Asymptotic inference for MLEs under the special exponential family with double-truncation

Takeshi Emura（江村剛志）
Graduate Institute of Statistics
National Central University

Joint work with
Ya-Hsuan Hu, Yoshihiko Konno

11/5/2015

Outline

- Presentation is based on our papers:

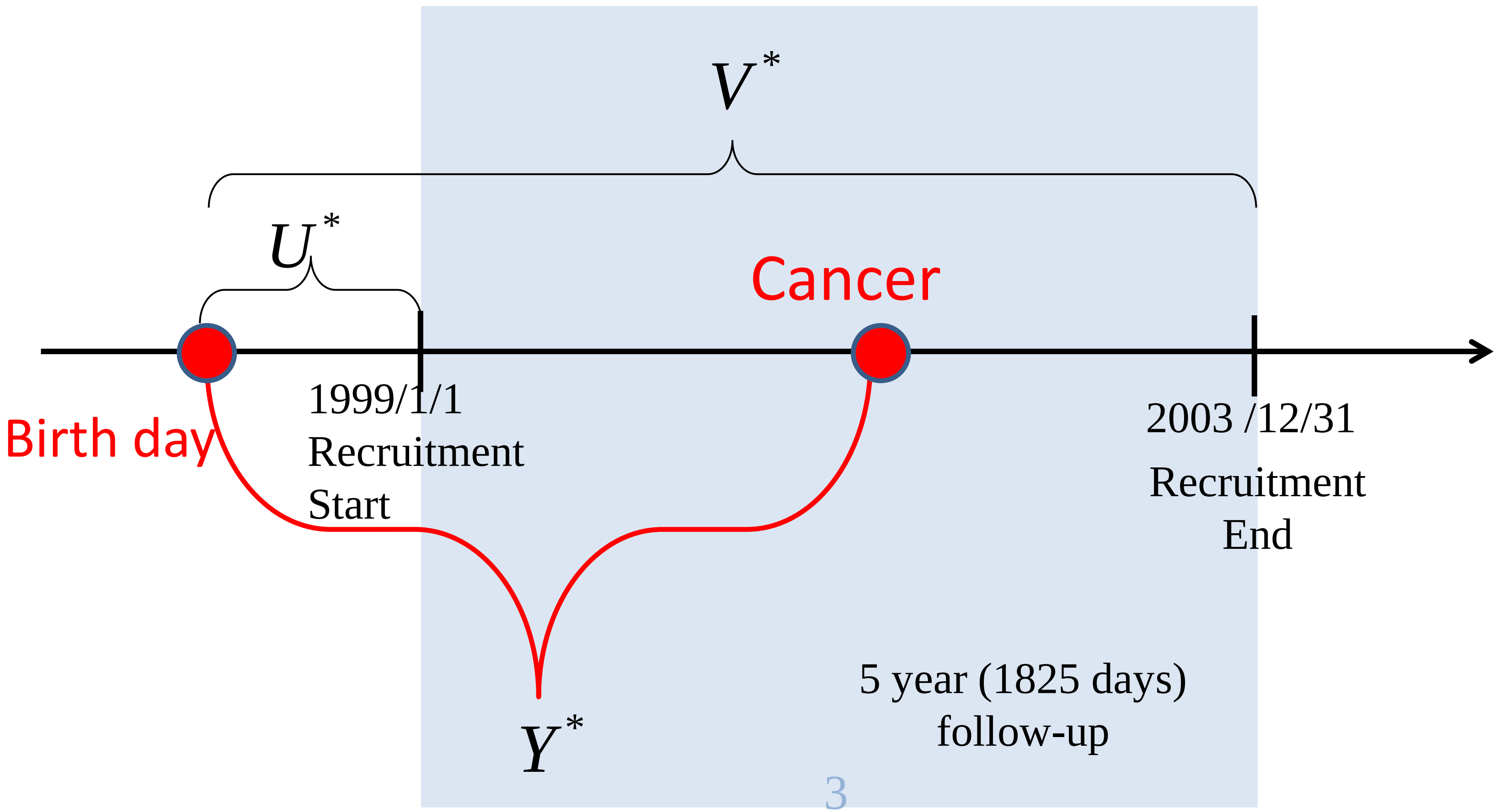
[1] Emura T*, Hu YH, & Konno Y (2015+), Asymptotic inference for maximum likelihood estimators for a cubic exponential family under random double-truncation (major revision submitted to *Statistical Papers*)

[2] Hu YH & Emura T* (2015), Maximum likelihood estimation for a special exponential family under random double-truncation, *Computational Statistics*. DOI 10.1007/s00180-015-0564-z

[3] Hu YH (2014). Maximum likelihood estimation for double-truncation data under a special exponential family. NCU, Master Thesis

Childhood cancer data
(Moreira and de Uña-Álvarez 2010)

Y^* : Age at cancer (in days)
U^* : Age at recruitment start (in days)
$V^* = U^* + 1825$: Age at recruitment end (in months)



- Double-truncation:

$$(U^*, Y^*, V^*) \quad \text{with} \quad (U^*, V^*) \perp Y^*$$

$$* \text{If} \quad U^* \le Y^* \le V^* \quad \Rightarrow \quad \text{observed;}$$

$$\text{otherwise, nothing is available !}$$

- Observation: $y_1, y_2, \ldots, y_n$

$$\text{subject to} \quad u_i \le y_i \le v_i, i = 1, 2, \ldots, n$$

- Target of Estimation: $S(y) = P(Y^* > y)$

$$f(y) = \frac{d}{dy} P(Y^* \le y)$$

Nonparametric approaches to double-truncation

Efron and Petrosian (1999, JASA)

- Nonparametric maximum likelihood estimator (NPMLE).

Shen (2010, AISM)

- Uniform consistency and weak convergence of NPMLE

Moreira and Uña-Á lvarez (2010, J of Nonpar)

- Bootstrap confidence interval

Moreira and Keliegom (2013, CSDA)

-A kernel density estimation

Shen (2012 J. Nonpar), Austin et al. (2014 LIDA)

- Independence test $(U^*, V^*) \perp Y^*$

Emura et al. (2015 LIDA)

- Explicit formula of asymptotic variance of NPMLE

Parametric approaches to double-truncation

Efron and Petrosian (1999, JASA)

-Maximum likelihood estimator (MLE) under the series exponential family (SEF).

Hu and Emura (2015, Computational Statistics)

-Newton Raphson algorithm to obtain the MLE under the SEF

Motivation:

•Asymptotic theory of the MLE under double-truncation has not been studied

• Asymptotic theory does not follows from the usual central limit theorems due to double-truncation

Why? ➔ Data are **not i.i.d**

Special Exponential family (SEF) introduced by Efron and Petrosian (1999, JASA)

Lifetime Y^* follows a continuous distribution with a density

$$f_{\boldsymbol{\eta}}(y) = \exp\{\,\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y) - \phi(\,\boldsymbol{\eta}\,)\,\}, \quad y \in \mathcal{Y}.$$

