Some algorithms to fit some reliability mixture models under censoring

Laurent Bordes — University of Pau
Didier Chauveau — University of Orléans

COMPSTAT August 22-27, 2010

Table of contents

Plan

About lifetimes

The lifetime data are assumed to come from a finite mixture of m component densities f_j, $j = 1, \ldots, m$, where $f_j(\cdot) = f(\cdot|\xi_j) \in \mathcal{F}$ a parametric family indexed by a Euclidean parameter ξ. The lifetime density of an observation X may be written

$$X \sim g(x|\boldsymbol{\theta}) = \sum_{j=1}^{m} \lambda_j f(x|\xi_j),$$

where $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\xi}) = (\lambda_1, \ldots, \lambda_m, \xi_1, \ldots, \xi_m)$.
Latent variable representation: $X = Y_Z$ where $Z \sim \mathcal{M}\mathrm{ult}(1, \boldsymbol{\lambda})$ and $(Y_Z|Z = j) \sim f(\cdot|\xi_j)$. For references on the broad literature of mixture models McLachlan and Peel (2000).

Right censored data

The censoring process is described by a random variable C with density function q, distribution function Q and survival function $\bar{Q}$. In the right censoring setup the only available information is

$$T = \min(X, C), \quad D = \mathbb{I}(X \leq C).$$

The n lifetime data are $\mathbf{x} = (x_1, \ldots, x_n)$ iid$\sim g$, associated to n censoring times $\mathbf{c} = (c_1, \ldots, c_n)$ iid$\sim C$. The observations are thus

$$(\mathbf{t}, \mathbf{d}) = ((t_1, d_1), \ldots, (t_n, d_n)),$$

where $t_i = \min(x_i, c_i)$ and $d_i = \mathbb{I}(x_i \leq c_i)$.

Complete data choice

- The observed data $(\mathbf{t}, \mathbf{d})$ depends on $\mathbf{x}$ which comes from a finite mixture $\Rightarrow$ missing data are naturally associated to it.
- To these *incomplete* data are associated *complete* data which correspond to the situation where the component of origin $z_i \in \{1, \ldots, m\}$ of each individual lifetime x_i is known.
- The complete model at the level of (X, Z) is given by $\mathbb{P}_{\boldsymbol{\theta}}(Z = z) = \lambda_z$ and $(X|Z = z) \sim f_z$.
- With the right censoring process the complete data are $(\mathbf{t}, \mathbf{d}, \mathbf{z})$, where $\mathbf{z} = (z_1, \ldots, z_n)$.

Remark.

As in Chauveau (1995) the complete data can be $(\mathbf{x}, \mathbf{z})$ instead of $(\mathbf{t}, \mathbf{d}, \mathbf{z})$.

Complete data pdf

Because we have:

$$\begin{aligned} f^c_{\boldsymbol{\theta}}(T=t, D=1, Z=z) &= \mathbb{P}_{\boldsymbol{\theta}}(Z=z)\, f_{\boldsymbol{\theta}}(D=1, T=t|Z=z) \\ &= \lambda_z\, f_{\boldsymbol{\theta}}(C \geq X, X=t|z) \\ &= \lambda_z\, \mathbb{P}_{\boldsymbol{\theta}}(C \geq t)\, f_{\boldsymbol{\theta}}(X=t|z) \\ &= \lambda_z\, f_z(t)\bar{Q}(t), \end{aligned}$$

and similarly $f^c_{\boldsymbol{\theta}}(t, 0, z) = \lambda_z \bar{F}_z(t) q(t)$, the complete data pdf is summarized by

$$f^c(t, d, z|\boldsymbol{\theta}) = \left[\lambda_z f(t|\xi_z)\bar{Q}(t)\right]^d \left[\lambda_z \bar{F}(t|\xi_z) q(t)\right]^{1-d}.$$

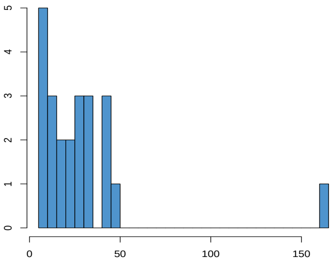
Plan

Acute Myelogenous Leukemia survival data (Miller, 1997)

