应用概率统计　第二十四卷
第二期　2008年4月

Chinese Journal of Applied Probability
and Statistics Vol.24 No.2 Apr. 2008

右随机截断下混合分布系数的非参数推断 *

陆　福　忠

(嘉兴学院数学与信息工程学院, 嘉兴, 314001)

摘　要

对于混合分布模型$H=\lambda F+(1-\lambda)G$, 构造了在右随机截断下混合分布系数λ的估计量$\widehat{\lambda}$, 并证明了$\widehat{\lambda}$的渐进正态性.

关键词: 混合分布, 渐进正态, Hadamard可微, 影响函数.

学科分类号: O212.1.

§1. 引　　言

考虑非参数混合分布模型$H=\lambda F+(1-\lambda)G$, 其中F, G为两个不同的单变量分布, $0<\lambda<1$, 本文的目标是对混合分布系数λ作出估计和推断.

如果可以假定F, G为参数分布, 则对未知参数λ的推断就有很多方法可循, 具体可见[6, 第4页]. 本文不假定F, G为参数分布, 事实上, 不切实际的参数分布假定将不会得到混合分布系数λ的相合估计.

非参数混合分布模型的研究相对较少. 原因在于如果没有对于分布F, G的适当限制或是得到分布F, G的训练样本, 模型是不可识别的.

因而, 对于非参数混合分布模型$H=\lambda F+(1-\lambda)G$中λ的估计, 有两种方法. 一是对分布F, G作适当限制, 如[2]假定F, G为单变量对称分布, 且两者只相差一个位置参数. [1]和[3]讨论了如下m维变量混合分布模型$F(x)=\sum\limits_{g=1}^{G}\lambda_g P_g(x)$, 其中$G$已知, 且假定$P_g(x)=\prod\limits_{j=1}^{m}F_g(x_j)$, F_g为单变量分布, $g=1,2,\cdots,G$. 通过离散化观测数据将原混合分布模型转化为与之相应的一个混合多项分布模型, 两个模型中的λ一样, 从而通过研究混合多项分布模型得到λ的估计. 应当指出, [2]的假定使得模型可以归结为参数模型, 应用最小距离方法得到了λ的估计, [1, 3]的方法要求$m\geq 2G-1$以保证模型可以识别. 他们的方法不能应用到本文的单变量混合情况.

另一种方法是得到分布F,G的训练样本来得到F, G的经验分布. 如[4, 5, 11]. 再用最小距离方法得到λ的估计.

*嘉兴学院博士基金资助(70507016).
本文2005年10月10日收到, 2006年8月9日收到修改稿.

本文要求有分布F, G的训练样本, 从分布F中抽取n_1个独立样本, 从分布G中抽取n_2个独立样本, 从分布H中抽取$n_3 = n - n_1 - n_2$个独立样本, 目标是λ的估计. 本文与[4, 5, 11]等不同的是所有样本均是经过右随机截断后的样本. 在生存分析和可靠性实验时, 通常都只能得到截断样本. 一个实际应用是: 一批产品中有合格产品和不合格产品, 但是不知道它们的比例. 我们要估计这一比例. 假设我们可以通过随机抽样预先得到合格产品和不合格产品的生存时间大致的性状, 但是不假设它们的生存时间服从参数分布. 对于此问题, 据我所知, 目前尚无文献涉及.

本文构造了混合分布系数λ的强相合估计量$\widehat{\lambda}$, 并证明了$\widehat{\lambda}$的渐近正态性. 渐近正态性的证明所用方法是经典的可微统计量函数的渐近展开方法, 利用影响函数这一工具得到了$\sqrt{n}(\widehat{\lambda} - \lambda)$的渐近方差.

§2. 混合分布系数λ的估计量的构造

本文的模型为

$$H^0 = \lambda F^0 + (1 - \lambda)G^0. \tag{2.1}$$

F^0和G^0为混合模型H^0的组成成分分布, 设随机变量X^0服从分布F^0, Y^0服从分布G^0, Z^0服从分布H^0, 我们想要得到分别来自分布F^0, G^0, H^0的样本, 但由于受到随机右截断, 实际上得到如下观测量:

$$\begin{aligned}
&(X_i = \min(X_i^0, X_i^c),\ \delta_i^X = \mathbb{I}(X_i^0 \le X_i^c)), && i = 1, 2, \cdots, n_1,\\
&(Y_j = \min(Y_j^0, Y_j^c),\ \delta_j^Y = \mathbb{I}(Y_j^0 \le Y_j^c)), && j = 1, 2, \cdots, n_2,\\
&(Z_k = \min(Z_k^0, Z_k^c),\ \delta_k^Z = \mathbb{I}(Z_k^0 \le Z_k^c)), && k = 1, 2, \cdots, n_3.
\end{aligned}$$

其中, X_i^c服从分布C_F是独立于X_i^0的随机截断变量, Y_j^c是服从分布C_G独立于Y_j^0的截断变量, Z_k^c是服从分布C_H独立于Z_k^0的随机截断变量, $\mathbb{I}(\cdot)$为示性函数. 本文假定X_i, Y_j, Z_k两两相互独立.

由以上记号不难得到: 各X_i服从共同分布F满足$1 - F = (1 - F^0)(1 - C_F)$, 各$Y_j$服从共同分布$G$满足$1 - G = (1 - G^0)(1 - C_G)$, 各$Z_k$服从共同分布$H$满足$1 - H = (1 - H^0)(1 - C_H)$.

