An Empirical Saddlepoint Approximation Method for Smoothing Survival Functions Under Right-Censoring

Alex Trindade

Dept. of Mathematics & Statistics, *Texas Tech University*

Pratheepa Jeganathan, *Stanford University* (former student)

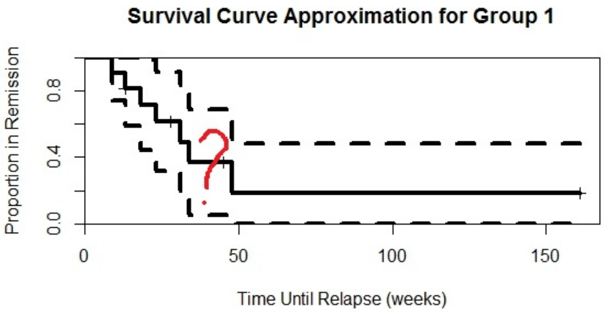
Rob Paige, *Missouri University of Science and Technology*

Noro Randrianampy, *HECMMA*, Madagascar (former student)

July 2017

1. Motivation
2. Construction of proposed method
3. Large-sample properties of method
4. Simulation study
5. Real data examples

- Goal: produce smooth survival function (nonparametrically...)
- Kaplan-Meier (KM) estimator: best method? (Maximizes NP likelihood, but is not smooth...)



- Let T be a continuous random variable
  - Unknown density function $f(t)$ and CDF $F(t)$
  - Known moment generating function (MGF) $M(s)$ and thus, cumulant generating function (CGF) $K(s) = \log M(s)$
- Saddlepoint approximation for density:

$$\hat{f}_S(t) = \left[\frac{1}{2\pi K^{(2)}(\hat{s}_t)}\right]^{1/2} \exp\left\{K(\hat{s}_t) - \hat{s}_t t\right\}$$

where $\hat{s}_t$ is such that $K^{(1)}(\hat{s}_t) = t$

- SPA for CDF:

$$\hat{F}_S(t) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})\left(\hat{w}^{-1} - \hat{u}^{-1}\right) & \textit{if} \quad \hat{s}_t \neq 0 \\ \frac{1}{2} + \frac{K^{(3)}(0)}{6\sqrt{2\pi}K^{(2)}(0)^{3/2}} & \textit{if} \quad \hat{s}_t = 0 \end{cases}$$

  - $\hat{w} = sgn(\hat{s}_t)\sqrt{2\left[\hat{s}_t t - K(\hat{s}_t)\right]}$
  - $\hat{u} = \hat{s}_t\sqrt{K^{(2)}(\hat{s}_t)}$
  - $\Phi(.)$ and $\phi(.)$: CDF and PDF of $N(0, 1)$
- Basic idea in both: invert characteristic function by approximating the integral over the complex plane (inversion formula)

- Start with empirical MGF

$$\tilde{M}_n(s) = \frac{1}{n}\sum_{i=1}^{n} e^{t_i s}$$

- Now replace known CGF by $\tilde{K}_n(s) = \log \tilde{M}_n(s)$ in SPA formulas:
  - $$\tilde{f}_n(t) = \left[\frac{1}{2\pi \tilde{K}_n^{(2)}(\hat{s}_t)}\right]^{1/2} exp\left\{\tilde{K}_n(\hat{s}_t) - \hat{s}_t t\right\}$$
  - Similarly, $\tilde{F}_n(t) = \cdots$

Notations

- $T_1, ..., T_n$: IID failure times
- $Y_1, ..., Y_n$: Censoring times, $T \perp\!\!\!\perp Y$
- $f(t)$, $F(t)$, $S(t)$: Density, CDF, survival function of the failure times
- $g(y)$, $G(y)$, and $S_g(y)$: Density, CDF, and survival function of the censoring times
- $Z_i = \min(T_i, Y_i)$
- $\Delta_i = I\{T_i \leq Y_i\}$
- Observe: (Z_i, Δ_i)
- $Z_{(i)}$-i^{th} order statistic of $Z_1, \ldots, Z_n$
- $\Delta_{(i)}$ - concomitant of $Z_{(i)}$
- $H(\cdot)$ distribution of Z
- $\tau = \sup\{z : H(z) \leq 1\}$

- For right censored failure times (Z_i, Δ_i), have KM survival estimate:

$$\hat{S}_K = \prod_{i=1}^{n} \left(1 - \frac{\Delta_{(i)}}{n-i+1}\right)^{I\left[Z_{(i)} \leq t\right]}$$

- Corresponding KM PDF approximation (discrete):

$$\hat{f}_{KM}\left(t = z_{(i)}\right) = \frac{\Delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{1}{n-j+1}\right)^{\Delta_{(j)}}, \quad i = 1, ..., n$$

- But: $\sum_{i=1}^{n} \hat{f}_{KM}\left(z_{(i)}\right) \neq 1$ when last obs is censored...
- Brown et al (1974): allocate remaining mass meaningfully over $[z_{(n)}, \infty)$ with an exponential distribution

