國立政治大學統計學系 演 講
2013, 4/22

Statistical inference based on the nonparametric MLE under double-truncation

(joint work with Konno, Y., Japan Women's University)

(accepted by Lifetime Data Analysis)

Takeshi Emura 江村剛志
（國立中央大學統計研究所）

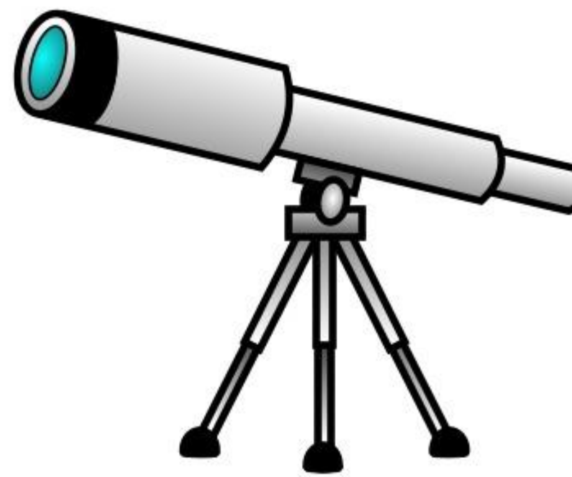
Outlines

- Doubly Truncated data: Review
- The nonparametric estimator: Review
- Proposed variance estimator
- Statistical Inference

  Confidence interval, goodness-of-fit, confidence band
- Simulation (compare with Bootstrap & Jackknife)
- Data analysis

• Example (Efron & Petrosian, 1992 JASA):

T^* : Quasar's luminosity (brightness)



* Telescope cannot detect quaser if

Too dim : $T^* < U^*$

Too bright : $T^* > V^*$

If $U^* \leq T^* \leq V^* \Rightarrow$ observed;

otherwise, nothing is observed !

Example 2: Moreira and Álvarez(2010)

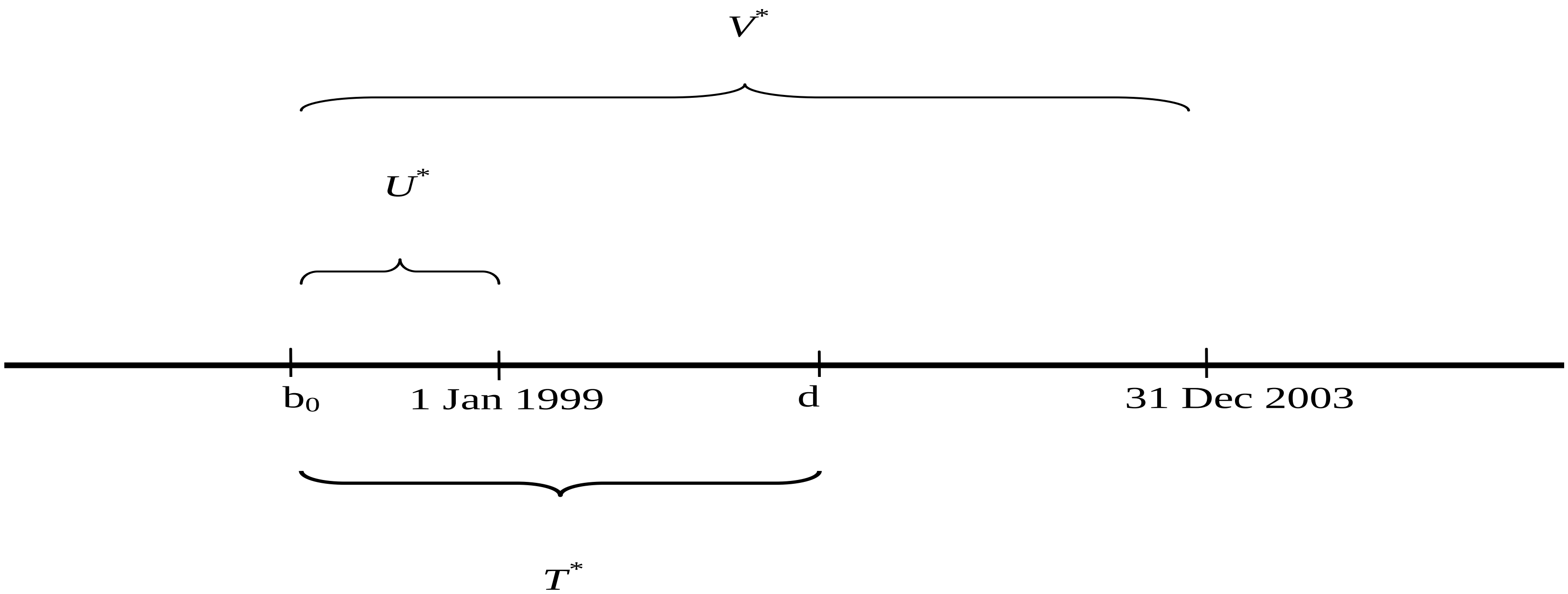


Fig. 1: The childhood cancer cases in North Portugal (Moreira and Á lvarez 2010)

b_0 : birth date

d : date of diagnosis

T^*: age at diagnosis

U^*: age at 1 January 1999

$V^* = U^*+1825$: age at 31 Dec 2003

If $U^* \leq T^* \leq V^* \Rightarrow$ observed;

otherwise, nothing is observed!

