

proof_attempt

1 Attempt at Proof

1.1 Asymptotic Unbiasedness

I now focus on describing the conditions in which the treatment effect is asymptotically unbiased. The asymptotic bias of the treatment effect can be written as:

$$\begin{aligned}\lim_{T_0 \rightarrow \infty} \mathbb{E}[\hat{\tau}_{0,T_0+i} - \tau_{0,T_0+i}] &= \lim_{T_0 \rightarrow \infty} \mathbb{E} \left[Y_{0,T_0+i}(1) - \hat{Y}_{0,T_0+i}(0) - (Y_{0,T_0+i}(1) - Y_{0,T_0+i}(0)) \right] \\ &= \lim_{T_0 \rightarrow \infty} \mathbb{E} \left[Y_{0,T_0+i}(0) - \hat{Y}_{0,T_0+i}(0) \right]\end{aligned}$$

Prior to proving asymptotic unbiasedness of $\mathbb{E} [\alpha_{0,t}]$, some mechanisms must be introduced.

Definition 1.1. The Kullback-Leibler Divergence of the proposed distribution $Pr(y_{0,T_0+i}(0)|Y_0)$ with respect to the true distribution $f(y_{0,T_0+i}(0))$ is defined as:

$$KL_f(\tilde{v}) = \mathbb{E} \left(\log \left(\frac{f(y_{0,T_0+i}(0))}{Pr(y_{0,T_0+i}(0)|Y_0)} \right) \right)$$

where \tilde{v} denotes the parameters and f the true distribution. The $KL_f(v) \geq 0$ with equality when the true data generating process is $Pr(y_{0,T_0+i}(0)|Y_0)$ [gelman_bayesian_2014].

Define v^0 as the unique set of parameters such that:

$$v^0 = \operatorname{argmin}_{\tilde{v}} \mathbb{E} \left(\log \left(\frac{f(y_{0,T_0+i}(0))}{Pr(y_{0,T_0+i}(0)|Y_0)} \right) \right)$$

If the true data generating process is included in the model specification (i.e. $f(y_{0,T_0+i}(0)) = Pr(y_{0,T_0+i}(0)|Y_0)$), then $KL_f(v^0) = 0$. Otherwise, v^0 is the value of the parameters that minimizes the distance between the model and the true data generating process with respect to the Kullback-Leibler Divergence.

The treatment effect can be written as $\mathbb{E}[\hat{\alpha}_{0,t}] = \mathbb{E}[y_{0,t}(1) - \hat{y}_{0,t}(0)] = \mathbb{E}[\alpha_{0,t} + y_{0,t}(0) - \hat{y}_{0,t}(0)]$. In order for $\mathbb{E}[\hat{\alpha}_{0,t}] \xrightarrow{p} \mathbb{E}[\alpha_{0,t}]$ as $T_0 - 1 \rightarrow \infty$, $\mathbb{E}[y_{0,t}(0) - \hat{y}_{0,t}(0)] \xrightarrow{p} 0$. It is sufficient to show $Pr(\hat{y}_{0,T_0+i}(0)|Y_0) \xrightarrow{p} Pr(y_{0,T_0+i}(0)|Y_0)$ as $T_0 - 1 \rightarrow \infty$. In order to show this, it is sufficient to show that the posterior of the parameters converge to the true values (i.e. $Pr(v^0|Y_0) \xrightarrow{p} 1$).

Theorem 1.1 (Asymptotic Unbiasedness). Assume that:

- 1) the model specification is the true data generating process.
- 2) v is defined on a finite parameter space V and $Pr(v = v^0) > 0$.
- 3) $\{y_{0,1}(0), \dots, y_{0,T}(0)\}$ is a stationary ergodic process.
- 4) The poster distribution $Pr(v|Y_0)$ is unimodal.

If all the above assumptions are met, then $\mathbb{E}[\hat{\alpha}_{0,t}] \xrightarrow{p} \mathbb{E}[\alpha_{0,t}]$ as $T_0 - 1 \xrightarrow{p} \infty$.

I will prove this result in three steps. the first step is based off of @gelman_bayesian_2014 (appendix B).

