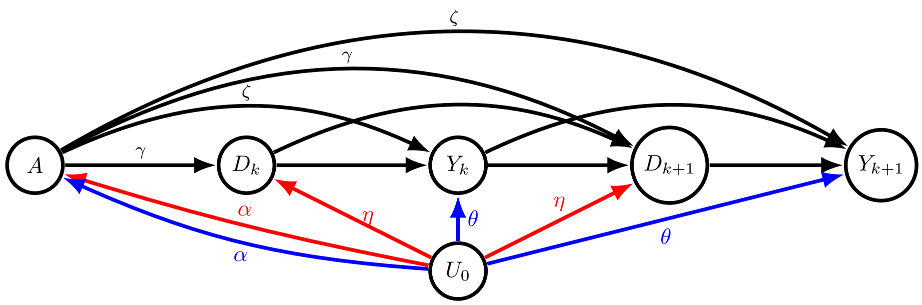
Analytical derivation of Bias

Jost Viebrock

March 2025



Let us assume for the given DAG the parametric distributions:

$$\begin{aligned}
&U_0 \sim Ber(p) \\
&A|U_0 = u_0 \sim Ber\left(\text{logit}^{-1}\left(\alpha u_0\right)\right) \\
&T_D|U_0 = u_0, A = a \sim Exp\left(\lambda_D \exp \gamma a \exp \eta u_0\right) \\
&T_Y|U_0 = u_0, A = a \sim Exp\left(\lambda_Y \exp \zeta a \exp \theta u_0\right)
\end{aligned}$$

We assume exponetial distribution of T_D and T_Y, i.e. $s_D = s_Y = 1$. Therefore, we know about the hazards for the event of interest and the competing event, given $A = a, U_0 = u_0$:

$$\begin{aligned}
h_{T_Y}(k) &= \lambda_Y \exp(\zeta a) \exp(\theta u_0) \\
h_{T_D}(k) &= \lambda_D \exp(\gamma a) \exp(\eta u_0)
\end{aligned}$$

This means, for given $U_0 = u_0$ and $A = a$, we obtain for the cumulative incidence of Y setting u_0 and a

$$\begin{aligned}
F_{T_Y}^{a;u_0}(k) &= 1 - \exp\left(-\sum_{i=0}^{k} h_{T_Y}(i)\right) \\
&= 1 - \exp\left(-\sum_{i=0}^{k} \lambda_Y \exp(\zeta a) \exp(\theta u_0)\right) \\
&= 1 - \exp\left(-\lambda_Y \exp(\zeta a) \exp(\theta u_0) k\right)
\end{aligned}$$

and $F_{T_D}^{a;u_0}(k)$ analogically in order to compute the observed minimum. Therefore, the cumulative incidence of the observed minimum T is

$$\begin{aligned}
F_T(k) &= 1-(1-(1-\exp(-\lambda_Y\exp(\zeta a)\exp(\theta u_0)k)))\,(1-(1-\exp(-\lambda_D\exp(\gamma a)\exp(\eta u_0)k))) \\
&= 1-\exp(-\lambda_Y\exp(\zeta a)\exp(\theta u_0)k)\exp(-\lambda_D\exp(\gamma a)\exp(\eta u_0)k) \\
&= 1-\exp((-\lambda_Y\exp(\zeta a)\exp(\theta u_0)-\lambda_D\exp(\gamma a)\exp(\eta u_0))\,k)\,.
\end{aligned}$$

For two exponentially distributed random variables we know

$$\mathbb{P}[T_Y < T_D] = \frac{\lambda_{T_Y}}{\lambda_{T_Y}+\lambda_{T_D}} = \frac{\lambda_Y\exp(\zeta a)\exp(\theta u_0)}{\lambda_Y\exp(\zeta a)\exp(\theta u_0)+\lambda_D\exp(\gamma a)\exp(\eta u_0)}.$$