- **Population:** Tri-variate random variables

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

$$* \text{If} \quad U^* \le T^* \le V^* \quad \Rightarrow \quad \text{observed;}$$
$$\text{otherwise,} \quad \text{nothing is observed!}$$

- **Observation**:

$$\{ (U_j, T_j, V_j) : j = 1, \ldots, n \} \quad \text{subject to} \quad U_j \le T_j \le V_j$$

- **Parameter of interest**:

$$F(t) = \Pr(T^* \le t)$$

Nonparametric MLE (NPMLE) by Efron & Petrosian (1999 JASA)

- Probability masses at $(T_1, \ldots, T_n)$:

$$\mathbf{f} = (f_1, \ldots, f_n)^{\mathrm{T}}, \qquad \sum_{j=1}^{n} f_j = 1$$

- Nonparametric likelihood

$$L_n(\mathbf{f}) = \prod_{j=1}^{n} \frac{f_j}{F_j}$$

← Horvitz-Thompson Weighting

where $F_j = \sum_{m=1}^{n} f_m J_{jm}$, and $J_{jm} = \mathbf{I}\{ U_j \le T_m \le R_j \}$

- Likelihood equation:

$$\partial \log L_n(\mathbf{f})/\partial \mathbf{f} = \mathbf{f}^{-1} - J^T \mathbf{F}^{-1} = 0$$

Self-consistency algorithm (Efron & Petrosian, 1999)

Step 0: Set $\hat{\mathbf{f}}_{(0)} = (1/n, \ldots, 1/n)^T$ and $\hat{\mathbf{F}}_{(0)} = J\hat{\mathbf{f}}_{(0)}$,

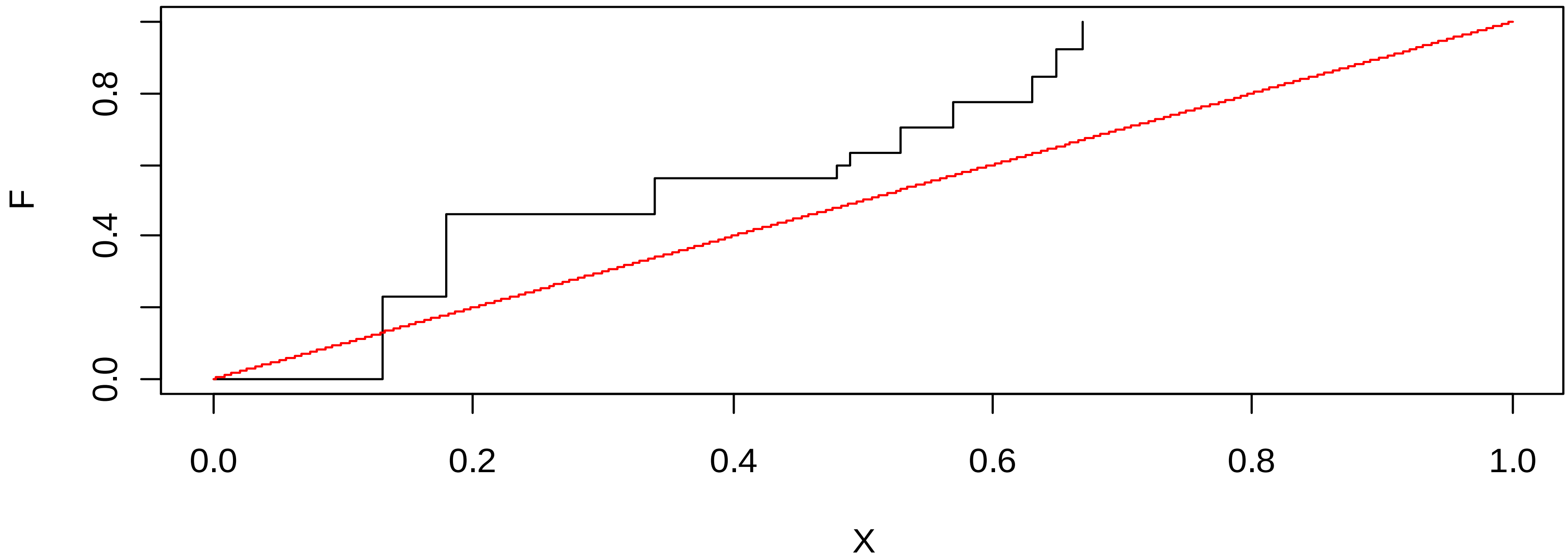
Step 1: Obtain $\hat{\mathbf{f}}_{(1)}$ by $\mathbf{f}_{(1)}^{-1} = J^T \mathbf{F}_{(0)}^{-1}$, set $\hat{\mathbf{F}}_{(1)} = J\hat{\mathbf{f}}_{(1)}$,

Step 2: Repeat step 1 to update $\hat{\mathbf{f}}_{(l+1)}$ stop algorithm when $\| \hat{\mathbf{f}}_{(l+1)} - \hat{\mathbf{f}}_{(l)} \|^2 < \varepsilon$

