Heteroskedastic BART and time-to-event outcomes

Rodney Sparapani
Associate Professor of Biostatistics
Medical College of Wisconsin

September 17, 2025

This research funded, in part, by the US Office of Naval Research

Heteroskedastic BART (HBART)

Pratola, Chipman, George & McCulloch 2020 *JCGS*

$$y_i = f(x_i) + s(x_i)\epsilon_i \qquad \epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0,\ \sigma^2)$$

$$f \overset{\text{prior}}{\sim} \mathrm{BART}\ (H, \mu, \kappa, \tau, \alpha, \beta)$$

$$s^2 \overset{\text{prior}}{\sim} \mathrm{HBART}\ (\widetilde{H}, \tilde{\lambda}, \tilde{\nu}, \tilde{\alpha}, \tilde{\beta})$$

$$s^2(x_i) \equiv \prod_{h=1}^{\widetilde{H}} g(x_i; \widetilde{\mathcal{T}}_h, \widetilde{\mathcal{M}}_h) \qquad \widetilde{H} \approx H/5$$

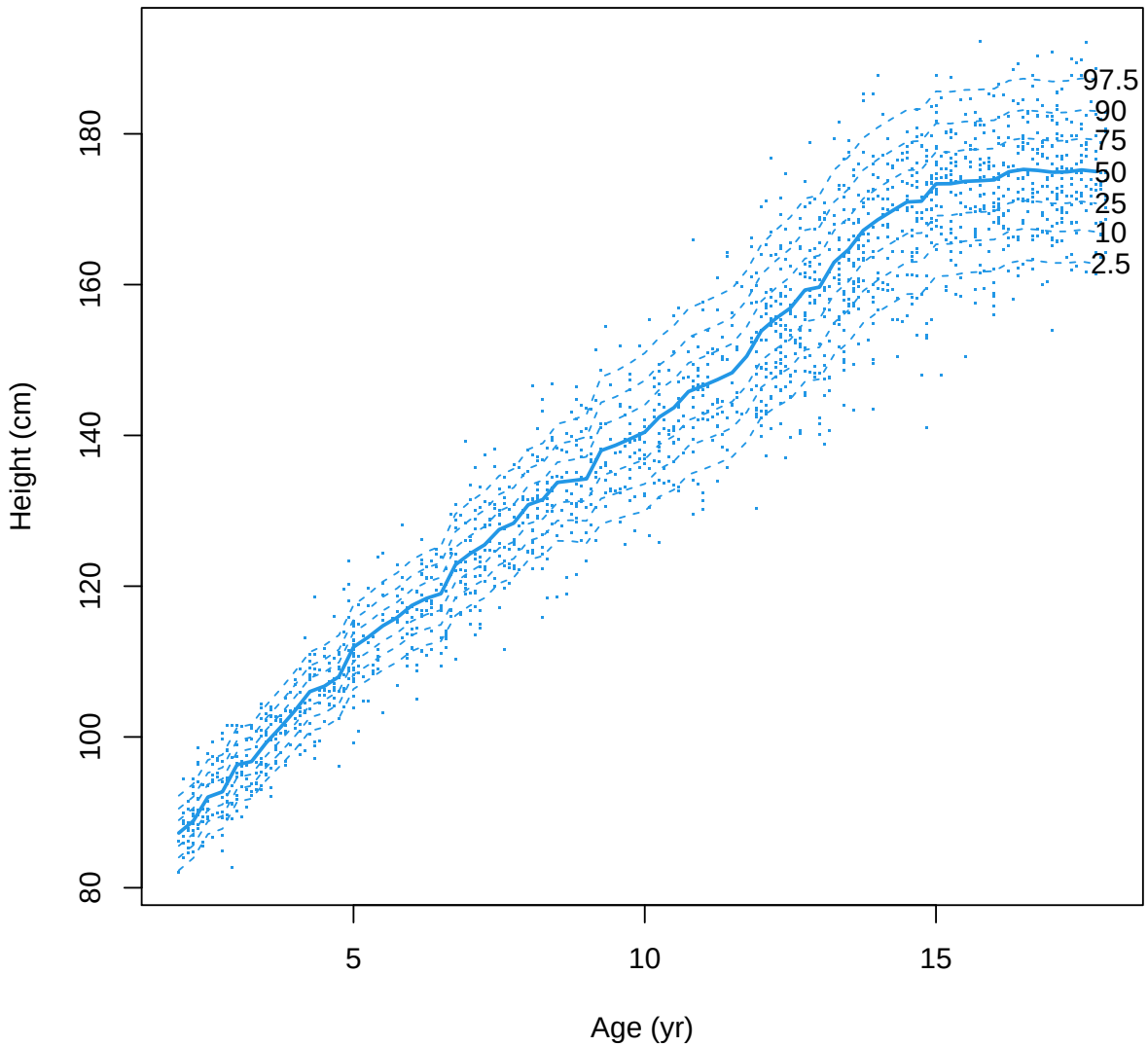
$$\sigma^2_{hl} | \widetilde{\mathcal{T}}_h \overset{\text{prior}}{\sim} \lambda\nu\chi^{-2}(\nu) \text{ leaves of } \widetilde{\mathcal{T}}_h \qquad \lambda = \tilde{\lambda}^{1/\widetilde{H}}$$

$$\overset{\text{prior}}{\sim} \mathrm{Gamma}^{-1}(\nu/2,\ \lambda\nu/2) \qquad \mathrm{E}\left[\sigma^2_{hl}\right] = \lambda\nu/(\nu-2)$$

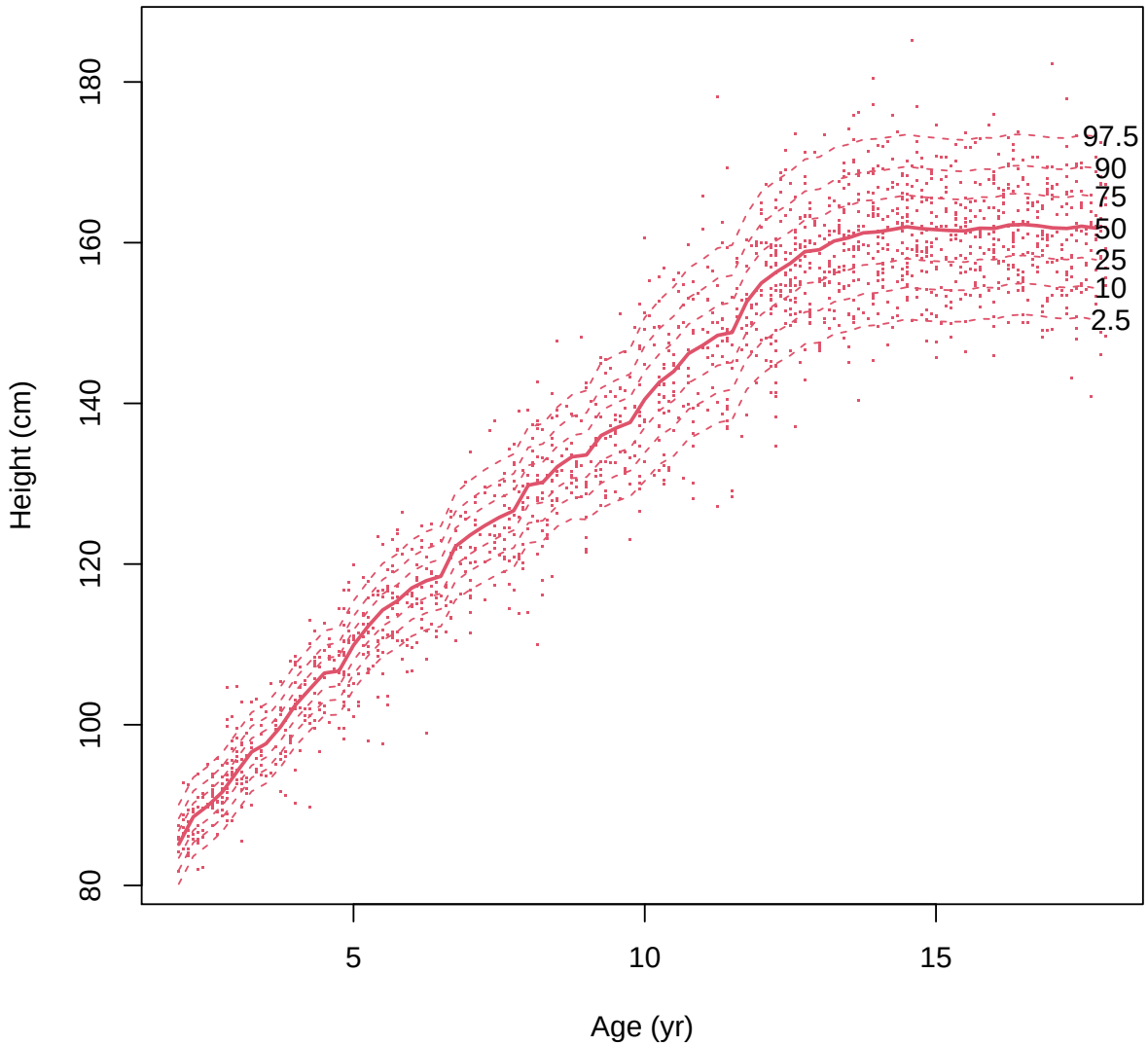
$$\in \widetilde{\mathcal{M}}_h \qquad \nu = 2\left[1 - \left(1 - \frac{2}{\tilde{\nu}}\right)^{1/\widetilde{H}}\right]^{-1}$$

Marginal effect of age: HBART predictions for M

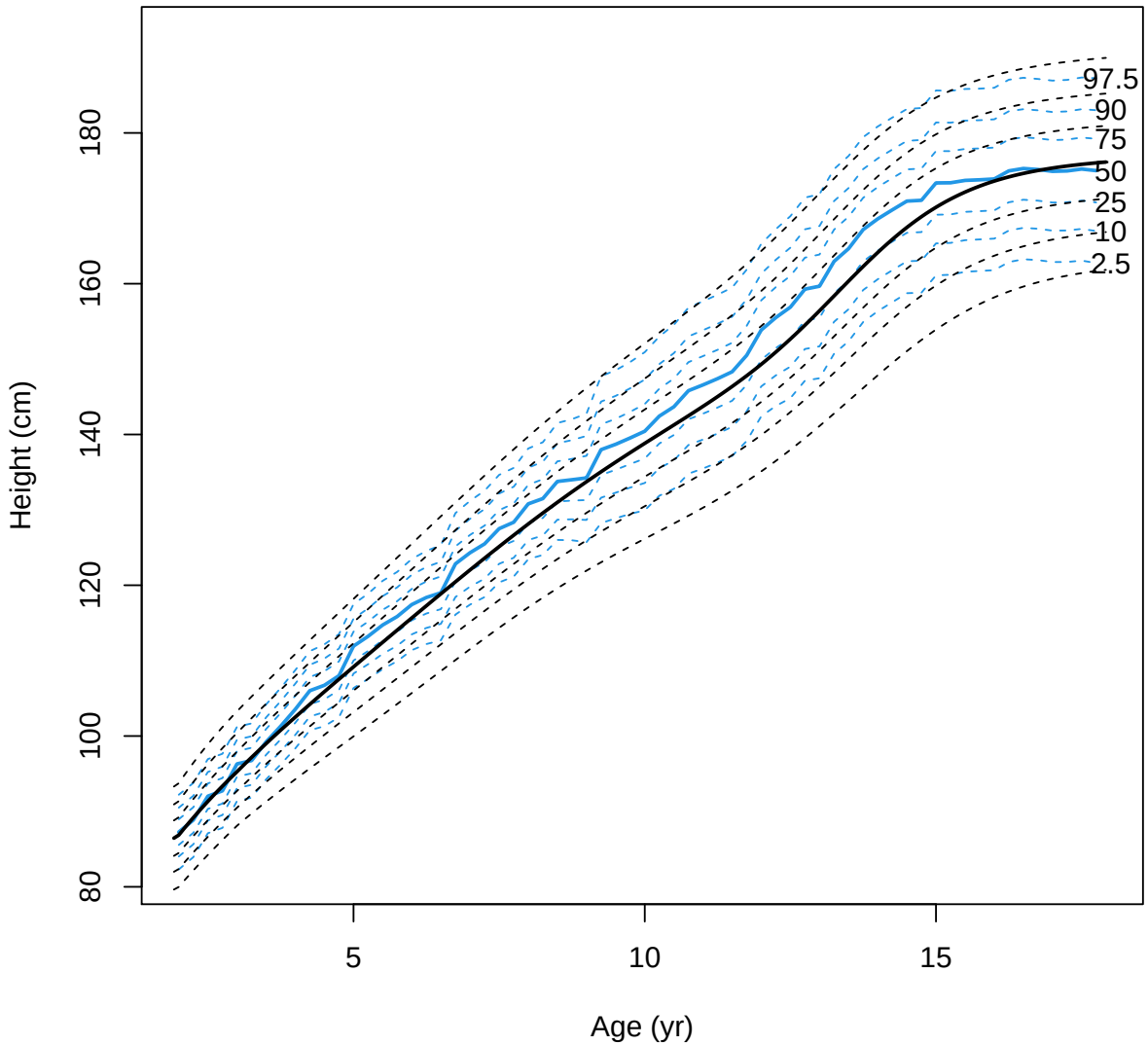
$$H = 300, \widetilde{H} = 60, \texttt{numcut} = 200$$



