Incorporating unobserved heterogeneity in Weibull survival models: A Bayesian approach

Catalina Vallejos Mark F.J. Steel

Department of Statistics, University of Warwick

Wirtschaftsuniversität Wien, May 14, 2014

Motivation

Parametric survival models such as the Weibull or log-normal

- Do not allow **unobserved heterogeneity** between observations,
- Do not produce robust inference under the presence of **outliers**.

Motivation

Parametric survival models such as the Weibull or log-normal

- Do not allow **unobserved heterogeneity** between observations,
- Do not produce robust inference under the presence of **outliers**.

Mixtures of life distributions

Definition

The distribution of T_i is defined as a mixture of life distributions, if and only if its density function is given by

$$f(t_i|\psi,\theta) \equiv \int_{\mathcal{L}} f^*(t_i|\psi,\Lambda_i=\lambda_i)\, dP_{\Lambda_i}(\lambda_i|\theta),$$

where $f^*(\cdot|\psi,\Lambda_i=\lambda_i)$ is the density of a lifetime distribution and $P_{\Lambda_i}(\cdot|\theta)$ is a distribution function on $\mathcal{L}$ possibly depending on a parameter $\theta \in \Theta$.

Mixtures of life distributions

- **Unobserved heterogeneity** is incorporated via λ_i (frailty),
- The influence of **outlying observations** is attenuated,
- **Flexible** distributions are generated on the basis of well-known distributions,
- The **intuition** behind the underlying model is preserved.

Mixtures of life distributions

- **Unobserved heterogeneity** is incorporated via λ_i (frailty),
- The influence of **outlying observations** is attenuated,
- **Flexible** distributions are generated on the basis of well-known distributions,
- The **intuition** behind the underlying model is preserved.

Mixtures of life distributions

- **Unobserved heterogeneity** is incorporated via λ_i (frailty),
- The influence of **outlying observations** is attenuated,
- **Flexible** distributions are generated on the basis of well-known distributions,
- The **intuition** behind the underlying model is preserved.

Mixtures of life distributions

- **Unobserved heterogeneity** is incorporated via λ_i (frailty),
- The influence of **outlying observations** is attenuated,
- **Flexible** distributions are generated on the basis of well-known distributions,
- The **intuition** behind the underlying model is preserved.

CRiSM

Rate Mixtures of Weibull distributions

Definition

A random variable T_i has distribution in the family of Rate Mixtures of Weibull distributions (RMW) iff

$$f(t_i|\alpha,\gamma,\theta) = \int_{\mathcal{L}} \gamma\alpha\lambda_i e^{-\alpha\lambda_i t_i^{\gamma}} t_i^{\gamma-1}\, dP_{\Lambda_i}(\lambda_i|\theta), \qquad t_i > 0, \alpha, \gamma > 0, \theta \in \Theta, \tag{1}$$

with $P_{\Lambda_i}(\cdot|\theta)$ defined on $\mathcal{L} \subseteq (0,\infty)$ (possibly discrete). Denote $T_i \sim \text{RMW}_P(\alpha,\gamma,\theta)$. Alternatively, (1) can be expressed as the hierarchical representation

$$T_i|\alpha,\gamma,\Lambda_i = \lambda_i \sim \text{Weibull}\,(\alpha\lambda_i,\gamma)\,, \qquad \Lambda_i|\theta \sim P_{\Lambda_i}(\cdot|\theta). \tag{2}$$

Rate Mixtures of Weibull distributions

- Relates to existing literature, where usually $\gamma = 1$ and mixing distribution is gamma (Lomax distribution)
- Case $\gamma = 1$: Rate Mixtures of Exponentials $T_i \sim \text{RME}_P(\alpha, \theta)$
- RMW and RME linked by simple power transformation If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.
- For $\gamma \leq 1$: decreasing hazard rate for any P
- For $\gamma > 1$: hazard rate can be non-monotone
- Identifiability precludes separate unknown scale parameters in P

Rate Mixtures of Weibull distributions

- Relates to existing literature, where usually $\gamma = 1$ and mixing distribution is gamma (Lomax distribution)
- Case $\gamma = 1$: Rate Mixtures of Exponentials $T_i \sim \text{RME}_P(\alpha, \theta)$
- RMW and RME linked by simple power transformation If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.
- For $\gamma \leq 1$: decreasing hazard rate for any P
- For $\gamma > 1$: hazard rate can be non-monotone
- Identifiability precludes separate unknown scale parameters in P

CRiSM

Rate Mixtures of Weibull distributions

- Relates to existing literature, where usually $\gamma = 1$ and mixing distribution is gamma (Lomax distribution)
- Case $\gamma = 1$: Rate Mixtures of Exponentials
  $T_i \sim \text{RME}_P(\alpha, \theta)$
- RMW and RME linked by simple power transformation
  If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.
- For $\gamma \leq 1$: decreasing hazard rate for any P
- For $\gamma > 1$: hazard rate can be non-monotone
- Identifiability precludes separate unknown scale parameters in P