- NPMLE of $F(t) = \Pr(T^* \leq t)$:

$$\hat{F}(t) = \sum_{j=1}^{n} \mathbf{I}(T_j \leq t)\, \hat{f}_j$$

- NPMLE: $\hat{F}(t) = \sum_{j=1}^{n} \mathbf{I}(T_j \le t)\,\hat{f}_j$

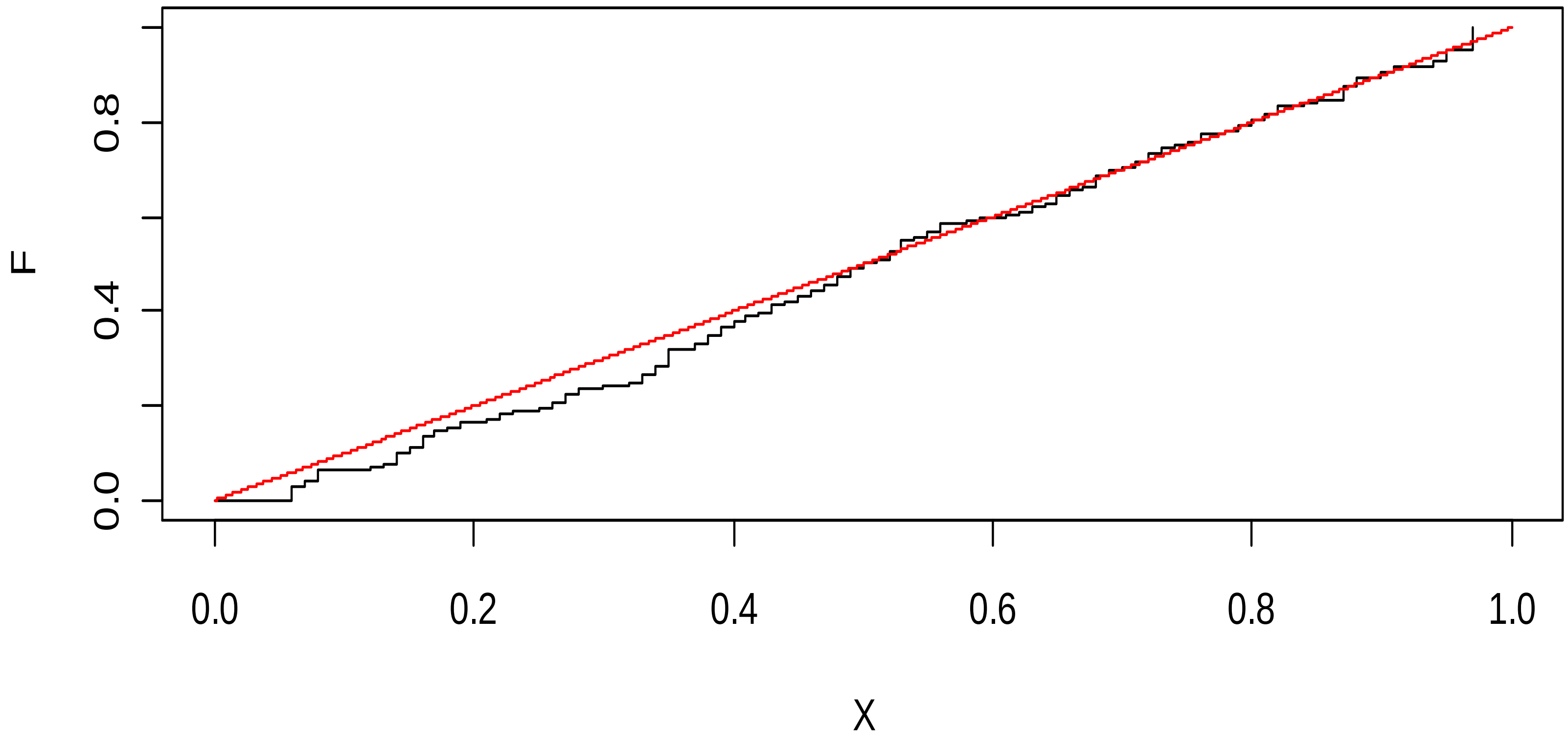


Example: n=10,

$$\mathbf{f} = (\ f_1, \ldots, \ f_n\ )$$

=(0.13 0.18 0.34 0.48 0.49 0.53 0.57 0.63 0.65 0.67)

- NPMLE: $\hat{F}(t) = \sum_{j=1}^{n} \mathbf{I}(T_j \le t)\, \hat{f}_j$



The NPMLE $\hat{F}$ when n=250

- Shen (2010 *AISM*) shows
$$\sqrt{n}\{\hat{F}(t)-F(t)\}\rightarrow_d N(0,V(t))$$
where the form of $V(t)$ is not given explicitly.

- Moreira and Álvarez(2010 *J. of Nonpar*.) suggest Bootstrap methods to construct the confidence interval (C.I.) of $F(t)$

- Shen (2012 *J. of App. Stat.*) invert the empirical likelihood ratio test to derive C.I. of $F(t)$

- Motivation of my work is to derive an alternative estimator of $V(t)$ to the Bootstrap

Proposal 1:

- New estimator of $Var\{\hat{F}(t)\}$ based on analytical approach
- Also, new confidence interval of $F(t)$

• Revisit: Efron & Petrosian's Self-consistency equation

$$\begin{cases} \partial \log L_n(\mathbf{f}) / \partial \mathbf{f} = \mathbf{f}^{-1} - J^T \mathbf{F}^{-1} = 0 \\ \qquad \text{subject to} \quad \mathbf{1}_n \mathbf{f} = 1 \end{cases}$$

$$\text{where} \quad \mathbf{1}_n = (1, \ldots, 1)^T$$

Remarks:

*They treat $\mathbf{f} = (f_1, \ldots, f_n)^T$ as *n* parameters when taking derivative

$$\frac{\partial}{\partial f_j} \log L_n(\mathbf{f}) = \frac{1}{f_j} - \sum_{i=1}^{n} \frac{J_{ij}}{F_i}$$

* In fact, only *n-1* parameters are unknown

Idea: Modify the likelihood Equation

- I treat $\mathbf{f}_{(-n)} = (f_1, \ldots, f_{n-1})^{\mathrm{T}}$ as *n-1* parameters

