Workshop on Empirical Likelihood Methods in Survival Analysis

Ian McKeague
Department of Statistics
Florida State University
Tallahassee, FL 32306-4330, USA
<http://stat.fsu.edu/∼mckeague>

Outline

- Background on empirical likelihood (EL)
- Background on survival analysis
- EL methods in one-sample problems with censoring
- Two-sample problems

Classical likelihood ratio method

$\{F_\eta\}$ a parametric model

$\theta = \theta(\eta)$ a q-dimensional parameter.

Likelihood ratio statistic:

$$R(\theta_0) = \frac{\sup\{L(\eta) : \theta(\eta) = \theta_0\}}{\sup\{L(\eta)\}}$$

Accept $\theta = \theta_0$ if $R(\theta_0)$ is large.

Theorem (Wilks, 1938). Under mild regularity conditions, if $\theta = \theta_0$ then

$$-2 \log R(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_q.$$

Likelihood ratio confidence region for θ:

$$\{\theta : -2 \log R(\theta) \leq \chi^2_{q,\alpha}\}$$

where $\chi^2_{q,\alpha}$ is the upper α-quantile.

Improvement over Wald-type confidence regions.

Background on Empirical likelihood

- Thomas and Grunkemeier (1975) for survival function estimation. Owen (1988, 1990, . . . , 2001).

- First developed for finite-dimensional features $\theta = \theta(F)$ of a cdf (e.g., mean, median, cdf at a single point).

Advantages	Disadvantages
reflects emphasis on the observed data (cf. bootstrap)	computational problems more severe than in Wald type procedures (Lagrange multipliers)
better small sample performance than approaches based on asymptotic normality (uses Neyman–Pearson critical regions)	asympt of LR statistics can be difficult to develop beyond the classical parametric setting, e.g., Cox model with interval censoring
confidence bands reflect the range of the parameter	
often yields distribution-free tests (no need for simulation)	
regularity conditions are weak and natural (smoothness conditions often not needed)	
confidence regions are Bartlett correctable (unlike bootstrap) and transformation preserving	

Empirical cdf

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \le x\}$$

Nonparametric likelihood

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-)).$$

F_n is the NPMLE:

$$F_n = \arg\max_F L(F)$$

EL ratio

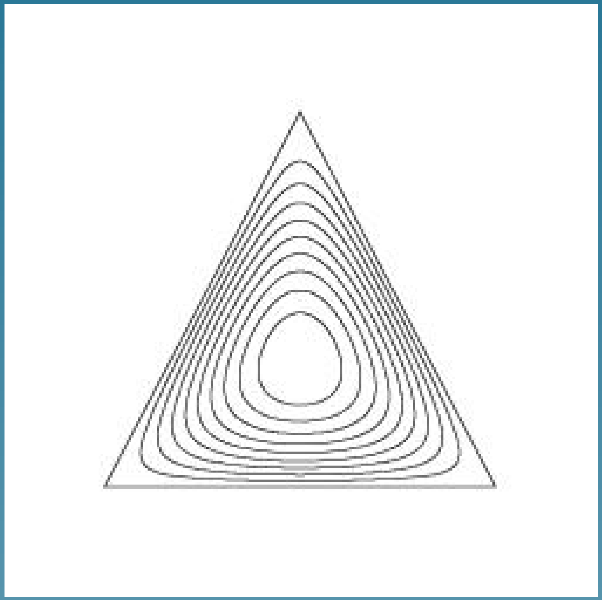
$$\tilde{R}(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i$$

where (part of) the mass on X_i is $p_i \geq 0$, $\sum_{i=1}^{n} p_i \leq 1$.

To maximize $\tilde{R}(F)$, only need consider F supported on the data, i.e.,

$$\sum_{i=1}^{n} p_i = 1.$$

Contours of EL ratio for $n = 3$



Simplex

$$\{(p_1, p_2, p_3) : p_i \geq 0, p_1 + p_2 + p_3 = 1\}$$

Lemma If $\tilde{R}(F) \geq r_0 > 0$ then F places mass $m_n = O(1/n)$ outside $\{X_1, \ldots, X_n\}$.