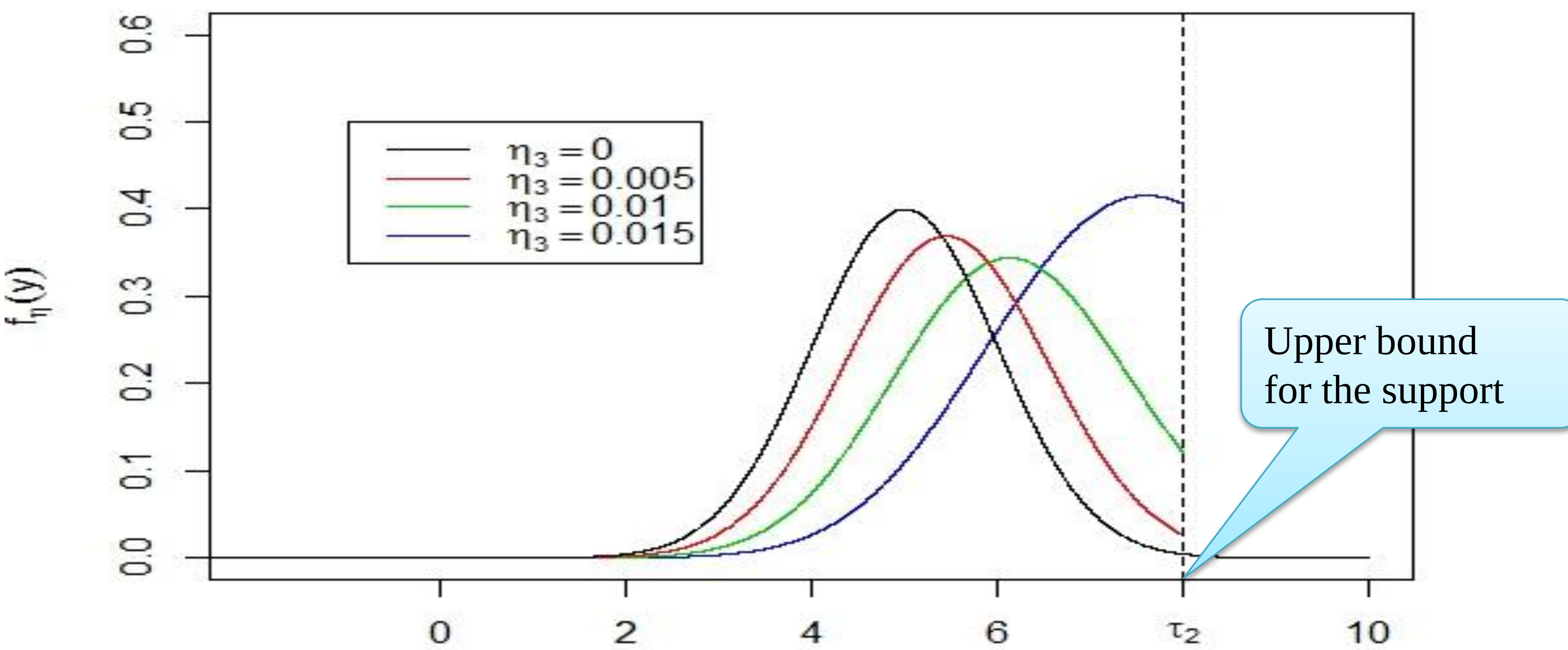
- $\mathcal{Y} \subset \Re$: the support of Y^*
- $\mathbf{t}(\,y\,) = (\,y, y^2, \cdots, y^k\,)^{\mathrm{T}}$
- $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_k\,)^{\mathrm{T}} \in \Theta \subset \Re^{\mathrm{k}}$
- $\phi(\,\boldsymbol{\eta}\,) = \log[\,\int_{\mathcal{Y}} \exp\{\,\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(\,y\,)\,\}\,dy\,]$

I focus on $k = 3$

Special Exponential family $(k=3)$

$$f_{\boldsymbol{\eta}}(y) = \exp[\,\eta_1 y + \eta_2 y^2 + \eta_3 y^3 - \phi(\boldsymbol{\eta})\,], \quad y \in \mathcal{Y} = (-\infty, \tau_2\,],$$

Parameter space $\Theta^{+} = \{\,(\eta_1, \eta_2, \eta_3) : \eta_1 \in \Re, \eta_2 \in \Re, \ \eta_3 > 0\,\}$

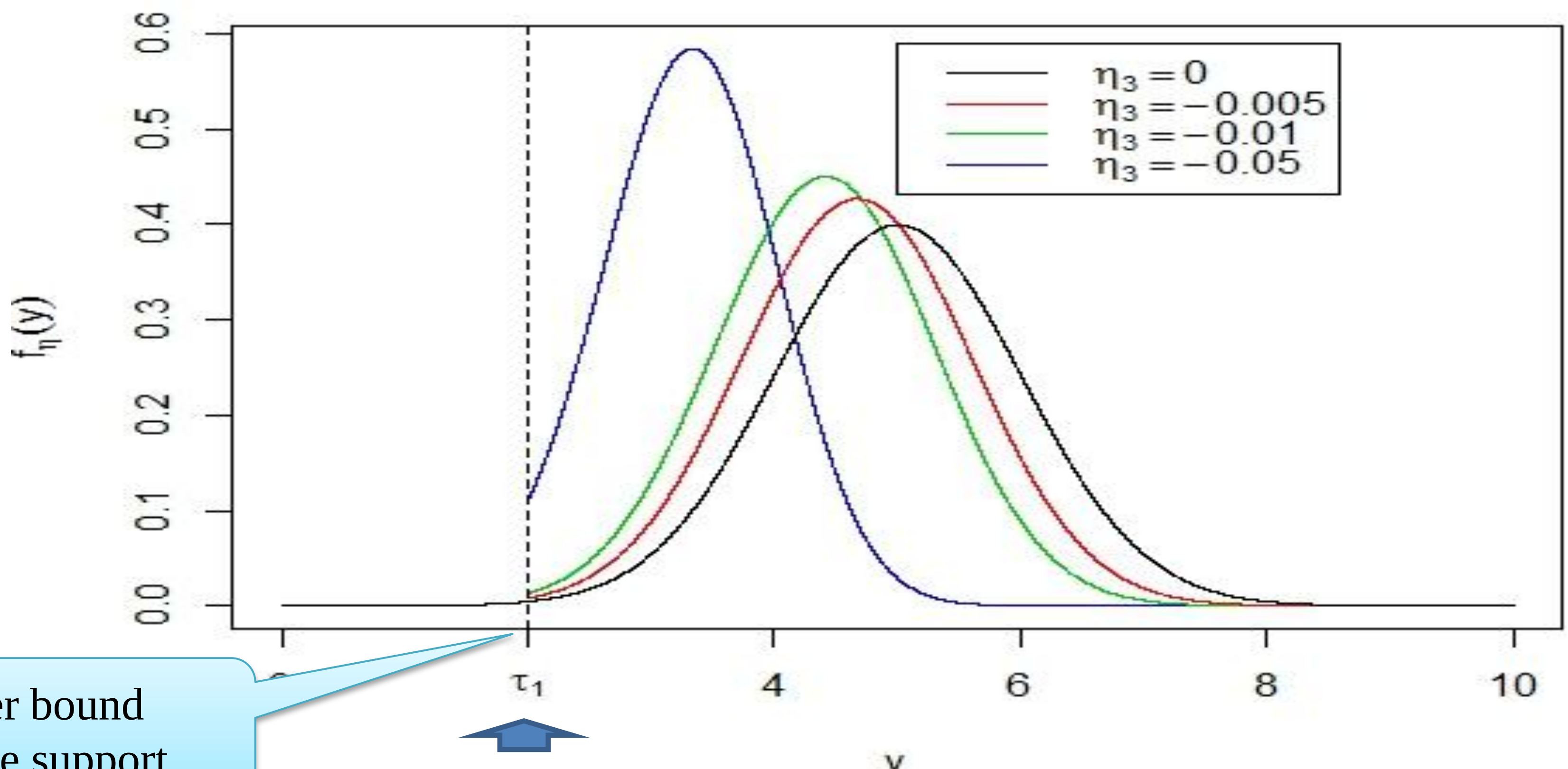


$$y_{(1)} = \max(y_i) = \text{maximum observed lifetime}$$

Special Exponential family $(k=3)$

$$f_{\boldsymbol{\eta}}(\,y\,) = \exp[\,\eta_1 y + \eta_2 y^2 + \eta_3 y^3 - \phi(\,\boldsymbol{\eta}\,)\,], \quad y \in \mathcal{Y} = [\tau_1, -\infty),$$

