Robust Bayesian Causal Inference for High-Dimensional Problems

Tathagata Basu TATHAGATABASUMATHS@GMAIL.COM
Matthias C. M. Troffaes MATTHIAS.TROFFAES@DURHAM.AC.UK
Jochen Einbeck JOCHEN.EINBECK@DURHAM.AC.UK
Department of Mathematical Sciences, Durham University, UK

Consider an observational study comprising of n subjects, where subject i has output $y_i \in \mathbb{R}$, input $x_i \in \mathbb{R}^p$, and treatment indicator $T_i \in \{0,1\}$. In causal inference, we are interested in the effect of T_i on y_i, for instance, to determine the effects of, say, a drug or a medical procedure. In many cases, we are also interested in the association between the inputs (x_i) and the output variable and we consider a regressional model $Y_i = T_i\alpha + x_i^T\beta$ where α is the causal effect. Often, there is also correlation between the treatment indicator T_i and the input x_i, and to avoid bias in the inferences, this needs to be explicitly modelled. Koch et al. [2] suggested a probit model $P(T_i = 1|x_i) = \Phi(x_i^T\gamma)$ (where Φ is the standard normal cumulative distribution function) through a latent variable T_i^*:

$$Y_i|T_i, x_i \sim N(T_i\alpha + x_i^T\beta, \sigma^2) \qquad T_i^*|x_i \sim N(x_i^T\gamma, 1) \qquad T_i := \begin{cases} 1 \text{ if } T_i^*>0 \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

with parameters $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^p$. We are interested in inference about $E(Y_i|T_i = 1) - E(Y_i|T_i = 0) = \alpha$.

For high dimensional problems, that is when $p > n$, we wish to perform variable selection as well as estimating the average causal effect. However, this becomes problematic as we often lack the necessary information. This motivates us to perform a robust Bayesian analysis. We consider spike and slab priors to specify β and γ, so that for $1 \le j \le p$,

$$\beta_j, \gamma_j \mid \pi_j, \sigma^2 \sim \pi_j \mathcal{N}\left(0, \tau_1^2 \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}\right) + (1-\pi_j)\mathcal{N}\left(0, \tau_0^2 \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}\right); \quad \pi_j \sim \text{Beta}\left(sq_j, s(1-q_j)\right). \tag{2}$$

We fix sufficiently small τ_0^2 ($1 \gg \tau_0^2 > 0$) so that (β_j, γ_j) has its probability mass concentrated around zero and consider τ_1^2 to be large so that $\tau_1^2 \gg \tau_0^2$. This allows the prior for $(\beta_j, \gamma_j) \neq (0,0)$ to be flat. We use a set of beta priors to specify the selection probability π_j of the j-th group where q_j represents our prior expectation of the selection probability (π_j) and 's' represents a concentration parameter. We perform our robust Bayesian analysis on $q := (q_1, \ldots, q_p) \in \mathscr{P} := (\mathscr{P}_1, \cdots, \mathscr{P}_p) \subseteq (0,1)^p$. For α, we use a normal distribution with mean zero and a large variance to get an estimate of the causal effect. We can also consider a set of priors for α in a more generalised setting. However, for very large values of p, this can be computationally expensive.

Our hierarchical model allows us to compute the posteriors through Gibbs sampling method. We use a robust decision rule on the posterior expectation of π_j for variable selection. This type of variable selection includes both β and γ simultaneously. However, in some cases, the association with treatment or the association with outcome can be zero or negligible. Therefore we need to perform an ad-hoc rule to recover sparse effects. To do so, we use "decoupled shrinkage and selection" method [1] to enable an adjusted sparse estimate of a potentially weak effect in a selected group. This way, we get a set of robust Bayesian estimates for β and γ as well as a set of robust Bayesian estimates for the average causal effect α.

References

[1] P. Richard Hahn and Carlos M. Carvalho. Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *Journal of the American Statistical Association*, 110(509):435–448, 2015. doi: 10.1080/01621459.2014.993077.

[2] Brandon Koch, David M. Vock, and Julian Wolfson. Covariate selection with group lasso and doubly robust estimation of causal effects. *Biometrics*, 74(1):8–17, 2018. doi: 10.1111/biom.12736.