记各子分布函数

$$\begin{aligned}
&F^u(t) = \mathsf{P}(X_i \le t,\ \delta_i^X = 1), && F^c(t) = \mathsf{P}(X_i \le t,\ \delta_i^X = 0).\\
&G^u(t) = \mathsf{P}(Y_j \le t,\ \delta_j^Y = 1), && G^c(t) = \mathsf{P}(Y_j \le t,\ \delta_j^Y = 0).\\
&H^u(t) = \mathsf{P}(Z_k \le t,\ \delta_k^Z = 1), && H^c(t) = \mathsf{P}(Z_k \le t,\ \delta_k^Z = 0).
\end{aligned}$$

其各自对应的子经验分布函数为:

$$F^u_{n_1}(t)=\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{I}(X_i\le t,\delta_i^X=1),\qquad F^c_{n_1}(t)=\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{I}(X_i\le t,\delta_i^X=0).$$
$$G^u_{n_2}(t)=\frac{1}{n_2}\sum_{j=1}^{n_2}\mathbb{I}(Y_j\le t,\delta_j^Y=1),\qquad G^c_{n_2}(t)=\frac{1}{n_2}\sum_{j=1}^{n_2}\mathbb{I}(Y_j\le t,\delta_j^Y=0).$$
$$H^u_{n_3}(t)=\frac{1}{n_3}\sum_{k=1}^{n_3}\mathbb{I}(Z_k\le t,\delta_k^Z=1),\qquad H^c_{n_3}(t)=\frac{1}{n_3}\sum_{k=1}^{n_3}\mathbb{I}(Z_k\le t,\delta_k^Z=0).$$

应用Peterson展开表达式[7]中的记号, 记生存函数

$$S^u_{F^0}(t)=1-F^0(t),\qquad S^u_{G^0}(t)=1-G^0(t),\qquad S^u_{H^0}(t)=1-H^0(t).$$
$$S^u_F(t)=1-F(t),\qquad S^u_G(t)=1-G(t),\qquad S^u_H(t)=1-H(t).$$

则子生存函数表示为:

$$S^u_F(t)=\mathsf{P}(X_i>t,\,\delta_i^X=1),\qquad S^c_F(t)=\mathsf{P}(X_i>t,\,\delta_i^X=0).$$
$$S^u_G(t)=\mathsf{P}(Y_j>t,\,\delta_j^Y=1),\qquad S^c_G(t)=\mathsf{P}(Y_j>t,\,\delta_j^Y=0).$$
$$S^u_H(t)=\mathsf{P}(Z_k>t,\,\delta_k^Z=1),\qquad S^c_H(t)=\mathsf{P}(Z_k>t,\,\delta_k^Z=0).$$

而经验子生存函数表示为:

$$S^u_{F_{n_1}}(t)=\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{I}(X_i>t,\delta_i^X=1),\qquad S^c_{F_{n_1}}(t)=\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{I}(X_i>t,\delta_i^X=0).$$
$$S^u_{G_{n_2}}(t)=\frac{1}{n_2}\sum_{j=1}^{n_2}\mathbb{I}(Y_j>t,\delta_j^Y=1),\qquad S^c_{G_{n_2}}(t)=\frac{1}{n_2}\sum_{j=1}^{n_2}\mathbb{I}(Y_j>t,\delta_j^Y=0).$$
$$S^u_{H_{n_3}}(t)=\frac{1}{n_3}\sum_{k=1}^{n_3}\mathbb{I}(Z_k>t,\delta_k^Z=1),\qquad S^c_{H_{n_3}}(t)=\frac{1}{n_3}\sum_{k=1}^{n_3}\mathbb{I}(Z_k>t,\delta_k^Z=0).$$

于是S_{F^0},S_{G^0},S_{H^0}的Kaplan-Meier估计量定义为:

$$\widehat{S}_{F^0}(t)=1-\widehat{F}^0(t)\triangleq\prod_{i:X_i\le t}\Big(\frac{n_1-i}{n_1-i+1}\Big)^{\delta^X_{(i)}},\qquad\forall t\in\mathbf{R},$$
$$\widehat{S}_{G^0}(t)=1-\widehat{G}^0(t)\triangleq\prod_{j:Y_j\le t}\Big(\frac{n_2-j}{n_2-j+1}\Big)^{\delta^Y_{(j)}},\qquad\forall t\in\mathbf{R},$$
$$\widehat{S}_{H^0}(t)=1-\widehat{H}^0(t)\triangleq\prod_{k:Z_k\le t}\Big(\frac{n_3-k}{n_3-k+1}\Big)^{\delta^Z_{(k)}},\qquad\forall t\in\mathbf{R}.$$

这里, $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n_1)}$, $Y_{(1)}\le Y_{(2)}\le\cdots\le Y_{(n_2)}$, $Z_{(1)}\le Z_{(2)}\le\cdots\le Z_{(n_3)}$为次序统计量.

由模型(2.1)得知$\lambda=(F^0-G^0)^{-1}\times(H^0-G^0)$. 由于$F^0,G^0$为不同分布, 故存在$x_0$, 使得$F^0(x_0)\ne G^0(x_0)$. 本文用Kaplan-Meier估计$\widehat{F}^0,\widehat{G}^0,\widehat{H}^0$分别估计分布$F^0,G^0,H^0$, 由K-M估计的性质知道这些估计是强相合的. 之后用$\widehat{\lambda}=(\widehat{F}^0-\widehat{G}^0)^{-1}(x_0)\times(\widehat{H}^0-\widehat{G}^0)(x_0)$作为未知参数$\lambda$的估计量.

记$\widehat{\lambda}=T(\widehat{F}^0,\widehat{G}^0,\widehat{H}^0)\triangleq(\widehat{F}^0-\widehat{G}^0)^{-1}(x_0)\times(\widehat{H}^0-\widehat{G}^0)(x_0)$, 则$\lambda=T(F^0,G^0,H^0)$.

由Peterson展式[7]:

$$\begin{aligned}S_{F^0}(t) &= 1-F^0(t)\triangleq T^0(S_F^u,S_F^c,t)\\ &= \exp\Big(\int_{-\infty}^{t}\frac{\mathrm{d}S_F^u(s)}{(S_F^u+S_F^c)(s)}\Big)\times\exp\Big(\sum_{s\le t}\ln\Big(\frac{(S_F^u+S_F^c)(s+)}{(S_F^u+S_F^c)(s-)}\Big)\Big).\end{aligned}$$

其中积分区域为属于$(-\infty,t)$的$S_F^u(\cdot)$为连续的所有开区间之并, 求和区域为$S_F^c(\cdot)$不连续的点集.