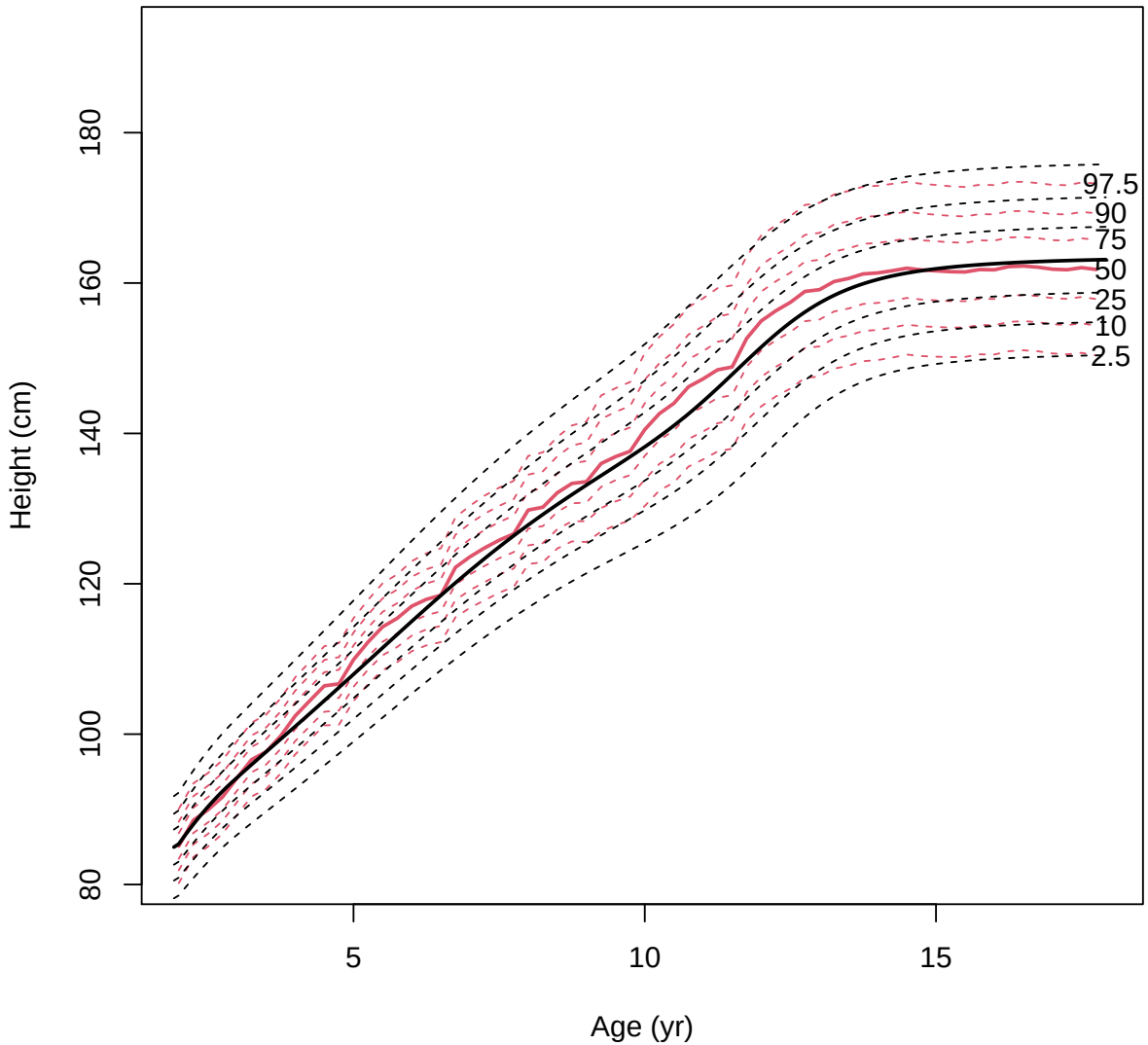
Marginal effect of age: HBART predictions for F



Marginal effect of age: HBART vs. CDC for M



Marginal effect of age: HBART vs. CDC for F



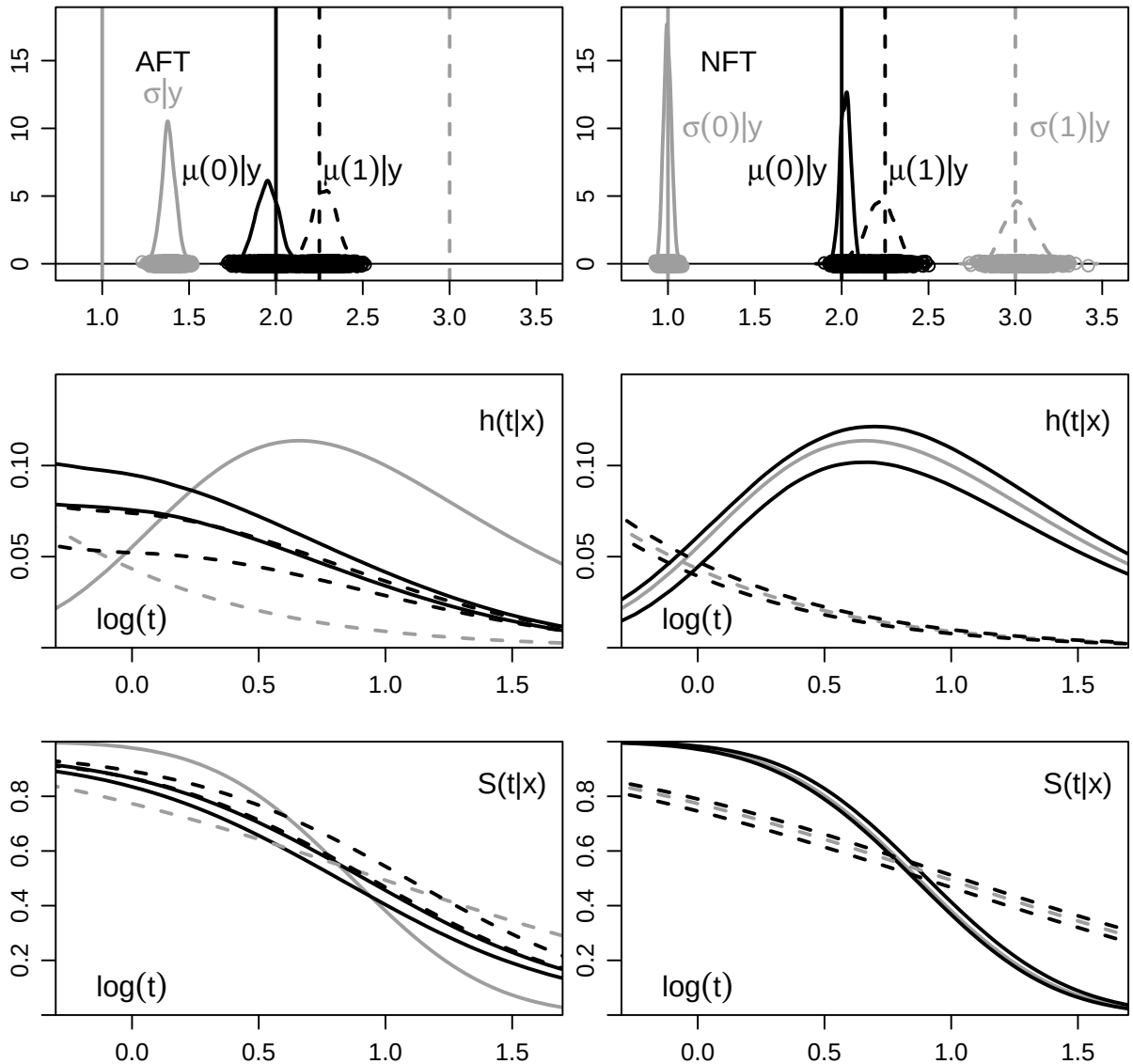
Personalized Hematopoietic Stem Cell Transplant (HSCT)

- HSCT is a treatment for white blood/bone marrow cancers
- Here we are concerned with unrelated donors that are human leukocyte antigen (HLA) 8/8 matched to the recipients transplanted from 2016:2019
- Goal: optimal donor matching for better recipient outcomes
- The outcome here is time to an event, i.e., event-free survival with both right and left censoring
- Events include death, relapse, graft failure/rejection or moderate/severe chronic graft vs. host disease (GVHD): whichever comes first
- There are $\boldsymbol{P = 45}$ covariates that may have an impact
- $\mathbf{5}$ are donor-related characteristics: age, sex/childbearing, HLA DPB1 match, HLA DQB1 match and CMV match
- We wanted to *learn* the (likely complex) functional relationship between these covariates and the outcome with BART
- The cohort has $\mathbf{10016}$ for training and $\mathbf{1802}$ for validation
- A bit too large for our Discrete Time BART approach
- For this application, we developed NFT BART methodology

Methodological/Computational Pros and Cons

	Comparison of BART survival analysis methods				
Property	Hier-archical	Discrete Time	AFT	Mod-ulated	NFT
Flexible assumptions	Con	Pro	Con	Pro	Pro
Non-parametric	Con	Pro	Pro	Pro	Pro
Left-censoring	Con	Con	Pro	Con	Pro
Time-dep. covariates	Con	Pro	Con	Pro	Con
Friendly to compute	Pro	Con	Pro	Con	Pro
First-author Year	Bonato 2011	Sparapani 2016	Henderson 2018	Linero 2021	Sparapani 2023

Two groups: AFT BART vs. NFT BART



Bayesian Additive Regression Trees (BART)
NFT notation

Sparapani, Logan, Laud & McCulloch 2023 *Biometrics*

$$\mu \overset{\text{prior}}{\sim} \mathbf{BART}\ (a = 0.95, b = 2, H = 200, \kappa = 2, \tilde{\mu} = \bar{y})$$

$$y_i = \mu(x_i) + \epsilon_i \text{ where } \epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}\left(0,\ \sigma^2\right)$$

$$\mu(x_i) \equiv \tilde{\mu} + \sum_h g(x_i; \mathcal{T}_h, \mathcal{M}_h)$$

the **BART** prior implies the following priors (among others)

$$\mu_{hl}|\mathcal{T}_h \overset{\text{prior}}{\sim} \mathrm{N}\left(0,\ \frac{0.25\,\mathrm{range}(y)^2}{H\kappa^2}\right) \text{ leaves of } \mathcal{T}_h$$

$$\in \mathcal{M}_h$$

$$\sigma^2 \overset{\text{prior}}{\sim} \lambda\nu\chi^{-2}(\nu)$$

Heteroskedastic BART (HBART)
NFT notation

Pratola, Chipman, George & McCulloch 2019 *JCGS*

$$\mu \overset{\text{prior}}{\sim} \textbf{BART}\ (a, b, H = 200, \kappa = 5, \tilde{\mu})$$

$$\sigma^2 \overset{\text{prior}}{\sim} \textbf{HBART}\ (\tilde{a} = 0.95, \tilde{b} = 2, \widetilde{H} = 40, \tilde{\lambda}, \tilde{\nu})$$

$$y_i = \mu(x_i) + \epsilon_i \text{ where } \epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}\left(0,\ \sigma^2(x_i)\right)$$

$$\sigma^2(x_i) \equiv \prod_{h=1}^{\widetilde{H}} g(x_i; \widetilde{\mathcal{T}}_h, \widetilde{\mathcal{M}}_h) \text{ where } \widetilde{H} \approx H/5$$

the **HBART** prior implies the following priors (among others)

$$\sigma^2_{hl} | \widetilde{\mathcal{T}}_h \overset{\text{prior}}{\sim} \lambda\nu\chi^{-2}(\nu) \text{ leaves of } \widetilde{\mathcal{T}}_h$$

$$\in \widetilde{\mathcal{M}}_h$$

The Accelerated Failure Time (AFT) model: part 1

- Time-to-event data notation: (t_i, δ_i) $i = 1, \ldots, N$ subjects
  if $\delta_i = 0$, then t_i is a right censoring time
  else if $\delta_i = 1$, then a failure time
  else if $\delta_i = 2$, then left censoring
- How is failure time explained by a vector of covariates x_i?
- take logarithms $y_i = \log t_i$ and use a linear model (Con)
  $y_i = [1, x_i']\beta + \sigma\epsilon_i = \beta_0 + x_i'\beta_x + \sigma\epsilon_i$
  where β and σ are unknown coefficients to be estimated
  with $\epsilon_i \overset{\text{iid}}{\sim} F_\epsilon(\mu_\epsilon = 0, \sigma_\epsilon^2 = 1)$
  which is typically parametric (Con)