Rate Mixtures of Weibull distributions

- Relates to existing literature, where usually $\gamma = 1$ and mixing distribution is gamma (Lomax distribution)
- Case $\gamma = 1$: Rate Mixtures of Exponentials $T_i \sim \text{RME}_P(\alpha, \theta)$
- RMW and RME linked by simple power transformation If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.
- For $\gamma \leq 1$: decreasing hazard rate for any P
- For $\gamma > 1$: hazard rate can be non-monotone
- Identifiability precludes separate unknown scale parameters in P

CRiSM

Rate Mixtures of Weibull distributions

- Relates to existing literature, where usually $\gamma = 1$ and mixing distribution is gamma (Lomax distribution)
- Case $\gamma = 1$: Rate Mixtures of Exponentials
  $T_i \sim \text{RME}_P(\alpha, \theta)$
- RMW and RME linked by simple power transformation
  If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.
- For $\gamma \leq 1$: decreasing hazard rate for any P
- For $\gamma > 1$: hazard rate can be non-monotone
- Identifiability precludes separate unknown scale parameters in P

Rate Mixtures of Weibull distributions

- Relates to existing literature, where usually $\gamma = 1$ and mixing distribution is gamma (Lomax distribution)
- Case $\gamma = 1$: Rate Mixtures of Exponentials $T_i \sim \text{RME}_P(\alpha, \theta)$
- RMW and RME linked by simple power transformation If $T_i \sim \text{RME}_P(\alpha, \theta)$ then $T_i^{1/\gamma} \sim \text{RMW}_P(\alpha, \gamma, \theta)$.
- For $\gamma \leq 1$: decreasing hazard rate for any P
- For $\gamma > 1$: hazard rate can be non-monotone
- Identifiability precludes separate unknown scale parameters in P

CRiSM

Some distributions in the RME family

Table : Some distributions included in the RME family. $K_p(\cdot)$ is the modified Bessel function

Mixing density	$\mathrm{E}(\Lambda_i\|\theta)$	$f(t_i\|\alpha,\theta)$	$S(t_i\|\alpha,\theta)$
Exponential(1)	1	$\alpha(\alpha t_i + 1)^{-2}$	$(\alpha t_i + 1)^{-1}$
Gamma(θ,θ)	1	$\alpha([\alpha/\theta]\, t_i + 1)^{-(\theta+1)}$, $\theta > 2$	$\alpha([\alpha/\theta]\, t_i + 1)^{-1}$
Inv-Gamma($\theta, 1$)	$\frac{1}{\theta-1}$	$2\alpha K_{-(\theta-1)}(2\sqrt{\alpha t_i})(\alpha t_i)^{(\theta-1)/2}, \theta > 1$	$2K_{-\theta}(2\sqrt{\alpha t_i})(\alpha t_i)^{\theta/2}$
Inv-Gaussian($\theta, 1$)	θ	$\alpha\, e^{1/\theta} \left[\frac{1}{\theta^2} + 2\alpha t_i\right]^{-1/2} e^{-\left[\frac{1}{\theta^2}+2\alpha t_i\right]^{1/2}}$	$e^{1/\theta}\, e^{-\left[\frac{1}{\theta^2}+2\alpha t_i\right]^{1/2}}$
Log-Normal($0, \theta$)	$e^{\theta/2}$	$\frac{\alpha}{\sqrt{2\pi\theta}} \int_0^\infty e^{-\alpha\lambda_i t_i}\, e^{-\frac{(\log(\lambda_i))^2}{2\theta}}\, d\lambda_i$	No closed form

CRiSM

Some examples of RMW

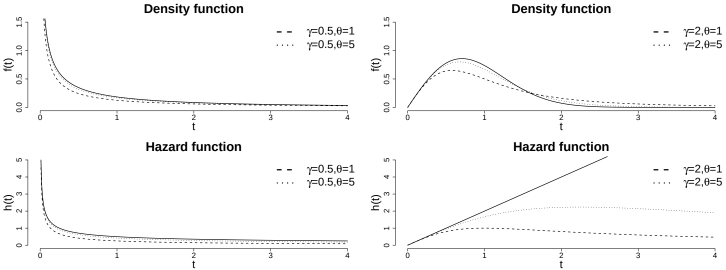

Figure : Some RMW models ($\alpha = 1$). The mixing distribution is Gamma(θ, θ) (Exponential(1) for $\theta = 1$). The solid line is the Weibull($1, \gamma$) density and hazard function.

