Large sample distribution of estimators in a mixture of semiparametric models

Yuichi Hirose
Joint project with
K. Preedalikit, I. Liu, N. Sibanda, and D. Fernández

School of Mathematics and Statistics,
Victoria University of Wellington,
New Zealand

November 17, 2015

Outline

(1) Introduction- mixture model and EM
(2) Profile likelihood estimation
(3) Example: Joint mixture model

Introduction: mixture model and EM (1)

We consider a mixture of semiparametric models

$$\sum_{k=1}^{K} \pi_k p_k(x; \theta_k, \eta_k)$$

where

- for each $k = 1, \ldots, K$,

  $$p_k(x; \theta_k, \eta_k)$$

  is a semiparametric model with finite dimensional parameter θ_k and infinite dimensional parameter η_k;
- $\pi_1, \ldots, \pi_K$ are mixture probabilities: $\pi_k > 0$ for each k and $\sum_{k=1}^{K} \pi_k = 1$.

Introduction: mixture model and EM (2)

Once we observe iid data $X_1, \ldots, X_n$ from the mixture model, the joint probability function of data $\mathbf{X} = (X_1, \ldots, X_n)$ is given by

$$p(\mathbf{X}|\theta, \eta, \pi) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k p_k(X_i; \theta_k, \eta_k). \tag{1}$$

where $\theta = (\theta_1, \ldots, \theta_K)$, $\eta = (\eta_1, \ldots, \eta_K)$ and $\pi = (\pi_1, \ldots, \pi_K)$.
We consider θ is the parameters of interest, and η and π are nuisance parameters.
We aim to establish large sample properties of the parameter θ using EM-algorithm and profile likelihood approach.

Introduction: mixture model and EM (3)

To discuss the EM-algorithm, we further introduce notations. Let

$$Z_i = (Z_{i1}, \ldots, Z_{iK})$$

be group indicator variable for the subject i: for each k,

$$Z_{ik} = 0 \text{ or } = 1 \text{ with } \sum_{k=1}^{K} Z_{ik} = 1.$$

Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$. The joint probability function of the complete data $(\mathbf{X}, \mathbf{Z})$ is

$$p(\mathbf{X}, \mathbf{Z} | \theta, \eta, \pi) = \prod_{i=1}^{n} \prod_{k=1}^{K} [\pi_k p_k(X_i; \theta_k, \eta_k)]^{Z_{ik}}. \tag{2}$$

Introduction: mixture model and EM (4)

Then the EM-algorithm utilize the identity

$$\begin{aligned} \log p(\mathbf{X};\theta,\eta,\pi) &= \sum_{\mathbf{Z}} q(\mathbf{Z}) \log p(\mathbf{X},\mathbf{Z};\theta,\eta,\pi) - \sum_{\mathbf{Z}} q(\mathbf{Z}) \log q(\mathbf{Z}) \\ &\quad - \sum_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{p(\mathbf{Z}|\mathbf{X};\theta,\eta,\pi)}{q(\mathbf{Z})}, \end{aligned} \tag{3}$$

where $q(\mathbf{Z})$ is a distribution of $\mathbf{Z}$ (Bishop (2006), Equation (9.70), section 9.4).

Introduction: mixture model and EM (5)

In the E-step put

$$q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X}; \theta, \eta, \pi),$$

then the third term in the right hand side is zero:

$$\sum_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{p(\mathbf{Z}|\mathbf{X}; \theta, \eta, \pi)}{q(\mathbf{Z})} = 0.$$

Introduction: mixture model and EM (6)

Under this condition, the equation (3) implies that maximizing the mixture log likelihood function

$$\log p(\mathbf{X}; \theta, \eta, \pi)$$

with respect to θ, η and π is the same as maximizing the expectation of complete data log likelihood function

$$\sum_{\mathbf{Z}} q(\mathbf{Z}) \log p(\mathbf{X}, \mathbf{Z} | \theta, \eta, \pi)$$

with respect to θ, η and π. Therefore in the M-step, we maximize the expectation of complete data log likelihood function.
In practice we must repeat E-step and M-step iteratively until we achieve the maximum.