  Then, the likelihood equation become

$$\frac{\partial \log L_n(\mathbf{f})}{\partial \mathbf{f}_{(-n)}} = D\left[\mathbf{f}^{-1} - J^{\mathrm{T}}\mathbf{F}^{-1}\right]_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\tilde{\mathbf{f}}}$$

$$\text{where} \quad D = \left[I_{n-1} \quad -1_{n-1}\right] = \begin{bmatrix} 1 & \cdots & 0 & -1 \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & 1 & -1 \end{bmatrix}$$

is an adjustment factor regarding the constraint

- Information matrix

$$i_n(\mathbf{f}) = -\frac{\partial^2 \log L_n(\mathbf{f})}{\partial \mathbf{f}_{(-n)} \partial \mathbf{f}_{(-n)}^{\mathrm{T}}}$$

$$= D\left\{\mathrm{diag}\left(\frac{1}{\mathbf{f}^2}\right) - J^{\mathrm{T}}\mathrm{diag}\left(\frac{1}{\mathbf{F}^2}\right)J\right\}\Bigg|_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\tilde{\mathbf{f}}} D^{\mathrm{T}}$$

$: \ (n-1)\times(n-1)$ matrix

*This is not usual Fisher info. matrix since the number of parameters grow with *n*

Heuristic

$$\hat{F}(t) = \sum_{j=1}^{n} \mathbf{I}(T_j \le t)\,\hat{f}_j = w_t'\hat{\mathbf{f}}$$

where $w_t = (\mathbf{I}(T_1 \le t), \ldots, \mathbf{I}(T_n \le t))^T$

$$\Downarrow$$

$$Cov\{\hat{F}(s), \hat{F}(t)\} = w_s' Cov(\hat{\mathbf{f}}) w_t \approx w_s' i_n(\hat{\mathbf{f}})^{-1} w_t$$

Proposed Variance Estimator

$$\hat{V}_{\text{Info}}\{\,\hat{F}(t)\,\} = w_t^{\mathrm{T}} \left[D \left\{ \text{diag}\left(\frac{1}{\hat{\mathbf{f}}^{2}} \right) - J^{\mathrm{T}} \text{diag}\left(\frac{1}{\hat{\mathbf{F}}^{2}} \right) J \right\} D^{\mathrm{T}} \right]^{-1} w_t$$

- Asymptotic normality:

$$\log \hat{F}(t) - \log F(t) \sim N(\,0, \hat{V}_{\text{Info}}\{\,\hat{F}(t)\,\}/\,\hat{F}(t)^2\,)$$

- Invert the Wald test:

➔a log-transformed confidence interval of F(t):

$$(\,\hat{F}(t)\exp[\,-z_{\alpha/2}\hat{V}_{\text{Info}}^{1/2}\{\hat{F}(t)\}/\,\hat{F}(t)\,],\ \hat{F}(t)\exp[\,z_{\alpha/2}\hat{V}_{\text{Info}}^{1/2}\{\hat{F}(t)\}/\,\hat{F}(t)\,]\,)$$

Competitor 1:

Simple Bootstrap algorithm (Moreira and Álvarez, 2010):

Step 1: For each $b = 1, \ldots, B$, draw Bootstrap resamples $\{(U_{jb}^*, T_{jb}^*, V_{jb}^*) : j = 1, \ldots, n\}$

from $\{(U_j, T_j, V_j) : j = 1, \ldots, n\}$, and then compute the NPMLE $\hat{F}_b^*(t)$ from them.

Step 2: Compute the Bootstrap variance estimator

$$\hat{V}_{\text{Boot}}\{\hat{F}(t)\} = \frac{1}{B-1}\sum_{b=1}^{B}\{\hat{F}_b^*(t) - \overline{F}^*(t)\}^2 ,$$

where $\overline{F}^*(t) = \frac{1}{B}\sum_{b=1}^{B}\hat{F}_b^*(t)$, and take the $(\alpha/2)\times 100\%$ and $(1-\alpha/2)\times 100\%$ points of

$\{\hat{F}_b^*(t) : b = 1, \ldots, B\}$ for the $(1-\alpha)\times 100\%$ confidence interval.

Competitor 2:

Jackknife algorithm

Step 1: For each $i = 1, \ldots, n$, delete i th sample from $\{ (U_j, T_j, V_j) : j = 1, \ldots, n \}$, and then compute the NPMLE $\hat{F}_{(-i)}(t)$ from the remaining n-1 samples.

Step 2: Compute the jackknife variance estimator

$$\hat{V}_{\text{Jack}}\{\hat{F}(t)\} = \frac{n-1}{n} \sum_{i=1}^{n} \{\hat{F}_{(-i)}(t) - \overline{F}_{(\cdot)}(t)\}^2 \,,$$

where $\overline{F}_{(\cdot)}(t) = \frac{1}{n} \sum_{i=1}^{n} \hat{F}_{(-i)}(t)$, and the log-transformed $(1-\alpha)\times 100\,\%$ confidence interval

$$(\, \hat{F}(t)\exp[\,-z_{\alpha/2}\hat{V}_{\text{Jack}}^{1/2}\{\hat{F}(t)\}/\hat{F}(t)\,],\;\; \hat{F}(t)\exp[\,z_{\alpha/2}\hat{V}_{\text{Jack}}^{1/2}\{\hat{F}(t)\}/\hat{F}(t)\,]\,)\,.$$

- Simulations

1) Generate samples from

$$U^* \sim Uniform(0, 0.5)$$
$$T^* \sim Uniform(0,1)$$
$$V^* \sim Uniform(0.5, 1)$$

2) Get $\{(U_j, T_j, V_j) : j = 1, \ldots, n\}$ subject to $U_j \leq T_j \leq V_j$

3) Get $\hat{V}_{\text{Info}}\{\hat{F}(t)\}$, $\hat{V}_{\text{Boot}}\{\hat{F}(t)\}$ and $\hat{V}_{\text{Jack}}\{\hat{F}(t)\}$

4) Repeat this 1000 times. Compare the 3 methods in terms of Bias, MSE and coverage prob.