Coefficient of variation

Corollary

If all the required moments exist, the coefficient of variation (cv) of the survival distributions in (1) is

$$cv(\gamma,\theta) = \sqrt{\frac{\Gamma(1+2/\gamma)}{\Gamma^2(1+1/\gamma)} \underbrace{\frac{var_{\Lambda_i}(\Lambda_i^{-1/\gamma}|\theta)}{E^2_{\Lambda_i}(\Lambda_i^{-1/\gamma}|\theta)}}_{(cv^*(\gamma,\theta))^2} + \underbrace{\frac{\left[\Gamma(1+2/\gamma) - \Gamma^2(1+1/\gamma)\right]}{\Gamma^2(1+1/\gamma)}}_{(cv^W(\gamma))^2}}. \quad (3)$$

Simplifies to $\sqrt{2\frac{var_{\Lambda_i}(\Lambda_i^{-1}|\theta)}{E^2_{\Lambda_i}(\Lambda_i^{-1}|\theta)}+1}$ *when* $\gamma = 1$.

We restrict the range of (γ,θ) such that cv is finite (not required when θ does not appear).

Coefficient of variation inflation

- cv of the Weibull $cv^W(\gamma)$ is a lower bound for $cv(\gamma,\theta)$
- $cv(\gamma,\theta) = cv^W(\gamma)$ iff $\Lambda_i = \lambda_0$ with probability 1.
- Evidence of unobserved heterogeneity:

$$R_{cv}(\gamma,\theta) = \frac{cv(\gamma,\theta)}{cv^W(\gamma)}, \tag{4}$$

  i.e. the cv inflation that the mixture induces (w.r.t. Weibull with the same γ).
- If $\gamma \to 0$, $cv^W(\gamma)$ and, thus, $cv(\gamma,\theta)$ become unbounded. Then $R_{cv}(\gamma,\theta)$ behaves as $\sqrt{[cv^*(\gamma,\theta)]^2+1}$.

CRiSM

Coefficient of variation inflation

- cv of the Weibull $cv^W(\gamma)$ is a lower bound for $cv(\gamma, \theta)$
- $cv(\gamma, \theta) = cv^W(\gamma)$ iff $\Lambda_i = \lambda_0$ with probability 1.
- Evidence of unobserved heterogeneity:

$$R_{cv}(\gamma, \theta) = \frac{cv(\gamma, \theta)}{cv^W(\gamma)}, \quad (4)$$

  i.e. the cv inflation that the mixture induces (w.r.t. Weibull with the same γ).
- If $\gamma \to 0$, $cv^W(\gamma)$ and, thus, $cv(\gamma, \theta)$ become unbounded. Then $R_{cv}(\gamma, \theta)$ behaves as $\sqrt{[cv^*(\gamma, \theta)]^2 + 1}$.

CRiSM

Coefficient of variation inflation

- cv of the Weibull $cv^W(\gamma)$ is a lower bound for $cv(\gamma,\theta)$
- $cv(\gamma,\theta) = cv^W(\gamma)$ iff $\Lambda_i = \lambda_0$ with probability 1.
- Evidence of unobserved heterogeneity:

$$R_{cv}(\gamma,\theta) = \frac{cv(\gamma,\theta)}{cv^W(\gamma)}, \tag{4}$$

i.e. the cv inflation that the mixture induces (w.r.t. Weibull with the same γ).

- If $\gamma \to 0$, $cv^W(\gamma)$ and, thus, $cv(\gamma,\theta)$ become unbounded. Then $R_{cv}(\gamma,\theta)$ behaves as $\sqrt{[cv^*(\gamma,\theta)]^2+1}$.

Coefficient of variation inflation

- cv of the Weibull $cv^W(\gamma)$ is a lower bound for $cv(\gamma, \theta)$
- $cv(\gamma, \theta) = cv^W(\gamma)$ iff $\Lambda_i = \lambda_0$ with probability 1.
- Evidence of unobserved heterogeneity:

$$R_{cv}(\gamma, \theta) = \frac{cv(\gamma, \theta)}{cv^W(\gamma)}, \tag{4}$$

  i.e. the cv inflation that the mixture induces (w.r.t. Weibull with the same γ).
- If $\gamma \to 0$, $cv^W(\gamma)$ and, thus, $cv(\gamma, \theta)$ become unbounded. Then $R_{cv}(\gamma, \theta)$ behaves as $\sqrt{[cv^*(\gamma, \theta)]^2 + 1}$.

Regression Model

Weibull regression model can equivalently be written in terms of Accelerated Failure Times (AFT) and Proportional Hazard (PH) specifications

AFT-RMW (covariates affect the time scale through α):

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i' \beta}, \quad i = 1, \ldots, n, \tag{5}$$

or

$$\log(T_i) = x_i' \beta + \log(\Lambda_i T_0), \tag{6}$$

where $\Lambda_i \sim dP_{\Lambda_i}(\theta)$ and $T_0 \sim$ Weibull$(1, \gamma)$.