The Accelerated Failure Time (AFT) model: part 2

- Consider a *baseline* survival function for a *standard* subject where the covariates are all zero, i.e., $S_0(t) = S(t|x=0)$.
- We can define the survival function for any given subject with a standard subject by accelerating, or decelerating, failure time

$$\begin{aligned} S(t|x_i) =& \mathrm{P}[s_i > t|x_i] = \mathrm{P}[y_i > \log t|x_i] \\ =& \mathrm{P}\left[\beta_0 + x_i'\beta_x + \sigma\epsilon_i > \log t|x_i\right] \\ =& \mathrm{P}\left[\beta_0 + \sigma\epsilon_i > \log t - x_i'\beta_x|x_i\right] \\ =& S_0(t\exp\{-x_i'\beta_x\}) \end{aligned}$$

- however, AFT is a precarious restrictive assumption (Con)
  $S(t|x) = \mathrm{P}[\log s > \log t] = 1 - F_\epsilon\left(\log t; x'\beta, \sigma^2\right)$
  the covariates can only explain a log-linear location shift

Survival analysis with AFT BART

NFT notation

Henderson, Louis et al. 2018 *Biostatistics*

- $y_i = \mu(x_i) + \epsilon_i$ where $\epsilon_i | \mu_i \sim \mathrm{N}\left(\mu_i,\ \sigma^2\right)$: Pro
  $\mu \overset{\text{prior}}{\sim}$ **BART**
- To ensure identifiability, constrain $\frac{1}{N} \sum_i \mu_i = 0$
- $\mu_i | G \sim G$
  $G | \alpha \overset{\text{prior}}{\sim} \mathrm{DP}\left(\alpha,\ F_0\right)$
- $S(t, x) = 1 - \frac{1}{N} \sum_i \Phi\left(\frac{\log t - \mu_i - \mu(x)}{\sigma}\right)$
  Con: the covariates still only explain a log-linear location shift

Survival analysis with NFT BART

Sparapani et al. 2023 *Biometrics*

- $y_i = \mu(x_i) + \epsilon_i$ where $\epsilon_i | (\mu_i, \sigma_i) \sim \mathrm{N}\left(\mu_i,\ \sigma_i^2 \sigma^2(x_i)\right)$: Pro
  $\mu \overset{\text{prior}}{\sim} \mathbf{BART}$
  $\sigma^2 \overset{\text{prior}}{\sim} \mathbf{HBART}$
- To ensure identifiability: $\frac{1}{N}\sum_i \mu_i = 0$ and $\frac{1}{N}\sum_i \sigma_i^2 = 1$
- if $\delta_i = 1$, then $y_i = \log t_i$
  else draw

$$y_i \sim \mathrm{N}\left(\mu_i + \mu(x_i),\ \sigma_i^2 \sigma^2(x_i)\right) \begin{cases} \mathrm{I}(\log t_i, \infty) & \text{if } \delta_i = 0 \\ \mathrm{I}(-\infty, \log t_i) & \text{if } \delta_i = 2 \end{cases}$$

- $(\mu_i, \sigma_i) | G \sim G$
  $G | \alpha \overset{\text{prior}}{\sim} \mathrm{DP}\left(\alpha,\ F_0\right)$
- $S(t, x) = 1 - \frac{1}{N}\sum_i \Phi\left(\frac{\log t - \mu_i - \mu(x)}{\sigma_i \sigma(x)}\right)$
  Pro: the covariates can explain a location shift and rescaling!

Dirichlet Process Mixtures (DPM): infinite mixtures

Ferguson 1973 & Antoniak 1974 *Annals of Statistics*;
Escobar & West 1995 *JASA*; Neal 2000 *JCGS*
DPM-like finite mixture clustering: Miller & Harrison 2017 *JASA*

$$
\begin{aligned}
y_i|\theta_i &\sim F(\theta_i) && \text{usual notation}\\
&\text{where } i = 1,\ldots,N &&\\
y_i|\theta^*_{c_i} &\sim F(\theta^*_{c_i}) && \text{ephemeral clusters}\\
&\text{where } c_i \in \{1,\ldots,k\} && k \text{ is random}\\
\theta_i|G &\sim G && \text{nonparametric (Pro)}\\
G|\alpha &\overset{\text{prior}}{\sim} \text{DP}(\alpha,\ F_0) && G \text{ "centered" on } F_0\\
\alpha &\overset{\text{prior}}{\sim} \text{Gamma}(a,\ b) && \text{concentration parameter}\\
&\propto k &&\\
\theta_1 &\sim F_0 && \text{integrating over } G\\
\theta_2|\theta_1 &\sim \frac{1}{1+\alpha}\delta_K(\theta_1) + \frac{\alpha}{1+\alpha}F_0 && \text{mixture}
\end{aligned}
$$

Constrained DPM

Yang, Dunson & Baird 2010
Computational Statistics & Data Analysis

- How do we constrain $\frac{1}{N}\sum_i \boldsymbol{\mu_i} = \mathbf{0}$?
- Simply sample $(\tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\sigma}}_i)|\boldsymbol{G} \sim \boldsymbol{G}$ as usual
  Let $\tilde{\boldsymbol{\mu}}_0 = \frac{1}{N}\sum_i \tilde{\boldsymbol{\mu}}_i$
  And $\boldsymbol{\mu_i} = \tilde{\boldsymbol{\mu}}_i - \tilde{\boldsymbol{\mu}}_0$
- Similarly, if we need to constrain $\frac{1}{N}\sum_i \boldsymbol{\sigma_i^2} = \mathbf{1}$
  Let $\tilde{\boldsymbol{\sigma}}_0 = \sqrt{\frac{1}{N}\sum_i \tilde{\boldsymbol{\sigma}}_i^2}$
  And $\boldsymbol{\sigma_i} = \tilde{\boldsymbol{\sigma}}_i / \tilde{\boldsymbol{\sigma}}_0$

Low Information Omnibus (LIO) Dirichlet Process Mixtures prior hierarchy

Shi, Martens, Banerjee, Laud 2018 *Bayesian Analysis*
Sparapani et al. 2023 *Biometrics*

- With either DPM or Constrained DPM
- For convenience, re-parameterize in terms of $\tau_i = \sigma_i^{-2}$
  $F_0(\mu_0, k_0, a_0, b_0)$ is a Normal-Gamma prior
  $[\mu_i, \tau_i | k_0, b_0] = [\tau_i | b_0]\,[\mu_i | \tau_i, k_0]$
  with $\mu_i | \tau_i, k_0 \overset{\text{prior}}{\sim} \mathrm{N}\left(\mu_0,\ (\tau_i k_0)^{-1}\right)$
  and $\tau_i | b_0 \overset{\text{prior}}{\sim} \mathbf{Gamma}\,(a_0,\ b_0)$
- NFT LIO prior parameter settings
  $\mu_0 = 0$, $k_0 \overset{\text{prior}}{\sim} \mathbf{Gamma}\,(1.5,\ 7.5)$
  $a_0 = 3$, $b_0 \overset{\text{prior}}{\sim} \mathbf{Gamma}\,(2,\ 1)$

NFT model: prediction intervals

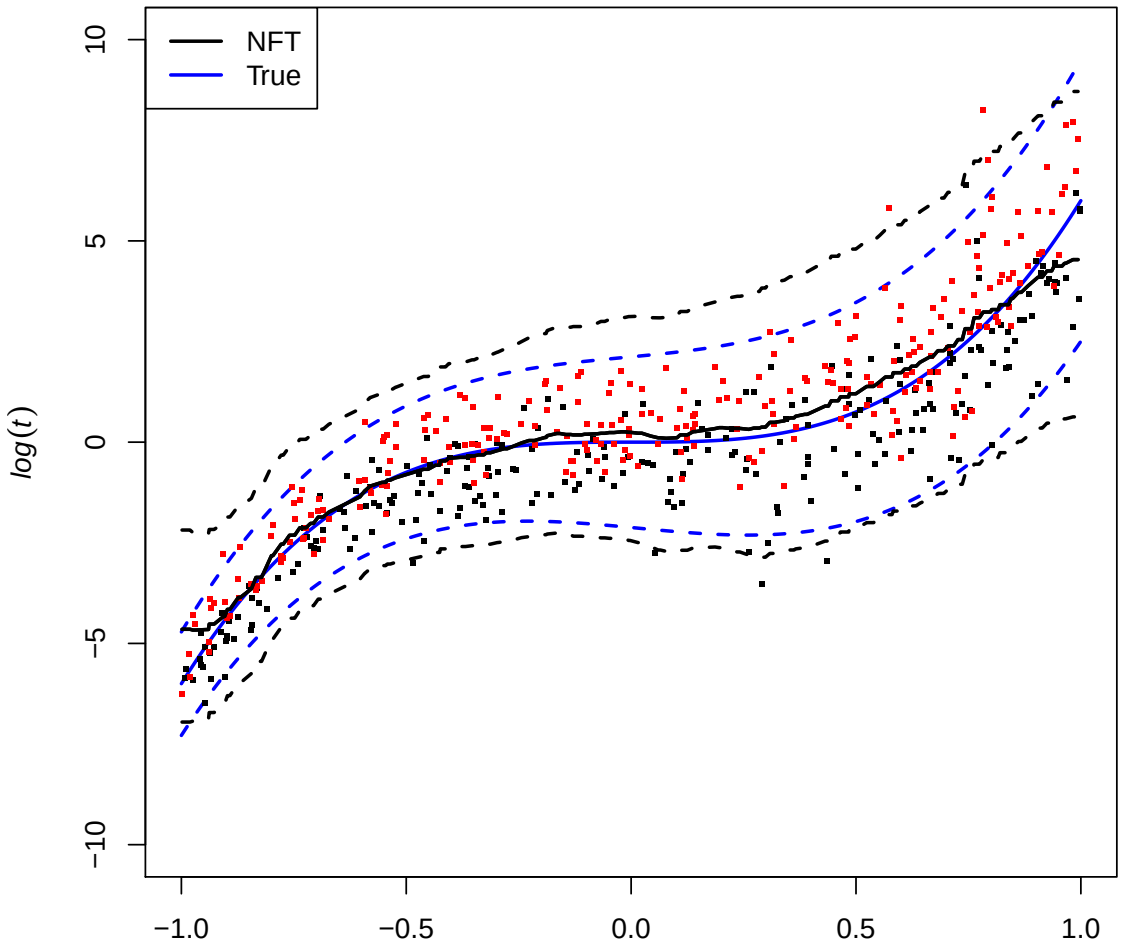
- $\log t_i = y_i = \mu(x_i) + \epsilon_i$ where $\epsilon_i \sim \mathrm{N}\left(\mu_i,\ \sigma_i^2 \sigma^2(x_i)\right)$
  To ensure identifiability: $\frac{1}{N}\sum_i \mu_i = 0$ and $\frac{1}{N}\sum_i \sigma_i^2 = 1$
- $F_\epsilon = \frac{1}{N}\sum_i \mathrm{N}\left(\mu_i,\ \sigma_i^2\right)$: nonparametric mixture of Normals
- $(1-\alpha)\times 100\%$ Prediction Interval
  $(\mu(x) + c_{\alpha/2}\sigma(x),\ \mu(x) + c_{1-\alpha/2}\sigma(x))$
  where $c_\pi = F_\epsilon^{-1}(\pi)$

NFT scenario $t(16)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$,

$\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(16)$

and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 84.8\%$ uncensored, $R^2 = 85.1\%$ censored