This means, we have cumulative incidence of the observed event Y

$$\begin{aligned}
&F_Y^{a;u_0}(k) \\
&= (1-\exp((-\lambda_Y\exp(\zeta a+\theta u_0)-\lambda_D\exp(\gamma a+\eta u_0))\,k))\cdot\frac{\lambda_Y\exp(\zeta a+\theta u_0)}{\lambda_Y\exp(\zeta a+\theta u_0)+\lambda_D\exp(\gamma a+\eta u_0)}.
\end{aligned}$$

For the estimation, we would only condition on $A=a$, so U_0 is integrated out. We need to compute $\mathbb{P}[U_0=u_0|A=a]$. For a Bernoulli distributed random variable, we know:

$$\begin{aligned}
\mathbb{P}[U_0=u_0] &= p^{u_0}\cdot(1-p)^{1-u_0} \\
\mathbb{P}[A=a|U_0=u_0] &= (\text{logit}^{-1}(\alpha u_0)^a)((1-\text{logit}^{-1}(\alpha u_0)))^{1-a} \\
\mathbb{P}[A=a] &= \sum_{i=0}^{1}\mathbb{P}(A=a\mid U_0=i)\cdot\mathbb{P}(U_0=i) \\
&= (p\cdot\text{logit}^{-1}(\alpha))^a)\cdot(p\cdot(1-\text{logit}^{-1}(\alpha)))^{1-a}+\frac{1-p}{2}
\end{aligned}$$

By the Bayes' theorem, we obtain:

$$\begin{aligned}
\mathbb{P}[U_0=u_0|A=a] &= \frac{\mathbb{P}[A=a|U_0=u_0]\cdot\mathbb{P}[U_0=u_0]}{\mathbb{P}[A=a]} \\
&= \frac{\text{logit}^{-1}(\alpha\cdot u_0)^a\left(1-\text{logit}^{-1}(\alpha\cdot u_0)\right)^{1-a}\cdot p^{u_0}\cdot(1-p)^{1-u_0}}{\left(p\cdot\text{logit}^{-1}(\alpha)\right)^a\left(p\cdot\left(1-\text{logit}^{-1}(\alpha)\right)\right)^{1-a}+\frac{1-p}{2}}
\end{aligned}$$

We get the cumulative incidence of the estimation, omitting U_o, by

$$\mathbb{E}\left[\hat{F}_Y^a(k)\right] = \sum_{i=0}^{1} F_{\min(T_Y,T_D)}^{a;u_0}(k) \cdot \mathbb{P}\left[U_0 = i | A = a\right]$$

$$= \left(1 - \exp\left(\left(-\lambda_Y \exp(\zeta a) - \lambda_D \exp(\gamma a)\right) k\right)\right) \cdot \frac{\lambda_Y \exp(\zeta a)}{\lambda_Y \exp(\zeta a) + \lambda_D \exp(\gamma a)}$$

$$\cdot \frac{\frac{(1-p)}{2}}{\left(p \cdot \operatorname{logit}^{-1}(\alpha)\right)^a \left(p \cdot \left(1 - \operatorname{logit}^{-1}(\alpha)\right)\right)^{1-a} + \frac{1-p}{2}}$$

$$+ \left(1 - \exp\left(\left(-\lambda_Y \exp(\zeta a + \theta) - \lambda_D \exp(\gamma a + \eta)\right) k\right)\right) \cdot \frac{\lambda_Y \exp(\zeta a + \theta)}{\lambda_Y \exp(\zeta a + \theta) + \lambda_D \exp(\gamma a + \eta)}$$

$$\cdot \frac{\operatorname{logit}^{-1}(\alpha)^a \left(1 - \operatorname{logit}^{-1}(\alpha)\right)^{1-a} \cdot p}{\left(p \cdot \operatorname{logit}^{-1}(\alpha)\right)^a \left(p \cdot \left(1 - \operatorname{logit}^{-1}(\alpha)\right)\right)^{1-a} + \frac{1-p}{2}}$$

