Data-Generating Process (Simulator)

Subgroups.

$$G \sim \text{Categorical}(\pi_1, \dots, \pi_K), \qquad \Pr(G = g) = \pi_g.$$

Covariates.

$$\begin{aligned}
&L_1 \sim \mathcal{N}(0,1),\\
&L_2 = 0.6L_1 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1),\\
&\text{(Both } L_1, L_2 \text{ are standardized to mean 0, sd 1, and are independent of } G.\text{)}\\
&L_3 \mid i \sim \text{Bernoulli}(p_i), \quad p_i = \text{clip}\big(0.5 + 0.3\sin(\phi_i), 0.1, 0.9\big),
\end{aligned}$$

with ϕ_i a smooth index along the sample; $L_3 \perp G$ by construction.

Treatment assignment (propensity).

$$\Pr(A = 1 \mid L, G) \;=\; \text{logit}^{-1}\big(0.15\,L_1 + 0.15\,L_2 + 0.1\,L_3\big), \quad A \sim \text{Bernoulli}\big(\Pr(A = 1 \mid L, G)\big).$$

Subgroup treatment effects.

$$\tau_g \text{ is fixed for each } g \in \{1, \dots, K\} \quad \text{(from a smooth pattern over } g\text{)}.$$

Event time (AFT–Weibull; potential-outcomes view). Let $k_T > 0$ be the Weibull shape for the event time. Define the AFT mean-time link

$$\log \mu_T(a; L, G) \;=\; \alpha \;+\; b_{L1}L_1 \;+\; b_{L2}L_2 \;+\; b_{L3}L_3 \;+\; b_{G,G} \;+\; \tau_G\, a.$$

Then the potential outcome

$$T^{(a)} \;\sim\; \text{Weibull}\Big(k_T,\; \lambda_T(a; L, G)\Big), \qquad \lambda_T(a; L, G) = \frac{\mu_T(a; L, G)}{\Gamma(1 + 1/k_T)}.$$

Survival and 1-year risk:

$$S_T(t \mid a, L, G) \;=\; \exp\Big(-\big(\tfrac{t\,\Gamma(1+1/k_T)}{\mu_T(a; L, G)}\big)^{k_T}\Big), \qquad \Pr\big(Y_t = 1 \mid a, L, G\big) = 1 - S_T(t \mid a, L, G).$$

Censoring time (informative in A, L). With $k_C > 0$ and

$$\log \mu_C(A, L) \;=\; \alpha_c + c_{L1}L_1 + c_{L2}L_2 + c_{L3}L_3 + c_A A,$$

we simulate

$$C \;\sim\; \text{Weibull}\Big(k_C,\; \lambda_C(A, L)\Big), \qquad \lambda_C(A, L) = \frac{\mu_C(A, L)}{\Gamma(1 + 1/k_C)}.$$

Observed time and indicator.

$$\text{time} \;=\; \min(T, C), \qquad \delta \;=\; \mathbf{1}\{T \le C\}.$$

A 1-year endpoint is $Y_1 = \mathbf{1}\{T \le 1\}$, with partial observation under censoring.

Stan Estimation Model (Parametric AFT–Weibull with Informative Censoring)

Centered linear predictors. Let $\tilde{L}_j = L_j - \bar{L}_j$ and $\tilde{A} = A - \bar{A}$ (sample means). For each i,

$$\eta_{T,i} = \text{Intercept} + b_{L1}\tilde{L}_{1i} + b_{L2}\tilde{L}_{2i} + b_{L3}\tilde{L}_{3i} + b_{G,G_i} + \tau_{G_i}\tilde{A}_i,$$
$$\eta_{C,i} = \text{Intercept}_c + c_{L1}\tilde{L}_{1i} + c_{L2}\tilde{L}_{2i} + c_{L3}\tilde{L}_{3i} + c_A\tilde{A}_i.$$

Weibull parameterization in Stan:

$$T_i \mid A_i, L_i, G_i \sim \text{Weibull}\Big(k_T,\, \exp(\eta_{T,i})/\Gamma(1+1/k_T)\Big), \quad C_i \mid A_i, L_i \sim \text{Weibull}\Big(k_C,\, \exp(\eta_{C,i})/\Gamma(1+1/k_C)\Big).$$

Joint observed-data likelihood. Writing $t_i = \text{time}_i$ and δ_i as above,

$$\log \mathscr{L} \;=\; \sum_{\delta_i=1} \big[\log f_T(t_i \mid A_i, L_i, G_i) + \log S_C(t_i \mid A_i, L_i)\big] \;+\; \sum_{\delta_i=0} \big[\log f_C(t_i \mid A_i, L_i) + \log S_T(t_i \mid A_i, L_i, G_i)\big].$$