- This tail-completed KM PDF defines a modified empirical MGF through Riemann-Stieltjes integral :

$$\tilde{M}_n(s) = \int_0^{z_{(n)}} e^{ts} d\hat{F}_K(t) + \int_{z_{(n)}}^{\infty} e^{st} \phi_n e^{-\phi_n t} dt$$

- And associated j-th derivative:

$$\tilde{M}_n^{(j)}(s) = \int_0^{z_{(n)}} t^j e^{st} d\hat{F}_K(t) + \frac{\phi_n}{\phi_n - s} \left(z_{(n)} + \frac{1}{\phi_n - s} \right)^j e^{-(\phi_n - s) z_{(n)}}$$
$$:= \hat{M}_n^{(j)}(s) + g_n^{(j)}(s)$$

$\forall j \in Z^* = \{\text{non-negative integers}\}$

Do what ? (?) did but using our $\tilde{M}_n(s)$!!!

- Modified Empirical Saddlepoint Approximation for Density:
  Replace $K(s)$ with $\tilde{K}_n(s)$ in saddlepoint density approximation

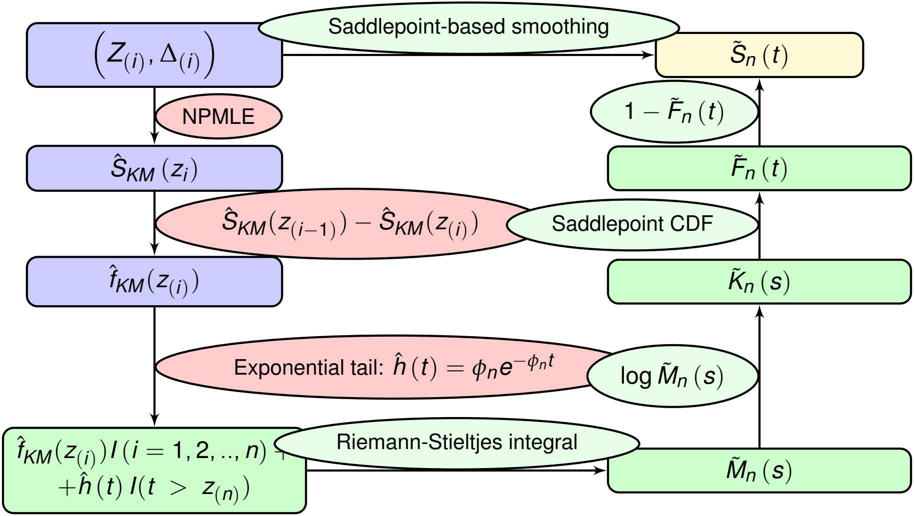
$$\tilde{f}_n(t) = \cdots$$

- Modified Empirical Saddlepoint Approximation for Distribution:
  Replaces known $K(s)$ with $\tilde{K}_n(s)$ in saddlepoint distribution approximation

$$\tilde{F}_n(t) = \cdots$$

- Smoothed Survival Function:

$$\tilde{S}_n(t) = 1 - \tilde{F}_n(t)$$

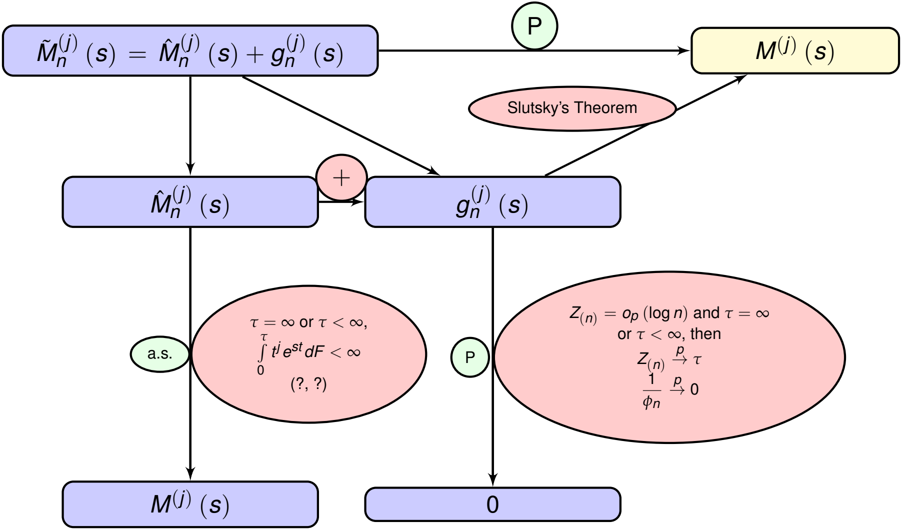

$\left(Z_{(i)}, \Delta_{(i)}\right)$
Saddlepoint-based smoothing
$\tilde{S}_n(t)$
NPMLE
$1 - \tilde{F}_n(t)$
$\hat{S}_{KM}(z_i)$
$\tilde{F}_n(t)$
$\hat{S}_{KM}(z_{(i-1)}) - \hat{S}_{KM}(z_{(i)})$
Saddlepoint CDF
$\hat{f}_{KM}(z_{(i)})$
$\tilde{K}_n(s)$
Exponential tail: $\hat{h}(t) = \phi_n e^{-\phi_n t}$
$\log \tilde{M}_n(s)$
$\hat{f}_{KM}(z_{(i)}) I(i = 1, 2, .., n) +$
$+\hat{h}(t) I(t > z_{(n)})$
Riemann-Stieltjes integral
$\tilde{M}_n(s)$

