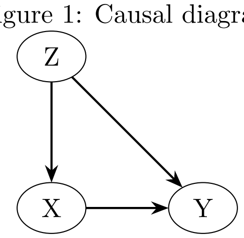
Notes on Evans & Didelez (2023)

Gregor Steiner

July 24, 2023

This document collects my notes on Evans and Didelez (2023). Consider the causal structure in Figure 1. They propose a new parameterization, termed **frugal parameterization**, which consists of three pieces: the joint distribution of the treatment and confounders $p_{ZX}(z,x)$ (the 'past'), the causal distribution of interest $p^*_{Y|X}(y|x)$, and a dependence measure between the outcome and the confounders conditional on the treatment $\phi^*_{YZ|X}$ (could be a conditional odds-ratio or a copula). In sequential treatment models, this parameterization circumvents the so-called **g-null paradox** (Robins and Wasserman, 1997). Their main result shows that a frugal parametrization $\theta = (\theta_{ZX}, \theta_{Y|X}, \phi_{YZ|X})$ of the observational distribution induces a corresponding parameterization $\theta^* = (\theta_{ZX}, \theta^*_{Y|X}, \phi^*_{YZ|X})$ that is also frugal. Replacing θ_{ZX} in θ^* with $\eta_{ZX}(\theta_{ZX})$, where η_{ZX} is a twice differentiable function with a Jacobian of constant rank, yields a parameterization of the causal joint distribution p^*_{ZXY}. Using this, they propose a rejection sampling algorithm to sample from p_{ZXY} (implemented in the R-package causl). Furthermore, they show that under certain assumptions we can obtain consistent parameter estimates for the model $p^*_{Y|X}$ by maximizing the likelihood with respect to the observational data from p_{ZXY}.

Figure 1: Causal diagram



My understanding is that the main advantage of the frugal parameterization is that it provides a convenient characterization of the joint distribution p_{ZXY}, which is usually unknown or difficult to handle. Instead, we can specify a model for p_{ZX} and $p^*_{Y|X}$ and choose a dependence measure of Y and Z conditional on X. Then, we can use these three to perform likelihood-based inference or simulate from the joint distribution. The joint distribution can be factorized into the frugal parameterization:

$$\begin{aligned} p_{ZXY}(z,x,y|\theta^*) &= p_{ZX}(z,x|\theta_{ZX}) p^*_{Y|ZX}(y|z,x) \\ &= p_{ZX}(z,x|\theta_{ZX}) p^*_{Y|X}(y|x;\theta^*_{Y|X}) c(y,z|x), \end{aligned}$$

where $c(y,z|x) =: \phi^*_{YZ|X}$ is a copula density for continuous Y and Z given X. Given observations for (Y,X,Z), we can estimate θ^* by maximizing the expression above.

In their abstract, they mention that this can be used to conduct Bayesian inference, but do not discuss this in the main text. I assume this would work as follows: We can obtain a joint posterior distribution

$$p(\theta^*|z,x,y) \propto p_{ZXY}(z,x,y|\theta^*) p(\theta^*),$$

where $p(\theta^*)$ is some prior distribution. Then, we can find the posterior distribution for the parameters of the marginal model $p^*_{Y|X}$ by integrating out θ_{ZX} and $\phi^*_{YZ|X}$,

$$p(\theta^*_{Y|X}|z,x,y) = \int\int p(\theta^*|z,x,y) d\theta_{ZX} d\phi^*_{YZ|X}.$$

Comments/Questions:

- I don't quite understand what the role of η_{ZX} is and how we obtain it. Is it chosen by the analyst?
- It seems that for the MLE to be consistent the joint model has to be correctly specified. However, the propensity score model can be misspecified. So one only has to correctly specify the marginal model and the dependence measure? This still seems much more restrictive than a doubly robust estimator, while the performance is comparable in their simulations.

Further reading:

- Saarela et al. (2015).
- Robins and Wasserman (1997) and McGrath et al. (2022) to better understand the g-null paradox. Also, problem 29.1 in Ding (2023) is similar to their example R2.
- Bergsma and Rudas (2002) to better understand the discrete case and for some of the proofs.

References

Bergsma, W. P. and Rudas, T. (2002). Marginal models for categorical data. *The Annals of Statistics*, 30(1):140 – 159.

Ding, P. (2023). A first course in causal inference.

Evans, R. J. and Didelez, V. (2023). Parameterizing and Simulating from Causal Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkad058.

McGrath, S., Young, J. G., and Hernán, M. A. (2022). Revisiting the g-null paradox. *Epidemiology (Cambridge, Mass.)*, 33(1):114.

Robins, J. M. and Wasserman, L. (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, UAI'97, page 409–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Saarela, O., Stephens, D. A., Moodie, E. E. M., and Klein, M. B. (2015). On bayesian estimation of marginal structural models. *Biometrics*, 71(2):279–288.