$S_{F^0}(t)$的经验估计为

$$\widehat{S}_{F_{n_1}^0}(t)=\exp\Big(\int_{-\infty}^{t}\frac{\mathrm{d}S_{F_{n_1}}^u(s)}{(S_{F_{n_1}}^u+S_{F_{n_1}}^c)(s)}\Big)\times\exp\Big(\sum_{s\le t}\ln\Big(\frac{(S_{F_{n_1}}^u+S_{F_{n_1}}^c)(s+)}{(S_{F_{n_1}}^u+S_{F_{n_1}}^c)(s-)}\Big)\Big).$$

积分区域及求和区域类似上式规定.

同样类似定义$S_{G^0}(t),S_{H^0}(t)$及其估计量$\widehat{S}_{G_{n_2}^0}(t),\widehat{S}_{H_{n_3}^0}(t)$.

则λ和$\widehat{\lambda}$可以分别表示为

$$\begin{aligned}\widehat{\lambda} &= T(\widehat{F}^0,\widehat{G}^0,\widehat{H}^0)\\ &= T(T^0(S_{F_{n_1}}^u,S_{F_{n_1}}^c,t),T^0(S_{G_{n_2}}^u,S_{G_{n_2}}^c,t),T^0(S_{H_{n_3}}^u,S_{H_{n_3}}^c,t))\\ &\triangleq V(S_{F_{n_1}}^u,S_{F_{n_1}}^c,S_{G_{n_2}}^u,S_{G_{n_2}}^c,S_{H_{n_3}}^u,S_{H_{n_3}}^c),\\ \lambda &= V(S_F^u,S_F^c,S_G^u,S_G^c,S_H^u,S_H^c).\end{aligned}$$

由于估计量$\widehat{\lambda}$是关于Kaplan-Meier估计$\widehat{F}^0,\widehat{G}^0,\widehat{H}^0$的函数, 为保证$\widehat{\lambda}$的存在及要证明的渐近正态性, 必须对随机截断变量作些限制, [10]有如下结果(将其应用到本文中的F^0及C_F得到如下定理).

定理 2.1 (Stute and Wang Theorem) 设F^0和C_F没有公共跳跃点, $\tau_{F^0}\le\tau_{C_F}$, 其中$\tau_{F^0}\triangleq\min\{t:F^0(t)=1\}$, $\phi(x)$为直线上Borel可测函数, 满足$\int|\phi(x)|\mathrm{d}F^0(x)<\infty$. 如果有$F^0(\tau_{F^0}-)=1$或$C_F(\tau_{F^0}-)<1$, 则有

$$\lim_{n_1\to\infty}\int_{-\infty}^{\infty}\phi(x)\mathrm{d}\widehat{F}^0(x)\to\int_{-\infty}^{\infty}\phi(x)\mathrm{d}F^0(x)\qquad\text{a.s..}$$

因此本文总假设:

1. F^0和C_F无公共跳跃点, $\tau_{F^0}\le\tau_{C_F}$, $F^0(\tau_{F^0}-)=1$或$C_F(\tau_{F^0}-)<1$.
2. G^0和C_G无公共跳跃点, $\tau_{G^0}\le\tau_{C_G}$, $G^0(\tau_{G^0}-)=1$或$C_G(\tau_{G^0}-)<1$.
3. H^0和C_H无公共跳跃点, $\tau_{H^0}\le\tau_{C_H}$, $H^0(\tau_{H^0}-)=1$或$C_H(\tau_{H^0}-)<1$.

一般而言, 以上条件是容易满足的, $\tau_{F^0}\le\tau_{C_F}$, $\tau_{G^0}\le\tau_{C_G}$, $\tau_{H^0}\le\tau_{C_H}$是在右截断情况下自然的条件, 其余的条件如当F^0,G^0,H^0连续时显然成立.

§3. 统计量函数T, V的可微性证明及其相应的影响函数的表达

本节证明统计量函数T的Fréchet可微性及统计量函数V的Hadamard可微性, 然后推广[8]中对于多元复合映射的影响函数的链法则公式, 求出统计量函数V的影响函数.

先证明$\lambda = T(F^0, G^0, H^0)$中的函数$T$在$(F^0, G^0, H^0)$处是Fréchet可微的. (Fréchet可微的定义见[8, 9])

命题 3.1 统计量函数$T = (F^0 - G^0)^{-1} \times (H^0 - G^0)$在$(F^0, G^0, H^0)$处关于最大范数$\|\cdot\|_\infty$是Fréchet可微的.

证明: 只需验证T满足Fréchet可微的定义. 即: 当$\|F_m - F^0\|_\infty \to 0$, $\|G_n - G^0\|_\infty \to 0$, $\|H_l - H^0\|_\infty \to 0$时, 存在线性函数$d_1T, d_2T, d_3T$, 使得$T(F_m, G_n, H_l) - T(F_0, G_0, H_0) - d_1T(F_m - F^0, G^0, H^0) - d_2T(F^0, G_n - G^0, H^0) - d_3T(F^0, G^0, H_l - H^0)$是比$\|F_m - F^0\|_\infty$, $\|G_n - G^0\|_\infty$, $\|H_l - H^0\|_\infty$更高阶的无穷小.