Consistency of $\tilde{M}_n^{(j)}(s)$

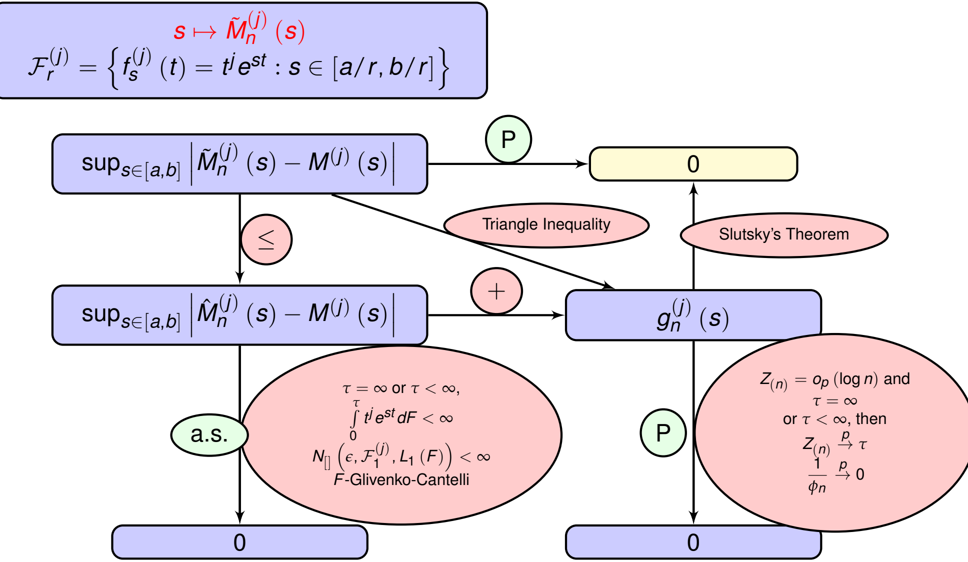
- **Weak consistency:** for $s \in [a, b] \subset (-\infty, \phi_n)$, and regularity conditions
$$\tilde{M}_n^{(j)}(s) \xrightarrow{p} M^{(j)}(s)$$

- **Uniform consistency:** under additional regularity conditions,
  - (Uniform) weak consistency,
$$\sup_{s\in[a,b]} \left| \tilde{M}_n^{(j)}(s) - M^{(j)}(s) \right| \xrightarrow{P} 0$$
  - (Uniform) strong consistency (under constant right censoring),
$$\sup_{s\in[a,b]} \left| \tilde{M}_n^{(j)}(s) - M^{(j)}(s) \right| \xrightarrow{a.s.} 0$$

Proof: (Pointwise) Consistency



Proof: Uniform Weak Consistency



Under similar regularity conditions as for consistency, have:

- Weak convergence:

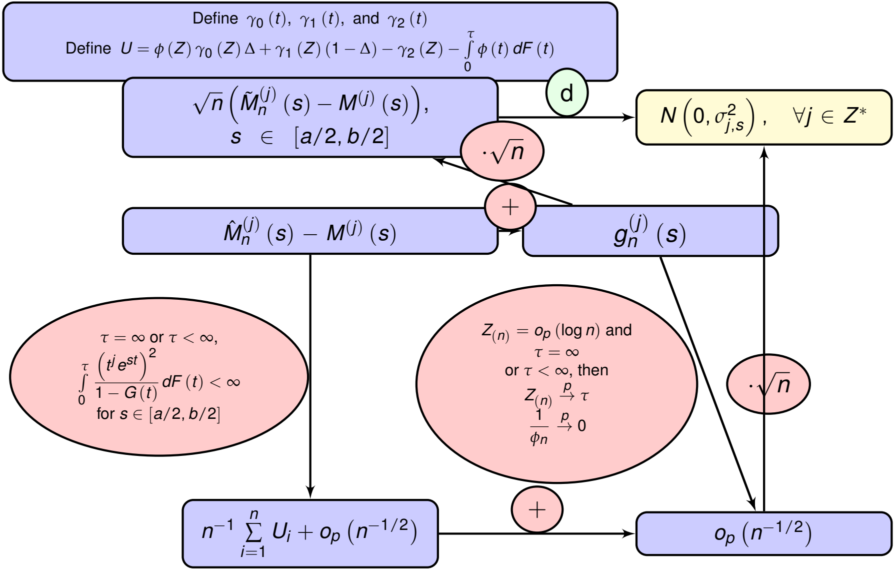
$$\sqrt{n}\left(\tilde{M}_n^{(j)}(s) - M^{(j)}(s)\right) \xrightarrow{d} N\left(0, \sigma_{j,s}^2\right)$$