- Group effect with two groups
- Sample size: 23
- Censored lifetimes: 5

group	scale estimation
`Maintained`	63.3
`Nonmaintained`	25.1

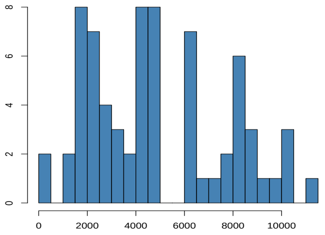
Variables	Description
`time`	survival or censoring time
`status`	censoring status
`x`	maintenance chemotherapy given

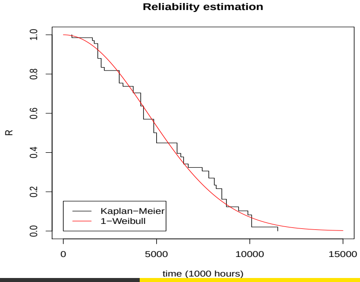


Lifetimes of diesel engines fans (Nelson, 1982)

- Time scale (1000s of hours)
- Sample size: 70
- Censored lifetimes: 12

Variables	Description
`time`	survival or censoring time
`status`	censoring status





Plan

Parametric EM-algorithm: complete data $= (\mathbf{t}, \mathbf{d}, \mathbf{z})$

Usual missing data framework (Dempster, Laird and Rubin, 1977) $\Rightarrow$ define an EM algorithm that generates a sequence $(\boldsymbol{\theta}^k)_{k=1,2,\ldots}$ (with arbitrary initial value $\boldsymbol{\theta}^0$) by iteratively maximize

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) &= \mathbb{E}\left[\log f^c(\mathbf{t}, \mathbf{d}, \mathbf{Z}|\boldsymbol{\theta}) \mid \mathbf{t}, \mathbf{d}, \boldsymbol{\theta}^k\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\log f^c(t_i, d_i, Z_i|\boldsymbol{\theta}) \mid t_i, d_i, \boldsymbol{\theta}^k\right].
\end{aligned}
$$

Calculation of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$ requires calculation of the following *posterior* probabilities

$$
\begin{aligned}
p_{ij}^k &:= \mathbb{P}(Z_i = j | t_i, d_i, \boldsymbol{\theta}^k) \\
&= \lambda_j^k \left(\frac{f(t_i|\xi_j^k)}{\sum_{\ell=1}^p \lambda_\ell^k f(t_i|\xi_\ell^k)}\right)^{d_i} \left(\frac{\bar{F}(t_i|\xi_j^k)}{\sum_{\ell=1}^p \lambda_\ell^k \bar{F}(t_i|\xi_\ell^k)}\right)^{1-d_i}. \qquad (1)
\end{aligned}
$$

Exponential lifetimes: complete data $= (\mathbf{t}, \mathbf{d}, \mathbf{z})$

EM algorithm: $\boldsymbol{\theta}^k \to \boldsymbol{\theta}^{k+1}$

1. **E-step:** Calculate the posterior probabilities p_{ij}^k as in Equation (1), for all $i = 1, \dots, n$ and $j = 1, \dots, m$.
2. **M-step:** Set

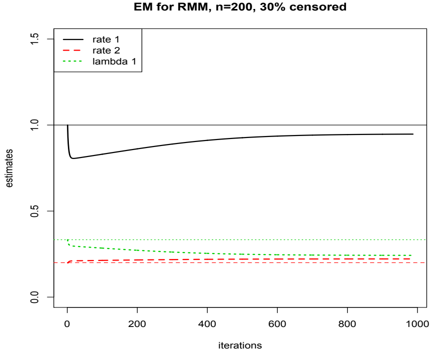
$$\lambda_j^{k+1} = \frac{1}{n} \sum_{i=1}^{n} p_{ij}^k \quad \text{for } j = 1, \dots, m$$