- AFT-RMW is itself an AFT model
- e^{β_j} can be interpreted as the proportional change of the median survival time after a unit change in covariate j
- PH-RMW model is not PH model, and interpretation of coefficients is less clear

Regression Model

Weibull regression model can equivalently be written in terms of Accelerated Failure Times (AFT) and Proportional Hazard (PH) specifications

AFT-RMW (covariates affect the time scale through α):

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i' \beta}, \quad i = 1, \dots, n, \tag{5}$$

or

$$\log(T_i) = x_i'\beta + \log(\Lambda_i T_0), \tag{6}$$

where $\Lambda_i \sim dP_{\Lambda_i}(\theta)$ and $T_0 \sim$ Weibull$(1, \gamma)$.

- AFT-RMW is itself an AFT model
- e^{β_j} can be interpreted as the proportional change of the median survival time after a unit change in covariate j
- PH-RMW model is not PH model, and interpretation of coefficients is less clear

Regression Model

Weibull regression model can equivalently be written in terms of Accelerated Failure Times (AFT) and Proportional Hazard (PH) specifications

AFT-RMW (covariates affect the time scale through α):

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i' \beta}, \quad i = 1, \ldots, n, \tag{5}$$

or

$$\log(T_i) = x_i'\beta + \log(\Lambda_i T_0), \tag{6}$$

where $\Lambda_i \sim dP_{\Lambda_i}(\theta)$ and $T_0 \sim$ Weibull$(1, \gamma)$.

- AFT-RMW is itself an AFT model
- e^{β_j} can be interpreted as the proportional change of the median survival time after a unit change in covariate j
- PH-RMW model is not PH model, and interpretation of coefficients is less clear

CRiSM

Regression Model

Weibull regression model can equivalently be written in terms of Accelerated Failure Times (AFT) and Proportional Hazard (PH) specifications

AFT-RMW (covariates affect the time scale through α):

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i' \beta}, \quad i = 1, \ldots, n, \tag{5}$$

or

$$\log(T_i) = x_i'\beta + \log(\Lambda_i T_0), \tag{6}$$

where $\Lambda_i \sim dP_{\Lambda_i}(\theta)$ and $T_0 \sim$ Weibull$(1, \gamma)$.

- AFT-RMW is itself an AFT model
- e^{β_j} can be interpreted as the proportional change of the median survival time after a unit change in covariate j
- PH-RMW model is not PH model, and interpretation of coefficients is less clear

Regression Model

Weibull regression model can equivalently be written in terms of Accelerated Failure Times (AFT) and Proportional Hazard (PH) specifications

AFT-RMW (covariates affect the time scale through α):

$$T_i \sim RMW_P(\alpha_i, \gamma, \theta), \quad \alpha_i = e^{-\gamma x_i' \beta}, \quad i = 1, \ldots, n, \tag{5}$$

or

$$\log(T_i) = x_i'\beta + \log(\Lambda_i T_0), \tag{6}$$

where $\Lambda_i \sim dP_{\Lambda_i}(\theta)$ and $T_0 \sim$ Weibull$(1, \gamma)$.

- AFT-RMW is itself an AFT model
- e^{β_j} can be interpreted as the proportional change of the median survival time after a unit change in covariate j
- PH-RMW model is not PH model, and interpretation of coefficients is less clear

Bayesian inference for the AFT-RMW model
Prior

First consider RME ($\gamma = 1$)
Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta), \tag{7}$$

but they are complicated to derive and $\pi(\theta)$ need not be proper (no comparison through BF).

Approach:

- Keep structure in (7), but use a proper $\pi(\theta)$
- Match priors through common proper prior for $c\theta$, say $\pi^*(c\theta)$
- Using (3), derive the functional relationship between c and θ: Table does this for some RME distributions

CRiSM

Bayesian inference for the AFT-RMW model

Prior

First consider RME ($\gamma = 1$)
Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta), \tag{7}$$

but they are complicated to derive and $\pi(\theta)$ need not be proper (no comparison through BF).

Approach:

- Keep structure in (7), but use a proper $\pi(\theta)$
- Match priors through common proper prior for cv, say $\pi^*(cv)$
- Using (3), derive the functional relationship between cv and θ. Table does this for some RME distributions

Bayesian inference for the AFT-RMW model

Prior

First consider RME ($\gamma = 1$)
Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta), \tag{7}$$

but they are complicated to derive and $\pi(\theta)$ need not be proper (no comparison through BF).