- Uniform weak convergence:

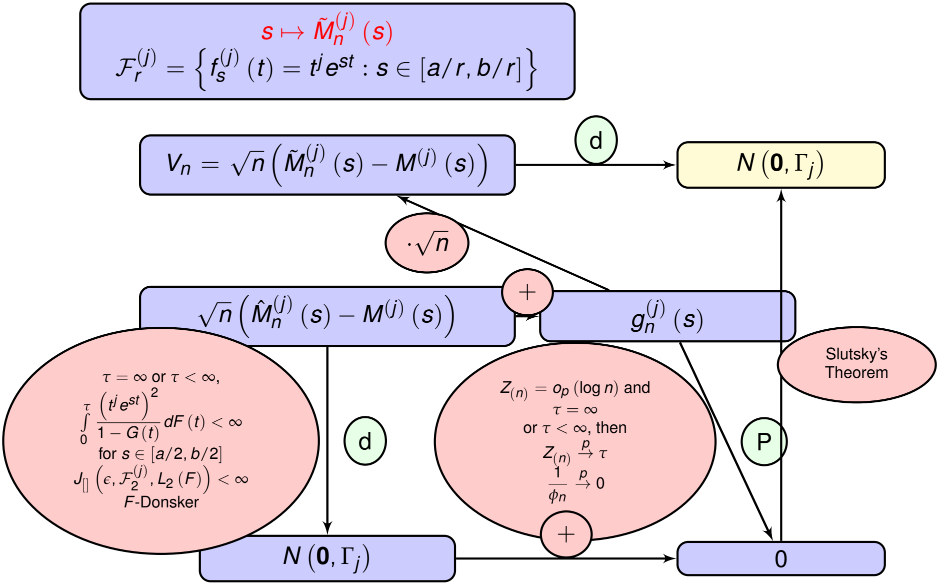
$$\sqrt{n}\left(\tilde{M}_n^{(j)}(s) - M^{(j)}(s)\right) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_j\right)$$

for all $j \in Z^*$ on $s \in [a/2, b/2]$

Proof: Weak Convergence



Proof: Uniform Weak Convergence



- $$\hat{f}_s(y) = h\left(M(s), M^{(j)}(s), \hat{s}_t\right), \quad j = 1, 2$$

- M-estimation and continuous mapping theorem: weak consistency of $\tilde{f}_n(t)$ to $\hat{f}_S(t)$
$$\tilde{f}_n(y) \xrightarrow{p} \hat{f}_s(y)$$

- M-estimation and delta methods: weak convergence of $\tilde{f}_n(t)$ to a normal distribution
$$\sqrt{n}\left(\tilde{f}_n(y) - \hat{f}_s(y)\right) \xrightarrow{d} N(0, \nu)$$
for all y such that $s \in [a/2, b/2]$

Proof: Set Up Notation

$$\hat{f}_s(y) = h(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\left(\frac{\theta_1\theta_3 - \theta_2^2}{\theta_1^2}\right)}} e^{[\ln\theta_1 - \theta_4 y]}$$

$$\theta_j = M^{(j-1)}(s), j = 1, 2, 3$$
$$\theta_4 = s_y$$

$$\hat{\theta}_{n,j} = \tilde{M}_n^{(j-1)}(s), j = 1, 2, 3$$
$$\hat{\theta}_{n,4} = \hat{s}_y$$

$$\psi_j(t, \theta_j) = t^{j-1}e^{st} - \theta_j, j = 1, 2, 3$$
$$\psi_4(t, \theta_4) = te^{\theta_4 t} - ye^{\theta_4 t}$$

$\boldsymbol{\theta}_0$- true value of $\boldsymbol{\theta}$

$$\lambda_{j,F}(\theta_j) = \int_0^\tau \left(t^{j-1}e^{st} - \theta_j\right) dF(t), j = 1, 2, 3$$
$$\lambda_{4,F}(\theta_4) = \int_0^\tau \left(te^{\theta_4 t} - ye^{\theta_4 t}\right) dF(t)$$

$$\lambda_{j,n}(\theta_j) = \int_0^\tau \left(t^{j-1}e^{st} - \theta_j\right) d\hat{F}_{KM}(t) + g_n^{(j)}(s), j = 1, 2, 3$$
$$\lambda_{4,n}(\theta_4) = \int_0^\tau \left(te^{\theta_4 t} - ye^{\theta_4 t}\right) d\hat{F}_{KM}(t) + g_n(s) + g_n^{(1)}(s)$$