取线性函数d_1T, d_2T, d_3T, 使得

$$
\begin{aligned}
&d_1T(F_m - F^0, G^0, H^0) = -\lambda \times (F_m - F^0)(x_0)/(F^0 - G^0)(x_0),\\
&d_2T(F^0, G_n - G^0, H^0) = (\lambda - 1) \times (G_n - G^0)(x_0)/(F^0 - G^0)(x_0),\\
&d_3T(F^0, G^0, H_l - H^0) = (H_l - H^0)(x_0)/(F^0 - G^0)(x_0).
\end{aligned}
$$

显然d_1T, d_2T, d_3T均是相应变量的线性函数. 将其表达式代入, 计算后整理得余项为

$$
\begin{aligned}
\mathrm{Rem} &= T(F_m, G_n, H_l) - T(F^0, G^0, H^0)\\
&\quad - d_1T(F_m - F^0, G^0, H^0) - d_2T(F^0, G_n - G^0, H^0) - d_3T(F^0, G^0, H_l - H^0)\\
&= \frac{((F^0 - G^0)(H_l - H^0) + (G^0 - H^0)(F_m - F^0) + (H^0 - F^0)(G_n - G^0))(x_0)}{(F^0 - G^0)(x_0)}\\
&\quad \times \frac{(F_m - F^0)(x_0) + (G_n - G^0)(x_0)}{(F_m - G_n)(x_0) \times (F^0 - G^0)(x_0)}.
\end{aligned}
$$

由于余项的表达式是若干个两项无穷小的乘积之和, 显然它是较之$\|F_m - F^0\|_\infty$, $\|G_n - G^0\|_\infty$, $\|H_l - H^0\|_\infty$更高阶的无穷小. □

由于Fréchet可微蕴涵Hadamard可微, 因而函数T也是Hadamard可微的. 且d_1T, d_2T, d_3T分别是T对于三个变量的微分. 根据[8]中影响函数的定义, 可以得到如下关系表达式

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\varepsilon} T(F^0 + \varepsilon(F_m - F^0), G^0, H^0)|_{\varepsilon=0} &= d_1T(F_m - F^0, G^0, H^0)\\
&= \int IC_1(T, F^0, G^0, H^0, x)\mathrm{d}(F_m - F^0)(x),\\
\frac{\mathrm{d}}{\mathrm{d}\varepsilon} T(F^0, G^0 + \varepsilon(G_n - G^0), H^0)|_{\varepsilon=0} &= d_2T(F^0, G_n - G^0, H^0)\\
&= \int IC_2(T, F^0, G^0, H^0, x)\mathrm{d}(G_n - G^0)(x),
\end{aligned}
$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}T(F^0,G^0,H^0+\varepsilon(H_l-H^0))|_{\varepsilon=0} &= d_3T(F^0,G^0,H^l-H^0)\\ &= \int IC_3(T,F^0,G^0,H^0,x)\mathrm{d}(H_l-H^0)(x).\end{aligned}$$

由于已经求出了d_1T,d_2T,d_3T, 从而可以解出T对于三个变量的影响函数分别为:

$$\begin{aligned}&IC_1(T,F^0,G^0,H^0,x)=-\lambda\times(F^0-G^0)^{-1}(x_0)\times\mathbb{I}(x\le x_0),\\ &IC_2(T,F^0,G^0,H^0,x)=(\lambda-1)\times(F^0-G^0)^{-1}(x_0)\times\mathbb{I}(x\le x_0),\\ &IC_3(T,F^0,G^0,H^0,x)=(F^0-G^0)^{-1}(x_0)\times\mathbb{I}(x\le x_0).\end{aligned}\tag{3.1}$$

统计量函数T^0的Hadamard可微性已在[8]中获得, 其对应的影响函数为:

$$\begin{aligned}&IC_1(T^0,S_F^u,S_F^c,s)(t)=S_{F^0}(t)\times\Big(\frac{\mathbb{I}(s\le t)}{S_F(s)}+\int_{-\infty}^{s\wedge t}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\Big),\\ &IC_2(T^0,S_F^u,S_F^c,s)(t)=S_{F^0}(t)\times\int_{-\infty}^{s\wedge t}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)},\\ &IC_1(T^0,S_G^u,S_G^c,s)(t)=S_{G^0}(t)\times\Big(\frac{\mathbb{I}(s\le t)}{S_G(s)}+\int_{-\infty}^{s\wedge t}\frac{\mathrm{d}S_G^u(u)}{S_G^2(u)}\Big),\\ &IC_2(T^0,S_G^u,S_G^c,s)(t)=S_{G^0}(t)\times\int_{-\infty}^{s\wedge t}\frac{\mathrm{d}S_G^u(u)}{S_G^2(u)},\\ &IC_1(T^0,S_H^u,S_H^c,s)(t)=S_{H^0}(t)\times\Big(\frac{\mathbb{I}(s\le t)}{S_H(s)}+\int_{-\infty}^{s\wedge t}\frac{\mathrm{d}S_H^u(u)}{S_H^2(u)}\Big),\\ &IC_2(T^0,S_H^u,S_H^c,s)(t)=S_{H^0}(t)\times\int_{-\infty}^{s\wedge t}\frac{\mathrm{d}S_H^u(u)}{S_H^2(u)}.\end{aligned}$$

由Hadamard可微统计量函数的链法则易知,

$$V(S_F^u,S_F^c,S_G^u,S_G^c,S_H^u,S_H^c)=T(T^0(S_F^u,S_F^c,t),T^0(S_G^u,S_G^c,t),T^0(S_H^u,S_H^c,t))$$

