arg\max_{\sigma} \frac{1}{\sigma^n} \prod_{i=1}^{n} f_0 \left(\frac{x_i}{\sigma} \right).$$

Review of robust statistics: M-estimator of scale

Taking logs and differentiating with respect to σ yields

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{x_i}{\sigma_n}\right) = 1$$

where $\rho(t) = t\psi(t)$, with $\psi = -f_0'/f_0$. If f_0 is $\mathrm{N}(0,1)$ then $\rho(t) = t^2$. In general, any estimator satisfying an equation of the form

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{x_i}{\sigma_n}\right) = \delta, \tag{2.1}$$

where ρ is a ρ-function and δ is a positive constant, will be called an *M*-estimator of scale. Note that in order for (2.1) to have a solution we must have $0 < \delta < \rho(\infty)$. Hence if ρ is bounded it will be assumed without loss of generality that

$$\rho(\infty) = 1, \quad \delta \in (0,1).$$

Review of robust statistics: Breakdown point

Consider an estimator $\theta_n = \theta_n(x)$ depending on a sample $\mathbf{x} = \{x_1, \ldots, x_n\}$ of size n of i.i.d. variables with distribution F. In all cases of practical interest, there is a value depending on F, $\theta_\infty = \theta_\infty(F)$, such that

$$\theta_n \to_p \theta_\infty(F).$$

Because in real world F is usually approximately known, we are interested in the behavior of $\theta_\infty(F)$ when F ranges over a "neighborhood" of a distribution F_0. There are several ways to characterize neighborhoods.

Review of robust statistics: Breakdown point

The easiest to deal with are contamination neighborhoods:

$$\mathcal{F}(F, \varepsilon) = \{(1-\varepsilon)F + \varepsilon G : G \in \mathcal{G}\}$$

where $\mathcal{C}$ is a suitable set of distributions, often the set of all distributions but in some cases the set of point-mass distributions, where the "point mass" δ_{x_0} is the distribution such that $\mathrm{P}(x = x_0) = 1$.

Definition 2.1

The asymptotic contamination BP of the estimator θ_n at F, denoted by $\varepsilon^*(\theta_n, F)$, is the largest $\varepsilon^* \in (0,1)$ such that for $\varepsilon < \varepsilon^*, \theta_\infty((1-\varepsilon)F + \varepsilon G)$ remains bounded away from the boundary of Θ for all G.

Example: The BP of the M-estimator of scale defined by (2.1) is $\min(\delta, 1-\delta)$, thus δ is often set to $1/2$.

Review of robust statistics: Finite-sample breakdown point

The replacement finite-sample breakdown point (FBP) of $\widehat{\theta}_n$ at x is the largest proportion $\varepsilon_n^*(\theta_n, \mathbf{x})$ of data points that can be arbitrarily replaced by outliers without θ_n leaving a set which is bounded, and also bounded away from the boundary of Θ (Donoho and Huber, 1983). More formally, call $\mathcal{X}_m$ the set of all datasets $\mathbf{y}$ of size n having $n - m$ elements in common with $\mathbf{x}$:

$$\mathcal{X}_m = \{\mathbf{y} : \#(\mathbf{y}) = n, \#(\mathbf{x} \cap \mathbf{y}) = n - m\}.$$

Then

$$\varepsilon_n^*(\theta_n, \mathbf{x}) = \frac{m^*}{n}$$

where $m^* = \max\{m \geq 0 : \theta_n(\mathbf{y}) \text{ bounded and also bounded away from } \partial\Theta \; \forall \mathbf{y} \in \mathcal{X}_m\}$.

Review of robust statistics: S-estimate

Let $\sigma_n = \sigma_n(\mathbf{r})$ be a scale M-estimator based on a vector of residuals

$$\mathbf{r}(\boldsymbol{\beta}) = (r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta}))\,.$$

and ρ is a bounded ρ-function. Then a regression S-estimator can be defined as (Rousseeuw and Yohai, 1984)

$$\boldsymbol{\beta}_n = \arg\min_{\boldsymbol{\beta}} \sigma_n(\mathbf{r}(\boldsymbol{\beta})).$$

Although the maximum BP and FBP for S-estimator are approximately 1/2 when $\delta = 1/2$, it was shown by Hössjer (1992) that an S-estimator with $\mathrm{BP} = 0.5$ has an asymptotic efficiency under normally distributed errors that is not larger than 0.33.

Review of robust statistics: MM-estimate

The MM-estimate is defined in three stages as follows.
Stage 1. Take an estimate $\theta_{0,n}$ of θ_0 with high breakdown-point, possibly 0.5.
Stage 2. Compute the residuals

$$r_i\left(\theta_{0,n}\right) = y_i - \theta_{0,n}' \mathbf{x}_i, \quad 1 \leq i \leq n,$$

and compute the M-estimator of scale $s_n = s\left(\mathbf{r}\left(\theta_{0,n}\right)\right)$ defined by (2.1), using a function ρ_0 satisfying assumption (A1) and $\delta = 1/2$.

Review of robust statistics: MM-estimate

Stage 3. Let ρ_1 be another function satisfying assumption (A1) and such that

$$\rho_1(u) \leq \rho_0(u).$$

Let $\psi_1 = \rho_1'$. Then the MM-estimate $\theta_{1,n}$ is defined as any solution of

$$\sum_{i=1}^{n} \psi_1\left(r_i(\theta)/s_n\right)\mathbf{x}_i = 0,$$

which verifies

$$\sum_{i=1}^{n} \rho_1\left(r_i(\theta_{1,n})/s_n\right) \leq \sum_{i=1}^{n} \rho_1\left(r_i(\theta_{0,n})/s_n\right).$$

Generally, the $\theta_{0,n}$ and s_n in Stage 1&2 are S-estimators. Therefore, the high-breakdown point property of MM-estimate come from S-estimators and the efficiency of MM-estimate can be controlled by ρ_1 in Stage 3.