Estimated SD = $\frac{1}{1000}\sum_{r=1}^{1000}\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}$; MSE = $\frac{1}{1000}\sum_{r=1}^{1000}\left(\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}} - \mathrm{SD}\{\hat{F}(t)\}\right)^2$

			n = 100	n = 150	n = 250
$F(t)=0.5$	Sample SD		0.0999	0.0855	0.0686
	Estimated SD	Proposed	0.0862	0.0724	0.0571
		Bootstrap	0.0864	0.0726	0.0576
		Jackknife	0.0935	0.0785	0.0609
	MSE	Proposed	0.00385	0.00304	0.00177
		Bootstrap	0.00151	0.00132	0.00088
		Jackknife	0.00462	0.00415	0.00242
	95% Coverage	Proposed	0.926	0.937	0.951
		Bootstrap	0.936	0.948	0.951
		Jackknife	0.932	0.946	0.958

Estimated SD = $\frac{1}{1000}\sum_{r=1}^{1000}\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}$; MSE = $\frac{1}{1000}\sum_{r=1}^{1000}\left(\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}-\mathrm{SD}\{\hat{F}(t)\}\right)^2$

			n = 100	n = 150	n = 250
$F(t)=0.2$	Sample SD		0.0839	0.0664	0.0552
	Estimated SD	Proposed	0.0680	0.0570	0.0458
		Bootstrap	0.0698	0.0589	0.0474
		Jackknife	0.0742	0.0619	0.0487
	MSE	Proposed	0.00262	0.00100	0.00065
		Bootstrap	0.00193	0.00126	0.00084
		Jackknife	0.00427	0.00275	0.00147
	95% Coverage	Proposed	0.942	0.939	0.930
		Bootstrap	0.906	0.909	0.918
		Jackknife	0.952	0.948	0.938

Simulation Results

- Bootstrap is best in terms of MSE
- Jackknife is best in terms of Bias
- Proposed method is best in terms of computational cost
- Proposed and jackknife has better 95% coverage probability than the Bootstrap at the tail (i.e., F(t)=0.2).
- Bootstrap has serious under-coverage at the tail (also, reported in Moreira and Álvarez 2010)

Proposal 2:

- Goodness-of-fit $H_0:\ F = F_0$ vs. $H_1:\ F \neq F_0$
- Confidence band for $F(\cdot)$

* These become possible by estimating covariance structure by

$$Cov\{\hat{F}(s), \hat{F}(t)\} \approx w_s' i_n(\hat{\mathbf{f}})^{-1} w_t$$

- Goodness-of-fit test

$$H_0: F = F_0 \quad \text{vs. } H_1: F \neq F_0$$

where F_0 is specified

- Cramér-von Mises type statistic

$$C = n\int_0^\infty \{\hat{F}(x) - F_0(x)\}^2 dF_n(x) = \sum_{j=1}^n \{\hat{F}(T_j) - F_0(T_j)\}^2$$

- **Reject H0 when *C* is greater some value.**

- Cramér-von Mises statistic

$$C = \sum_{j=1}^{n} \{ \hat{F}(T_j) - F_0(T_j) \}^2$$

with the covariance estimated as

$$Cov\{\hat{F}(T_k), \hat{F}(T_j)\} \approx w'_{T_k} i_n(\hat{\mathbf{f}})^{-1} w_{T_j}$$

- Approximate C by the sum of multivariate normal variables

$$C^* = \sum_{j=1}^{n-1} \{ G_j \}^2$$

with $Cov\{G_k, G_j\} = w'_{T_k} i_n(\hat{\mathbf{f}})^{-1} w_{T_j}$

Cramér-von Mises test for $H_0:\ F = F_0$;

Step 1: Calculate $C = \sum_{j=1}^{n} \{ \hat{F}(T_j) - F_0(T_j) \}^2$ and the observed information matrix $i_n(\hat{\mathbf{f}})$.

Step 2: For each $b = 1, \ldots, B$, generate $\mathbf{G}^{(b)} = (G_1^{(b)}, \ldots, G_{n-1}^{(b)}) \sim N(\mathbf{0}_{n-1}, \mathbf{W}_{n-1} i_n(\hat{\mathbf{f}})^{-1} \mathbf{W}_{n-1}^{\mathrm{T}})$,

and compute $C^{(b)} = (\mathbf{G}^{(b)})^{\mathrm{T}} \mathbf{G}^{(b)}$.

Step 3: Reject $H_0:\ F = F_0$ with level α if $\sum_{b=1}^{B} \mathbf{I}(C^{(b)} > C) / B < \alpha$.

Kolmogorov-Smirnov test for $H_0: F = F_0$;

Step 1: Calculate $K = \max_i \{ \hat{F}(T_j) - F_0(T_j) \}$ and the observed information matrix $i_n(\hat{\mathbf{f}})$.

Step 2: For each $b = 1, \ldots, B$, generate $\mathbf{G}^{(b)} = (G_1^{(b)}, \ldots, G_{n-1}^{(b)}) \sim N(\mathbf{0}_{n-1}, \mathbf{W}_{n-1} i_n(\hat{\mathbf{f}})^{-1} \mathbf{W}_{n-1}^{\mathrm{T}})$,

and compute $K^{(b)} = \max_{i=1,..,n-1} G_i^{(b)}$.