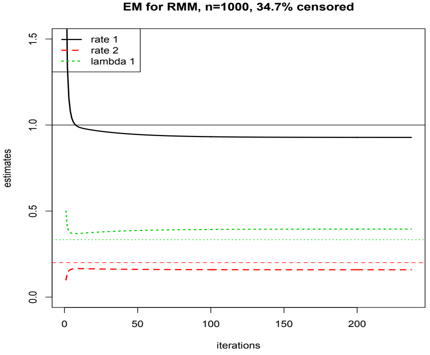
$$\xi_j^{k+1} = \frac{\sum_{i=1}^{n} p_{ij}^k \, d_i}{\sum_{i=1}^{n} p_{ij}^k \, t_i} \quad \text{for } j = 1, \dots, m.$$

Simulation example

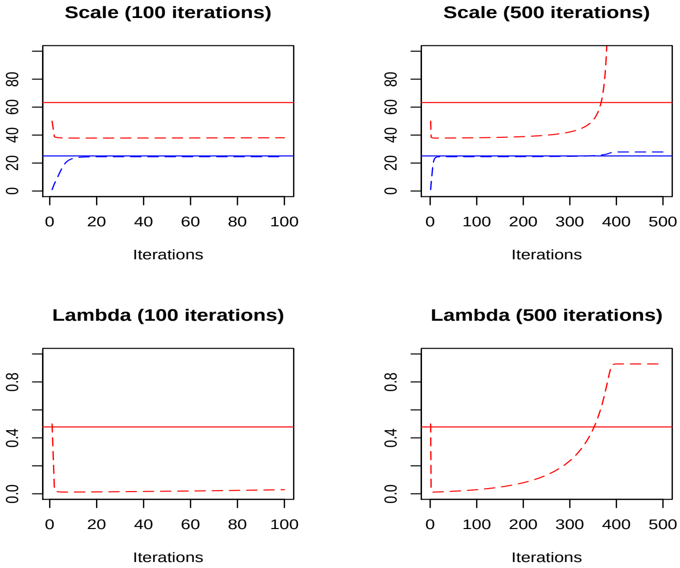
$$g(x) = \lambda_1 \xi_1 \exp(-\xi_1 x) + \lambda_2 \xi_2 \exp(-\xi_2 x) \quad x > 0,$$

with $\xi_1 = 1$ — — —, $\xi_2 = 0.2$ — — — and $\lambda_1 = 1/3$ — — —.





Application to AML data: be careful!



Parametric EM-algorithm: complete data $= (\mathbf{x}, \mathbf{z})$

Complete data pdf

$$f^c(x,z) = \lambda_z f_z(x).$$

Then

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) &= \mathbb{E}\left[\log f^c(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) \mid \mathbf{t},\mathbf{d},\boldsymbol{\theta}^k\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[\log f^c(X_i,Z_i|\boldsymbol{\theta}) \mid t_i,d_i,\boldsymbol{\theta}^k\right]. \end{aligned}$$

Calculation of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$ requires calculation of the following *posterior* pdf

$$\begin{aligned} f_i^k(x,j) &:= f(X_i = x, Z_i = j|t_i,d_i,\boldsymbol{\theta}^k) \\ &= \lambda_j^k \left(\frac{\mathbb{I}(x=t_i)f(t_i|\xi_j^k)}{\sum_{\ell=1}^p \lambda_\ell^k f(t_i|\xi_\ell^k)}\right)^{d_i} \left(\frac{\mathbb{I}(x>t_i)f(x|\xi_j^k)}{\sum_{\ell=1}^p \lambda_\ell^k \bar{F}(t_i|\xi_\ell^k)}\right)^{1-d_i}. \end{aligned}$$