Two stages of MM-estimate for MR

Define the residuals of MR, $\mathrm{r}_i(\beta)$ as

$$\mathrm{r}_i(\beta) = \frac{\widehat{\Gamma}_i - \beta\hat{\gamma}_i}{\sqrt{\beta^2\sigma_{X_i}^2 + \sigma_{Y_i}^2}},$$

The MM-estimate for MR is defined as follows.
Stage 1. Find the S-estimate $(\hat{\beta}_0, \hat{\tau}_0)$ for (β_0, τ_0), using a function ρ_0 satisfying assumption (A1) and $\delta = 1/2$.
Stage 2. Let ρ_1 be another function satisfying assumption (A1) and such that

$$\rho_1(u) \leq \rho_0(u).$$

.
The MM-estimate for MR is defined as the global minimum $\hat{\beta}_1$ of

$$\frac{1}{p}\sum_{i=1}^{p}\rho_1\left(\frac{r_i(\beta)}{\hat{\tau}_0}\right) \tag{2.2}$$

S-estimator for MR

The S-estimator for MR is defined by:

$$\arg\min_{\beta} \tau(\beta), \text{ where } \tau(\beta) \text{ satisfies } \frac{1}{p}\sum_{i=1}^{p} \rho_0\left(\frac{r_i(\beta)}{\tau}\right) - \frac{1}{2} = 0 \qquad (2.3)$$

S-estimator for MR

Lemma 2.2

For each fixed $\beta \in \mathbb{R}$*, there exists a unique solution* $\tau = \tau(\beta)$ *of the equation (2.2).*

Lemma 2.3

Let M_0 *be any positive number. There exist positive numbers* M_1 *and* M_2 *such that:*
(i) $\tau(\beta) \geqslant M_1$ *for* $\beta \in \mathbb{R}$*;*
(ii) $\tau(\beta) \leqslant M_2$ *if* $|\beta| \leqslant M_0$*.*

Corollary 2.4

$\tau(\beta)$ *is a continuous function of* β*.*

S-estimator for MR

Theorem 2.5

Suppose ρ_0 satisfies (A1) and (A4)-(A7) hold, then (β_0, τ_0) is the only solution of the problem:

$$\min_{\beta \in B} \tau(\beta),$$

where B is a compact set contains β_0 and $\tau(\beta)$ satisfies

$$\frac{1}{p} \sum_{i=1}^{p} E\rho_0 \left(\frac{r_i(\beta)}{\tau} \right) - \frac{1}{2} = 0,$$

where ρ_0 is selected such that $E\rho_0(u) = \frac{1}{2}$ and $u \sim N(0,1)$.

S-estimator for MR

Theorem 2.6

Suppose ρ_0 satisfies (A1) and (A4)-(A7) hold, then (β_0, τ_0) is the only solution of the problem:

$$\min_{\beta \in B} \tau(\beta),$$

where B is a compact set contains β_0 and $\tau(\beta)$ satisfies

$$\frac{1}{p} \sum_{i=1}^{p} E\rho_0\left(\frac{r_i(\beta)}{\tau}\right) - \frac{1}{2} = 0,$$

where ρ_0 is selected such that $E\rho_0(u) = \frac{1}{2}$ and $u \sim N(0,1)$.

S-estimator for MR

Theorem 2.7 (Strongly consistency of S-estimator)

Under the same assumptions of Theorem 2.6, all solutions $\left(\hat{\beta}_0, \hat{\tau}_0\right)$ of the problem:

$$\min_{\beta \in B} \hat{\tau}(\beta),$$

almost surely converge to (β_0, τ_0), where B is a sufficiently large compact set that contains β_0 and $\hat{\tau}(\beta)$ satisfies

$$\frac{1}{p} \sum_{i=1}^{p} \rho_0 \left(\frac{r_i(\beta)}{\tau} \right) - \frac{1}{2} = 0.$$

Estimates under winner curse

BMI-BMI:

RAPS	dIVW	robMR
0.9682	0.9650	0.9890
0.9553	0.9549	0.9831
0.9439	0.9448	0.9812
0.9453	0.9469	0.9815
0.9513	0.9544	0.9853
0.9646	0.9657	0.9945
0.9707	0.9699	0.9942
0.9959	0.9932	1.0096
1.0066	1.0064	1.0070

Estimates under winner curse

BMI-CAD:

RAPS	dIVW	robMR
0.4076	0.4059	0.4236
0.4329	0.4288	0.4478
0.3860	0.3675	0.4094
0.3767	0.3518	0.3977
0.3848	0.3626	0.4113
0.3918	0.3698	0.4196
0.3568	0.3329	0.4032
0.3815	0.3537	0.4142
0.3977	0.3647	0.4047

Estimates under winner curse

BMI-AIS:

RAPS	dIVW	robMR
0.2069	0.1704	0.2133
0.1941	0.1672	0.1994
0.2140	0.1831	0.2144
0.1926	0.1616	0.1953
0.1866	0.1616	0.1959
0.1816	0.1591	0.1941
0.1603	0.1393	0.1819
0.1519	0.1342	0.1713
0.1661	0.1476	0.1659

Unfinished tasks

- Consistency of MM-estimate;
- Breakdown point of S- and MM-estimate;
- Account for winner curse (asymptotic properties of MM-estimate after debiasing);
- Simulation;
- Real data analysis;

The End