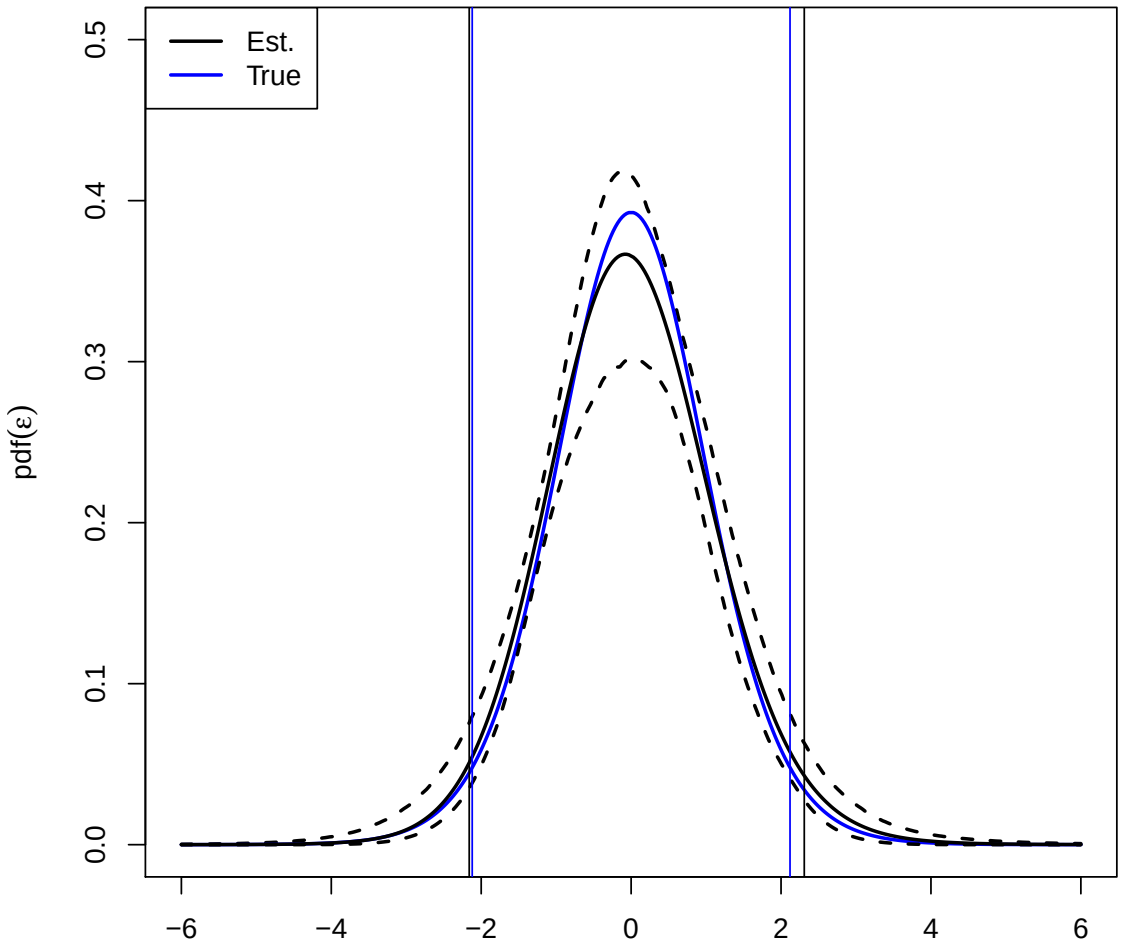


NFT scenario $t(16)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$,

$\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(16)$

and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 84.8\%$ uncensored, $R^2 = 85.1\%$ censored

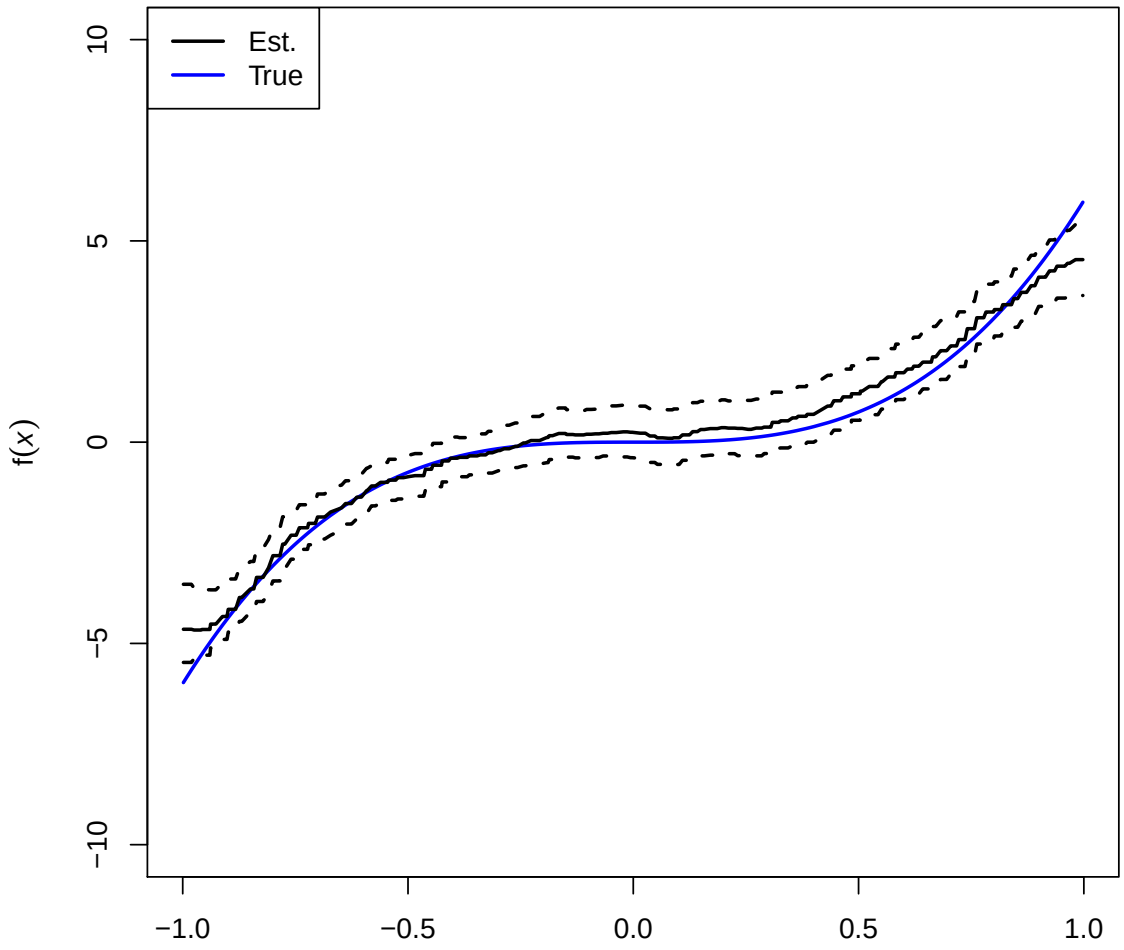


NFT scenario $t(16)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$,

$\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(16)$

and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 84.8\%$ uncensored, $R^2 = 85.1\%$ censored

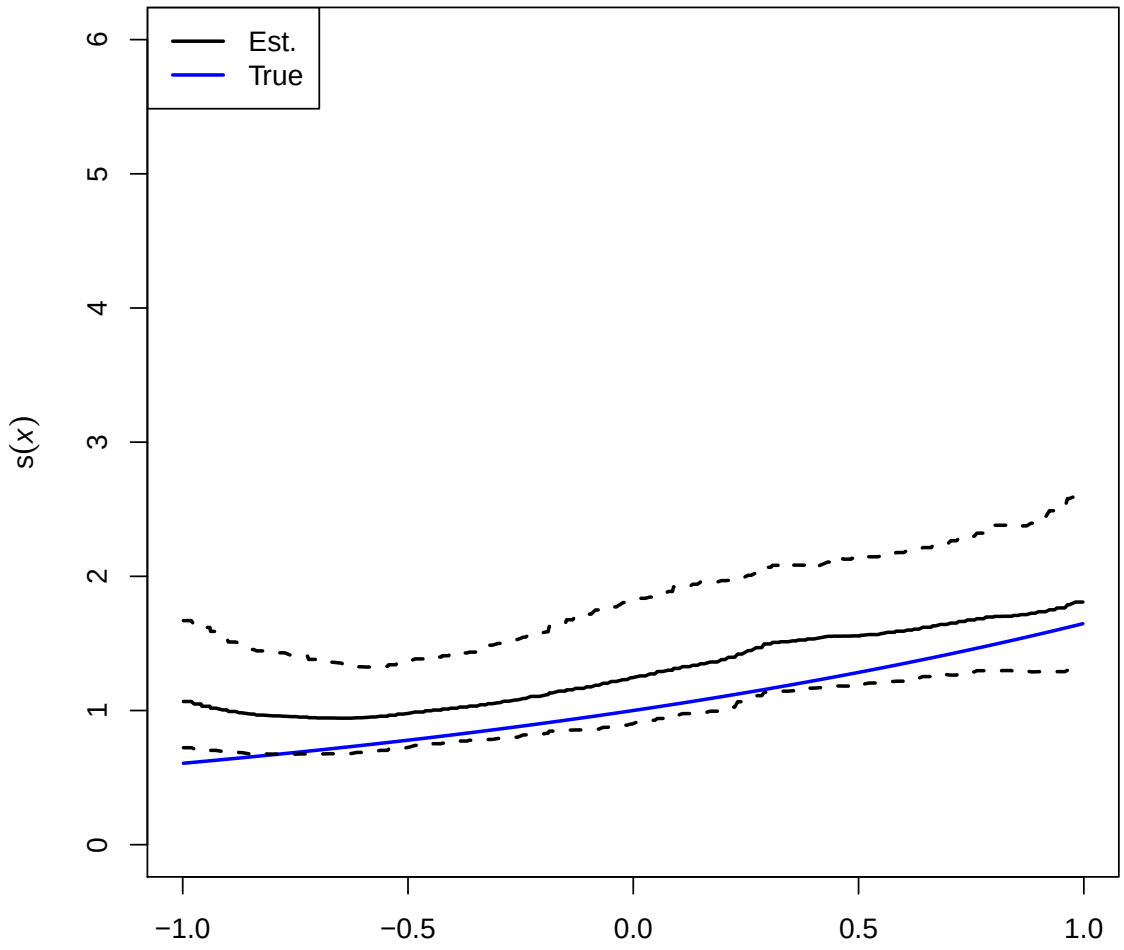


NFT scenario $t(16)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$,

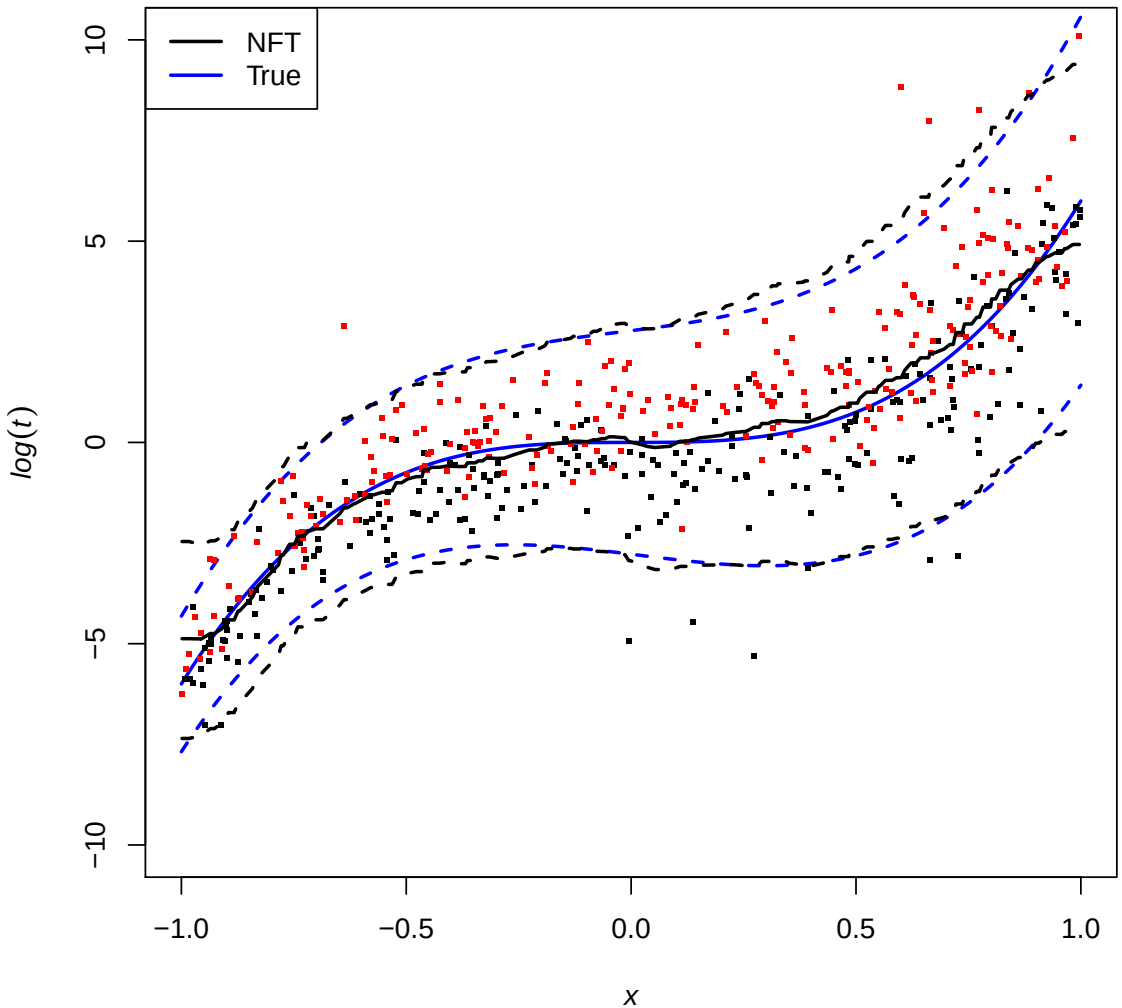
$\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(16)$

and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 84.8\%$ uncensored, $R^2 = 85.1\%$ censored