Hierarchical priors.

$$b_{G,g} \sim \mathcal{N}(0, \sigma_G^2), \qquad \tau_g \sim \mathcal{N}(\mu_\tau, \sigma_\tau^2),$$
$$\sigma_G, \sigma_\tau \sim \mathcal{N}^+(0, 0.7), \quad k_T \sim \log\mathcal{N}(\log 1.5, 0.3), \quad k_C \sim \log\mathcal{N}(\log 1.2, 0.3),$$

Intercept, b_{Lj}, Intercept_c, c_{Lj}, c_A have weakly-informative Normal priors.

Posterior and reported quantities. The posterior is proportional to likelihood $\times$ priors. In `generated quantities`:

$$\texttt{subgroup\_cate} = \tau_{1:K}, \qquad b\_\text{Intercept} = \text{Intercept} - \big(\bar{L}_1 b_{L1} + \bar{L}_2 b_{L2} + \bar{L}_3 b_{L3} + \bar{A}\,\mu_\tau\big).$$

An overall AFT-scale effect can be formed as

$$\bar{\tau} \;=\; \sum_{g=1}^{K} p_g\,\tau_g, \qquad p_g = \Pr(G=g) \text{ (empirical share).}$$

HDP–HBB g-Computation (Nonparametric Integration over $p(L \mid G)$)

Goal (subgroup-marginal causal curves). For treatment level $a \in \{0,1\}$ and time t,

$$E\big[Y_t^{(a)} \mid G=g\big] \;=\; \int \big(1 - S_T(t \mid a, L, G=g)\big)\; p(L \mid G=g)\, dL.$$

Define the survival-based ATE curve and the risk-difference (RD) curve by

$$\text{ATE}_S^{(g)}(t) \;=\; S_T(t \mid a{=}1, G{=}g) - S_T(t \mid a{=}0, G{=}g), \qquad \text{RD}^{(g)}(t) \;=\; \big(1 - S_1^{(g)}(t)\big) - \big(1 - S_0^{(g)}(t)\big) = S_0^{(g)}(t) - S_1^{(g)}(t).$$

HDP weighting over individuals in subgroup g. Let $\{i : G_i = g\}$ index the N_g members of subgroup g, and let N be the total sample size.

1. Draw global weights:
$$\pi_0 \sim \text{Dirichlet}\Big(\tfrac{\alpha_0}{N}, \ldots, \tfrac{\alpha_0}{N}\Big).$$

2. Restrict and renormalize to subgroup atoms:
$$b_g(i) \propto \pi_0(i) \text{ for } i \in \{G_i = g\}, \qquad \sum_{i \in g} b_g(i) = 1.$$

3. Draw subgroup weights:
$$w_g \sim \text{Dirichlet}\big(\alpha_g\, b_g\big), \quad \sum_{i \in g} w_g(i) = 1.$$

HDP–HBB g-computation of subgroup survival. For a fixed posterior draw of (η_T, k_T), compute individual survivals

$$S_i^{(a)}(t) \;=\; \Pr\big(T_i^{(a)} > t \,\big|\, L_i, G_i{=}g\big) \;=\; \exp\Big(-\big(\tfrac{t\,\Gamma(1+1/k_T)}{\mu_{T,i}(a)}\big)^{k_T}\Big), \quad \mu_{T,i}(a) = \exp\big(\eta_{T,i}(a)\big).$$

Aggregate with HDP weights:

$$S_a^{(g)}(t) \;=\; \sum_{i \in g} w_g(i)\, S_i^{(a)}(t), \qquad \text{ATE}_S^{(g)}(t) = S_1^{(g)}(t) - S_0^{(g)}(t), \quad \text{RD}^{(g)}(t) = S_0^{(g)}(t) - S_1^{(g)}(t).$$

Repeating over posterior draws and HDP weight draws yields posterior summaries (means/CrIs) for the curves.

Overall effects (optional). With subgroup prevalence p_g, combine to an overall curve:

$$\text{RD}^{\text{overall}}(t) \;=\; \sum_{g=1}^{K} p_g\, \text{RD}^{(g)}(t).$$

Notes. In the current simulator, $(L_1, L_2, L_3) \perp G$, so $p(L \mid G = g) = p(L)$ and HDP weights tend toward uniform within subgroups; the HDP–HBB procedure nonetheless provides a principled nonparametric integration (and enables borrowing strength if subgroup covariate distributions differ in other datasets).