Parameter space $\Theta^- = \{\,(\,\eta_1, \eta_2, \eta_3\,) : \eta_1 \in \Re,\, \eta_2 \in \Re,\, \eta_3 < 0\,\}$

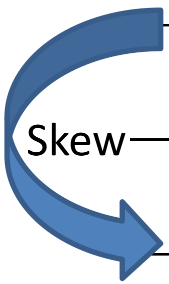


Lower bound
for the support

$$y_{(1)} = \min(\,y_i\,) = \text{minimum observed lifetime}$$

Special exponential family (SEF): Summary

$$f_{\boldsymbol{\eta}}(y) = \exp[\eta_1 y + \eta_2 y^2 + \cdots + \eta_k y^k - \phi(\boldsymbol{\eta})]$$



	Distribution
1 parameter SEF ($\eta_1 > 0$)	Location-exponential tail distribution
1 parameter SEF ($\eta_1 < 0$)	Location-Scale exponential distribution
2 parameter SEF ($\eta_2 < 0$)	Normal distribution
2 parameter SEF ($\eta_2 > 0$)	U-shaped distribution
3 parameter (cubic) SEF ($\eta_3 < 0$)	Positively skewed normal distribution
3 parameter (cubic) SEF ($\eta_3 > 0$)	Negatively skewed normal distribution

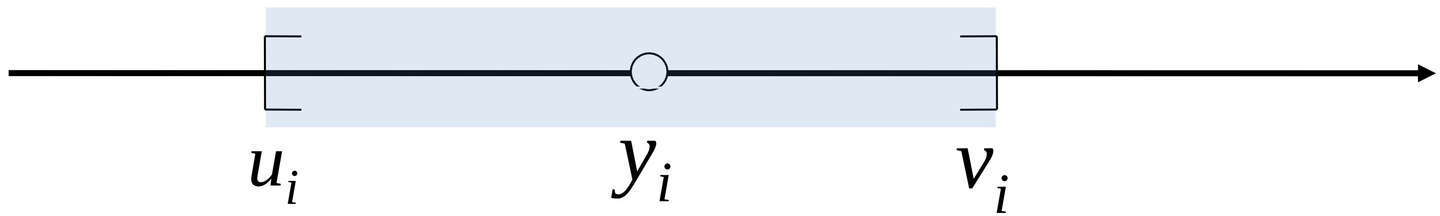
Resemble to **flexible skew normal class**

Ref: Chang and Genton (2007 Comm. Stat. T&M)

, and many other skew normal distributions

Likelihood under double-truncation

- Truncation interval: $R_i = [\, u_i , v_i \,]$



No likelihood to be observed outside the interval !
(Trunbull, 1976 JRSSB; Efron and Petrosian, 1999 JASA)

- Truncated density
  (conditional density
  given $y \in R_i$)

$$f_i(\, y \,|\, \boldsymbol{\eta}\,) \equiv \frac{f_{\eta}(\, y\,)}{F_i(\,\boldsymbol{\eta}\,)} 1\{\, y \in R_i \,\}$$

where $F_i = \int_{u_i}^{v_i} f(\, y\,)\, dy$

Likelihood under double-truncation

- Log-likelihood:

$$\ell_n(\boldsymbol{\eta}) = \log\left\{\prod_{i=1}^{n} f_i(y_i \mid \boldsymbol{\eta})\right\} = \sum_{i=1}^{n}\{\log f_{\boldsymbol{\eta}}(y_i) - \log F_i(\boldsymbol{\eta})\}$$

- Under the cubic SEF (k=3): $\eta_3 < 0$; Positive skew

$$\ell_n(\boldsymbol{\eta}) = \sum_{i=1}^{n}(\eta_1 y_i + \eta_2 y_i^2 + \eta_3 y_i^3)$$

$$-\sum_{i=1}^{n}\delta_i \log\left\{\int_{u_i}^{v_i}\exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\,dy\right\}$$

$$-\sum_{i=1}^{n}(1-\delta_i)\log\left\{\int_{\tau_1}^{v_i}\exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\,dy\right\}.$$

Maximum likelihood estimation- Cubic SEF $(k=3)$

Newton-Raphson Algorithm

Step 1: Choose the initial value $\boldsymbol{\eta}=(\eta_1^{(0)},\eta_2^{(0)},\eta_3^{(0)})^{\mathrm{T}}$.

Step 2: Repeat

$$\begin{bmatrix}\eta_1^{(p+1)}\\ \eta_2^{(p+1)}\\ \eta_3^{(p+1)}\end{bmatrix}=\begin{bmatrix}\eta_1^{(p)}\\ \eta_2^{(p)}\\ \eta_3^{(p)}\end{bmatrix}-\left.\begin{bmatrix}\frac{\partial^2}{\partial\eta_1^2}\ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial\eta_1\partial\eta_2}\ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial\eta_1\partial\eta_3}\ell(\boldsymbol{\eta})\\ \frac{\partial^2}{\partial\eta_2\partial\eta_1}\ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial\eta_2^2}\ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial\eta_2\partial\eta_3}\ell(\boldsymbol{\eta})\\ \frac{\partial^2}{\partial\eta_3\partial\eta_1}\ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial\eta_3\partial\eta_2}\ell(\boldsymbol{\eta}) & \frac{\partial^2}{\partial\eta_3^2}\ell(\boldsymbol{\eta})\end{bmatrix}^{-1}\right|_{(\eta_1^{(p)},\eta_2^{(p)},\eta_3^{(p)})}\left.\begin{bmatrix}\frac{\partial}{\partial\eta_1}\ell(\boldsymbol{\eta})\\ \frac{\partial}{\partial\eta_2}\ell(\boldsymbol{\eta})\\ \frac{\partial}{\partial\eta_3}\ell(\boldsymbol{\eta})\end{bmatrix}\right|_{(\eta_1^{(p)},\eta_2^{(p)},\eta_3^{(p)})}$$

until convergence, $|\eta_i^{(p+1)}-\eta_i^{(p)}|<\varepsilon_i \quad \forall i=1,2,3$

Then, $\hat{\boldsymbol{\eta}}_n=(\eta_1^{(p+1)},\eta_2^{(p+1)},\eta_3^{(p+1)})$ is the MLE

Stopping criterion : $\varepsilon_1=0.0001,\quad \varepsilon_2=0.0001,\quad \varepsilon_3=0.0000001$

Maximum likelihood estimation- Cubic SEF $(k=3)$

Randomized Newton-Raphson Algorithm
(Hu and Emura, 2015 Computational Statistics):

Step 4: If the Newton Raphson **diverges**, i.e.,

$$|\eta_1^{(p+1)} - \eta_1^{(p)}| > D_1 \quad \text{or} \quad |\eta_2^{(p+1)} - \eta_2^{(p)}| > D_2 \quad \text{or} \quad |\eta_3^{(p+1)} - \eta_3^{(p)}| > D_3$$

("Diameters" $D_1 = 20, \quad D_1 = 10, \quad D_3 = 1$)

Return to Step 1: by replacing $(\eta_1^{(0)}, \eta_2^{(0)}, \eta_3^{(0)})^{\mathrm{T}}$ with

$$(\eta_1^{(0)} + u_1, \eta_2^{(0)} + u_2, \eta_3^{(0)})^{\mathrm{T}}$$

where $u_1 \sim U(-d_1, d_1)$ and $u_2 \sim U(-d_2, d_2)$

We chose $d_1 = 6$ and $d_2 = 0.5$.