Step 3: Reject $H_0: F = F_0$ with level α if $\sum_{b=1}^{B} \mathbf{I}(K^{(b)} > K)/B < \alpha$.

- Confidence band:

$$1-\alpha = \Pr\left\{\sup_t |\hat{F}(t) - F_0(t)| \le c\right\}$$

$$= \Pr\{\, \hat{F}(t) - c \le F_0(x) \le \hat{F}(t) + c \quad \text{for all } \ t \,\}$$

where c is the $(1-\alpha)$ percentile of the Kolmogorov Smirnov statistics.

Confidence band for F ;

Step 1: Calculate the NPMLE $\hat{F}$ and the observed information matrix $i_n(\hat{\mathbf{f}})$.

Step 2: For each $b = 1, \ldots, B$, generate $\mathbf{G}^{(b)} = (G_1^{(b)}, \ldots, G_{n-1}^{(b)}) \sim N(\mathbf{0}_{n-1}, \mathbf{W}_{n-1} i_n(\hat{\mathbf{f}})^{-1} \mathbf{W}_{n-1}^{\mathrm{T}})$, and compute $K^{(b)} = \max_{i=1,\ldots,n-1} G_i^{(b)}$.

Step 3: Obtain the confidence band $\hat{F} \pm c$, where c is the $(1-\alpha)100\%$ point for $\{ K^{(b)}; b = 1, \ldots, B \}$.

Table 5: Simulation results for the proposed goodness-of-fit tests under the null hypothesis based on 1000 replications.

	Cramér-von Mises test (C)			Kolmogorov-Smirnov test (K)		
	n =100	n = 150	n =250	n = 100	n = 150	n = 250
Rejection rate at $\alpha = 0.10$	0.120	0.105	0.088	0.089	0.081	0.078
Rejection rate at $\alpha = 0.05$	0.063	0.055	0.045	0.037	0.033	0.030
Rejection rate at $\alpha = 0.01$	0.015	0.014	0.008	0.006	0.007	0.005
$E[C]$ or $E[K]$	1.078	1.206	1.306	0.143	0.121	0.098
$E[C^{(b)}]$ or $E[K^{(b)}]$	1.167	1.281	1.331	0.137	0.116	0.093

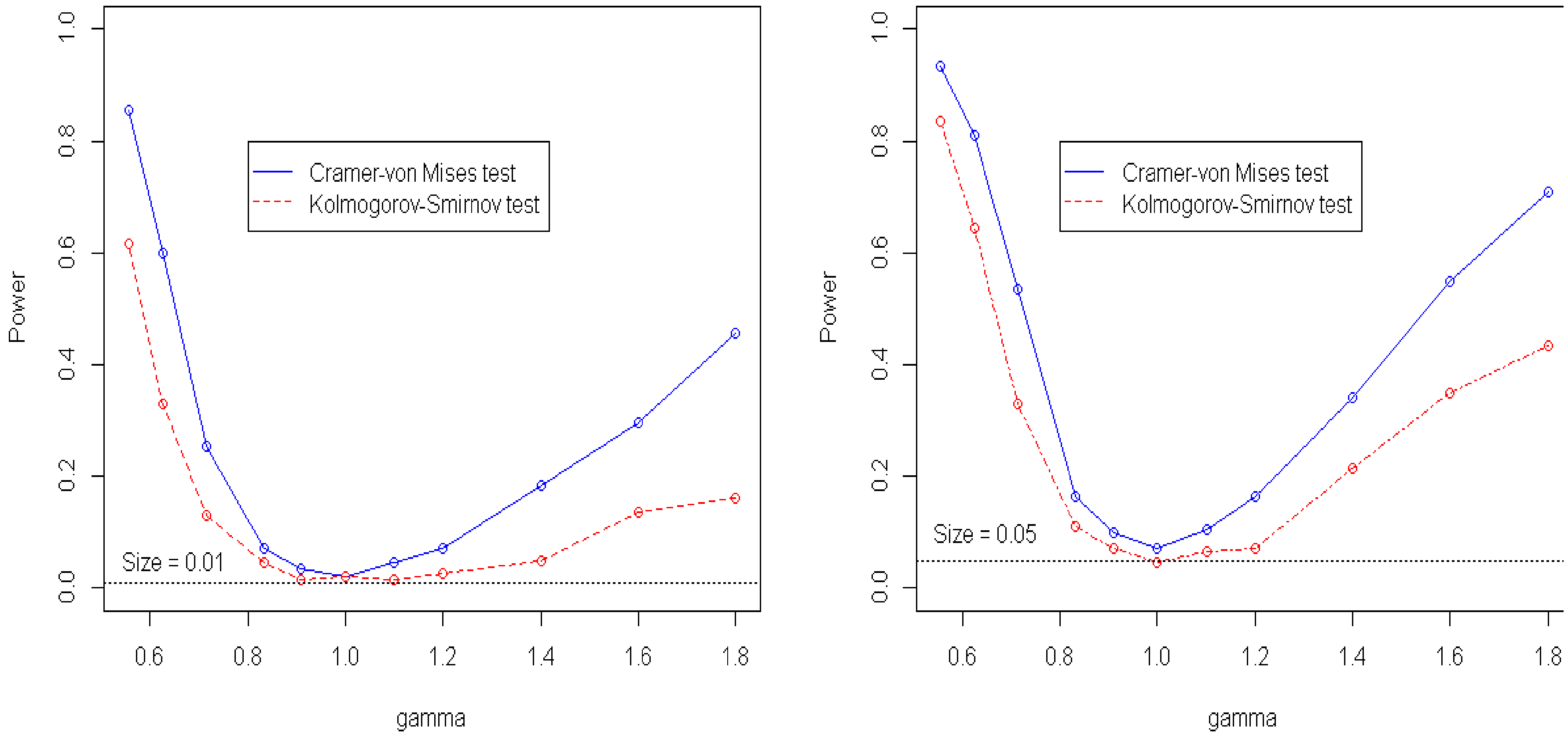


Fig.2: The power curves for the proposed goodness-of-fit tests with sizes $\alpha = 0.01$ (left panel) and $\alpha = 0.05$ (right panel) based on n = 150. The parameter γ =0 corresponds to the null model while $\gamma \neq 1$ corresponds to the alternative model.