在$(S_F^u,S_F^c,S_G^u,S_G^c,S_H^u,S_H^c)$处是Hadamard可微的, 用链法则求复合函数的影响函数, 则

$$\begin{aligned}&d_1V(S_{F_{n-1}}^u-S_F^u,S_F^c,\cdots,S_H^c)=d_1T(\widehat{S}_{F^0}-S_{F^0},S_{G^0},S_{H^0})\circ d_1T^0(S_{F_{n_1}}^u-S_F^u,S_F^c,s)(t),\\ &d_2V(S_F^u,S_{F_{n-1}}^c-S_F^c,\cdots,S_H^c)=d_1T(\widehat{S}_{F^0}-S_{F^0},S_{G^0},S_{H^0})\circ d_2T^0(S_F^u,S_{F_{n_1}}^c-S_F^c,s)(t),\\ &d_3V(\cdots,S_{G_{n-2}}^u-S_G^u,\cdots,S_H^c)=d_2T(S_{F^0},\widehat{S}_{G^0}-S_{G^0},S_{H^0})\circ d_1T^0(S_{G_{n_2}}^u-S_G^u,S_G^c,s)(t),\\ &d_4V(\cdots,S_{G_{n_2}}^c-S_G^c\cdots,S_H^c)=d_2T(S_{F^0},\widehat{S}_{G^0}-S_{G^0},S_{H^0})\circ d_2T^0(S_G^u,S_{G_{n_2}}^c-S_G^c,s)(t),\\ &d_5V(S_F^u,\cdots,S_{H_{n_2}}^u-S_H^u,S_H^c)=d_3T(\widehat{S}_{F^0},S_{G^0},\widehat{S}_{H^0}-S_{H^0})\circ d_1T^0(S_{H_{n_3}}^u-S_H^u,S_H^c,s)(t),\\ &d_6V(S_F^u,\cdots,S_H^u,S_{H_{n_3}}^c-S_H^c)=d_3T(\widehat{S}_{F^0},S_{G^0},\widehat{S}_{H^0}-S_{H^0})\circ d_2T^0(S_H^u,S_{H_{n_3}}^c-S_H^c,s)(t).\end{aligned}$$

其中, 记号"$\circ$"表示函数的复合. 由微分与影响函数的关系

$$d_1V(S)=\int IC_1(V,x)\mathrm{d}S(x)$$

知(简记$IC_1(V,S_F^u,S_F^c,S_G^u,S_G^c,S_H^u,S_H^c,x)$为$IC_1(V,x)$), 则

$$\begin{aligned}
IC_1(V,x) &= \int_{-\infty}^{\infty} IC_1(T^0,S_F^u,S_F^c,x)(y)\mathrm{d}IC_1(T,F^0,G^0,H^0,y),\\
IC_2(V,x) &= \int_{-\infty}^{\infty} IC_2(T^0,S_F^u,S_F^c,x)(y)\mathrm{d}IC_1(T,F^0,G^0,H^0,y),\\
IC_3(V,x) &= \int_{-\infty}^{\infty} IC_1(T^0,S_G^u,S_G^c,x)(y)\mathrm{d}IC_2(T,F^0,G^0,H^0,y),\\
IC_4(V,x) &= \int_{-\infty}^{\infty} IC_2(T^0,S_G^u,S_G^c,x)(y)\mathrm{d}IC_2(T,F^0,G^0,H^0,y),\\
IC_5(V,x) &= \int_{-\infty}^{\infty} IC_1(T^0,S_H^u,S_H^c,x)(y)\mathrm{d}IC_3(T,F^0,G^0,H^0,y),\\
IC_6(V,x) &= \int_{-\infty}^{\infty} IC_2(T^0,S_H^u,S_H^c,x)(y)\mathrm{d}IC_3(T,F^0,G^0,H^0,y).
\end{aligned}$$

以上计算用过一次分部积分, 因为$IC_i(T^0,x)(\pm\infty)=0,\ i=1,2,$ 故$IC_i(T^0,x)(\pm\infty)\times IC_j(T,y)=0,\ i=1,2,\ j=1,2,3.$

§4. 主要结果

上节已经得到$\widehat{\lambda}=V(S_{F_{n_1}}^u,S_{F_{n_1}}^c,S_{G_{n_2}}^u,S_{G_{n_2}}^c,S_{H_{n_3}}^u,S_{H_{n_3}}^c)$及$\lambda=V(S_F^u,S_F^c,S_G^u,S_G^c,S_H^u,S_H^c)$, 还证明了函数$V$的Hadamard可微性, 所以可以将$\widehat{\lambda}$展开为

$$\begin{aligned}
\widehat{\lambda} \;=\; & \lambda+\int_{-\infty}^{\infty} IC_1(V,y)\mathrm{d}(S_{F_{n_1}}^u-S_F^u)(y)+\int_{-\infty}^{\infty} IC_2(V,y)\mathrm{d}(S_{F_{n_1}}^c-S_F^c)(y)\\
&+\int_{-\infty}^{\infty} IC_3(V,y)\mathrm{d}(S_{G_{n_2}}^u-S_G^u)(y)+\int_{-\infty}^{\infty} IC_4(V,y)\mathrm{d}(S_{G_{n_2}}^c-S_G^c)(y)\\
&+\int_{-\infty}^{\infty} IC_5(V,y)\mathrm{d}(S_{H_{n_3}}^u-S_H^u)(y)+\int_{-\infty}^{\infty} IC_6(V,y)\mathrm{d}(S_{H_{n_3}}^c-S_H^c)(y)+\mathrm{Rem}.
\end{aligned}$$

类似命题3.1的讨论, 知道余项Rem为高阶无穷小, 故忽略Rem后得到

$$\begin{aligned}
\widehat{\lambda} \;=\; & \lambda-\frac{1}{n_1}\sum_{i:\delta_i^X=1} IC_1(V,X_i)-\int_{-\infty}^{\infty} IC_1(V,y)\mathrm{d}S_F^u(y)-\frac{1}{n_1}\sum_{i:\delta_i^X=0} IC_2(V,X_i)\\
&-\int_{-\infty}^{\infty} IC_2(V,y)\mathrm{d}S_F^c(y)-\frac{1}{n_2}\sum_{j:\delta_j^Y=1} IC_3(V,Y_j)-\int_{-\infty}^{\infty} IC_3(V,y)\mathrm{d}S_G^u(y)\\
&-\frac{1}{n_2}\sum_{j:\delta_j^Y=0} IC_4(V,Y_j)-\int_{-\infty}^{\infty} IC_4(V,y)\mathrm{d}S_G^c(y)-\frac{1}{n_3}\sum_{k:\delta_k^Z=1} IC_5(V,Z_k)\\
&-\int_{-\infty}^{\infty} IC_5(V,y)\mathrm{d}S_H^u(y)-\frac{1}{n_3}\sum_{k:\delta_k^Z=0} IC_6(V,Z_k)-\int_{-\infty}^{\infty} IC_6(V,y)\mathrm{d}S_H^c(y). \quad (4.1)
\end{aligned}$$