Maximum likelihood estimation- Cubic SEF $(k=3)$

MLE $\hat{\boldsymbol{\eta}}_n$: $\quad \dfrac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) = 0, \qquad j = 1, 2, 3$

- Estimator of density

$$f_{\hat{\boldsymbol{\eta}}_n}(t) = \exp\{\hat{\eta}_{n1} t + \hat{\eta}_{n2} t^2 + \hat{\eta}_{n3} t^3 - \phi(\hat{\boldsymbol{\eta}}_n)\}, \quad t \in \mathcal{Y}.$$

- Estimator of survival function

$$S_{\hat{\boldsymbol{\eta}}_n}(y) = \int_y^\infty \exp\{\hat{\eta}_{n1} t + \hat{\eta}_{n2} t^2 + \hat{\eta}_{n3} t^3 - \phi(\hat{\boldsymbol{\eta}}_n)\} dt, \qquad y \ge \tau_1$$

where $\phi(\boldsymbol{\eta}) = \log\left\{ \int_{\tau_1}^\infty \exp(\eta_1 y + \eta_2 y^2 + \eta_3 y^3)\, dy \right\}$

We always need SE and confidence intervals for $f(t)$ and $S(y)$

Asymptotic analysis of the MLE

$$\text{MLE} \quad \hat{\boldsymbol{\eta}}_n : \quad \frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) = 0, \qquad j = 1, 2, \ldots, k$$

- Existence of the MLE
- Consistency, Asymptotic normality
- SE and confidence interval

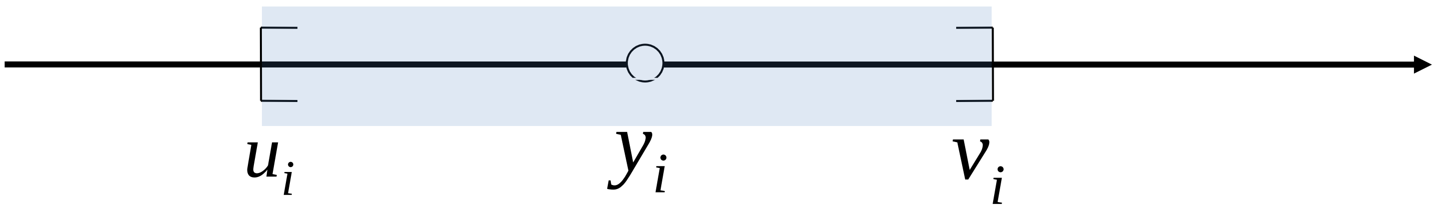
have not been studied

Our goal:

Asymptotic analysis based on independent but **not identically** distributed (i.n.i.d) data.

Likelihood under double-truncation

- **Likelihood:** $f_i(y\,|\,\boldsymbol{\eta}) \equiv \dfrac{f_{\eta}(y)}{F_i(\boldsymbol{\eta})} 1\{u_i \le y \le v_i\}$



Intervals are heterogeneous between samples

$$u_i \le y_i \le v_i,\ i = 1, 2, \ldots, n$$

- **Score function**: not the sum of i.i.d. terms

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\eta}} \ell_n(\boldsymbol{\eta}) = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\eta}} \log\left\{ \prod_{i=1}^{n} f_i(y_i\,|\,\boldsymbol{\eta}) \right\}$$

$$= \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\eta}} \{ \log f_{\eta}(y_i) - \log F_i(\boldsymbol{\eta}) \} \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

CLT ?

Asymptotic theories of the MLE under i.n.i.d.

- Bradley and Gart (1962 Biometrika):

  -Seminal work for the MLE under i.n.i.d.

  -Proofs are not rigorous (the literature in 1940's during which the probability theory was not established)

- Hoadley (1971 Ann. Math. Stat)

  - Regularity condition fairly technical and less intuitive

  - Example of lifetime model with fixed censored points

- Philippou and Roussas (1975 AISM)

  - Consistency of MLE is assumed at the beginning.

Conclusion:

- Above existing theorems are not reliable and helpful to our case.
- Yet, the idea of Bradley and Gart (1962) is appealing

Our strategy of establishing asymptotics

- **Regularity conditions**:
  Follow the styles of Bradley and Gart (1962 Biometrika):
  (but not follow their proofs)

- **Tools:** Textbooks of mathematical statistics:
  - Strong Law of Large Number for i.n.i.d. (Shao 2003)
  - Lindeberg-Feller Central Limit Theorem (van der Vaart, 1998)

- **Proofs:** Modify the proof of
  Lehmann and Casella (1998) for i.i.d. case
  - Handling multi-parameter cases
  - Rigorous + Clear
  - Simultaneously establish:
    Existence + Consistency + Asymptotic normality

Strong Law of Large Number (SLLN) for i.n.i.d.

Lemma 2 *Let* $Y_1, Y_2, \ldots$ *be independent random variables with* $E[\,|Y_i|\,] < \infty$

for $i = 1, 2, \ldots$. *If there is a constant* $p \in [1, 2]$ *s.t.* $\lim_{n\to\infty} \frac{1}{n^p} \sum_{i=1}^{n} E[\,|Y_i|^p\,] = 0$,

then $\frac{1}{n} \sum_{i=1}^{n} (Y_i - E[Y_i]) \xrightarrow{P} 0.$

Lindeberg-Feller Central Limit Theorem (CLT)

Lemma 3 *Let* $\mathbf{D}_{n,1}, \ldots, \mathbf{D}_{n,n}$ *be independent k-dimensional random vectors* *s.t.*

$$\sum_{i=1}^{n} E[\,\|\mathbf{D}_{n,i} - E[\mathbf{D}_{n,i}]\|^2 \, 1\{\|\mathbf{D}_{n,i} - E[\mathbf{D}_{n,i}]\| > \varepsilon\}\,] \to 0, \quad n \to \infty \quad \text{(Lindeberg Condition)}$$

for every $\varepsilon > 0$, *and* $\sum_{i=1}^{n} \mathbf{Cov}(\mathbf{D}_{n,i}) \to \Sigma$,

Then, $\sum_{i=1}^{n} (\mathbf{D}_{n,i} - E\mathbf{D}_{n,i}) \xrightarrow{d} N_k(\mathbf{0}, \Sigma)$ *as* $n \to \infty$.

Characterization of the MLE

Define $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \ldots, \hat{\eta}_{kn})^{\mathrm{T}}$ to be a solution to the score equations

$$\frac{\partial}{\partial \eta_j} \ell_n(\boldsymbol{\eta}) = 0, \qquad j = 1, 2, \ldots, k, \qquad (*)$$

Assumption (A) *The parameter space* Θ *is open and contains*

the true $\boldsymbol{\eta}^0 = (\eta_1^0, \eta_2^0, \ldots, \eta_k^0)^{\mathrm{T}}$. *Parameter space* Θ *is natural,*

i.e., $\int_y \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}\, dy < \infty$, $\boldsymbol{\eta} \in \Theta$.

Lemma 1 (Characterization of MLE)*: Under Assumption (A),*

if the solution $\hat{\boldsymbol{\eta}}_n = (\hat{\eta}_{1n}, \hat{\eta}_{2n}, \ldots, \hat{\eta}_{kn})^{\mathrm{T}}$ *exists, then it is the MLE,*

i.e., $\ell_n(\hat{\boldsymbol{\eta}}_n) \geq \ell_n(\boldsymbol{\eta})$ *for any* $\boldsymbol{\eta} \in \Theta$.