Exponential lifetimes: complete data $= (\mathbf{x}, \mathbf{z})$

EM algorithm: $\boldsymbol{\theta}^k \to \boldsymbol{\theta}^{k+1}$

1. **E-step:** Calculate the posterior probabilities p_{ij}^k as in Equation (1), for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$.
2. **M-step:** Set for $j = 1, \ldots, m$

$$
\begin{aligned}
\lambda_j^{k+1} &= \frac{1}{n} \sum_{i=1}^{n} p_{ij}^k, \\
\xi_j^{k+1} &= \frac{\sum_{i=1}^n p_{ij}^k}{\sum_{i=1}^n \left(d_i t_i p_{ij}^k + (1 - d_i) \frac{\lambda_j^k (1+\xi_j^k t_i) \exp(-\xi_j^k {}_i)}{\xi_j^k \sum_{\ell=1}^p \lambda_\ell^k \exp(-\xi_\ell^k t_i)} \right)}.
\end{aligned}
$$

Remarks about the parametric EM algorithms

- $+$ Whatever the choice of complete data the M-step for the λ_js always leads to explicit formula.
- $-$ $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$ depends strongly on the choice of the underlying parametric family $\mathcal{F}$.
- $-$ Except for exponential lifetimes, explicit maximizers of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$ are not reachable.
- $-$ Maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$ may be as complicated as maximizing the full likelihood function.

Plan

Parametric stochastic EM approach [1/2]

- Idea by Celeux and Diebolt (1985, 1986): at each iteration add a **stochastic step** where the missing data are simulated according to their posterior probability distribution given the current value $\boldsymbol{\theta}^k$ of the unknown parameter $\boldsymbol{\theta}$.
- What should be the *complete data*? It is enough to chose $(\mathbf{t}, \mathbf{d}, \mathbf{z})$.
- $\mathbf{p}_i^k = (p_{i1}^k, \ldots, p_{im}^k)$ is the posterior probability vector associated to observation i. Consider $Z \sim \mathcal{M}\mathrm{ult}(1, \mathbf{p}_i^k)$ a multinomial distributed random variable with parameters 1 and $\mathbf{p}_i^k$ (i.e., $Z \in \{1, \ldots m\}$ with probabilities $\mathbb{P}(Z = j) = p_{ij}^k$)).

Parametric stochastic EM approach [2/2]

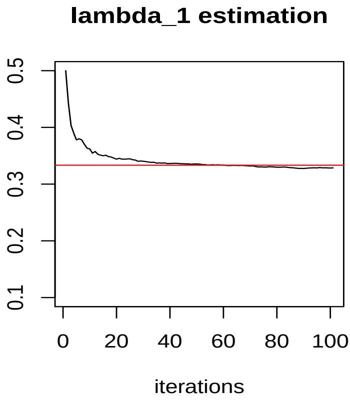
St-EM algorithm: $\boldsymbol{\theta}^k \to \boldsymbol{\theta}^{k+1}$

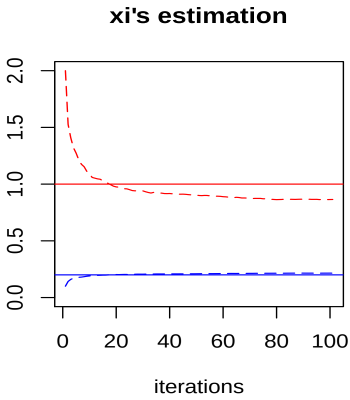
1. **E-step:** Calculate the posterior probabilities p_{ij}^k as in Equation (1), for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$.
2. **Stochastic step:** Simulate $Z_i^k \sim \mathcal{M}\mathrm{ult}(1, \mathbf{p}_i^k)$, $i = 1, \ldots, n$, and define the subsets
$$\chi_j^k = \{i \in \{1, \ldots, n\} : Z_i^k = j\}, \quad j = 1, \ldots, m. \tag{2}$$
3. **M-step:** For each component $j \in \{1, \ldots, m\}$
$$\lambda_j^{k+1} = \mathrm{Card}(\chi_j^k)/n,$$
and
$$\xi_j^{k+1} = \arg\max_{\xi \in \Xi} L_j(\xi),$$
where
$$L_j(\xi) = \prod_{i \in \chi_j^k} (f(t_i|\xi))^{d_i} (\bar{F}(t_i|\xi))^{1-d_i}.$$

Exponential mixture example ($n = 200$)