Approach:

- Keep structure in (7), but use a proper $\pi(\theta)$
- Match priors through common proper prior for cv, say $\pi^*(cv)$
- Using (3), derive the functional relationship between cv and θ: Table does this for some RME distributions

CRiSM

Bayesian inference for the AFT-RMW model

Prior

First consider RME ($\gamma = 1$)
Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta), \tag{7}$$

but they are complicated to derive and $\pi(\theta)$ need not be proper (no comparison through BF).

Approach:

- Keep structure in (7), but use a proper $\pi(\theta)$
- Match priors through common proper prior for cv, say $\pi^*(cv)$
- Using (3), derive the functional relationship between cv and θ: Table does this for some RME distributions

CRiSM

Bayesian inference for the AFT-RMW model

Prior

First consider RME ($\gamma = 1$)
Jeffreys and independence Jeffreys priors have structure

$$\pi(\beta, \theta) \propto \pi(\theta), \tag{7}$$

but they are complicated to derive and $\pi(\theta)$ need not be proper (no comparison through BF).

Approach:

- Keep structure in (7), but use a proper $\pi(\theta)$
- Match priors through common proper prior for cv, say $\pi^*(cv)$
- Using (3), derive the functional relationship between cv and θ: Table does this for some RME distributions

CRiSM

Relationship between cv and θ for RME

Table : Relationship between cv and θ for some distributions in the RME family.

Mixing density	Range of cv	$cv(\theta)$	$\left\lvert \frac{dcv(\theta)}{d\theta} \right\rvert$
Gamma(θ, θ)	$(1, \infty)$	$\sqrt{\frac{\theta}{\theta-2}}$	$\theta^{-1/2}(\theta-2)^{-3/2}$
Inverse-Gamma$(\theta, 1)$	$(1, \sqrt{3})$	$\sqrt{\frac{\theta+2}{\theta}}$	$\theta^{-3/2}(\theta+2)^{-1/2}$
Inverse-Gaussian$(\theta, 1)$	$(1, \sqrt{5})$	$\sqrt{\frac{5\theta^2+4\theta+1}{\theta^2+2\theta+1}}$	$\frac{3\theta+1}{(5\theta^2+4\theta+1)^{1/2}(\theta+1)^2}$
Log-Normal$(0, \theta)$	$(1, \infty)$	$\sqrt{2\,e^{\theta}-1}$	$e^{\theta}(2\,e^{\theta}-1)^{-1/2}$

CRiSM

Bayesian inference for the AFT-RMW model

Prior

For RMW we choose

$$\pi(\beta,\gamma,\theta) \propto \pi(\gamma,\theta) \equiv \pi(\theta|\gamma)\pi(\gamma), \tag{8}$$

where $\pi(\theta|\gamma)$ and $\pi(\gamma)$ are proper

- Define $\pi(\theta|\gamma)$ as before through $\pi^*(cv)$, given γ
- Choose a proper $\pi(\gamma)$

Bayesian inference for the AFT-RMW model
Prior

For RMW we choose

$$\pi(\beta,\gamma,\theta) \propto \pi(\gamma,\theta) \equiv \pi(\theta|\gamma)\pi(\gamma), \tag{8}$$

where $\pi(\theta|\gamma)$ and $\pi(\gamma)$ are proper

- Define $\pi(\theta|\gamma)$ as before through $\pi^*(cv)$, given γ
- Choose a proper $\pi(\gamma)$

Bayesian inference for the AFT-RMW model

Prior

For RMW we choose

$$\pi(\beta, \gamma, \theta) \propto \pi(\gamma, \theta) \equiv \pi(\theta|\gamma)\pi(\gamma), \tag{8}$$

where $\pi(\theta|\gamma)$ and $\pi(\gamma)$ are proper

- Define $\pi(\theta|\gamma)$ as before through $\pi^*(cv)$, given γ
- Choose a proper $\pi(\gamma)$

Bayesian inference for the AFT-RMW model

Posterior

- Improper prior: need to check propriety of posterior
- Some observations may be censored

Adding censored observations can not destroy posterior existence, so consider only non-censored ones for sufficient conditions:

Bayesian inference for the AFT-RMW model

Posterior

- Improper prior: need to check propriety of posterior
- Some observations may be censored

Adding censored observations can not destroy posterior existence, so consider only non-censored ones for sufficient conditions:

Theorem

Let $T_1, \ldots, T_n$ be the survival times of n independent individuals distributed as in (5). Define $X = (x_1 \cdots x_n)'$. Suppose $n \geq k$, $r(X) = k$ (full rank) and that the prior is proportional to $\pi(\gamma, \theta)$, which is proper for (γ, θ). If $t_i \neq 0$ for all $i = 1, \ldots, n$, the posterior distribution of (β, γ, θ) is proper.