Fisher information matrix for i.n.i.d. data

Assumption (B) There exist a $k \times k$ positive definite matrix

$I(\boldsymbol{\eta}) = \{ I_{js}(\boldsymbol{\eta}) \}_{j,s=1,2,\ldots,k}$ s.t.

$$\sum_{i=1}^{n} I_{i,\,js}(\boldsymbol{\eta}) / n \to I_{js}(\boldsymbol{\eta}), \qquad j, s \in \{1, 2, \ldots, k\}, \qquad \boldsymbol{\eta} \in \Theta, \qquad \text{as } n \to \infty.$$

where the Fisher information of the i th sample is

$$I_{i,\,js}(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}} \left\{ \frac{\partial}{\partial \eta_j} \log f_i(Y_i \mid \boldsymbol{\eta}) \cdot \frac{\partial}{\partial \eta_s} \log f_i(Y_i \mid \boldsymbol{\eta}) \right\},.$$

$$i = 1, 2, \ldots, n, \quad j, s = 1, 2, \ldots, k$$

Boundeness conditions:

Assumption (C) For $j, s, l \in \{1, 2, \ldots, k\}$, there exist $M_{jsl}(\cdot)$ s.t.

$$\left| \frac{\partial^3}{\partial \eta_j \partial \eta_s \partial \eta_l} \log f_i(y \mid \boldsymbol{\eta}) \right| \le M_{jsl}(y), \qquad m_{i,\,jsl} \equiv E_{\boldsymbol{\eta}^0}\{ M_{jsl}(Y_i) \} < \infty \quad \text{and} \quad m^2_{i,\,jsl} \equiv E_{\boldsymbol{\eta}^0}\{ M_{jsl}(Y_i)^2 \} < \infty .$$

Assumption (D) For $j, s \in \{1, 2, \ldots, k\}$, there exist $W_{js}(\cdot)$ s.t.

$$\left| \frac{\partial^2}{\partial \eta_j \partial \eta_s} \log f_i(y \mid \boldsymbol{\eta}) \right| \le W_{js}(y), \qquad w_{i,\,js} \equiv E_{\boldsymbol{\eta}^0}\{ W_{js}(Y_i) \} < \infty \quad \text{and} \quad w^2_{i,\,js} \equiv E_{\boldsymbol{\eta}^0}\{ W_{js}(Y_i)^2 \} < \infty .$$

Assumption (E) For $j \in \{1, 2, \ldots, k\}$, there exist $A_j(\cdot)$ s.t.

$$\left| \frac{\partial}{\partial \eta_j} \log f_i(y \mid \boldsymbol{\eta}) \right| \le A_j(y), \quad \text{with} \ \ \sup_y A_j^2(y) < \infty .$$

- Assumpiton (E): Similar to Bradley and Gart (1962 Biometrika) to regulate i.n.i.d. samples

Main result: Asymptotic theory

***Theorem 1**: If Assumptions (A)-(E) hold, then*

(a) Existence and consistency: There exists a solution $\hat{\boldsymbol{\eta}}_n$

with probability tending to one s.t. $\hat{\boldsymbol{\eta}}_n \xrightarrow{P} \boldsymbol{\eta}^0$ *as* $n \to \infty$.

(b) Asymptotic normality: $\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}^0) \xrightarrow{d} N_k(\mathbf{0}, I(\boldsymbol{\eta}^0)^{-1})$ *as* $n \to \infty$.

- Valid approximation to the Fisher information

$$I_{js}(\boldsymbol{\eta}^0) \approx \frac{1}{n}\sum_{i=1}^{n} I_{i,\,js}(\boldsymbol{\eta}^0) = \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\eta}^0}\left\{ \frac{\partial}{\partial \eta_j}\log f_i(Y_i \mid \boldsymbol{\eta}^0) \cdot \frac{\partial}{\partial \eta_s}\log f_i(Y_i \mid \boldsymbol{\eta}^0) \right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\eta}^0}\left\{ -\frac{\partial^2}{\partial \eta_j \partial \eta_s}\log f_i(Y_i \mid \boldsymbol{\eta}^0) \right\} \approx -\frac{1}{n}\frac{\partial^2}{\partial \eta_j \partial \eta_s}\ell_n(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n} \equiv \hat{I}_{js}(\hat{\boldsymbol{\eta}}_n).$$

Asymptotic inference for density

- Standard Error

$$SE\{\, f_{\hat{\boldsymbol{\eta}}_n}(y)\,\} = \sqrt{\left\{\frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(y)\right\}^{\mathrm{T}} \cdot \left\{-\frac{\partial^2}{\partial \boldsymbol{\eta}^2}\ell_n(\boldsymbol{\eta})\right\}^{-1} \cdot \frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(y)\Bigg|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n}}\ ,$$

Where $\dfrac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(y) = \begin{bmatrix} y - e^1(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \\ \vdots \\ y^k - e^k(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \end{bmatrix} \cdot f_{\boldsymbol{\eta}}(y)\,,$

and where $e^j(\boldsymbol{\eta}) = \int_{\mathcal{Y}} y^j \exp\{\boldsymbol{\eta}^{\mathrm{T}} \cdot \mathbf{t}(y)\}\,dy\,, \quad j \in \{0, 1, 2, \ldots, k\}\,.$

- $(1-\alpha)100\%$ confidence interval for $f_{\boldsymbol{\eta}}(y)$ is,

$$[\; f_{\hat{\boldsymbol{\eta}}_n}(y) - Z_{\alpha/2} \cdot SE\{\, f_{\hat{\boldsymbol{\eta}}_n}(y)\,\}, \quad f_{\hat{\boldsymbol{\eta}}_n}(y) + Z_{\alpha/2} \cdot SE\{\, f_{\hat{\boldsymbol{\eta}}_n}(y)\,\}\;]\,.$$

Asymptotic inference for survival function

- Standard Error

$$SE\{S_{\hat{\boldsymbol{\eta}}_n}(y)\} = \sqrt{\left\{\frac{\partial}{\partial \boldsymbol{\eta}} S_{\boldsymbol{\eta}}(y)\right\}^{\mathrm{T}} \cdot \left\{-\frac{\partial^2}{\partial \boldsymbol{\eta}^2} \ell_n(\boldsymbol{\eta})\right\}^{-1} \cdot \frac{\partial}{\partial \boldsymbol{\eta}} S_{\boldsymbol{\eta}}(y)\Big|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n}} ,$$

Where $\frac{\partial}{\partial \boldsymbol{\eta}} S_{\boldsymbol{\eta}}(y) = \int_{y \in \mathcal{Y}, t > y} \begin{bmatrix} t - e^1(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \\ \vdots \\ t^k - e^k(\boldsymbol{\eta})/e^0(\boldsymbol{\eta}) \end{bmatrix} \cdot f_{\boldsymbol{\eta}}(t)\, dt .$

- $(1-\alpha)100\%$ confidence interval for $S_{\boldsymbol{\eta}}(y)$ is

$$[\, S_{\hat{\boldsymbol{\eta}}_n}(y) - Z_{\alpha/2} \cdot SE\{S_{\hat{\boldsymbol{\eta}}_n}(y)\}, \quad S_{\hat{\boldsymbol{\eta}}_n}(y) + Z_{\alpha/2} \cdot SE\{S_{\hat{\boldsymbol{\eta}}_n}(y)\}\,] .$$

$v_0 - u_0$

Easy-to-check sufficient conditions

Lemma 4 *Assumptions (C), (D) and (E) hold under the following two conditions:*

Assumption (D) The parameter space Θ is bounded.

Assumption (G) The lower support of left-truncation $u_{\inf} \equiv \inf_i (u_i^*) < \infty$

The upper support of right-truncation $v_{\sup} \equiv \sup_i (v_i^*) < \infty$.