Introduction: mixture model and EM (7)

In the following, we aim to establish asymptotic properties of the maximum likelihood estimator of θ using the expectation of complete data log likelihood function

$$\sum_{\mathbf{Z}} q(\mathbf{Z}) \log p(\mathbf{X}, \mathbf{Z} | \theta, \eta, \pi)$$

with the condition that $q(\mathbf{Z}) = p(\mathbf{Z} | \mathbf{X}; \theta, \eta, \pi)$.

Introduction: mixture model and EM (8)

The conditional distribution $p(\mathbf{Z}|\mathbf{X};\theta,\eta,\pi)$ in our iid setting is

$$\begin{aligned} p(\mathbf{Z}|\mathbf{X};\theta,\eta,\pi) &= \frac{p(\mathbf{X},\mathbf{Z};\theta,\eta,\pi)}{\sum_{\mathbf{Z}} p(\mathbf{X},\mathbf{Z};\theta,\eta,\pi)} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\left[\frac{\pi_k p_k(X_i;\theta_k,\eta_k)}{\sum_{k=1}^{K}\pi_k p_k(X_i;\theta_k,\eta_k)}\right]^{Z_{ik}} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\gamma(Z_{ik})^{Z_{ik}}. \end{aligned} \tag{4}$$

where

$$\gamma(Z_{ik}) = E(Z_{ik}|X_i) = \frac{\pi_k p_k(X_i;\theta_k,\eta_k)}{\sum_{j=1}^{K}\pi_j p_j(X_i;\theta_j,\eta_j)}. \tag{5}$$

Introduction: mixture model and EM (9)

Using equations (2) and (4), the expected complete data log-likelihood is

$$\begin{aligned}
& \sum_{\mathbf{Z}} q(\mathbf{Z}) \log p(\mathbf{X}, \mathbf{Z} | \theta, \eta, \pi) \\
= & \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma(Z_{ik})[\log \pi_k + \log p_k(X_i; \theta_k, \eta_k)]. \qquad (6)
\end{aligned}$$

where $\gamma(Z_{ik})$ is given in (5).

Introduction: mixture model and EM (10)

Because π and (θ, η) are separated by $+$ in (6), the estimation of π is not related to the estimation of (θ, η). The method of Lagrange multiplier can be used to get the MLE $\hat{\pi}_k$ of π_k:

$$\hat{\pi}_k = \frac{\sum_{i=1}^{n} \gamma(Z_{ik})}{n}. \tag{7}$$

We require that, as $n \to \infty$,

$$\hat{\pi}_k \xrightarrow{P} \pi_{0k}$$

where π_{0k}, $k = 1, \ldots, K$, are the true mixture probability.

Profile likelihood estimation (1)

In the estimation of (θ, η) we use the profile likelihood approach: we maximize the expected complete data log-likelihood with respect to η to obtain $\hat{\eta}_{\theta,F_n} = (\hat{\eta}_{1,\theta,F_n}, \ldots, \hat{\eta}_{K,\theta,F_n})$:

$$\hat{\eta}_{\theta,F_n} = \mathrm{argmax}_\eta \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma(Z_{ik}) \log p_k(X_i; \theta_k, \eta_k)$$

where we dropped the $\log \pi_k$ term from the expected complete data log-likelihood (6).

The profile log-likelihood is then

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \gamma(Z_{ik}) \log p_k(X_i; \theta_k, \hat{\eta}_{k,\theta,F_n}). \tag{8}$$

Profile likelihood estimation (2)

Define the score function for the profile log-likelihood in the model

$$\phi(X_i;\theta,F_n) = \frac{\partial}{\partial\theta}\sum_{k=1}^{K}\gamma(Z_{ik})\log p_k(X_i;\theta_k,\hat{\eta}_{k,\theta,F_n}) \tag{9}$$

Profile likelihood estimation (3)

We assume that:

(R0) $\hat{\eta}_{\theta,F}$ satisfies $\hat{\eta}_{\theta_0,F_0} = \eta_0$ and the function

$$\dot{\ell}^*_\theta(x, \theta_0) = \phi(x, \theta_0, F_0)$$

is the efficient score function, i,e,

$$E(\dot{\ell}^*_\theta \dot{\ell}_\eta) = 0$$

where $\dot{\ell}_\eta$ the score operator of η.