Bayesian inference for the AFT-RMW model

Posterior

- Improper prior: need to check propriety of posterior
- Some observations may be censored

Adding censored observations can not destroy posterior existence, so consider only non-censored ones for sufficient conditions:

Theorem

Let $T_1, \ldots, T_n$ be the survival times of n independent individuals distributed as in (5). Define $X = (x_1 \cdots x_n)'$. Suppose $n \geq k$, $r(X) = k$ (full rank) and that the prior is proportional to $\pi(\gamma, \theta)$, which is proper for (γ, θ). If $t_i \neq 0$ for all $i = 1, \ldots, n$, the posterior distribution of (β, γ, θ) is proper.

Bayesian inference for the AFT-RMW model

Posterior

- Improper prior: need to check propriety of posterior
- Some observations may be censored

Adding censored observations can not destroy posterior existence, so consider only non-censored ones for sufficient conditions:

Theorem

Let $T_1, \dots, T_n$ be the survival times of n independent individuals distributed as in (5). Define $X = (x_1 \cdots x_n)'$. Suppose $n \geq k$, $r(X) = k$ (full rank) and that the prior is proportional to $\pi(\gamma, \theta)$, which is proper for (γ, θ). If $t_i \neq 0$ for all $i = 1, \dots, n$, the posterior distribution of (β, γ, θ) is proper.

Bayesian inference for the AFT-RMW model

Posterior

- Improper prior: need to check propriety of posterior
- Some observations may be censored

Adding censored observations can not destroy posterior existence, so consider only non-censored ones for sufficient conditions:

Theorem

Let $T_1, \ldots, T_n$ be the survival times of n independent individuals distributed as in (5). Define $X = (x_1 \cdots x_n)'$. Suppose $n \geq k$, $r(X) = k$ (full rank) and that the prior is proportional to $\pi(\gamma, \theta)$, which is proper for (γ, θ) . If $t_i \neq 0$ for all $i = 1, \ldots, n$, the posterior distribution of (β, γ, θ) is proper.

CRiSM

Bayesian inference for the AFT-RMW model

Model Comparison

We compare models on basis of:

- Bayes factors
- DIC
- Conditional Predictive Ordinate (CPO): for observation i,

$$\text{CPO}_i = f(t_i|t_{-i}), \qquad t_{-i} = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n),$$

  where $f(\cdot|t_{-i})$ is the predictive density given t_{-i}.
- $\text{PsML} = \prod_{i=1}^{n} \text{CPO}_i$

CRiSM

Bayesian inference for the AFT-RMW model

Model Comparison

We compare models on basis of:

- Bayes factors
- DIC
- Conditional Predictive Ordinate (CPO): for observation i,

$$\text{CPO}_i = f(t_i|t_{-i}), \qquad t_{-i} = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n),$$

  where $f(\cdot|t_{-i})$ is the predictive density given t_{-i}.
- PsML = $\prod_{i=1}^{n}$ CPO$_i$

Bayesian inference for the AFT-RMW model

Model Comparison

We compare models on basis of:

- Bayes factors
- DIC
- Conditional Predictive Ordinate (CPO): for observation i,

$$\text{CPO}_i = f(t_i|t_{-i}), \qquad t_{-i} = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n),$$

  where $f(\cdot|t_{-i})$ is the predictive density given t_{-i}.

- $\text{PsML} = \prod_{i=1}^{n} \text{CPO}_i$

Bayesian inference for the AFT-RMW model

Model Comparison

We compare models on basis of:

- Bayes factors
- DIC
- Conditional Predictive Ordinate (CPO): for observation i,

$$\text{CPO}_i = f(t_i | t_{-i}), \qquad t_{-i} = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n),$$

  where $f(\cdot | t_{-i})$ is the predictive density given t_{-i}.
- PsML = $\prod_{i=1}^{n}$ CPO$_i$

CRiSM

Bayesian inference for the AFT-RMW model

Outliers

- Outliers: extreme λ_i
- Effect of outliers on posterior of β is attenuated by mixing
- Identification of outliers through mixing variables: compare $H_0 : \Lambda_i = \lambda_{ref}$ with $H_1 : \Lambda_i \neq \lambda_{ref}$ (with all other $\Lambda_j, j \neq i$ free) BF can be computed by generalized Savage-Dickey density ratio

$$BF_{01}^{(i)} = \pi(\lambda_i | t, c) E\left(\frac{1}{dP(\lambda_i|\theta)} \right)\Bigg|_{\lambda_i=\lambda_{ref}}$$

(computationally intensive, but simplifies to SD density ratio when no θ). Choice of λ_{ref}?