There exist constants $u_0 < v_0$ such that

$$[u_0, v_0] \subset [u_i, v_i] \subset [u_{\inf}, v_{\sup}] \subset \mathcal{Y}, \qquad i = 1, 2, \ldots .$$

Interpretation of Assumption (G)

- Boundedness of truncation interval
- **Sufficient follow-up**: Interval cannot be too short (should be longer than $v_0 - u_0$)

$v_0 - u_0$

Simulation results

the cubic SEF based on 1000 repetitions (under the inclusion probability $P(U^* \le Y^* \le V^*) \approx 0.50$).

(η_1, η_2, η_3)	n	$E(\hat{\eta}_1)$	$SD(\hat{\eta}_1)$	$E[SE(\hat{\eta}_1)]$	95%Cov
$(5, -0.5, 0.005)$	100	5.856	7.282	7.436	0.936
	200	5.484	5.146	5.165	0.944
	300	5.378	4.125	4.207	0.951
(η_1, η_2, η_3)	n	$E(\hat{\eta}_2)$	$SD(\hat{\eta}_2)$	$E[SE(\hat{\eta}_2)]$	95%Cov
$(5, -0.5, 0.005)$	100	-0.622	1.397	1.437	0.940
	200	-0.573	0.995	0.998	0.945
	300	-0.561	0.797	0.813	0.946
(η_1, η_2, η_3)	n	$E(\hat{\eta}_3)$	$SD(\hat{\eta}_3)$	$E[SE(\hat{\eta}_3)]$	95%Cov
$(5, -0.5, 0.005)$	100	0.0101	0.089	0.091	0.944
	200	0.0084	0.063	0.063	0.950
	300	0.0081	0.051	0.052	0.949

$v_0 - u_0$

Simulation results

the cubic SEF based on 1000 repetitions (under the inclusion probability $P(U^* \le Y^* \le V^*) \approx 0.50$).

(η_1, η_2, η_3)	n	$E\{ S_{\hat{\boldsymbol{\eta}}}(t) \}$	$SD\{ S_{\hat{\boldsymbol{\eta}}}(t) \}$	$E[SE\{ S_{\hat{\boldsymbol{\eta}}}(t) \}]$	95%Cov
$(5, -0.5, 0.005)$	100	0.499	0.071	0.070	0.944
$S_{\boldsymbol{\eta}}(y) = 0.5$	200	0.500	0.049	0.049	0.939
	300	0.502	0.038	0.039	0.947
$(5, -0.5, -0.005)$	100	0.504	0.062	0.065	0.941
$S_{\boldsymbol{\eta}}(y) = 0.5$	200	0.503	0.044	0.045	0.957
	300	0.502	0.036	0.037	0.949

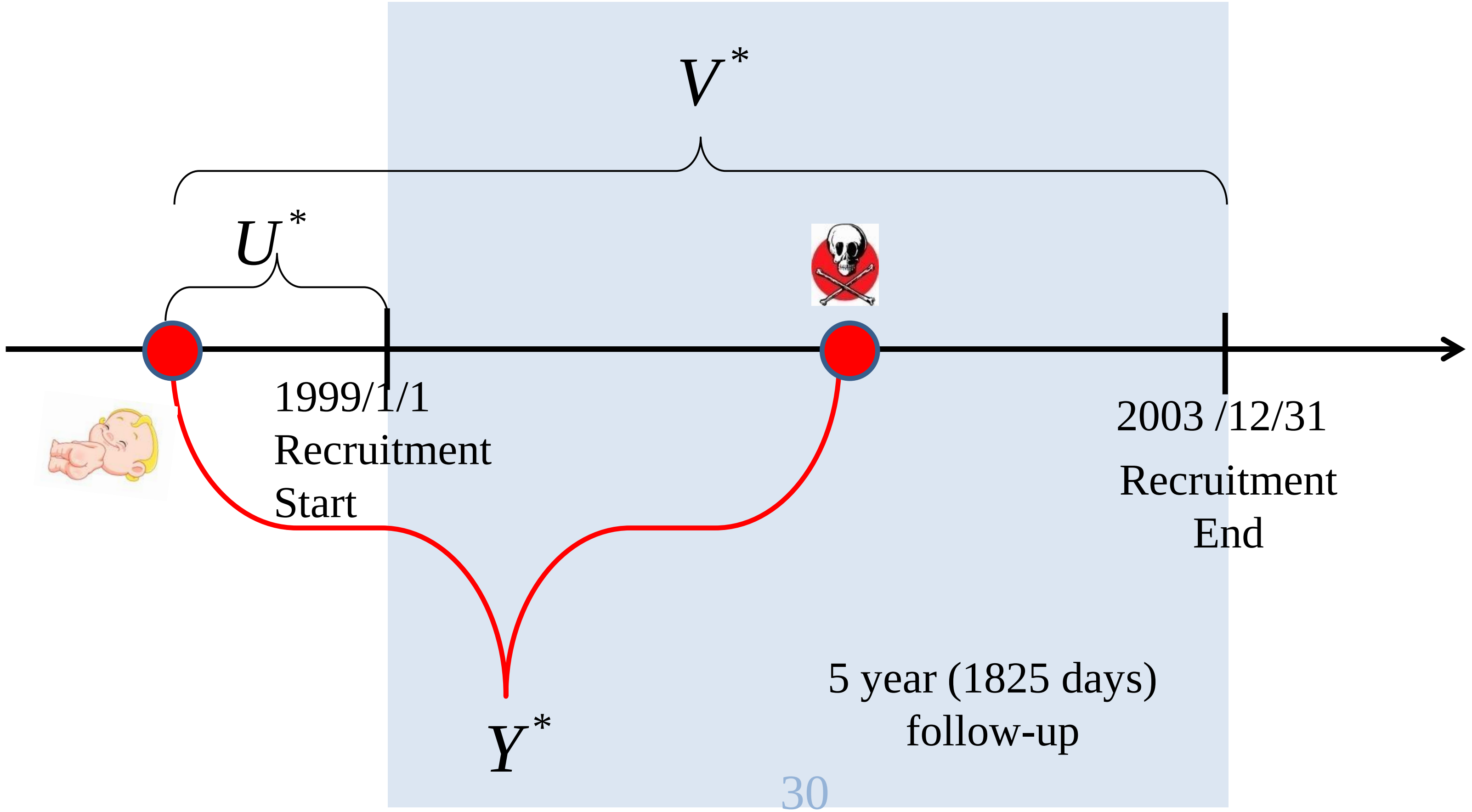
(η_1, η_2, η_3)	n	$E\{ f_{\hat{\boldsymbol{\eta}}}(t) \}$	$SD\{ f_{\hat{\boldsymbol{\eta}}}(t) \}$	$E[SE\{ f_{\hat{\boldsymbol{\eta}}}(t) \}]$	95%Cov
$(5, -0.5, 0.005)$	100	0.367	0.054	0.057	0.969
$f_{\boldsymbol{\eta}}(y) = 0.369$	200	0.367	0.036	0.039	0.967
	300	0.367	0.030	0.031	0.961
$(5, -0.5, -0.005)$	100	0.430	0.057	0.060	0.961
$f_{\boldsymbol{\eta}}(y) = 0.427$	200	0.428	0.040	0.041	0.947
	300	0.427	0.032	0.033	0.958

Data analysis of childhood cancer data (Moreira and de Uña-Álvarez 2010)

Y^* : Age at cancer (in days) ← **Estimation**

U^* : Age at recruitment start (in days)

$V^* = U^* + 1825$: Age at recruitment end (in months)