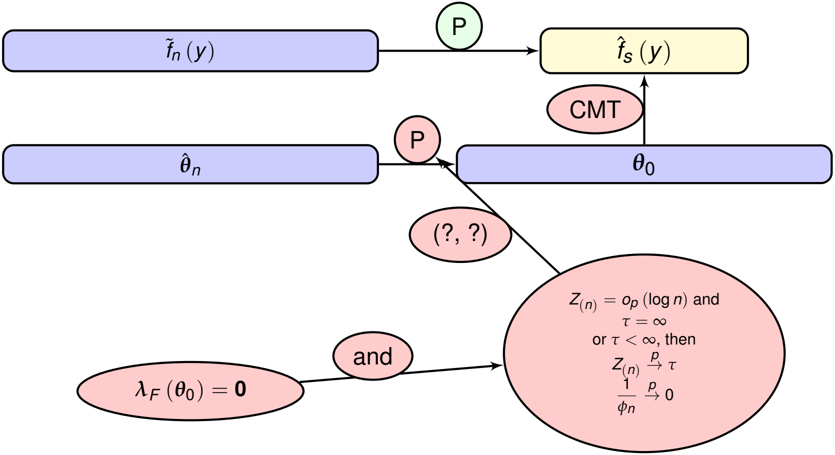
$$\boldsymbol{\psi}(t, \boldsymbol{\theta}) = \left[\psi_1(t, \boldsymbol{\theta}), \cdots, \psi_4(t, \boldsymbol{\theta})\right]^T$$
$$U = \phi(Z)\gamma_0(Z)\Delta + \gamma_1(Z)(1 - \Delta) - \gamma_2(Z) - \int_0^\tau \phi(t)\, dF(t)$$

Replace ϕ by $\psi_j(t, \boldsymbol{\theta})$

Get $\gamma_{i,j}, i = 0, 1, 2$ and $j = 1, 2, 3, 4$

Proof: Weak Consistency

$$\hat{f}_s(y) = h(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\left(\frac{\theta_1\theta_3 - \theta_2^2}{\theta_1^2}\right)}} e^{[\ln\theta_1 - \theta_4 y]}$$



$$\hat{f}_s(y) = h(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\left(\frac{\theta_1\theta_3 - \theta_2^2}{\theta_1^2}\right)}} e^{[\ln\theta_1 - \theta_4 y]}$$

$$\sqrt{n}\left(\tilde{f}_n(y) - \hat{f}_s(y)\right) \xrightarrow{d} N(0, \nu)$$

$h(\boldsymbol{\theta})$ real cts
Multivariate Delta Method
$\nu = \nabla h(\boldsymbol{\theta})^T \Sigma \nabla h(\boldsymbol{\theta})$

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

(?, ?)
M-estimation *and* $g_n^{(j)}(s) = o_p(1)$

$\lambda_F(\boldsymbol{\theta}_0) = \mathbf{0}$
$\psi_j(t, \boldsymbol{\theta})$ - cont at $\boldsymbol{\theta}_0$ uniformly in t $\forall j$
$\Longrightarrow$ $\psi_j(t, \boldsymbol{\theta})$ - bdd on $[a, b]$, $\forall j$
$\boldsymbol{\psi}$ -s differentiable w.r.t $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$.

and

$Z_{(n)} = o_p(\log n)$ and
$\tau = \infty$
or $\tau < \infty$, then
$Z_{(n)} \xrightarrow{p} \tau$
$\frac{1}{\phi_n} \xrightarrow{p} 0$

- The empirical saddlepoint approximation method (ESP) for density approximation is compared with the presmooth method (PS)
- Three models chosen for simulations: 500 runs
  1. Model 1: Weibull (shape=0.5, scale=10), exponential-like
  2. Model 2: Weibull (shape=3, scale=10), unimodal and almost symmetric
  3. Model 3: Weibull (shape=2, scale=$\sqrt{2/3}$), corresponds to Model 2 in ? (?), unimodal and positively skewed
- Use MISE to assess quality of approximation

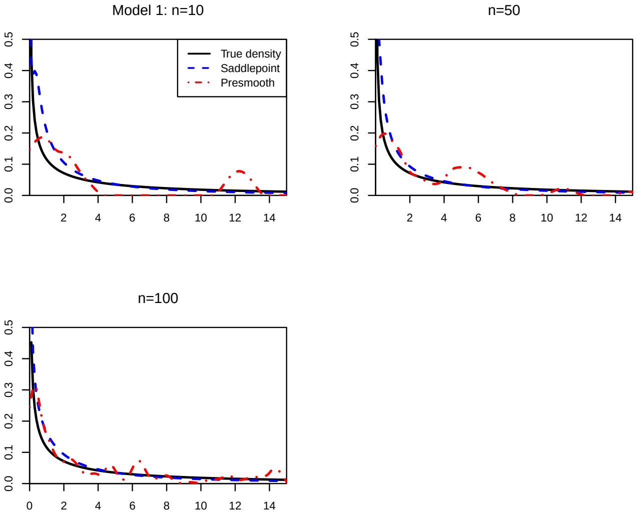
$$\begin{aligned} MISE\left(\tilde{f}_n(t)\right) &= \mathbb{E}\left[\int\left(\tilde{f}_n(t) - f(t)\right)^2\right] \\ &= \int Bias^2\left(\tilde{f}_n(t)\right) dt + \int Var\left(\tilde{f}_n(t)\right) dt \end{aligned}$$