经计算得出

$$\begin{aligned}
&\int_{-\infty}^{\infty} IC_1(V,y)\mathrm{d}S_F^u(y)+\int_{-\infty}^{\infty} IC_2(V,y)\mathrm{d}S_F^c(y)=0,\\
&\int_{-\infty}^{\infty} IC_3(V,y)\mathrm{d}S_G^u(y)+\int_{-\infty}^{\infty} IC_4(V,y)\mathrm{d}S_G^c(y)=0,\\
&\int_{-\infty}^{\infty} IC_5(V,y)\mathrm{d}S_H^u(y)+\int_{-\infty}^{\infty} IC_6(V,y)\mathrm{d}S_H^c(y)=0.
\end{aligned}$$

从而

$$\begin{aligned}
&\mathsf{E}\Big(\sum_{i:\delta_i^X=1} IC_1(V,X_i)+\sum_{i:\delta_i^X=0} IC_2(V,X_i)\Big)=0,\\
&\mathsf{E}\Big(\sum_{j:\delta_j^Y=1} IC_3(V,Y_j)+\sum_{j:\delta_j^Y=0} IC_4(V,Y_j)\Big)=0,\\
&\mathsf{E}\Big(\sum_{k:\delta_k^Z=1} IC_5(V,Z_k)+\sum_{k:\delta_k^Z=0} IC_6(V,Z_k)\Big)=0.
\end{aligned}\tag{4.2}$$

因而(4.1)式为

$$\begin{aligned}
\widehat{\lambda} \;=\; &\lambda-\frac{1}{n_1}\sum_{i:\delta_i^X=1} IC_1(V,X_i)-\frac{1}{n_1}\sum_{i:\delta_i^X=0} IC_2(V,X_i)\\
&-\frac{1}{n_2}\sum_{j:\delta_j^Y=1} IC_3(V,Y_j)-\frac{1}{n_2}\sum_{j:\delta_j^Y=0} IC_4(V,Y_j)\\
&-\frac{1}{n_3}\sum_{k:\delta_k^Z=1} IC_5(V,Z_k)-\frac{1}{n_3}\sum_{k:\delta_k^Z=0} IC_6(V,Z_k).
\end{aligned}$$

若$n=n_1+n_2+n_3\to\infty$时, 有

$$\frac{n_1}{n}\to C_1,\qquad \frac{n_2}{n}\to C_2,\qquad \frac{n_3}{n}\to C_3,$$

其中C_1, C_2, C_3为常数, 则由于$S_{F_{n_1}}^u$, $S_{F_{n_1}}^c$, $S_{G_{n_2}}^u$, $S_{G_{n_2}}^c$, $S_{H_{n_3}}^u$, $S_{H_{n_3}}^c$均为子经验分布, 应用中心极限定理知道$\sqrt{n}(S_{F_{n_1}}^u-S_F^u)$, $\sqrt{n}(S_{F_{n_1}}^c-S_F^c)$, $\sqrt{n}(S_{G_{n_2}}^u-S_G^u)$, $\sqrt{n}(S_{G_{n_2}}^c-S_G^c)$, $\sqrt{n}(S_{H_{n_3}}^u-S_H^u)$, $\sqrt{n}(S_{H_{n_3}}^c-S_H^c)$均为零均值的有限方差正态分布. 又因为$X_i,Y_j,Z_k$的两两独立性, 得到$\sqrt{n}(S_{F_{n_1}}^u-S_F^u, S_{F_{n_1}}^c-S_F^c, S_{G_{n_2}}^u-S_G^u, S_{G_{n_2}}^c-S_G^c, S_{H_{n_3}}^u-S_H^u, S_{H_{n_3}}^c-S_H^c)$服从零均值的多元正态分布. 由于已证明函数$V$的Hadamard可微性, 则此时函数$V$的作用实际上由其线性微分$\mathrm{d}V$所决定, 而多元正态分布的线性组合仍然是一正态分布. 故

$$\sqrt{n}(\widehat{\lambda}-\lambda)=\sqrt{n}(V(S_{F_{n_1}}^u,S_{F_{n_1}}^c,S_{G_{n_2}}^u,S_{G_{n_2}}^c,S_{H_{n_3}}^u,S_{H_{n_3}}^c)-V(S_F^u,S_F^c,S_G^u,S_G^c,S_H^u,S_H^c))$$

的渐近正态性由Function Delta Method所保证, 可参阅[12, 第296页]. 即有$\sqrt{n}(\widehat{\lambda}-\lambda)\to_{\mathfrak{D}} N(0,\sigma^2)$, 其中"$\mathfrak{D}$"表示依分布收敛. 渐近方差$\sigma^2$在下文给出. 显然, 由函数$V$的Hadamard可微性及上面的推导可知, $\widehat{\lambda}$是λ的强相合估计.

接下来求渐近方差σ^2. 应用式(4.2), 经计算得:

$$\begin{aligned}\mathsf{E}(\widehat{\lambda}-\lambda)^2 &= \mathsf{E}\Big(-\frac{1}{n_1}\sum_{i:\delta_i^X=1}IC_1(V,X_i)-\frac{1}{n_1}\sum_{i:\delta_i^X=0}IC_2(V,X_i)\\ &\quad -\frac{1}{n_2}\sum_{j:\delta_j^Y=1}IC_3(V,Y_j)-\frac{1}{n_2}\sum_{j:\delta_j^Y=0}IC_4(V,Y_j)\\ &\quad -\frac{1}{n_3}\sum_{k:\delta_k^Z=1}IC_5(V,Z_k)-\frac{1}{n_3}\sum_{k:\delta_k^Z=0}IC_6(V,Z_k)\Big)^2\\ &= \frac{1}{n_1}\Big(\int_{-\infty}^{\infty}IC_1^2(V,y)\mathrm{d}F^u(y)+\int_{-\infty}^{\infty}IC_2^2(V,y)\mathrm{d}F^c(y)\Big)\\ &\quad +\frac{1}{n_2}\Big(\int_{-\infty}^{\infty}IC_3^2(V,y)\mathrm{d}G^u(y)+\int_{-\infty}^{\infty}IC_4^2(V,y)\mathrm{d}F^c(y)\Big)\\ &\quad +\frac{1}{n_3}\Big(\int_{-\infty}^{\infty}IC_5^2(V,y)\mathrm{d}H^u(y)+\int_{-\infty}^{\infty}IC_6^2(V,y)\mathrm{d}H^c(y)\Big).\end{aligned}$$