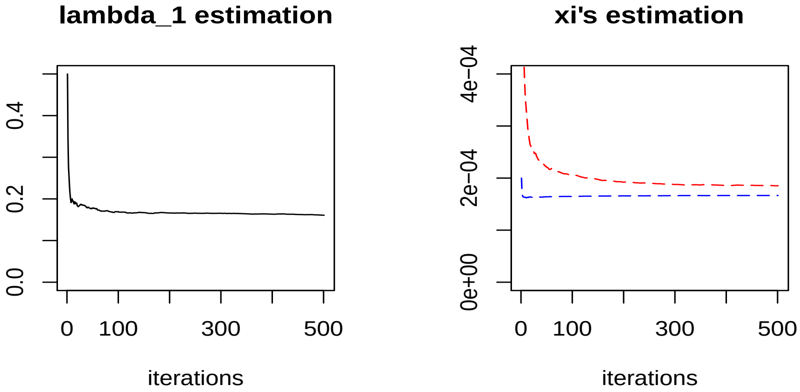
$$g(x) = \lambda\xi_1 \exp(-\xi_1 x) + (1-\lambda)\xi_2 \exp(-\xi_2 x) \quad x > 0,$$

with $\lambda = 1/3$, $\xi_1 = 1$ and $\xi_2 = 1/5$.

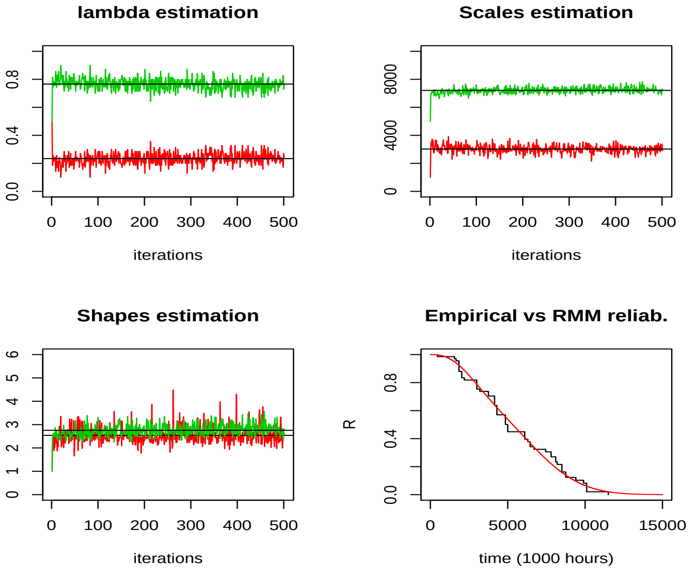




Application to engine fans (two exponentials)



Application to engine fans (two Weibulls)



Plan

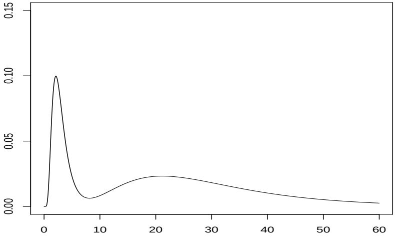
A semiparametric reliability mixture model (SRMM)

$$g(x|\boldsymbol{\theta}) = \lambda_1 f(x) + \lambda_2 \xi f(\xi x) \quad x > 0,$$

where $\boldsymbol{\theta} = (\lambda, \xi, f) \in (0,1) \times \mathbb{R}_*^+ \times \mathcal{F}$; $\mathcal{F}$ is a family of pdf.

Interpretation: accelerated lifetime model for grouped data with two groups and unobserved group label.

Example: $\lambda_1 = 0.3$, $\xi = 0.1$ and $f \sim \mathcal{LN}(1, 0.5)$.



Identifiabillity of $\boldsymbol{\theta}$

Question: how to chose $\mathcal{F}$ to obtain

$$\left[\forall x > 0 \quad g(x|\boldsymbol{\theta}) = g(x|\boldsymbol{\theta}')\right] \quad \Rightarrow \theta = \theta'?$$

Hard question... partial answer in Bordes, Mottelet and Vandekerkhove (2006) and in Hunter, Wang and Hettmansperger (2007): if $\mathcal{F}$ is a subset of pdf f such that $x \mapsto e^x f(e^x)$ is symmetric then identifiability holds!