Proof. Step 1: $Pr(v^0|Y_0) \xrightarrow{p} 1$

Define $v' \neq v^0$. Consider the following:

$$\log \left(\frac{Pr(v = v' | Y_0)}{Pr(v = v^0 | Y_0)} \right) = \log \left(\frac{\pi(v = v')}{\pi(v = v^0)} \right) + \log \left(\frac{Pr(Y_0 | v)}{Pr(Y_0 | v^0)} \right) \quad \text{Bayes Rule}$$

(1)

$$= \log \left(\frac{\pi(v = v')}{\pi(v = v^0)} \right) + \log \left(\frac{\prod_{t=1}^{T_0-1} Pr(y_{0,t}(0) | v = v')}{\prod_{t=1}^{T_0-1} Pr(y_{0,t}(0) | v = v^0)} \right) \quad \text{Model Specification}$$

(2)

$$= \log \left(\frac{\pi(v = v')}{\pi(v = v^0)} \right) + \sum_{t=1}^{T_0-1} \log \left(\frac{Pr(y_{0,t}(0) | v = v')}{Pr(y_{0,t}(0) | v = v^0)} \right) \quad \text{log rules}$$

(3)

$$= \log \left(\frac{\pi(v = v')}{\pi(v = v^0)} \right) + \frac{T_0 - 1}{T_0 - 1} \sum_{t=1}^{T_0-1} \log \left(\frac{Pr(y_{0,t}(0) | v = v')}{Pr(y_{0,t}(0) | v = v^0)} \right) \quad \text{multiply by 1}$$

(4)

Focus for a moment on the second term. Notice that as $T_0 - 1 \rightarrow \infty$:

$$\frac{1}{T_0 - 1} \sum_{t=1}^{T_0-1} \log \left(\frac{Pr(y_{0,t}(0) | v = v')}{Pr(y_{0,t}(0) | v = v^0)} \right) \rightarrow \mathbb{E} \left(\log \left(\frac{Pr(y_{0,t}(0) | v = v')}{Pr(y_{0,t}(0) | v = v^0)} \right) \right) \quad \text{ergodic stationary}$$

(5)

$$= \mathbb{E} \left(\log \left(\frac{Pr(y_{0,t}(0) | v = v') f(y_{0,t}(0))}{Pr(y_{0,t}(0) | v = v^0) f(y_{0,t}(0))} \right) \right) \quad \text{multiply by 1}$$

(6)

$$= KL_f(v^0) - KL_f(v') \quad (7)$$

$$< 0 \quad (8)$$

Combining this result and (10) yields:

$$\log \left(\frac{Pr(v = v' | Y_0)}{Pr(v = v^0 | Y_0)} \right) \rightarrow -\infty$$

as $T_0 - 1 \rightarrow \infty$. Rearranging concludes $Pr(v = v' | Y_0) \rightarrow 0$ for all $v \neq v^0$. Therefore, $Pr(v = v^0 | Y_0) \rightarrow 1$. Equivalently, $v \rightarrow v^0$ as $T_0 - 1 \rightarrow \infty$.

Step 2: $Pr(\hat{y}_{0,T_0+i}(0) | Y_0) \xrightarrow{p} Pr(y_{0,T_0+i}(0) | Y_0)$

Recall:

$$Pr(\hat{y}_{0,T_0+i}(0) | Y_0) = \sum_{v' \in V} Pr(y_{0,T_0+i}(0) | Y_0, v = v') Pr(v = v' | Y_0)$$

I need help here. I'm not sure if I can pull the limit into the integral. I feel like I can't, but I also know this should be an asymptotic unbiased estimator from other papers.

step 3: $\mathbb{E}[\hat{\alpha}_{0,t}] \xrightarrow{p} \mathbb{E}[\alpha_{0,t}]$ as $T_0 - 1 \rightarrow \infty$

In order for $\mathbb{E}[\hat{\alpha}_{0,t}] \xrightarrow{p} \mathbb{E}[\alpha_{0,t}]$ as $T_0 - 1 \rightarrow \infty$, $\mathbb{E}[y_{0,T_0+i}(0) - \hat{y}_{0,T_0+i}(0)] \xrightarrow{p} 0$

—Note for me—

Once I get step 2, I think I can say if $Pr(\hat{y}_{0,T_0+i}(0) | Y_0) \xrightarrow{p} Pr(y_{0,T_0+i}(0) | Y_0)$, then $\mathbb{E}[y_{0,t}(0) - \hat{y}_{0,t}(0) | Y_0] \xrightarrow{p} 0$ which then leads to $\mathbb{E}[\mathbb{E}[y_{0,t}(0) - \hat{y}_{0,t}(0) | Y_0]] \xrightarrow{p} 0$.

—

□

Intuition As the pre-treatment period gets larger and larger, the choice of priors will become less important. The data will converge to a point mass on the true parameter value, v^0 . In turn, the mean of the poster predictive distribution will converge to the true value $y_{0,T_0+i}(0)$ leading to an asymptotic unbiased estimator of the treatment effect.

Remark. So long as the choice of priors assigns non-zero probability to the true parameter values v^0 , then the choice of priors is irrelevant in the limit. However, this is not true for finite sample estimation. In application, the choice of priors can have massive effects on inference.

Remark. This proof was shown using a finite parameter space. However, the assumption can be relaxed by assuming a compact set on the parameter space. Should I do it for continuous parameter space? I don't think it adds anything

Remark. This result provides assurance of unbiasedness in the limit. However, inference is not drawn from asymptotic results.

#