Profile likelihood estimation (4)

We assume that:

(R1) The $\sqrt{n}$-consistency of F_n, $\sqrt{n}\|F_n - F_0\| = O_P(1)$, and for each $(\theta, F) \in \Theta \times \mathcal{F}$, the log-likelihood function for a observation X_i

$$\log p(X_i;\theta, F) = \sum_{k=1}^{K} \gamma(Z_{ik}) \log p_k(X_i;\theta_k, \hat{\eta}_{k,\theta,F})$$

is twice continuously differentiable with respect to β and Hadamard differentiable with respect to F for all x.

(R1)* The empirical process F_n satisfies $n^{1/4}\|F_n - F_0\| = o_P(1)$, and for each $(\theta, F) \in \Theta \times \mathcal{F}$, the log-likelihood function for a observation X_i, $\log p(X_i;\theta, F)$, is twice continuously differentiable with respect to θ and twice Hadamard differentiable with respect to F.
(Derivatives are denoted by $\phi(X_i;\beta, F) = \frac{\partial}{\partial\theta} \log p(X_i;\theta, F)$, $\frac{\partial}{\partial\theta}\phi(X_i;\theta, F)$, $d_F\phi(X_i;\theta, F)$, and $d_F^2\phi(X_i;\theta, F)$.)

Profile likelihood estimation (5)

We assume that:

(R2) The efficient information matrix $I^*_\theta = E_{\theta_0,\eta_0}\dot{\ell}^*_\theta \dot{\ell}^{*T}_\theta = E_{\theta_0,\eta_0}\phi\phi^T(X,\theta_0,F_0)$ is invertible.

(R3) There exists a $\rho > 0$ and a neighborhood Θ of θ_0 such that the class of functions $\{\phi(x,\theta,F): \ (\theta,F)\in\Theta\times\mathcal{C}_\rho\}$ is P_{θ_0,η_0}-Donsker with square integrable envelope function, and such that the class of functions $\{\frac{\partial}{\partial\beta}\phi(x,\theta,F): \ (\theta,F)\in\Theta\times\mathcal{C}_\rho\}$ is P_{θ_0,η_0}-Glivenko-Cantelli with integrable envelope function.
Here

$$\mathcal{C}_\rho = \{F\in\mathcal{F} : \|F-F_0\| < \rho\}.$$

Profile likelihood estimation (6)

A consistent solution $\hat{\beta}_n$ to the estimating equation

$$\sum_{i=1}^{n} \phi(X_i, \hat{\theta}_n, F_n) = 0 \tag{10}$$

is an asymptotically linear estimator for θ_0 with the efficient influence function

$$\tilde{\ell}^*_\theta(x, \theta_0) = (I^*_\theta)^{-1} \dot{\ell}^*_\theta(x, \theta_0)$$

so that

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\ell}^*_\theta(X_i, \theta_0) + o_P(1) \\ &\xrightarrow{d} N\left(0, (I^*_\theta)^{-1}\right) \end{aligned}$$

where $I^*_\theta = E_{\theta_0,\eta_0}(\dot{\ell}^*_\theta \dot{\ell}^{*T}_\theta)$ is the efficient information matrix.

Joint mixture model (1)

- Let Y_{ijm} be the ordered categorical response from 1 (poor) to L (excellent) on item (or question) j for subject i at the m^{th} protocol-specified time point.
- In total, there are J items in the questionnaire, collected at times $t_1, t_2, \ldots, t_M$.
- Given that subject i belongs to group r, a stereotype model can be written as

$$\log\left[\frac{P(Y_{ijm}=\ell \mid \theta_r)}{P(Y_{ijm}=1 \mid \theta_r)}\right] = a_\ell + \phi_\ell(b_j+\theta_r), \quad r=1,\ldots,R. \tag{11}$$

Joint mixture model(2)

The ordinal response part of likelihood function for the ith subject is

$$P(\boldsymbol{Y}_i \mid \theta_r, \boldsymbol{\alpha}) = \prod_{m=1}^{M_i} \prod_{j=1}^{J} \prod_{\ell=1}^{L} \left(\frac{\exp(a_\ell + \phi_\ell(b_j + \theta_r))}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k(b_j + \theta_r))} \right)^{y_{ijm\ell}} \quad (12)$$

where $\boldsymbol{\alpha} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\phi})$.