Stochastic EM algorithm for the SRMM [1/4]

Notations: f is the unknown pdf, write $\bar{F}$ the reliability function and $\alpha = f/\bar{F}$ the failure rate.
From $(\mathbf{t}, \mathbf{d})$:

- $\bar{F}$ is nonparametrically estimated by the Kaplan-Meier estimator,
- α is nonparametrically estimated by smoothing the Nelson-Aalen estimator.

Because λ^k, ξ^k, $\bar{F}^k$ and α^k are estimates of λ, ξ, $\bar{F}$ and α at step k we have:

$$\begin{aligned} p_{ij}^k &:= \mathbb{P}(Z_i = j | t_i, d_i, \boldsymbol{\theta}^k) \\ &= \left(\frac{\alpha^k(t_i)\bar{F}^k(t_i)}{\sum_{\ell=1}^p \lambda_\ell^k \alpha^k(t_i)\bar{F}^k(t_i)} \right)^{d_i} \left(\frac{\lambda_j^k \bar{F}^k(t_i)}{\sum_{\ell=1}^p \lambda_\ell^k \bar{F}^k(t_i)} \right)^{1-d_i}, \end{aligned}$$

where the pdf f is estimated by $f^k = \alpha^k \bar{F}^k$.

Stochastic EM algorithm for the SRMM [2/4]

1. Posterior probabilities calculation: for each item $i \in \{1, \dots, n\}$:
   if $d_i = 0$ then

$$p_{i1}^k = \frac{\lambda^k \bar{F}^k(t_i)}{\lambda^k \bar{F}^k(t_i) + (1-\lambda^k)\bar{F}^k(\xi^k t_i)},$$

   else

$$p_{i1}^k = \frac{\lambda^k \alpha^k(t_i)\bar{F}^k(t_i)}{\lambda^k \alpha^k(t_i)\bar{F}^k(t_i) + (1-\lambda^k)\xi^k \alpha^k(\xi^k t_i)\bar{F}^k(\xi^k t_i)}.$$

   Set $\mathbf{p}_i^k = (p_{i1}^k, 1 - p_{i1}^k)$.

2. Stochastic step: for each item $i \in \{1, \dots, n\}$ simulate $Z_i^k \sim \mathcal{M}\mathrm{ult}(1, \mathbf{p}_i^k)$. Then define subsets

$$\chi_j^k = \{i \in \{1, \dots, n\}; Z_i^k = j\} \quad \text{for } j = 1, 2.$$

Stochastic EM algorithm for the SRMM [3/4]

Facts: $\xi = \frac{E(X|Z=1)}{E(X|Z=2)}$ and if $S_j(s) = \mathbb{P}(X > s|Z = j)$ then $E(X|Z = j) = \int_0^{+\infty} S_j(s)ds$.

3. Update the euclidean parameters λ and ξ:

$$\begin{aligned} \lambda^{k+1} &= \frac{\mathrm{Card}(\chi_1^k)}{n}, \\ \xi^{k+1} &= \frac{\int_0^{\tau_1^k} \hat{S}_1^k(s)ds}{\int_0^{\tau_2^k} \hat{S}_2^k(s)ds}, \end{aligned}$$

where $\hat{S}_j^k$ is the Kaplan-Meier estimator for the subpopulation $\{(t_\ell, d_\ell); \ell \in \chi_j^k\}$ and $\tau_j^k = \max_{\ell \in \chi_j^k} t_\ell$.

Stochastic EM algorithm for the SRMM [4/4]

Fact: if X comes from component two (i.e. if $Z = 2$), then $\xi X \sim f$.