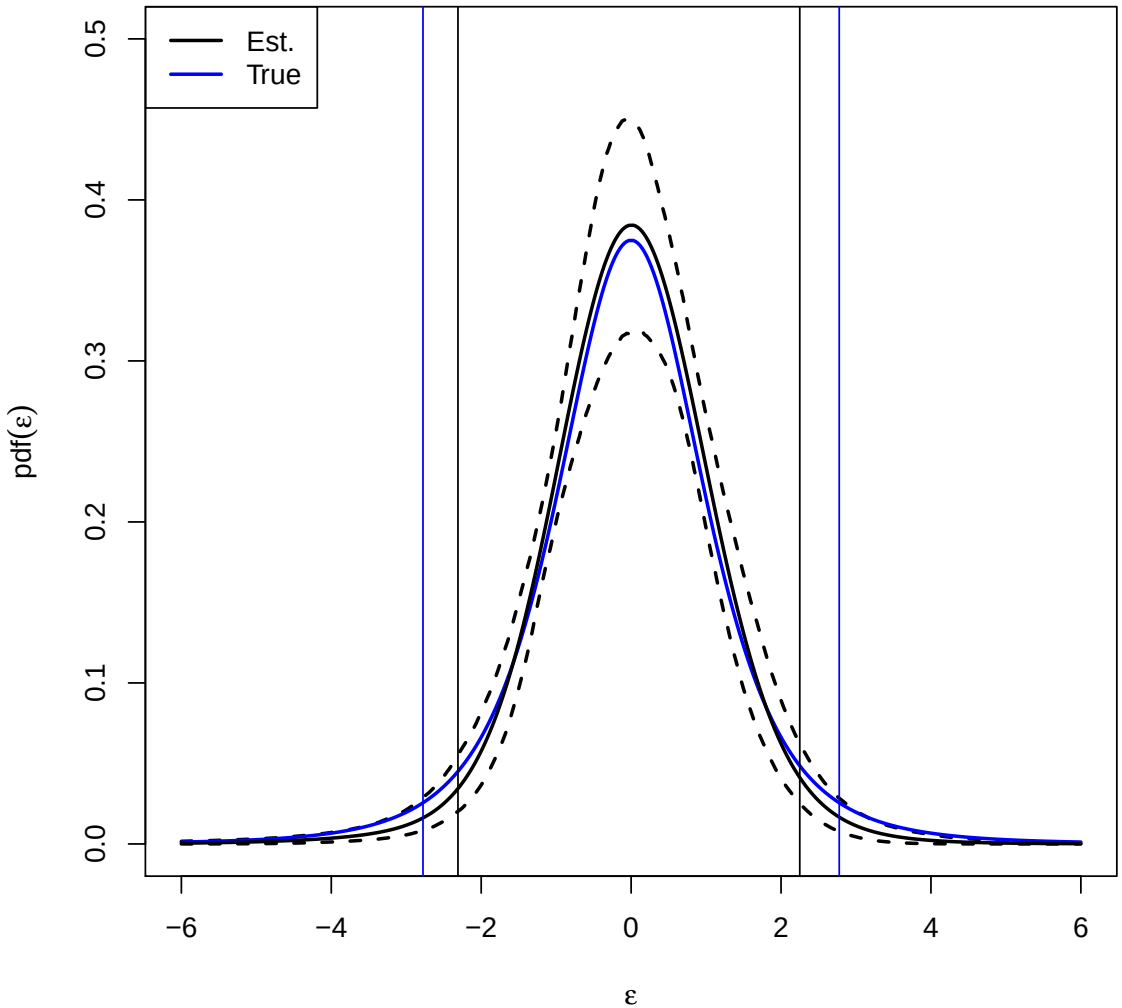
NFT scenario $t(4)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$, $\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(4)$ and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 80.7\%$ uncensored, $R^2 = 78.3\%$ censored



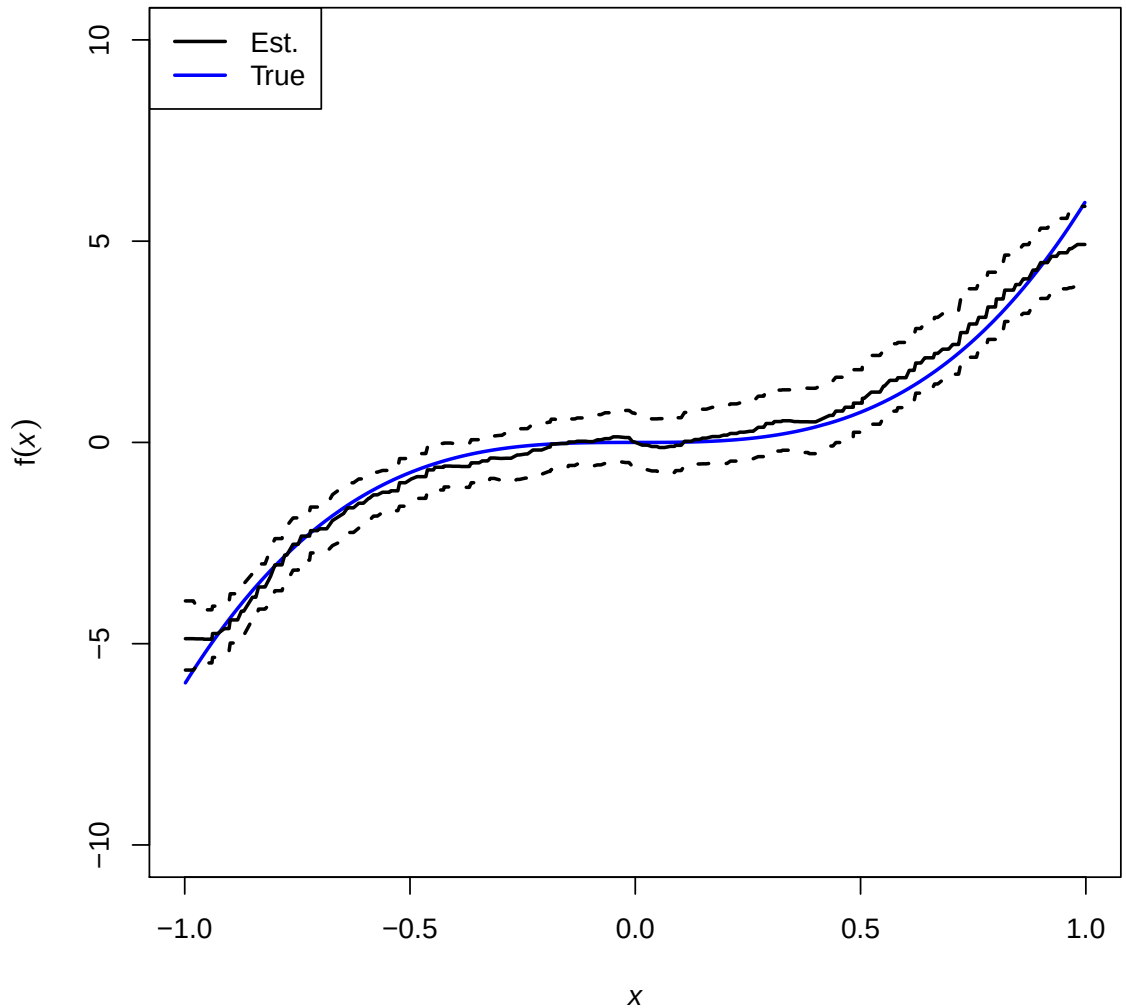
NFT scenario $t(4)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$, $\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(4)$ and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 80.7\%$ uncensored, $R^2 = 78.3\%$ censored



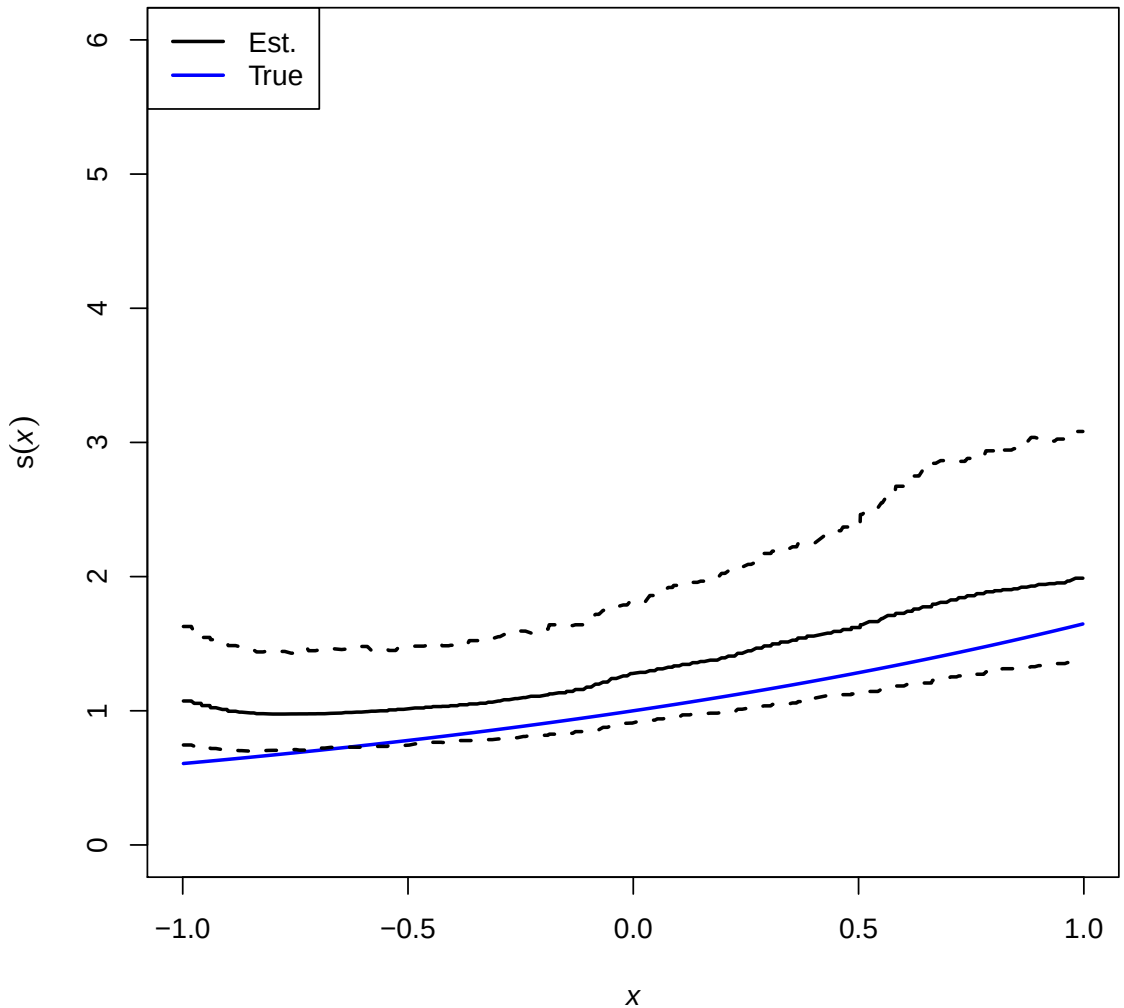
NFT scenario $t(4)$: $N = 500$ with 50% censoring