So the estimated total effect with ignoring U_0 is given by:

$$\mathbb{E}[\hat{F}^1(k) - \hat{F}^0(k)] := \mathbb{E}[\hat{\mu}(k)]$$

$$= \left(1 - \exp\left(-\lambda_Y \exp(\zeta)k - \lambda_D \exp(\gamma)k\right)\right) \cdot \frac{\lambda_Y \exp(\zeta)}{\lambda_Y \exp(\zeta) + \lambda_D \exp(\gamma)} \cdot \frac{\frac{1-p}{2}}{p \cdot \operatorname{logit}^{-1}(\alpha) + \frac{1-p}{2}}$$

$$+ \left(1 - \exp\left(-\lambda_Y \exp(\zeta + \theta)k - \lambda_D \exp(\gamma + \eta)k\right)\right) \cdot \frac{\lambda_Y \exp(\zeta + \theta)}{\lambda_Y \exp(\zeta + \theta) + \lambda_D \exp(\gamma + \eta)} \cdot \frac{\operatorname{logit}^{-1}(\alpha) \cdot p}{p \cdot \operatorname{logit}^{-1}(\alpha) + \frac{1-p}{2}}$$

$$- \left(1 - \exp\left(-\lambda_Y k - \lambda_D k\right)\right) \cdot \frac{\lambda_Y}{\lambda_Y + \lambda_D} \cdot \frac{\frac{1-p}{2}}{p \cdot \left(1 - \operatorname{logit}^{-1}(\alpha)\right) + \frac{1-p}{2}}$$

$$- \left(1 - \exp\left(-\lambda_Y \exp(\theta)k - \lambda_D \exp(\eta)k\right)\right) \cdot \frac{\lambda_Y \exp(\theta)}{\lambda_Y \exp(\theta) + \lambda_D \exp(\eta)} \cdot \frac{\left(1 - \operatorname{logit}^{-1}(\alpha)\right) \cdot p}{p \cdot \left(1 - \operatorname{logit}^{-1}(\alpha)\right) + \frac{1-p}{2}}$$

We obtaining the true total effect by randomization, which means we set $\alpha = 0$.

$$F^1(k) - F^0(k) := \mu(k)$$

$$= \left(1 - \exp\left(-\lambda_Y \exp(\zeta)k - \lambda_D \exp(\gamma)k\right)\right) \cdot \frac{\lambda_Y \exp(\zeta)}{\lambda_Y \exp(\zeta) + \lambda_D \exp(\gamma)} \cdot \frac{\frac{1-p}{2}}{\frac{p}{2} + \frac{1-p}{2}}$$

$$+ \left(1 - \exp\left(-\lambda_Y \exp(\zeta + \theta)k - \lambda_D \exp(\gamma + \eta)k\right)\right) \cdot \frac{\lambda_Y \exp(\zeta + \theta)}{\lambda_Y \exp(\zeta + \theta) + \lambda_D \exp(\gamma + \eta)} \cdot \frac{\frac{p}{2}}{\frac{p}{2} + \frac{1-p}{2}}$$

$$- \left(1 - \exp\left(-\lambda_Y k - \lambda_D k\right)\right) \cdot \frac{\lambda_Y}{\lambda_Y + \lambda_D} \cdot \frac{\frac{1-p}{2}}{\frac{p}{2} + \frac{1-p}{2}}$$

$$- \left(1 - \exp\left(-\lambda_Y \exp(\theta)k - \lambda_D \exp(\eta)k\right)\right) \cdot \frac{\lambda_Y \exp(\theta)}{\lambda_Y \exp(\theta) + \lambda_D \exp(\eta)} \cdot \frac{\frac{p}{2}}{\frac{p}{2} + \frac{1-p}{2}}.$$

The bias is

$$\mathbb{E}[\hat{\mu}(k)] - \mu(k).$$

A visualization program confirming that the simulated values for the exponential distribution align with the analytical results derived in the supplementary materials.