Bayesian inference for the AFT-RMW model
Outliers

- Outliers: extreme λ_i
- Effect of outliers on posterior of β is attenuated by mixing
- Identification of outliers through mixing variables: compare $H_0 : \Lambda_i = \lambda_{ref}$ with $H_1 : \Lambda_i \neq \lambda_{ref}$ (with all other $\Lambda_j, j \neq i$ free) BF can be computed by generalized Savage-Dickey density ratio

$$BF_{01}^{(i)} = \pi(\lambda_i | t, c) E\left(\frac{1}{dP(\lambda_i|\theta)}\right)\Bigg|_{\lambda_i=\lambda_{ref}}$$

(computationally intensive, but simplifies to SD density ratio when no θ). Choice of λ_{ref}?

Bayesian inference for the AFT-RMW model

Outliers

- Outliers: extreme λ_i
- Effect of outliers on posterior of β is attenuated by mixing
- Identification of outliers through mixing variables: compare $H_0 : \Lambda_i = \lambda_{ref}$ with $H_1 : \Lambda_i \neq \lambda_{ref}$ (with all other $\Lambda_j, j \neq i$ free)
  BF can be computed by generalized Savage-Dickey density ratio

$$BF_{01}^{(i)} = \pi(\lambda_i|t,c)E\left(\frac{1}{dP(\lambda_i|\theta)}\right)\Bigg|_{\lambda_i=\lambda_{ref}}$$

(computationally intensive, but simplifies to SD density ratio when no θ). Choice of λ_{ref}?

Bayesian inference for the AFT-RMW model

Outliers

Use $\lambda_{ref} = E(\Lambda_i|\theta)$, replacing θ by its posterior median

Mixing through scale parameter, so censoring very informative for mixing parameters. So for censored observations we use correction factor:

$$\lambda^c_{ref} = R_i(\beta, \gamma, \theta)\lambda^o_{ref}, \text{ with } R_i(\beta, \gamma, \theta) = \frac{E(\Lambda_i|t_i, c_i = 0, \beta, \gamma, \theta)}{E(\Lambda_i|t_i, c_i = 1, \beta, \gamma, \theta)}.$$

CRiSM

Bayesian inference for the AFT-RMW model

Outliers

Use $\lambda_{ref} = E(\Lambda_i|\theta)$, replacing θ by its posterior median

Mixing through scale parameter, so censoring very informative for mixing parameters. So for censored observations we use correction factor:

$$\lambda^c_{ref} = R_i(\beta,\gamma,\theta)\lambda^o_{ref}, \text{ with } R_i(\beta,\gamma,\theta) = \frac{E(\Lambda_i|t_i, c_i = 0, \beta, \gamma, \theta)}{E(\Lambda_i|t_i, c_i = 1, \beta, \gamma, \theta)}.$$

Application to cerebral palsy data

dataset

1,549 children affected by cerebral palsy, born 1966-1984 in Mersey region. Record survival times in years. Covariates: amount of severe impairments, birth weight. Only 242 recorded deaths, so 84.4% is right censored.

Analysed with AFT-RMW model as well as a Weibull model. Inference on β (see graph) is similar for most mixture distributions, but different from Weibull in β_1 (effect of no impairment)

Inference on γ clearly suggests $\gamma > 1$ (non-monotone hazard rate). Larger γ for mixture models (Weibull underestimates γ to accommodate data variability)

CRiSM

Application to cerebral palsy data

dataset

1,549 children affected by cerebral palsy, born 1966-1984 in Mersey region. Record survival times in years. Covariates: amount of severe impairments, birth weight. Only 242 recorded deaths, so 84.4% is right censored.

Analysed with AFT-RMW model as well as a Weibull model. Inference on β (see graph) is similar for most mixture distributions, but different from Weibull in β_1 (effect of no impairment)

Inference on γ clearly suggests $\gamma > 1$ (non-monotone hazard rate). Larger γ for mixture models (Weibull underestimates γ to accommodate data variability)

CRiSM

Application to cerebral palsy data

dataset

1,549 children affected by cerebral palsy, born 1966-1984 in Mersey region. Record survival times in years. Covariates: amount of severe impairments, birth weight. Only 242 recorded deaths, so 84.4% is right censored.

Analysed with AFT-RMW model as well as a Weibull model.

Inference on β (see graph) is similar for most mixture distributions, but different from Weibull in β_1 (effect of no impairment)

Inference on γ clearly suggests $\gamma > 1$ (non-monotone hazard rate). Larger γ for mixture models (Weibull underestimates γ to accommodate data variability)

CRiSM

Application to cerebral palsy data

dataset

1,549 children affected by cerebral palsy, born 1966-1984 in Mersey region. Record survival times in years. Covariates: amount of severe impairments, birth weight. Only 242 recorded deaths, so 84.4% is right censored.