Proof

$$r_0 \leq \tilde{R}(F) = \prod_{i=1}^{n} np_i \leq \prod_{i=1}^{n} n \left(\frac{1 - m_n}{n} \right) = (1 - m_n)^n$$

$$m_n \leq 1 - \exp(-n^{-1} \log(1/r_0)) \leq n^{-1} \log(1/r_0). \qquad \square$$

EL function

$$R(\theta_0) \quad = \quad \sup\{\tilde{R}(F) : \theta(F) = \theta_0\}$$

Equivalently

$$R(\theta_0) \quad = \quad \frac{\sup\{L(F) : \theta(F) = \theta_0\}}{\sup\{L(F)\}}$$

EL hypothesis tests

Accept $\theta(F) = \theta_0$ when $R(\theta_0) \geq r_0$ for some threshold r_0.

EL confidence regions

$$\{\theta \colon R(\theta) \geq r_0\}$$

with r_0 chosen via an EL analogue of Wilks's theorem.

EL for means

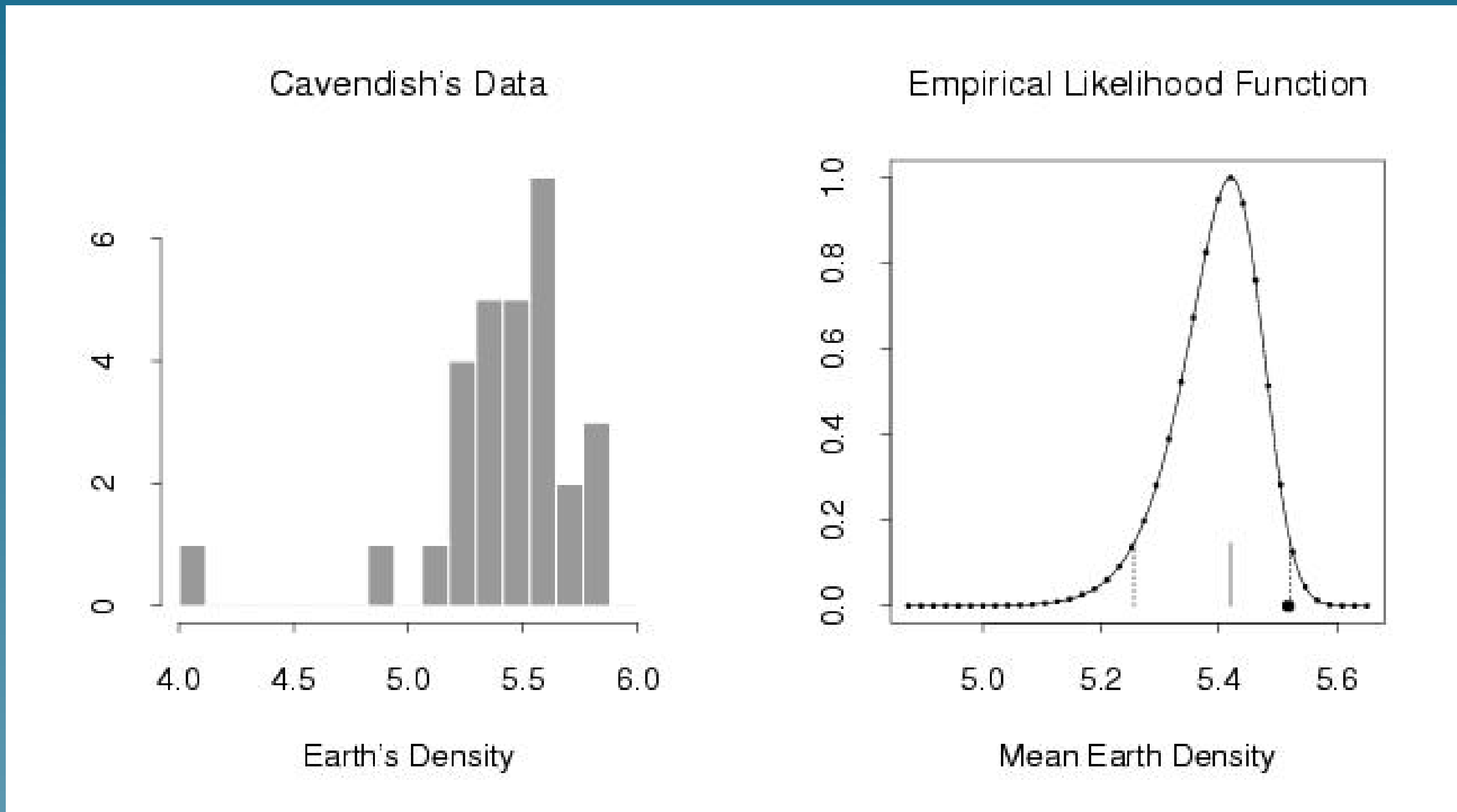
$\mu = E(X) \in \mathbb{R}^d$

$$R(\mu) = \max\left\{ \prod_{i=1}^{n} np_i \colon \sum_{i=1}^{n} p_i X_i = \mu, p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}$$