Table : ESP=empirical saddlepoint approximation and PS=presmooth

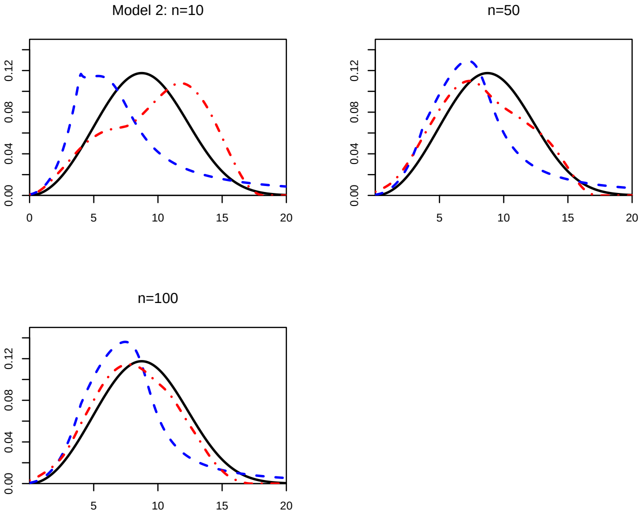
Sample Size	Model 1		Model 2		Model 3	
	ESP	PS	ESP	PS	ESP	PS
$n = 10$	0.06144	0.13130	0.02982	0.03216	0.36124	0.25387
$n = 50$	0.01619	0.03111	0.01559	0.00479	0.15294	0.10177
$n = 100$	0.01428	0.02214	0.01424	0.00227	0.13518	0.09081

- PS: Model 3 is contrived to be optimal
- ESP: very competitive for low n; really good for exp-type PDFs

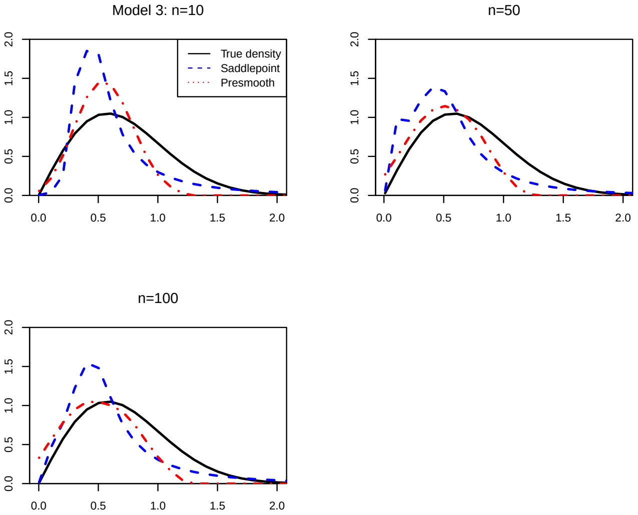
Density Approximation (Model 1)



Density Approximation (Model 2)



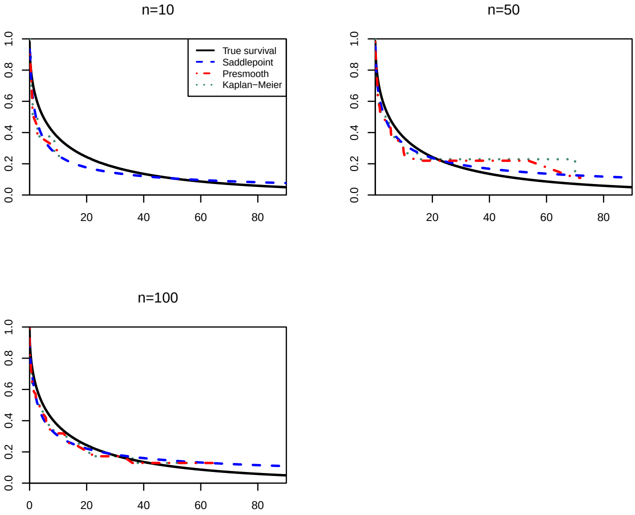
Density Approximation (Model 3)