Analysed with AFT-RMW model as well as a Weibull model. Inference on β (see graph) is similar for most mixture distributions, but different from Weibull in β_1 (effect of no impairment)

Inference on γ clearly suggests $\gamma > 1$ (non-monotone hazard rate). Larger γ for mixture models (Weibull underestimates γ to accommodate data variability)

Application to cerebral palsy data

dataset

1,549 children affected by cerebral palsy, born 1966-1984 in Mersey region. Record survival times in years. Covariates: amount of severe impairments, birth weight. Only 242 recorded deaths, so 84.4% is right censored.

Analysed with AFT-RMW model as well as a Weibull model. Inference on β (see graph) is similar for most mixture distributions, but different from Weibull in β_1 (effect of no impairment)

Inference on γ clearly suggests $\gamma > 1$ (non-monotone hazard rate). Larger γ for mixture models (Weibull underestimates γ to accommodate data variability)

Posterior results for cerebral palsy data

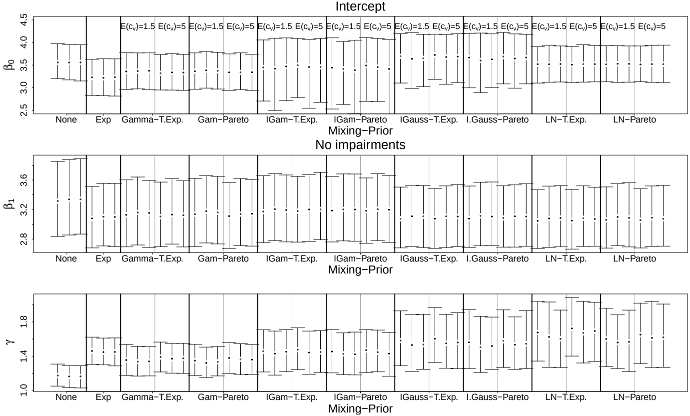

Figure : 95% HPD intervals and posterior medians. Model (5) and (8) with Gamma prior for γ and Trunc-Exp or Pareto prior for cv. From left to right: Gamma(4,1), Gamma(1,1) and Gamma(0.01,0.01) prior for γ. Values of E(cv) in top panel. β_0: intercept, β_1: no impairments.

Model comparison for cerebral palsy data

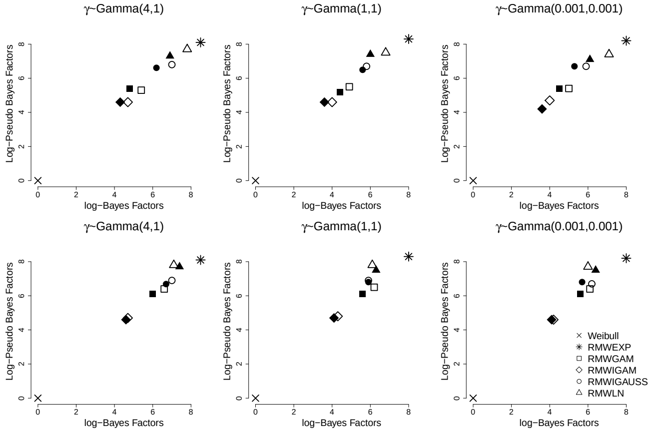

Figure : Cerebral palsy dataset. Model comparison in terms of BF and PsML. Unfilled and filled characters denote a truncated exponential and Pareto prior for cv. Upper panels use $E(cv) = 1.5$. Lower panels use $E(cv) = 5$

Outlier detection for cerebral palsy data

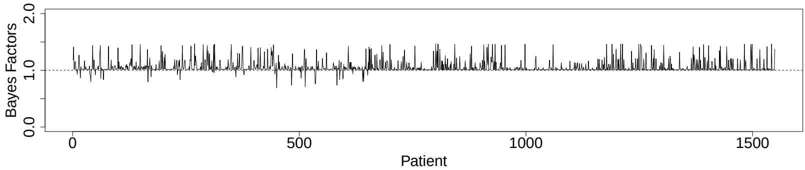

Figure : Cerebral palsy dataset using the exponential mixing distribution. BF in favour of the hypothesis $\lambda_i \neq \lambda_{ref}$, with $\lambda^o_{ref} = 1$ and $\lambda^c_{ref} = 1/2$

No individual outliers, but strong support for mixing. Corroborated by inference on R_{cv} (posterior median around 2).

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)

CRiSM

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)

CRiSM

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)

Conclusions

1. propose mixtures of life distributions (rate mixtures of Weibulls) to deal with unobserved heterogeneity and outliers
2. Obtain flexible classes in shape and tails
3. Covariates through AFT specification: retains AFT and β interpretable
4. Prior based on structure of Jeffreys prior, but allows meaningful BFs
5. Derive simple conditions for posterior existence
6. Outlier detection based on mixing parameters
7. Data support mixing; in particular exponential mixing distribution (easy to elicit and to implement, as no θ)