Joint mixture model(3)

We consider the Cox proportional hazards model for the survival part in the joint model. Let X be a time-independent covariate. The hazard function for the failure time T_i of the i^{th} subject is of the form

$$\lambda(t|X_i, \theta_r, \boldsymbol{\delta}) \quad = \quad \lambda_0(t) \exp(\theta_r \delta_0 + X_i \delta_1)$$

where $\lambda_0(t)$ is the baseline hazard function. The latent variable θ_r is linked with the ordinal response model and $\boldsymbol{\delta} = (\delta_0, \delta_1)$ are coefficients.

Joint mixture model(4)

Assume that the hazard is zero between adjacent times so that the survival time is discrete. Let λ_i be the hazard at time t_i, where $t_1 < t_2 < \ldots < t_n$ are the ordered observed times.
The cumulative hazard function $\Lambda_0(t_i) = \sum_{p \leq i} \lambda_p$ is a step function with jumps at the failure time t_i. Then the the survival part likelihood function of subject i is

$$\begin{aligned} P\big(T_i, d_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\big) &= \big(\lambda_i \exp(\theta_r \delta_0 + X_i \delta_1)\big)^{d_i} \\ &\quad \times \exp\Big(- \sum_{p \leq i} \lambda_p \exp(\theta_r \delta_0 + X_i \delta_1)\Big). \end{aligned}$$

The d_i is an indicator of censorship for individual i: if we observe failure time, then $d_i = 1$, otherwise $d_i = 0$.

Joint mixture model(5)

Let π_r be the unknown probability $(r = 1, \ldots, R)$ that a subject lies in group r, and $\boldsymbol{\Theta}$ be all the unknown parameters of the joint model. The mixture model likelihood function is

$$L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}) = \prod_{i=1}^{n} \left(\sum_{r=1}^{R} P\big(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\big) P\big(T_i, d_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\big)\pi_r \right). \tag{13}$$

Let Z_{ir} be the group indicator, where $Z_{ir} = 1$ if the i^{th} individual was from the r^{th} group and 0 otherwise. The complete data likelihood can be written as

$$L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{d}, \boldsymbol{Z}) = \prod_{i=1}^{n} \prod_{r=1}^{R} \Big(P\big(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\big) P\big(T_i, d_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\big)\pi_r \Big)^{Z_{ir}}. \tag{14}$$

Joint mixture model(6)

The expected complete data log likelihood under $q(\mathbf{Z}) = P(\mathbf{Z}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{d})$ is

$$
\begin{aligned}
& \sum_{\mathbf{Z}} q(\mathbf{Z}) \log L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{d}, \boldsymbol{Z}) \\
= \;& \sum_{i=1}^{n} \sum_{r=1}^{R} \gamma(Z_{ir}) \log \pi_r \\
& + \sum_{i=1}^{n} \sum_{r=1}^{R} \gamma(Z_{ir}) \left\{ \log P\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right) + \log P\left(T_i, d_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right) \right\}
\end{aligned}
$$

Joint mixture model (7)

Before starting the EM-step, we profile out the baseline hazard function $\lambda_0(t)$. The survival part of equation (15) can be separately maximized with respect to $\boldsymbol{\lambda}$.
We find the maximizer $\widehat{\lambda}_i$ by holding $(\boldsymbol{\theta}, \boldsymbol{\delta})$ fixed, and it is given by

$$\widehat{\lambda}_i(\boldsymbol{\theta}, \boldsymbol{\delta}) = \frac{d_i}{\sum_{p \geq i} \sum_{r=1}^{R} \gamma(Z_{pr}) \exp(\theta_r \delta_0 + X_p \delta_1)}. \tag{15}$$

Denote $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta}) = (\widehat{\lambda}_i(\boldsymbol{\theta}, \boldsymbol{\delta}), \ldots, \widehat{\lambda}_n(\boldsymbol{\theta}, \boldsymbol{\delta}))$.