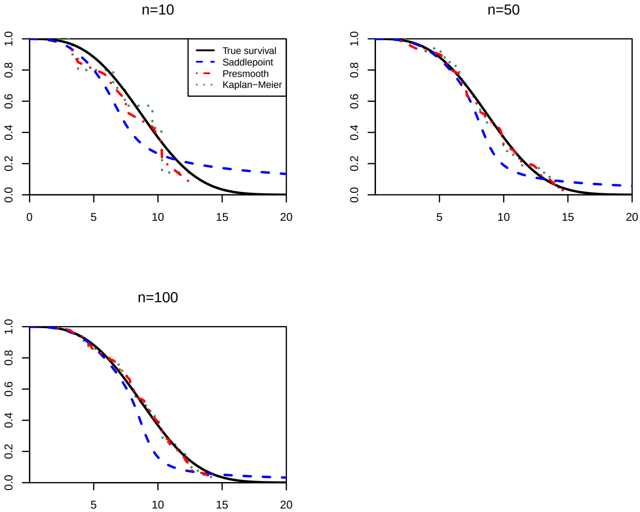
Sample Size	Model 1				Model 2			
	ESP		PS		ESP		PS	
	Bias	Variance	Bias	Variance	Bias	Variance	Bias	Variance
$n = 10$	0.04845	0.00674	0.06068	0.05487	0.02636	0.00346	0.01489	0.02638
$n = 50$	0.01613	0.00006	0.02897	0.00215	0.01508	0.00051	0.00324	0.00156
$n = 100$	0.01428	0.00001	0.02132	0.00082	0.01401	0.00023	0.00194	0.00032

- PS: kernel-density type smoother; larger variance & smaller bias
- ESP:
  - no tuning parameters
  - very competitive for low n
  - larger bias & smaller variance (but less boundary bias at $t = 0$)

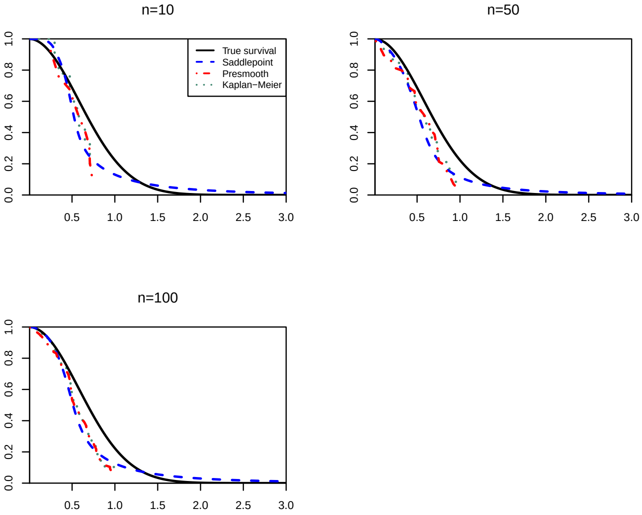
Survival Function Approximation (Model 1)



Survival Function Approximation (Model 2)



Survival Function Approximation (Model 3)



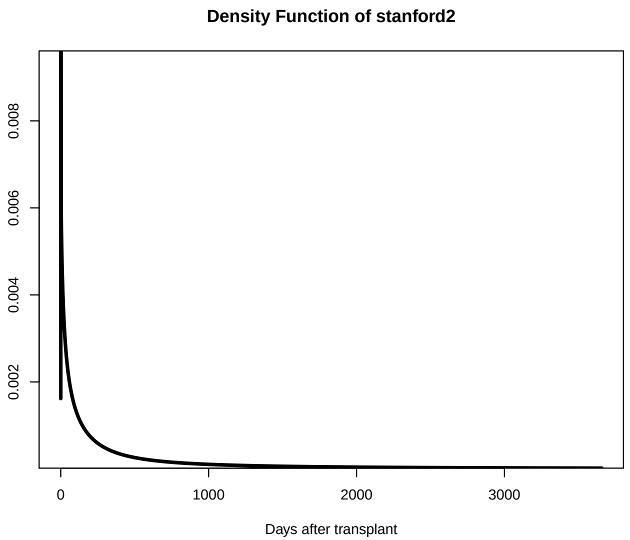
PRE's of Percentile Estimates (median values, based on 500 reps)

Model	Sample Size	10th Percentile			95th Percentile		
		ESP	KM	PS	ESP	KM	PS
Model 1	$n = 10$	85	1098	**80**	**307**	∞	∞
	$n = 50$	60	1295	**50**	**64**	∞	∞
	$n = 100$	45	2314	**36**	**47**	∞	∞
Model 2	$n = 10$	**15**	71	18	**428**	∞	∞
	$n = 50$	**6**	73	8	66	∞	**7**
	$n = 100$	7	68	**6**	17	48	**4**
Model 3	$n = 10$	**26**	96	33	**33**	∞	∞
	$n = 50$	**15**	100	30	**20**	∞	33
	$n = 100$	**15**	89	30	**13**	∞	33