上式成立是因为本文假设X_i,Y_j,Z_k两两相互独立, 所以各交差项求期望后都为零. 将$IC_i\cdot(V,x)$, $i=1,2,\cdots,6$的表达式代入上式, 有

$$\begin{aligned}\mathsf{E}(\widehat{\lambda}-\lambda)^2 &= \frac{1}{n_1}\Big(\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_1(T^0,x)(y)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x)\\ &\quad +\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_2(T^0,x)(y)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^c(x)\Big)\\ &\quad +\frac{1}{n_2}\Big(\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_1(T^0,x)(y)\mathrm{d}IC_2(T,y)\Big)^2\mathrm{d}G^u(x)\\ &\quad +\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_2(T^0,x)(y)\mathrm{d}IC_2(T,y)\Big)^2\mathrm{d}G^c(x)\Big)\\ &\quad +\frac{1}{n_3}\Big(\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_1(T^0,x)(y)\mathrm{d}IC_3(T,y)\Big)^2\mathrm{d}H^u(x)\\ &\quad +\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_2(T^0,x)(y)\mathrm{d}IC_3(T,y)\Big)^2\mathrm{d}H^c(x)\Big).\end{aligned}$$

把$IC_1(T^0,x)(y),IC_2(T^0,x)(y)$的表达式代入上式, 在这里只计算前两项和以说明计算方法及结果.

$$\begin{aligned}&\frac{1}{n_1}\Big(\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_1(T^0,x)(y)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x)\\ &+\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_2(T^0,x)(y)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^c(x)\Big)\\ =\ &\frac{1}{n_1}\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\Big(\frac{\mathbb{I}(x\le y)}{S_F(x)}+\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\Big)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x)\\ &+\frac{1}{n_1}\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^c(x)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n_1}\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\frac{\mathbb{I}(x\le y)}{S_F(x)}\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x)\\
&\quad+\frac{1}{n_1}\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}(F^u+F^c)(x)\\
&\quad+\frac{2}{n_1}\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\frac{\mathbb{I}(x\le y)}{S_F(x)}\Big)\mathrm{d}IC_1(T,y)\\
&\quad\times\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\Big)\mathrm{d}IC_1(T,y)\Big)\mathrm{d}F^u(x).
\end{aligned}
$$

对第二项做分部积分, 所得正好等于$-1\times$第三项, 因而

$$
\begin{aligned}
&\frac{1}{n_1}\Big(\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_1(T^0,x)(y)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x)\\
&+\int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty}IC_2(T^0,x)(y)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^c(x)\Big)\\
=\;&\frac{1}{n_1}\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\Big(\frac{\mathbb{I}(x\le y)}{S_F(x)}+\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\Big)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x).
\end{aligned}
$$

同理对另外四项求和, 最后得到

$$
\begin{aligned}
\mathsf{E}(\widehat{\lambda}-\lambda)^2 &= \frac{1}{n_1}\int_{-\infty}^{\infty}\Big(S_{F^0}(y)\Big(\frac{\mathbb{I}(x\le y)}{S_F(x)}+\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_F^u(u)}{S_F^2(u)}\Big)\mathrm{d}IC_1(T,y)\Big)^2\mathrm{d}F^u(x)\\
&\quad+\frac{1}{n_2}\int_{-\infty}^{\infty}\Big(S_{G^0}(y)\Big(\frac{\mathbb{I}(x\le y)}{S_G(x)}+\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_G^u(u)}{S_G^2(u)}\Big)\mathrm{d}IC_2(T,y)\Big)^2\mathrm{d}G^u(x)\\
&\quad+\frac{1}{n_3}\int_{-\infty}^{\infty}\Big(S_{H^0}(y)\Big(\frac{\mathbb{I}(x\le y)}{S_H(x)}+\int_{-\infty}^{x\wedge y}\frac{\mathrm{d}S_H^u(u)}{S_H^2(u)}\Big)\mathrm{d}IC_3(T,y)\Big)^2\mathrm{d}H^u(x).
\end{aligned}
$$

再把式(3.1)代入得

$$
\begin{aligned}
\mathsf{E}(\widehat{\lambda}-\lambda)^2 &= \frac{1}{n_1}\times\lambda^2\times(F^0-G^0)^{-2}(x_0)\times(1-F^0(x_0))^2\times\int_{-\infty}^{x_0}\frac{\mathrm{d}F^u(x)}{S_F^2(x)}\\
&\quad+\frac{1}{n_2}\times(\lambda-1)^2\times(F^0-G^0)^{-2}(x_0)\times(1-G^0(x_0))^2\times\int_{-\infty}^{x_0}\frac{\mathrm{d}G^u(x)}{S_G^2(x)}\\
&\quad+\frac{1}{n_3}\times(F^0-G^0)^{-2}(x_0)\times(1-H^0(x_0))^2\times\int_{-\infty}^{x_0}\frac{\mathrm{d}H^u(x)}{S_H^2(x)}.
\end{aligned}
$$