$f(x) = 6x^3$, $s(x) = \exp 0.5x$, $\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(4)$ and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 80.7\%$ uncensored, $R^2 = 78.3\%$ censored



NFT scenario $t(4)$: $N = 500$ with 50% censoring

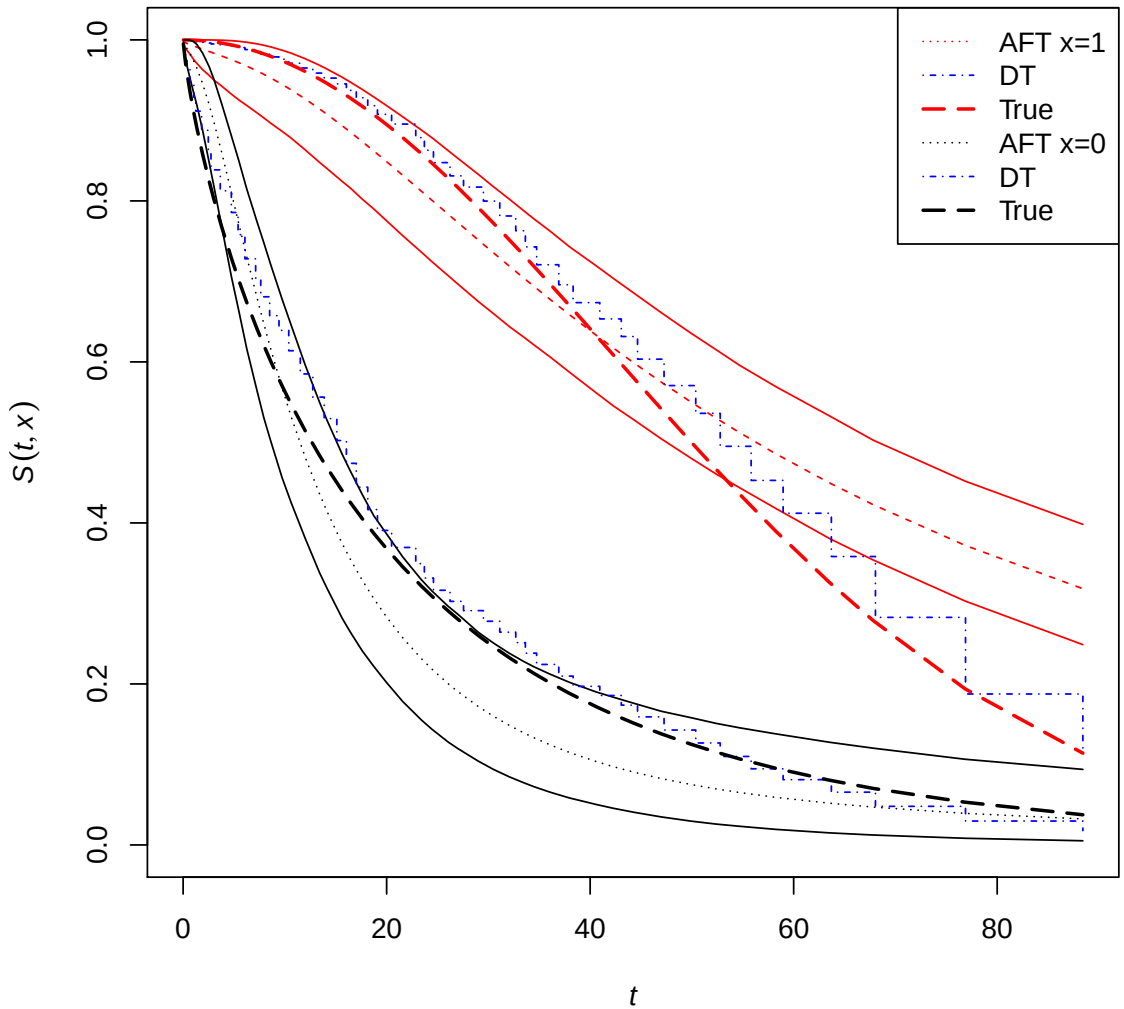
$f(x) = 6x^3$, $s(x) = \exp 0.5x$, $\log t = f(x) + s(x)\epsilon$ where $\epsilon \sim t(4)$ and $x \sim \mathrm{U}(-1, 1)$: $R^2 = 80.7\%$ uncensored, $R^2 = 78.3\%$ censored



Neither AFT nor NFT scenario: AFT failure!

$N = 500$ with 50% censoring

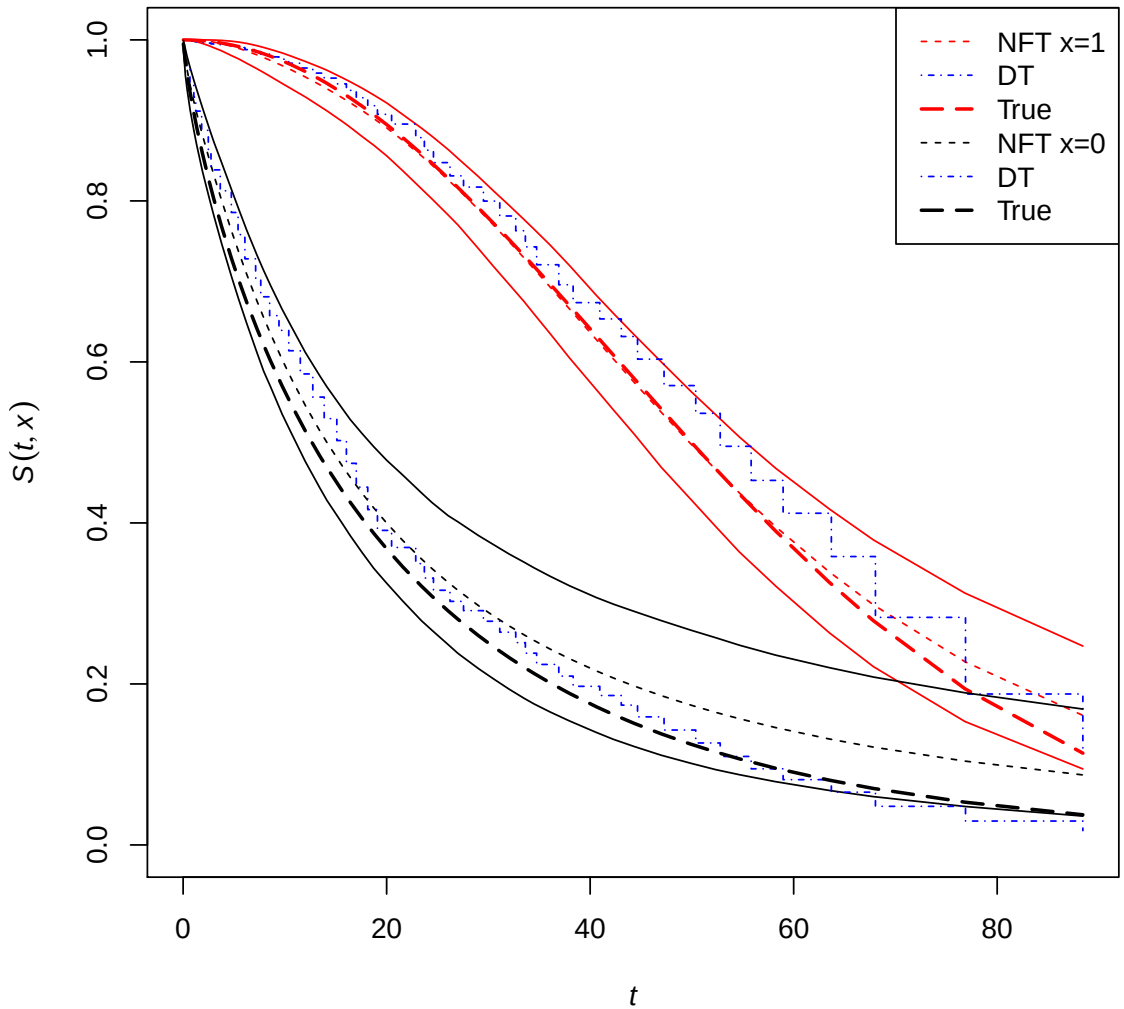
$\mathrm{Wei}\,(0.8 + 1.2x, 20 + 40x)$ where $x \sim \mathrm{B}(0.5)$



Neither AFT nor NFT scenario: NFT success!

$N = 500$ with 50% censoring

$\mathrm{Wei}\,(0.8 + 1.2x, 20 + 40x)$ where $x \sim \mathrm{B}(0.5)$



NFT BART posterior inference: the survival and distribution functions

$$S_m(t|x) = 1 - F_m(t|x)$$

$$F_m(t|x) = \int \Phi\left\{\frac{\log t - \mu_* - \mu_m(x)}{\sigma_* \sigma_m(x)}\right\} G_m(\mathrm{d}\mu_*, \mathrm{d}\sigma_*)$$

$$= \sum_{j=1}^{\infty} \omega_j \Phi\left\{\frac{\log t - \mu_j^* - \mu_m(x)}{\sigma_j^* \sigma_m(x)}\right\}$$

$$\approx \sum_{j=1}^{K_m} \omega_{jm} \Phi\left\{\frac{\log t - \mu_{jm}^* - \mu_m(x)}{\sigma_{jm}^* \sigma_m(x)}\right\}$$

where $(\mu_{jm}^*, \sigma_{jm}^*)$ are from the training set

NFT BART posterior inference: the survival function

$$\widehat{S}(t|x) = M^{-1} \sum_{m} S_m(t|x)$$

$1 - 2\pi$ level credible intervals from π and $1 - \pi$ quantiles $(\widehat{S}_\pi(t|x), \widehat{S}_{1-\pi}(t|x))$ such that $\widehat{S}_p(t|x) = S_{m_p}(t|x)$ where m_p corresponds to the $p = \pi$ or $p = 1 - \pi$

NFT BART posterior inference: the hazard and density functions

$$h_m(t|x) = f_m(t|x)/S_m(t|x)$$

$$f_m(t|x) = \int \frac{\phi\left\{\frac{\log t - \mu_* - \mu_m(x)}{\sigma_* \sigma_m(x)}\right\}}{t\sigma_* \sigma_m(x)} G_m(\mathrm{d}\mu_*, \mathrm{d}\sigma_*)$$

$$= \sum_{j=1}^{\infty} \frac{\omega_j \phi\left\{\frac{\log t - \mu_j^* - \mu_m(x)}{\sigma_j^* \sigma_m(x)}\right\}}{t\sigma_j^* \sigma_m(x)}$$

$$\approx \sum_{j=1}^{K_m} \frac{\omega_{jm} \phi\left\{\frac{\log t - \mu_{jm}^* - \mu_m(x)}{\sigma_{jm}^* \sigma_m(x)}\right\}}{t\sigma_{jm}^* \sigma_m(x)}$$

where $(\mu_{jm}^*, \sigma_{jm}^*)$ are from the training set

NFT BART posterior inference: marginal effects by Friedman's partial dependence function

Friedman 2001 *Annals of Statistics*

- The covariates of interest are fixed at settings of interest: a single setting denoted $\boldsymbol{x_A}$
- The complement take on the observed values found in the training data set denoted $\boldsymbol{x_{iB}}$ for subject $\boldsymbol{i}$
- So the setting for all covariates denoted as $\boldsymbol{(x_A, x_{iB})}$

$$F_{Am}(t|x_A) = N^{-1} \sum_i \Phi\left(\frac{\log t - \mu_{im} - \mu_m(x_A, x_{iB})}{\sigma_{im}\sigma_m(x_A, x_{iB})}\right)$$

where (μ_{im}, σ_{im}) are from the training set

$$\widehat{S}_A(t|x_A) = 1 - M^{-1} \sum_m F_{Am}(t|x_A)$$

Restricted Mean Survival Time (RMST)

- Nonparametric estimates of RMST are an alternative measure that may be more interpretable than survival probabilities Royston & Parmar 2013 *BMC Medical Research Methodology*
- The mean survival time is an infinite integral: $\int_0^\infty S(t|x)\mathrm{d}t$
- This is impractical, i.e., the observation period cannot possibly be lengthened until every participant passes away
- Rather, let's limit the observation period to, τ, a clinically meaningful follow-up period: $\mathbf{RMST}(\tau) = \int_0^\tau S(t|x)\mathrm{d}t$
- Consider a log-Normal time-to-event, e^y, where $y \sim \mathrm{N}\left(\mu,\ \sigma^2\right)$

$$\mathbf{E}\left[\mathrm{e}^y | y < \log(\tau)\right] = \mathrm{e}^{\mu+\sigma^2/2}\Phi\left(\frac{\log(\tau)-(\mu+\sigma^2)}{\sigma}\right)$$