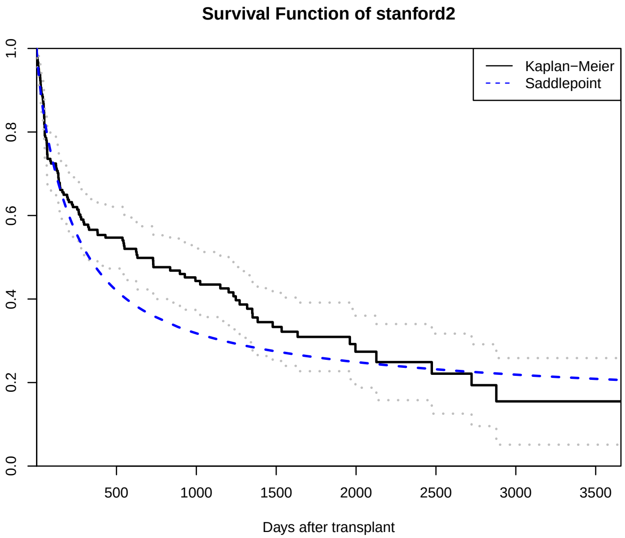
- KM & PS: discrete survival approximations, but ESP is smooth!
- **Thus:** no immediately obvious competitor for ESP...
- Support of KM & PS limited to obs times, but ESP produces approx over $[0, \infty)$
- ESP closer to truth for all n in models with convex (or mostly convex) survival function (Models 1 & 3)
- PS closer to truth for models with predominantly concave survival function (Model 2).
- **ESP: straightforward to generalize** to other types of censoring, e.g., interval-censored data...

- `stanford2` data set in the 'survival' package in R
- 184 transplant cases, time (measured from the date of transplant in days), status (dead or alive), age (patient age at first transplant, in years), mismatch score
- Death day 0 is recorded as 0.5 day
- $Z_{(184)} = 3695$ days
- Censoring percentage= 38.59% and $\Delta_{(184)} = 0$
- Support $t \in (0.005, 3658.224)$
- $Z_{(184)} = 3695 > 3658.224$ because the density approximation reaches zero at 3658.224

Empirical Saddlepoint Density Approximation

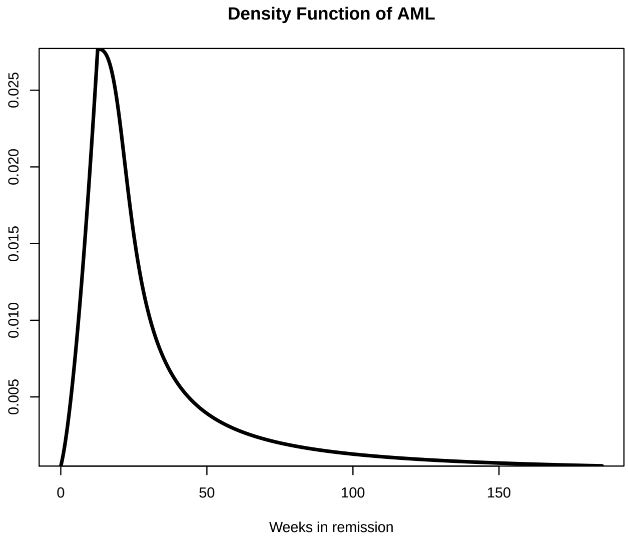


Smoothed Survival Function

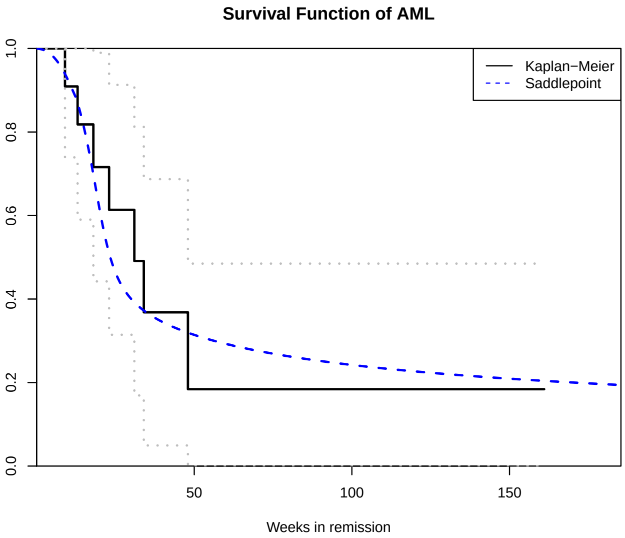


- `aml` data set in the 'survival' package in R
- Remission time of patients with AML, censoring status, and whether the chemotherapy treatment was maintained or not
- 23 patients were randomized into two groups
- First group received maintenance chemotherapy and the second did not
- Consider only 1st group, $n = 11$ obs with $\Delta_{(11)} = 0$

Density of the remission in weeks



Smoothed Survival Function



- Feuerverger, A. (1989). On the empirical saddlepoint approximation. Biometrika, 76, 457–464.
- Jacome, M., Gijbels, I., & Cao, R. (2008). Comparison of presmoothing methods in kernel density estimation under censoring. Computational Statistics, 23, 381–406.
- Lugannani, R., & Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. Advances in applied probability, 475–490.
- Stute, W., & Wang, J.-L. (1993). The strong law under random censorship. The Annals of Statistics, 1591–1607.
- Wang, J.-L. (1999). Asymptotic properties of m-estimators based on estimating equations and censored data. Scandinavian journal of statistics, 26, 297–318.