Computation of $R(\mu)$?

$$\{\mu \colon R(\mu) \geq r_0\} = \left\{ \sum_{i=1}^{n} p_i X_i \colon \prod_{i=1}^{n} np_i \geq r_0, p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}$$

Example



$R(\mu)$ (solid curve); 95% confidence limits (dotted bars); from Owen (2001).

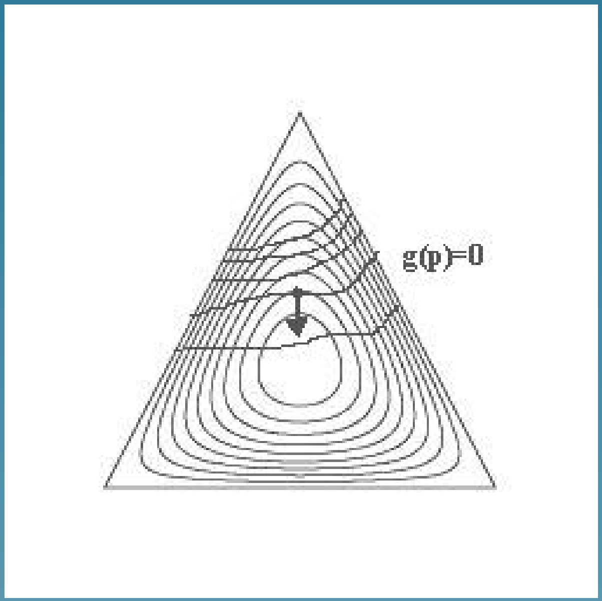
Method of Lagrange multipliers

Maximize $f(x)$ subject to the (multivariate) constraint $g(x) = 0$.

Find $x^* = x^*(\lambda)$ maximizing $f(x) - \lambda' g(x)$ such that $g(x^*) = 0$.
Then x^* solves the constrained problem.

Geometric intuition available when g is univariate

At the maximum, ∇f and ∇g must be parallel: $\nabla f = \lambda \nabla g$ for some constant λ (Lagrange multiplier).



Maximize

$$\log \tilde{R}(p_1, \ldots, p_n) = \sum_{i=1}^{n} \log(np_i)$$

under the constraints:

$$n \sum_{i=1}^{n} p_i(X_i - \mu) = 0, \quad 1 - \sum_{i=1}^{n} p_i = 0$$

Write

$$G = \sum_{i=1}^{n} \log(np_i) - n\lambda \sum_{i=1}^{n} p_i(X_i - \mu) - \gamma \left(1 - \sum_{i=1}^{n} p_i\right)$$

λ and γ are Lagrange multipliers.

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - n\lambda(X_i - \mu) + \gamma = 0$$

so

$$0 = \sum_{i=1}^{n} p_i \frac{\partial G}{\partial p_i} = n + \gamma$$

giving $\gamma = -n$. Thus

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(X_i - \mu)}$$

Plugging this back into the constraint:

$$g(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda(X_i - \mu)} = 0$$

This equation has a unique solution for $\lambda = \lambda(\mu)$.

Theorem (ELT, Owen 1990) $X_1, \ldots, X_n$ iid with finite mean μ_0, finite covariance matrix of rank $q > 0$. Then

$$-2 \log R(\mu_0) \xrightarrow{\mathcal{D}} \chi^2_q.$$

Sketch of proof Case $d = 1$. The Lagrange multiplier λ is the solution to

$$g(\lambda) = n^{-1} \sum_{i=1}^{n} \frac{X_i - \mu_0}{1 + \lambda(X_i - \mu_0)} = 0$$

and note that $g(0) = \bar{X} - \mu_0$. Denote $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \mu_0)^2$. Taylor expanding g gives

$$\begin{aligned} 0 &= g(\lambda) = g(0) + \lambda g'(0) + o_P(n^{-1/2}) \\ &= \bar{X} - \mu_0 - \lambda \hat{\sigma}^2 + o_P(n^{-1/2}) \end{aligned}$$