Table 6: Coverage rates of the proposed confidence bands at the $100(1-\alpha)\%$ level based on 1000 replications.

Nominal level	n = 100	n = 150	n = 250
$1-\alpha$ =0.900	0.912	0.919	0.922
$1-\alpha$ =0.950	0.963	0.967	0.970
$1-\alpha$ =0.990	0.994	0.993	0.995

- Data Analysis

  n = 409 childhood cancer cases:

  $$\{ (U_j, T_j, V_j) : j = 1, \ldots, 409 \}$$

  (available in Moreira and Álvarez, 2010)

- Objective:

  Inference on cancer appearance distribution

  $$F(t) = \Pr(T^* \leq t)$$

  T^* : Time to diagnosis of cancer (from birth)

- Goodness-of-fit test

$$H_{01}:\ F(t) = \frac{t}{5475}\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$$

: Cancer occurs *uniformly* below age 15 years

$$H_{02}:\ F(t) = \left(\frac{t}{5475}\right)^{3/4}\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$$

: Cancer occurs more frequently on early ages

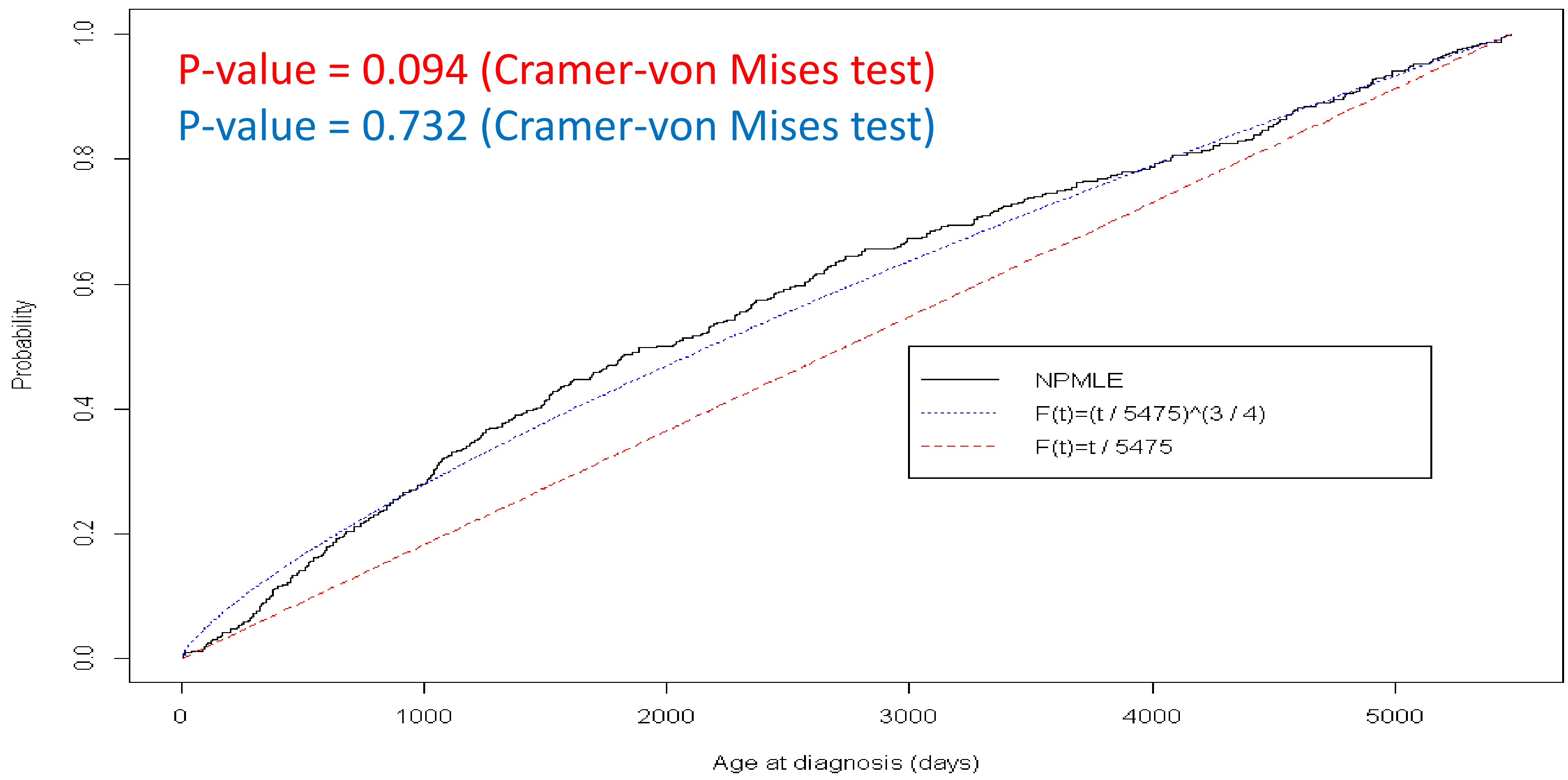


Fig.2: The NPMLE of the distribution of the age at diagnosis for the childhood cancer and the two hypothesized curves for $H_{01}: F(t) = (t/5475)\,\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$ and $H_{02}: F(t) = (t/5475)^{3/4}\,\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$.