Joint mixture model (8)

The E-step: In the E-step, we use the current parameter estimates $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ to find the expected values of Z_{ir} of the complete data log likelihood:

$$\begin{aligned}
\gamma(Z_{ir}) &= E\big(Z_{ir} | \boldsymbol{Y}_i, T_i, d_i\big) \\
&= \frac{\pi_r P\big(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\big) P\big(T_i, d_i \,|\, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta}), \theta_r, \boldsymbol{\delta}\big)}{\sum_{g=1}^{R} \pi_g P\big(\boldsymbol{Y}_i \,|\, \theta_g, \boldsymbol{\alpha}\big) P\big(T_i, d_i \,|\, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta}), \theta_g, \boldsymbol{\delta}\big)}.
\end{aligned}$$

Joint mixture model (9)

The M-step: In the M-step, we maximize equation (15) with respect to π_r and $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta})$. Due to the fact that there is no relationship between π_r and $\boldsymbol{\Theta}$, they can be estimated separately.

1. Calculate the estimates of π_r

$$\widehat{\pi_r} = \frac{\sum_{i=1}^{n} \gamma(Z_{ir})}{n}.$$

2. We maximize the second and third parts of equation (15) (with $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta})$ in the place of $\boldsymbol{\lambda}$)

$$\sum_{i=1}^{n} \sum_{r=1}^{R} \gamma(Z_{ir}) \left\{ \log P\big(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\big) + \log P\Big(T_i, d_i \,|\, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta}), \theta_r, \boldsymbol{\delta}\Big) \right\} \quad (16)$$

with respect to $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ to obtain $\widehat{\boldsymbol{\Theta}}$.

The estimated parameters from the M-step are returned into the E-step until convergence.

Joint mixture model (Score function) (10)

The score function for $\boldsymbol{\delta}$ is

$$
\begin{aligned}
\dot{\ell}_{\boldsymbol{\delta},\Lambda} &= \frac{\partial}{\partial \boldsymbol{\delta}} \log P\big(T, d \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\big) \\
&= \sum_{r=1}^{R} \gamma(Z_r) \begin{pmatrix} \theta_r \\ X \end{pmatrix} \{d - \Lambda(T) \exp(\theta_r \delta_0 + X \delta_1)\}
\end{aligned}
$$

Joint mixture model (Score operator) (11)

Let $h(t)$ be a function of t. The path defined by

$$d\Lambda_s = (1 + sh)d\Lambda$$

is a submodel passing through Λ at $s = 0$. The corresponding path for the λ is

$$\lambda_s(t) = (1 + sh)\lambda(t).$$

The score operator for Λ is the derivative of the log-likelihood function with respect to s at $s = 0$:

$$\begin{aligned} B_{\boldsymbol{\delta},\Lambda}h &= \left.\frac{d}{ds}\right|_{s=0} \sum_{r=1}^{R} \gamma(Z_r) \log P\big(T, d \mid \Lambda_s, \theta_r, \boldsymbol{\delta}\big) \\ &= \sum_{r=1}^{R} \gamma(Z_r) \left(dh(T) - \exp(\theta_r \delta_0 + X\delta_1) \int_0^T h(u) d\Lambda(u) \right). \end{aligned}$$

The adjoint of the score operator is denoted by $B^*_{\boldsymbol{\delta},\Lambda}$.

Joint mixture model (Efficient score function) (12)

The (survival part of) efficient score function for the survival part given by

$$\begin{aligned}
\tilde{\ell}_{\boldsymbol{\delta},\Lambda} &= \dot{\ell}_{\boldsymbol{\delta},\Lambda} - B_{\boldsymbol{\delta},\Lambda}(B^*_{\boldsymbol{\delta},\Lambda}B_{\boldsymbol{\delta},\Lambda})^{-1}B^*_{\boldsymbol{\delta},\Lambda}\dot{\ell}_{\boldsymbol{\delta},\Lambda} \\
&= \sum_{r=1}^{R}\gamma(Z_r)d\left[\begin{pmatrix}\theta_r \\ X\end{pmatrix} - \frac{M_{1,\boldsymbol{\delta},\Lambda}(T)}{M_{0,\boldsymbol{\delta},\Lambda}(T)}\right] \\
&- \sum_{r=1}^{R}\gamma(Z_r)\left(\exp(\theta_r\delta_0 + X\delta_1)\int_0^T\left[\begin{pmatrix}\theta_r \\ X\end{pmatrix} - \frac{M_{1,\boldsymbol{\delta},\Lambda}(u)}{M_{0,\boldsymbol{\delta},\Lambda}(u)}\right]d\Lambda(u)\right)
\end{aligned}$$