所以

$$
\begin{aligned}
\sigma^2 &= n\times\mathsf{E}(\widehat{\lambda}-\lambda)^2\\
&= \frac{1}{C_1}\times\lambda^2\times(F^0-G^0)^{-2}(x_0)\times(1-F^0(x_0))^2\times\int_{-\infty}^{x_0}\frac{\mathrm{d}F^u(x)}{S_F^2(x)}\\
&\quad+\frac{1}{C_2}\times(\lambda-1)^2\times(F^0-G^0)^{-2}(x_0)\times(1-G^0(x_0))^2\times\int_{-\infty}^{x_0}\frac{\mathrm{d}G^u(x)}{S_G^2(x)}\\
&\quad+\frac{1}{C_3}\times(F^0-G^0)^{-2}(x_0)\times(1-H^0(x_0))^2\times\int_{-\infty}^{x_0}\frac{\mathrm{d}H^u(x)}{S_H^2(x)}.
\end{aligned}
\tag{4.3}
$$

从而如下定理成立:

定理 4.1 对于混合模型$H^0 = \lambda F^0 + (1-\lambda)G^0$, 在随机右截断下得到$n_1$个服从分布$F$的独立样本, n_2个服从分布G的独立样本, n_3个服从分布H的独立样本, 各样本间任何个体两两相互独立, 分布F, G, H分别满足$1-F = (1-F^0)(1-C_F)$, $1-G = (1-G^0)(1-C_G)$, $1-H = (1-H^0)(1-C_H)$, C_F, C_G, C_H分别是对应于F, G, H的截断变量的分布函数, 满足第二节中的条件1, 2, 3, 若$n = n_1 + n_2 + n_3 \to \infty$时, 有

$$\frac{n_1}{n} \to C_1, \qquad \frac{n_2}{n} \to C_2, \qquad \frac{n_3}{n} \to C_3,$$

其中, C_1, C_2, C_3为常数, 则λ的相合估计量$\widehat{\lambda}$服从如下的渐近正态分布.

$$\sqrt{n}(\widehat{\lambda} - \lambda) \to_{\mathfrak{D}} N(0, \sigma^2).$$

其中, 渐近方差σ^2由(4.3)式给出.

注记 1 • 很明显, 可以选取适当的x_0, 使得$(F^0 - G^0)(x_0)$足够大, 从而可望得到λ的方差最小的强相合渐近正态估计量$\widehat{\lambda}$.

• 若无截断发生时, K-M估计退化为经验分布估计, 由(4.3)式得σ^2为

$$\begin{aligned}\sigma^2 \;=\; & \frac{1}{C_1} \times \lambda^2 \times (F^0 - G^0)^{-2}(x_0) \times F^0(x_0) \times (1 - F^0(x_0)) \\ & + \frac{1}{C_2} \times (\lambda - 1)^2 \times (F^0 - G^0)^{-2}(x_0) \times G^0(x_0) \times (1 - G^0(x_0)) \\ & + \frac{1}{C_3} \times (F^0 - G^0)^{-2}(x_0) \times H^0(x_0) \times (1 - H^0(x_0)).\end{aligned}$$

该结果与用经验分布估计F_m, G_n, H_l代替K-M估计$\widehat{F}^0, \widehat{G}^0, \widehat{H}^0$得出的结果一致.

• 对于多于两个分布的混合, 如模型$I^0 = \lambda_1 F^0 + \lambda_2 G^0 + (1 - \lambda_1 - \lambda_2)H^0$, 理论上本文的方法同样有效, 只须取不同的两点$x_0, x_1$, 建立$\lambda_1, \lambda_2$的表达式, 只是计算较繁.

参 考 文 献

[1] Cruz-Medina, I.R., Hettmansperger, T.P. and Thomas, H., Semiparametric mixtures models and repeated measures: the multinomial cut point model, *Appl. Statist.*, **53(3)**(2004), 463–474.

[2] Cruz-Medina, I.R., Hettmansperger, T.P., Nonparametric estimation in semi-parametric univariate mixture models, *J. Statistics Compt.*, **74(7)**(2004), 513–524.

[3] Elmore, R.T., Hettmansperger, T.P. and Thomas, H., Estimating component cumulative distribution functions in finite mixture models, *Communcations in Statistics: Theory and Methods*, **33(9)**(2004), 2075–2086.

[4] Hall, P., On the nonparametric estimation of mixture proportions, *J.R. Statist. Soc. B*, **43(2)**(1981), 147–156.

[5] Hall, P. and Titterington, D.M., Efficient nonparametric estimation of mixture proportions, *J.R. Statist. Soc. B*, **46(3)**(1984), 465–473.

[6] Mclachlan, G.J, Peel, D., *Finite Mixture Models*, John Wiley & Sons, Inc., 2000.

[7] Peterson, A.V., Expressing the Kaplan-Meier estimator as a function of empirical sub-survival functions, *J. Amer. Statist. Assoc.*, **72(360)**(1977), 854–858.

[8] Reid, N., Influence functions for censored data, *Ann. Statistics*, **9(1)**(1981), 78–92.

[9] Serfling, R.J., *Approxiamation Theorems of Mathematical Statistics*, John Wiley & Sons, Inc., 1980.

[10] Stute, W. and Wang, J.L., The strong law under random censorship, *Ann. Statistics*, **21(3)**(1993), 1591–1607.

[11] Titterington, D.M., Minimum distance non-parametric estimation of mixture proportions, *J.R. Statist. Soc. B*, **45(1)**(1983), 37–46.

[12] Van der Vaart, A.W., *Asymptotic Statistics*, Cambridge University Press, 1998.

Nonparametric Estimation of Mixture Proportion under Random Right Censoring

Lu Fuzhong

(*(School of Mathematics & Information Engineering, Jiaxing University, Jiaxing, 314001*)

On the mixture model $H = \lambda F + (1-\lambda)G$, we derive the consistent estimator $\widehat{\lambda}$ of the mixture proportion λ under random right censoring, we also discuss the asymptotic normality of the estimator $\widehat{\lambda}$.

Keywords: Mixture model, asymptotic normality, Hadamard differentiability, influence function.

AMS Subject Classification: 62G10.