Restricted Mean Survival Time (RMST)

- To extend this concept to NFT BART, we develop RMST for a mixture of log-Normals: $\log t = y^* \sim \sum_j \omega_j \phi(\mu_j, \sigma_j^2)$
- The probability that $\log(t)$ falls within the observation period

$$p = \mathrm{P}\left[y^* < \log(\tau)\right] = \sum_j \omega_j \int_{-\infty}^{\log(\tau)} \phi(y|\mu_j, \sigma_j^2)\mathrm{d}y = \sum_j \omega_j p_j$$

$$\text{where } p_j = \Phi\left(\frac{\log(\tau) - \mu_j}{\sigma_j}\right)$$

Restricted Mean Survival Time (RMST)

Derive the RMST as an expectation of random variable $\min(t, \tau)$

$$\begin{aligned}\text{RMST}(\tau) &= \text{E}\left[\min(t, \tau)\right] \\ &= \text{E}\left[e^{y^*} | y^* < \log(\tau)\right] + q\tau \\ &\text{where } q = 1 - p\end{aligned}$$

This result can be decomposed into its atoms: $y_j \sim \text{N}\left(\mu_j,\ \sigma_j^2\right)$

$$\text{E}\left[e^{y^*} | y^* < \log(\tau)\right] = \sum_j \omega_j \text{E}\left[e^{y_j} | y_j < \log(\tau)\right]$$

$$\text{E}\left[e^{y_j} | y_j < \log(\tau)\right] = \pi_j e^{\mu_j + \sigma_j^2/2} \text{ where } \pi_j = \Phi\left(\frac{\log(\tau) - (\mu_j + \sigma_j^2)}{\sigma_j}\right)$$

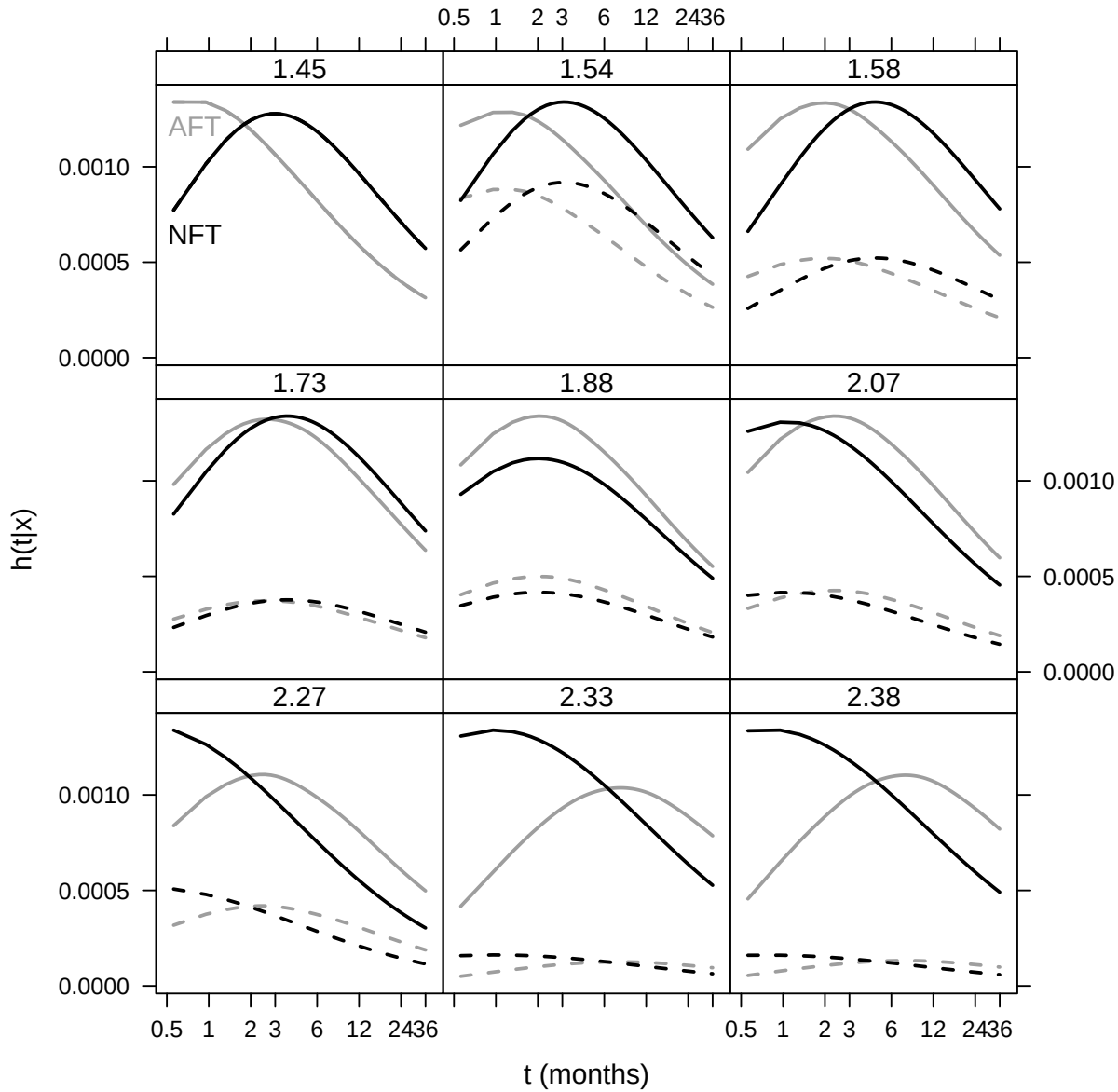
$$\text{E}\left[e^{y^*} | y^* < \log(\tau)\right] = \sum_j \omega_j \pi_j e^{\mu_j + \sigma_j^2/2}$$

Restricted Mean Survival Time (RMST)

- So, finally: $\mathbf{RMST}(\tau) = q\tau + \sum_j \omega_j \pi_j e^{\mu_j + \sigma_j^2/2}$
- It is a simple extension to NFT BART like so $\mathbf{E}\left[\mathbf{RMST}(\tau|x_i, y)\right] \approx M^{-1} \sum_m \mathbf{RMST}_m(\tau|x_i, y)$.
- And we can quantify uncertainty with posterior quantiles

Real data example: AFT BART vs. NFT BART

%-iles of $\widehat{\sigma}(x_i)$: 1, 5, 10, 30, 50, 70, 90, 95, 99



Thompson Sampling Variable Selection (TSVS)

Liu & Rockova 2023 *JASA*

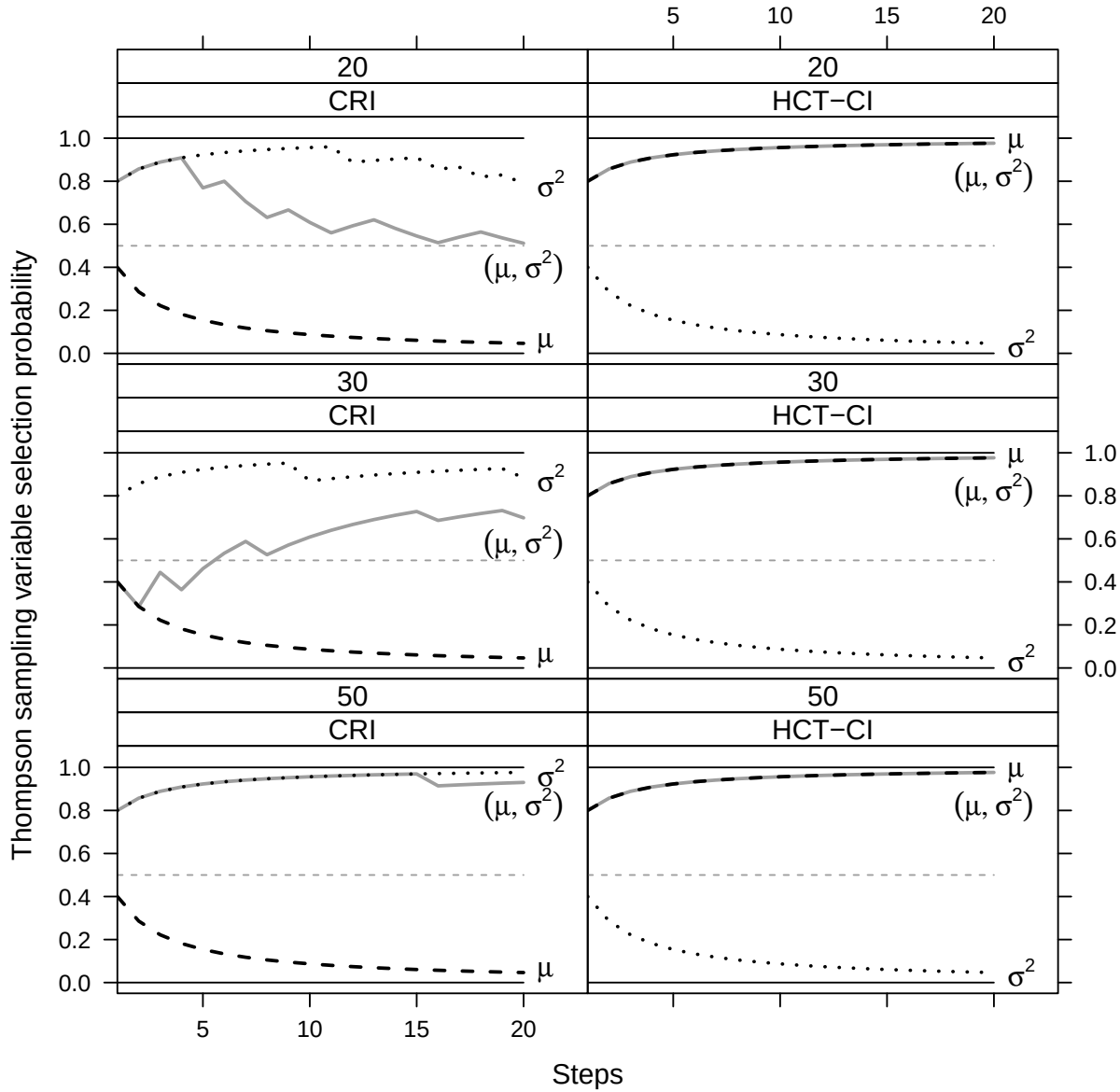
Set $\boldsymbol{H}$ *small*: 10, 20 or 40; smaller numbers engender more sparsity ($\widetilde{\boldsymbol{H}} \approx \boldsymbol{H}/\mathbf{5}$). TSVS is an iterative process: $\boldsymbol{k = 1, \ldots, K}$

Pseudo-Bayesian prior parameter defaults: $\boldsymbol{a_{j0} = 1}$ and $\boldsymbol{b_{j0} = 0.5}$

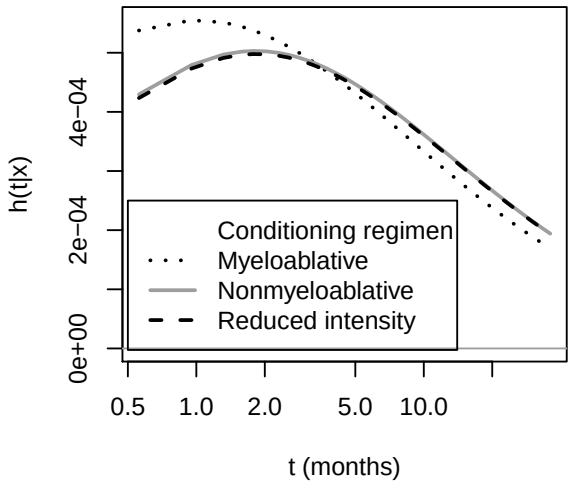
a. For $\boldsymbol{j = 1, \ldots, P}$: draw $\boldsymbol{\eta_{jk} \sim \mathbf{Beta}\left(a_{j,k-1},\ b_{j,k-1}\right)}$
b. Set $\boldsymbol{B_k = \{j : \eta_{jk} \geq 0.5\}}$: covariate subset selected at step $\boldsymbol{k}$
c. Fit an NFT BART model with covariates $\boldsymbol{x_{ij}}$ $\mathbf{where}$ $\boldsymbol{j \in B_k}$
d. For $\boldsymbol{j = 1, \ldots, P}$: do each sub-step
   (i) Reward: if $\boldsymbol{j \notin B_k}$, then $\boldsymbol{\gamma_{jk} = 0}$,
   else $\boldsymbol{\gamma_{jk} = \mathbf{I}\left(u_{jkM} + v_{jkM} > 0\right)}$ where $\boldsymbol{u_{jkM}}$ and $\boldsymbol{v_{jkM}}$ are the number of branches for variable $\boldsymbol{x_{.j}}$ in step $\boldsymbol{k}$ from $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$, respectively, at posterior draw $\boldsymbol{M}$
   (ii) Update via the reward: $\boldsymbol{a_{jk} = a_{j,k-1} + \gamma_{jk}}$ and $\boldsymbol{b_{jk} = b_{j,k-1} + 1 - \gamma_{jk}}$
   (iii) Calculate inclusion probabilities: $\boldsymbol{\pi_{jk} = \frac{a_{jk}}{a_{jk} + b_{jk}}}$
e. If $\boldsymbol{k < K}$, then return to a. and increment $\boldsymbol{k}$