First, model selection

1) **Kolmogorov-Smirnov distance**

$$D = \max_{y} \{ | \hat{S}_{NPMLE}(y) - S_{\hat{\boldsymbol{\eta}}}(y) | \}$$

- $\hat{S}_{NPMLE} = \hat{P}(Y > y) = \sum_{y_i > y} \hat{f}_i$: **Model-free** survival function
  where $\hat{\mathbf{f}} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n)$ is the NPMLE
  (Efron and Petrosian, 1999)
- $S_{\hat{\boldsymbol{\eta}}}(y) = P(Y > y)$: **Model-based** survival function

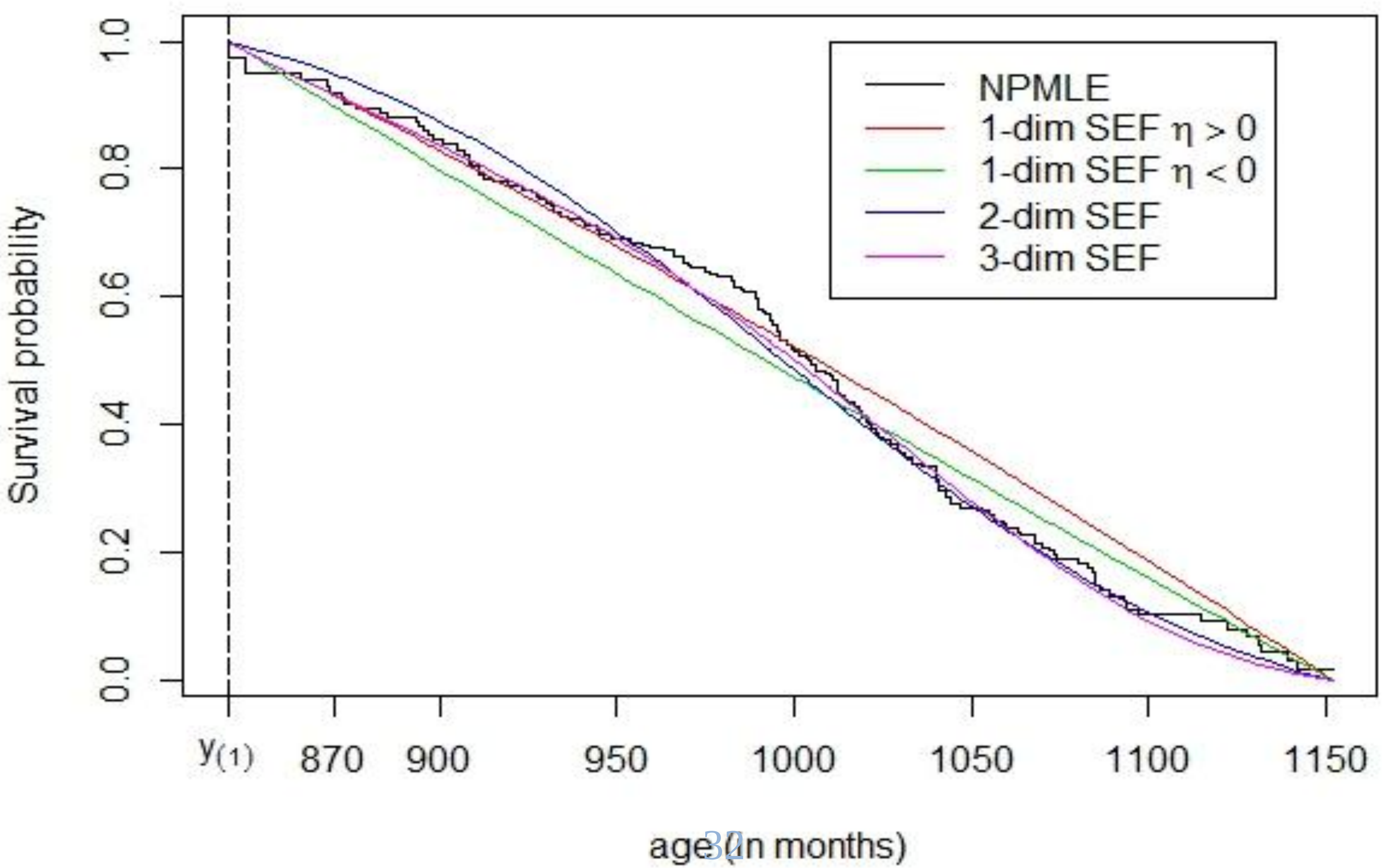
2) **AIC (Akaike Information Criterion)**

$$AIC = -2 \log L + 2k$$

- k : the number of unknown parameters
- $\log L$: maximized value of likelihood function

KS-distance between MLE and NPMLE

Best Model = Smallest KL distance = Cubic SEF with $\eta_3 < 0$



Data analysis -- Model selection --

The maximum likelihood inference for the childhood cancer data.

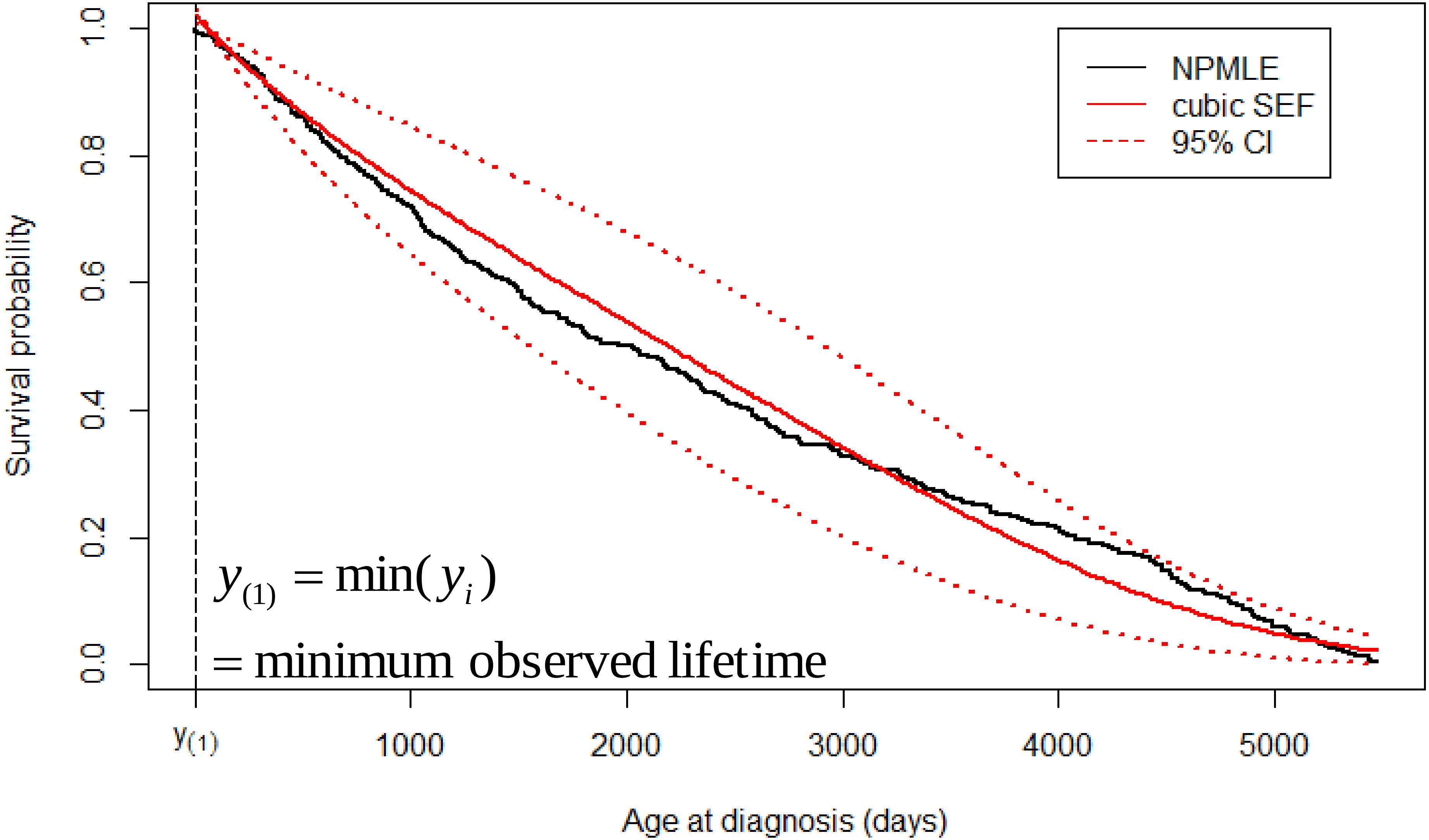
Model	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$	$\log L$	AIC	K-S statistic
(a) 1 par. SEF ($\eta_1 > 0$)	8.74×10^{-5}	0	0	-3013.6	6029.2	0.206
(b) 1 par. SEF ($\eta_1 < 0$)	-3.85×10^{-4}	0	0	-2999.6	6001.1	0.121
(c) 2 par. SEF	7.71×10^{-4}	-1.87×10^{-7}	0	-3027.6	6059.2	0.132
(d) Cubic SEF ($\eta_3 < 0$)	-7.90×10^{-4}	3.38×10^{-7}	-4.87×10^{-11}	-2991.6	5989.2	0.084