where

$$\begin{aligned}
M_{1,\boldsymbol{\delta},\Lambda}(t) &= E_{\boldsymbol{\delta},\Lambda}\sum_{r=1}^{R}\gamma(Z_r)\begin{pmatrix}\theta_r \\ X\end{pmatrix}\exp(\theta_r\delta_0 + X\delta_1)I(t \le T) \\
M_{0,\boldsymbol{\delta},\Lambda}(t) &= E_{\boldsymbol{\delta},\Lambda}\sum_{r=1}^{R}\gamma(Z_r)\exp(\theta_r\delta_0 + X\delta_1)I(t \le T)
\end{aligned}$$

Joint mixture model(13)

The log-profile likelihood function for one observation is

$$\begin{aligned}
&\log p(\boldsymbol{Y}, T, d; \boldsymbol{\Theta}, F_n) \\
= \quad & \sum_{r=1}^{R} \gamma(Z_{ir}) \left\{ \log P\big(\boldsymbol{Y} \,|\, \theta_r, \boldsymbol{\alpha}\big) + \log P\big(T, d \,|\, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta}), \theta_r, \boldsymbol{\delta}\big) \right\}
\end{aligned}$$

The score function is

$$\begin{aligned}
&\phi(\boldsymbol{Y}, T, d; \boldsymbol{\Theta}, F_n) = \frac{\partial}{\partial \boldsymbol{\Theta}} \log p(\boldsymbol{Y}, T, d; \boldsymbol{\Theta}, F_n) \\
= \quad & \sum_{r=1}^{R} \gamma(Z_r) \left\{ \frac{\partial}{\partial \boldsymbol{\Theta}} \log P\big(\boldsymbol{Y} \,|\, \theta_r, \boldsymbol{\alpha}\big) + \frac{\partial}{\partial \boldsymbol{\Theta}} \log P\big(T, d \,|\, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta}, \boldsymbol{\delta}), \theta_r, \boldsymbol{\delta}\big) \right\}
\end{aligned}$$

Joint mixture model (14)

Need to verify

- $$\phi(\boldsymbol{Y}, T, d; \boldsymbol{\Theta}, F_0)$$

  is the efficient score function.

- There exists a $\rho > 0$ and a neighborhood Θ of θ_0 such that the class of functions

  $$\{\phi(\boldsymbol{Y}, T, d; \boldsymbol{\Theta}, F) : \ (\boldsymbol{\Theta}, F) \in \Theta \times \mathcal{C}_\rho\}$$

  is Donsker with square integrable envelope function, and such that the class of functions

  $$\{\frac{\partial}{\partial \boldsymbol{\Theta}} \phi(\boldsymbol{Y}, T, d; \boldsymbol{\Theta}, F) : \ (\boldsymbol{\Theta}, F) \in \Theta \times \mathcal{C}_\rho\}$$

  is Glivenko-Cantelli with integrable envelope function.

Joint mixture model (15)

Then from the general result, we have

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\partial \phi}{\partial \boldsymbol{\Theta}} \right)^{-1} \phi(\boldsymbol{Y}_i, T_i, d_i; \boldsymbol{\Theta}_0, F_0) + o_P(1) \\ &\xrightarrow{d} N\left(0, \left(\frac{\partial \phi}{\partial \boldsymbol{\Theta}} \right)^{-1} \right)\end{aligned}$$

Reference

Hirose, Y. (2010). Efficiency of profile likelihood in semi-parametric models, Ann. Inst. Statist. Math.

Preedalikit, K. Liu, I. Hirose, Y. Sibanda, N. and Fernández, D. (2015) Joint modeling of survival and longitudinal ordered data using a semiparametric approach, Aust. N. Z. J.

Bishop, C.M. (2006) Pattern recognition and machine learning, Springer.