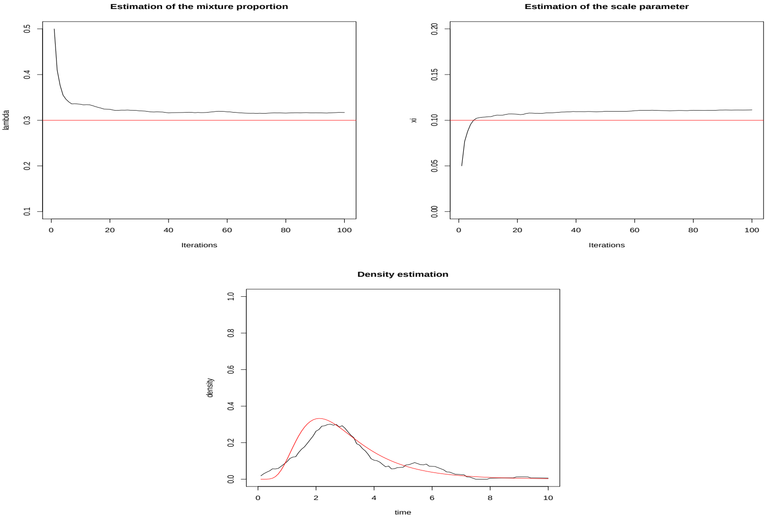
4. Update the functional parameters α and $\bar{F}$: set $\mathbf{t}^k = (t_1^k, \ldots, t_n^k)$ be the order statistic from $\{t_i; i \in \chi_1^k\} \cup \{\xi^k t_i; i \in \chi_2^k\}$; write $\mathbf{d}^k = (d_1^k, \ldots, d_n^k)$ the corresponding censoring indicators.

$$\begin{aligned}
\alpha^{k+1}(s) &= \sum_{i=1}^{n} \frac{1}{h}\mathcal{K}\left(\frac{s - t_i^k}{h}\right)\frac{d_i^k}{n-i+1}, \\
\bar{F}^{k+1}(s) &= \prod_{i:t_i^k \leq s}\left(1 - \frac{d_i^k}{n-i+1}\right),
\end{aligned}$$

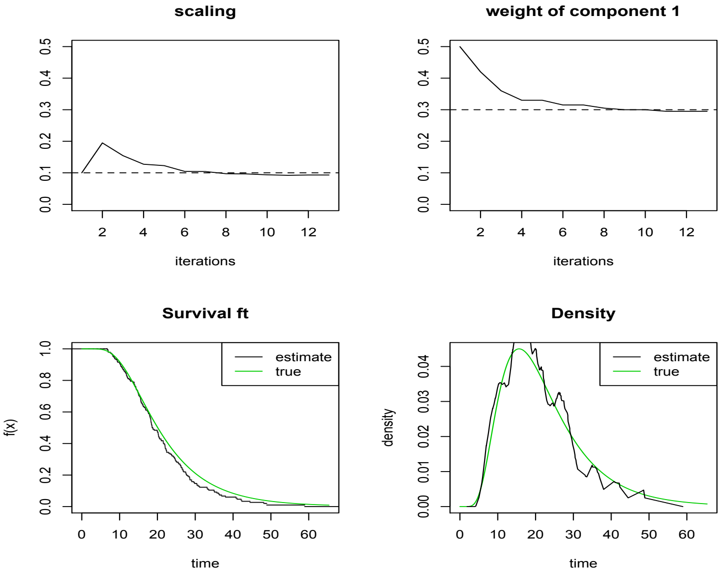
where $\mathcal{K}$ is a kernel function and h a bandwidth.

Remark: in practice the choice of both $\mathcal{K}$ and h is important!

Example: $g(x) = 0.3f(x) + 0.7\xi f(\xi x)$, $f \sim \mathcal{LN}(1, 0.5)$. Simulated sample: $n = 100$ with 0% of censoring.



Example (continued): $n = 200$ and 10% of censoring.



Conclusions

- All the algorithms introduced here have been/will be implemented in the publicly available package `mixtools` by Benaglia, Chauveau, Hunter and Young (2009) for the `R` statistical software (`R` Development Core Team, 2009).
- Asymptotic variances of the parametric St-EM estimators can be derived following Nielsen (2000).
- Many *tuning parameters* to improve. As an example, a local bandwidth choice should improve the semiparametric St-EM algorithm.

Thanks!