Best model

- Model (a) = The one-parameter SEF ($\eta_1 > 0$)
- Model (b) = The one-parameter SEF ($\eta_1 < 0$)
- Model (c) = The two-parameter SEF
- Model (d) = The cubic SEF ($\eta_3 < 0$)
- $\log L$ = The maximized log-likelihood
- AIC = Akaike information criterion, defined as $\text{AIC} = -2\log L + 2k$
- K-S statistic = The Kolmogorov-Smirnov distance between the MLE and the NPMLE

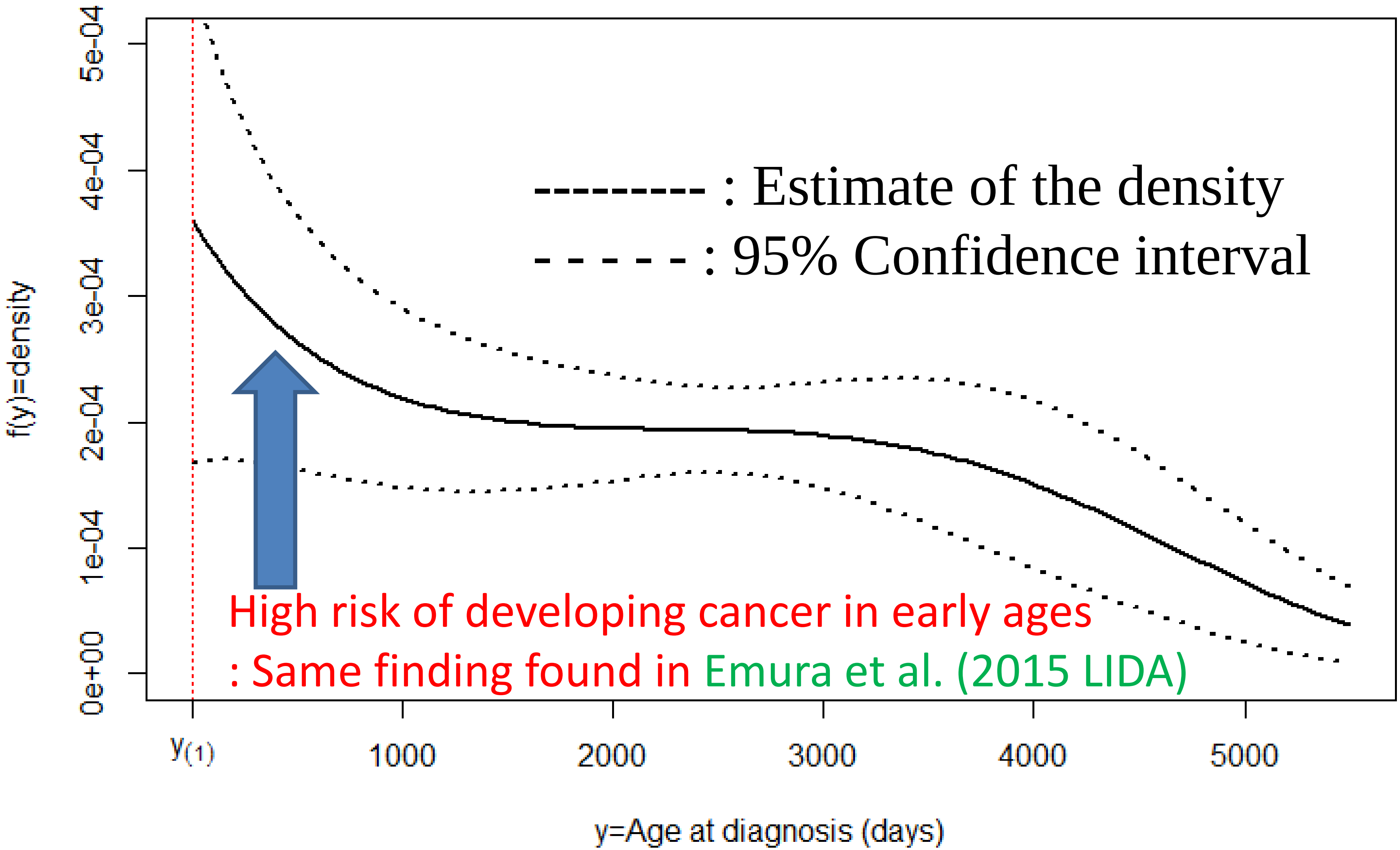
Data analysis under the cubic SEF (best model)

$$S_{\hat{\boldsymbol{\eta}}}(t) = \int_t^{\infty} f_{\hat{\boldsymbol{\eta}}}(y)dy = \int_t^{\infty} \exp[\,\hat{\eta}_1 y + \hat{\eta}_2 y^2 + \hat{\eta}_3 y^3 - \phi(\hat{\boldsymbol{\eta}})\,]dy$$



Asymptotic inference under the cubic SEF

$$f_{\hat{\boldsymbol{\eta}}}(\,y\,)=\exp[\,\hat{\eta}_1 y+\hat{\eta}_2 y^2+\hat{\eta}_3 y^3-\phi(\,\hat{\boldsymbol{\eta}}\,)\,]$$



Pick up referees' comments : Statistical Papers

- **Q1: Assumption (G)** *hold for real data example ?*

  Answer

  Follow-up length: Fixed at 5 years $d_0 = 1825$ (days).

  But **Assumption (G)** requires $d_0 > 7300$ (days)

  So Assumption (G) **does not hold**.

  But it is **easy to be checked by user**

Other target quantities under double-truncation

- Predictive survival $S(t+w \mid t) = S(Y > t+w \mid Y > t)$

  (Klein & Moeschberger, 2003, **with left-truncation only)**
- Mean / median residual life $m(u) = E(Y-u \mid Y > u)$

  (Chi et al. 2014 Com.Stat-Simulations**, with left-truncation only)**

$$m(u,v) \equiv E(Y - v \mid u > Y > v)$$

with double-truncation (Sankaran & Sunoj, 2004 Stat Papers**)**

Thank you for your listening