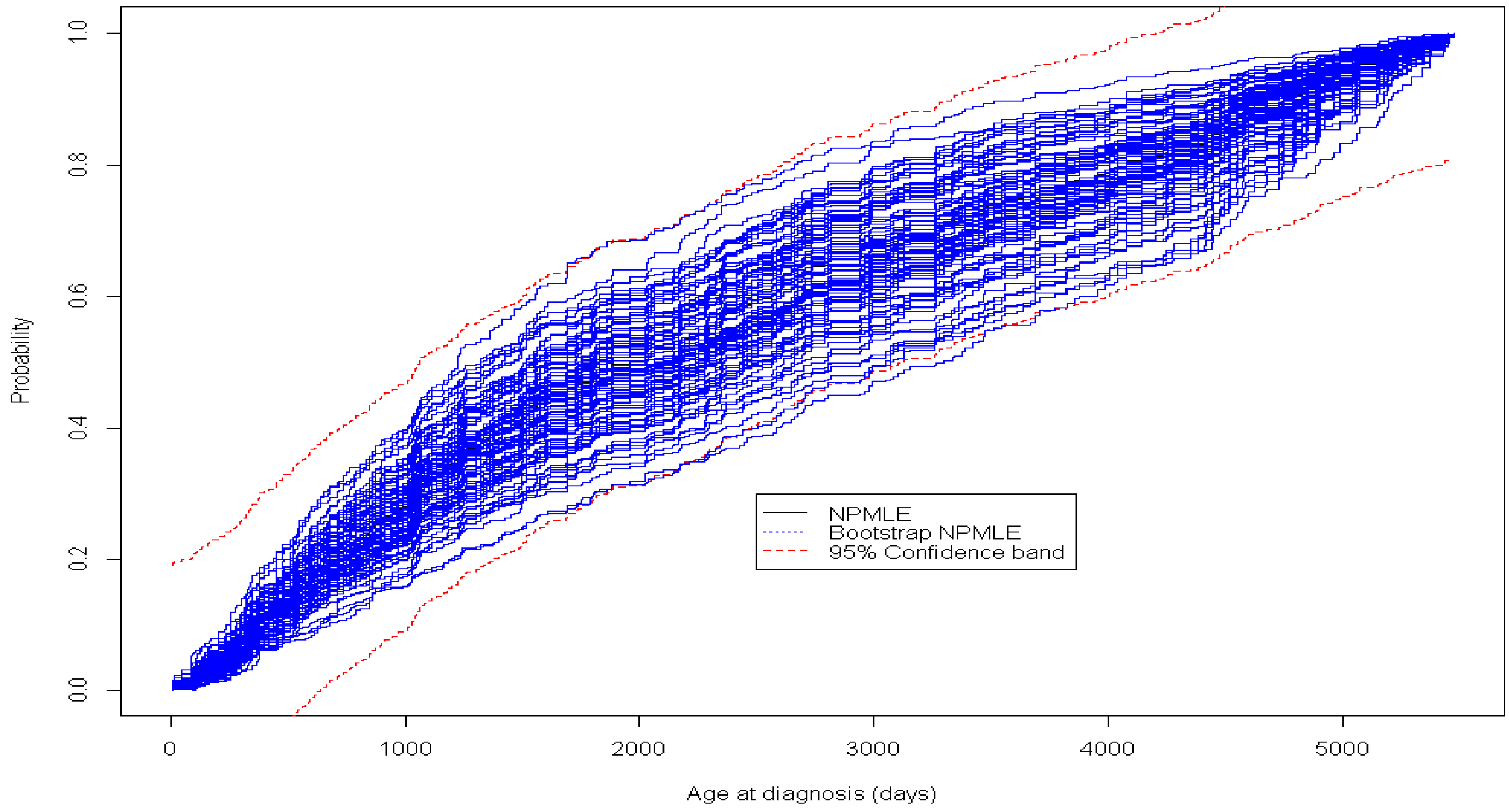


Fig.3: 95% confidence bands (red color, dotted lines). Validation of the 95% confidence bands are based on the 1000 Bootstrap NPMLE's, which result in 96.4% coverage. The first 100 Bootstrap NPMLE's (blue color) are displayed, which result in 97% coverage.

Summary

- We derive a simple analytical variance-covariance estimator of the NPMLE
- Reduces computational cost over the Bootstrap and jackknife

Table 7: Variance estimates of the NPMLE based the childhood cancer data.

	Proposed: $\sqrt{\hat{V}_{\text{Info}}\{\hat{F}(t)\}}$	Bootstrap: $\sqrt{\hat{V}_{\text{Boot}}\{\hat{F}(t)\}}$	Jackknife: $\sqrt{\hat{V}_{\text{Jack}}\{\hat{F}(t)\}}$
Variance estimate at t = 750.0	0.0469	0.0458	0.0473
Variance estimate at t = 2083.5	0.0817	0.0828	0.0815
Variance estimate at t = 4251.0	0.0599	0.0665	0.0644
Computation time (sec)	0.52	1158.87	438.85

- However, the Bootstrap variance estimator is most accurate in terms of MSE.
- Estimated covariance structure allows various inference, including goodness-of-fit and confidence bands, etc.

Thank you for your kind attention