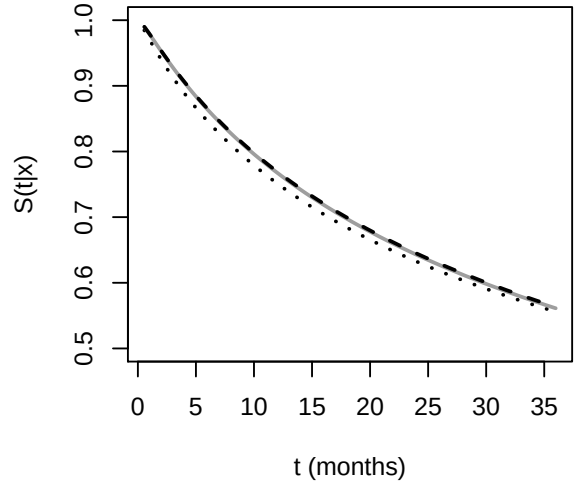
Important variables have trajectories of $\boldsymbol{\pi_{jk}}$ exceeding $\mathbf{0.5}$ by $\boldsymbol{K}$

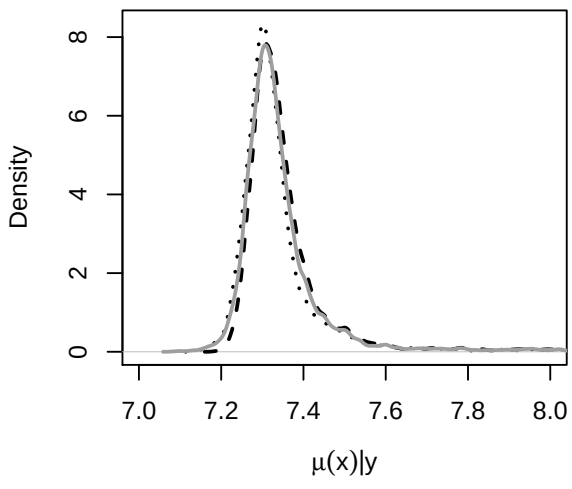
Real data example: TSVS with $\boldsymbol{H = 20, 30, 50}$ Conditioning regimen (CRI) and Comorbidity (HCT-CI)

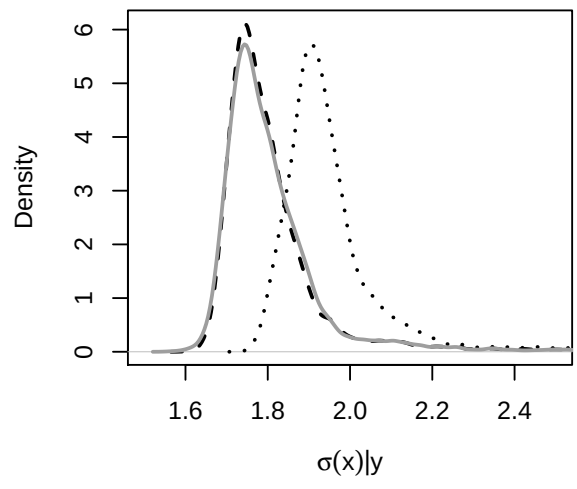


Real data example: Heteroskedasticity of the Conditioning Regimen Intensity (CRI)

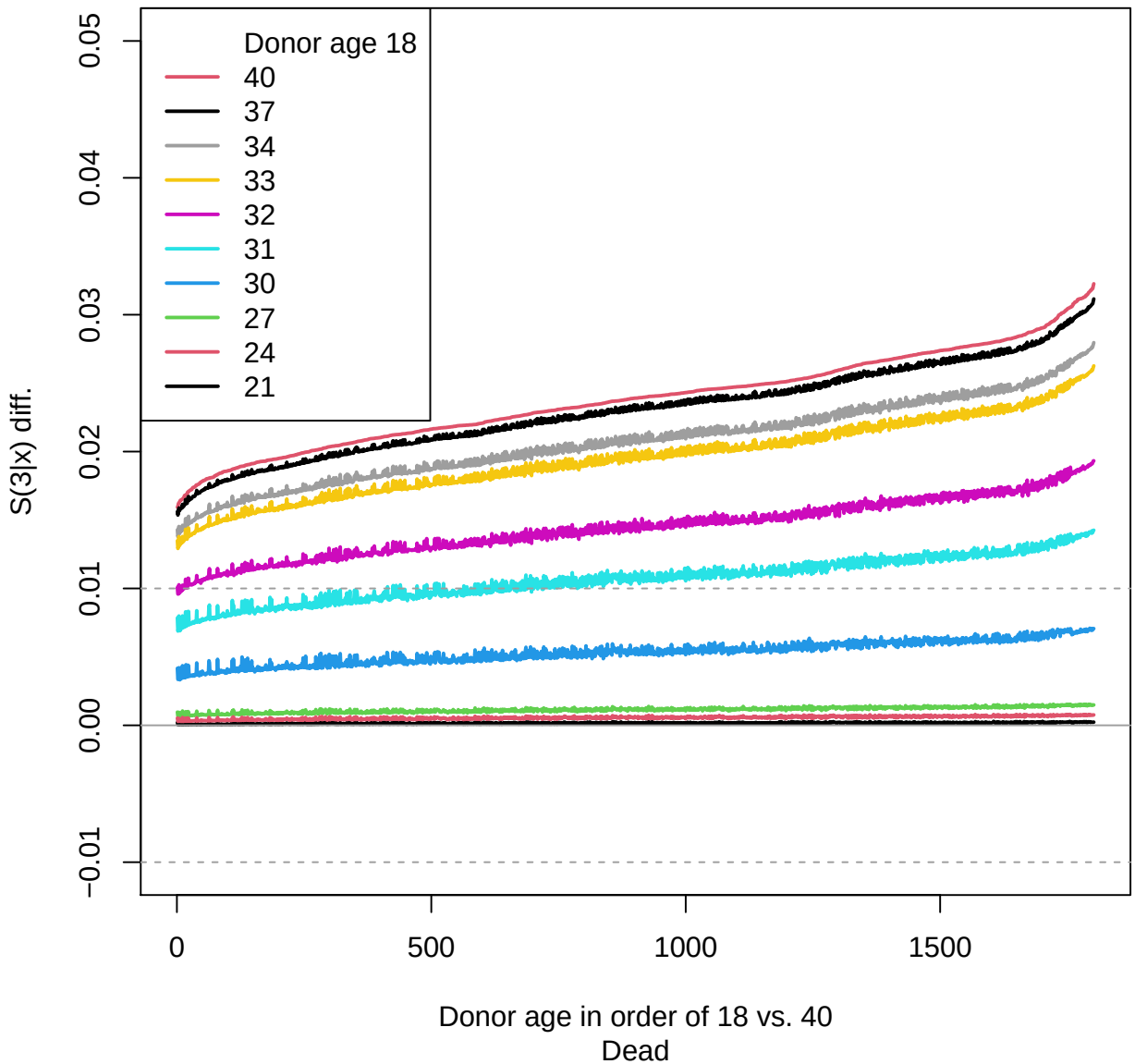




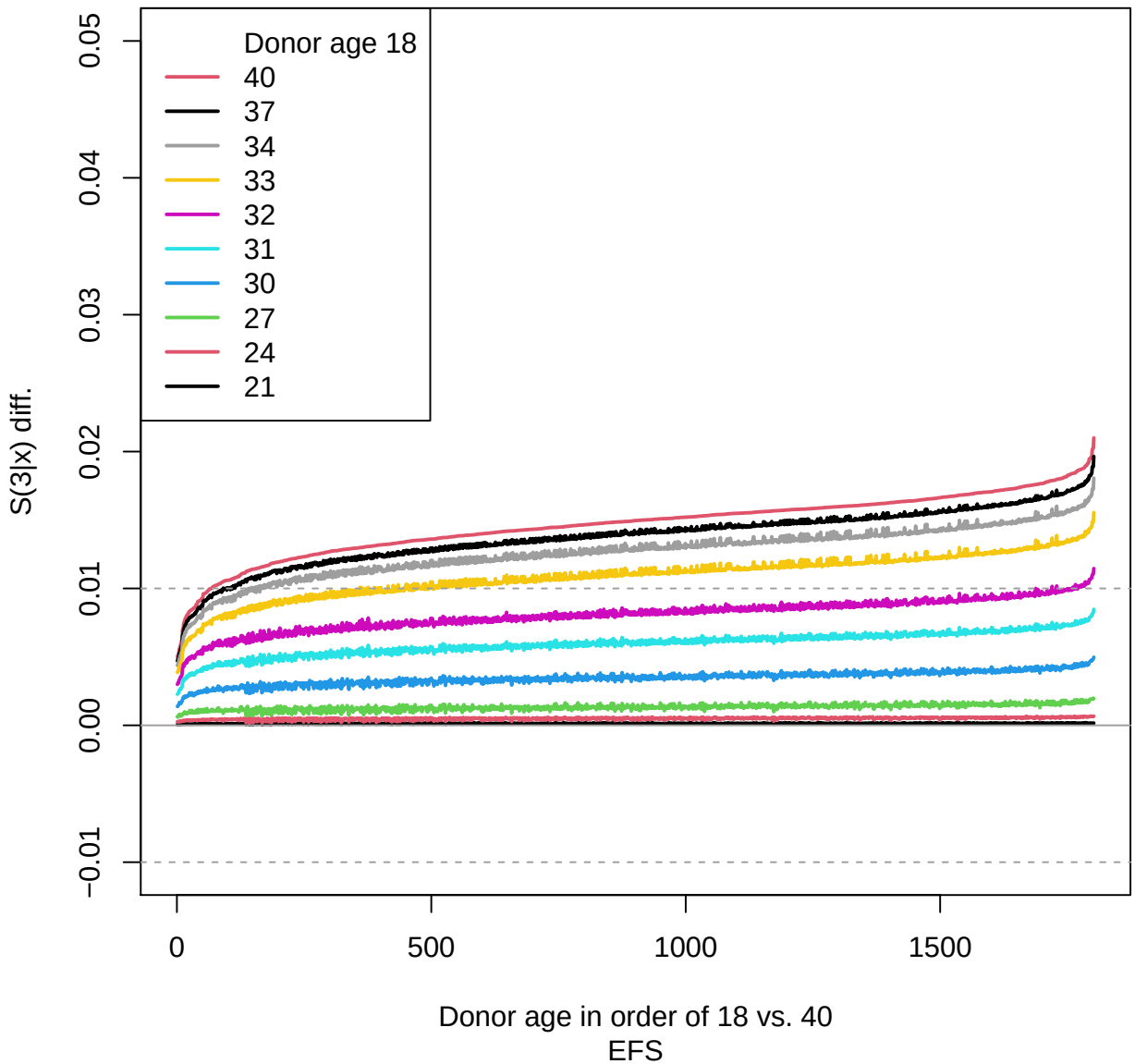




Real data example: Death and donor age



Real data example: EFS and donor age



Conclusions: part 1

- We constructed our new Nonparametric Failure Time (NFT) approach from robust Bayesian Nonparametric building blocks
  - Bayesian Additive Regression Trees (BART) and Heteroskedastic BART (HBART)
  - Constrained Dirichlet Process Mixtures (DPM) with the Low Information Omnibus (LIO) prior hierarchy
- along with the **nftbart** v2.1 R package available on the Comprehensive R Archive Network (CRAN)

Conclusions: part 2

- NFT has desirable properties
  - computationally friendly via MCMC
  - very flexible model which does not resort to precarious restrictive assumptions
  - default prior parameter settings that work well without computationally expensive cross-validation
  - natural extensions to variable selection via Thompson Sampling and marginal effects by Friedman's partial dependence function
- Personalized Hematopoietic Stem Cell Transplant (HSCT)
  - For Event-free Survival of HSCT recipients younger male donors likely result in better outcomes