Thus $\lambda = (\bar{X} - \mu_0)/\hat{\sigma}^2 + o_P(n^{-1/2}) = O_P(n^{-1/2})$. Recall

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(X_i - \mu_0)}$$

so, using the Taylor expansion $\log(1 + x) = x - x^2/2 + O(x^3)$,

$$\begin{aligned}
-2\log R(\mu_0) &= -2\sum_{i=1}^{n} \log(np_i) = 2\sum_{i=1}^{n} \log(1 + \lambda(X_i - \mu_0)) \\
&= 2n\lambda(\bar{X} - \mu_0) - n\lambda^2\hat{\sigma}^2 + o_P(1) \\
&= 2n(\bar{X} - \mu_0)^2/\hat{\sigma}^2 - n(\bar{X} - \mu_0)^2/\hat{\sigma}^2 + o_P(1) \\
&= n(\bar{X} - \mu_0)^2/\hat{\sigma}^2 + o_P(1) \\
&\xrightarrow{\mathcal{D}} \chi_1^2
\end{aligned}$$

This suggests the χ^2-calibration with threshold

$$r_0 = \exp(-\chi^2_{q,\alpha}/2)$$

for a $100(1-\alpha)\%$ confidence region; actual coverage $1-\alpha+O(n^{-1})$.

Fisher calibration

$$\frac{d(n-1)}{n-d} F_{d,n-d,\alpha}$$

Bartlett correction

$$\left(1+\frac{a}{n}\right)\chi^2_{q,\alpha}$$

a involves higher-order moments of X, and needs to be estimated. Coverage improves to $1-\alpha+O(n^{-2})$.

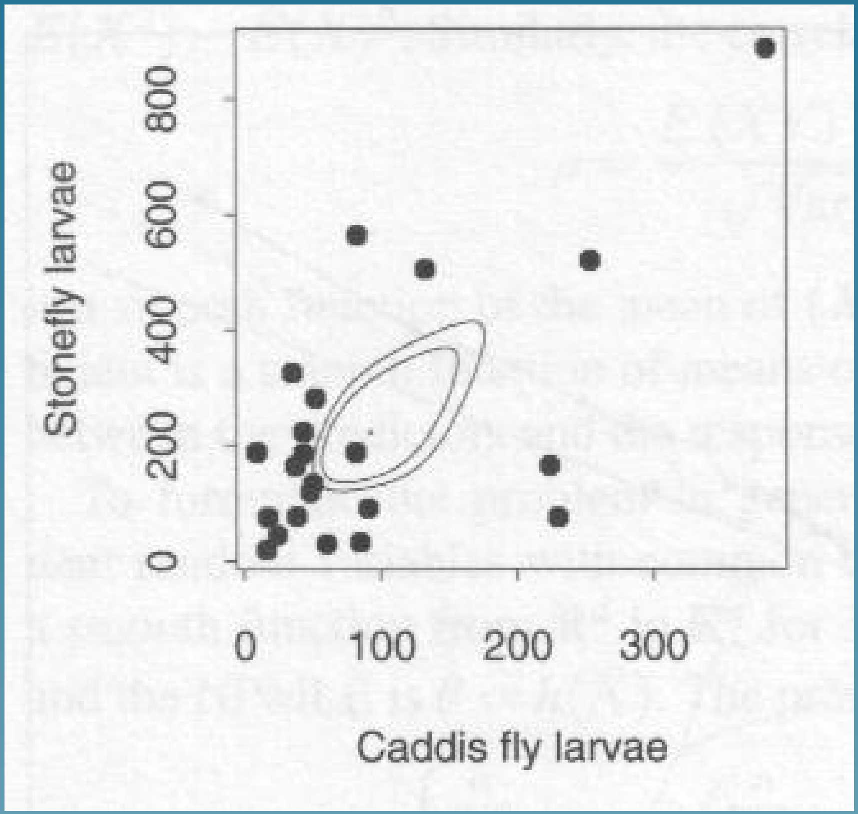
Bootstrap calibration

$X_1^*, \ldots, X_n^*$ iid from F_n. Simulation used to find the upper α-quantile of $-2\log R^*(\bar{X})$, where

$$R^*(\bar{X}) = \max\left\{\prod_{i=1}^{n} np_i \colon \sum_{i=1}^{n} p_i X_i^* = \bar{X}, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\right\}$$

Example

Counts of two types of aquatic larvae at 22 locations in Wales.



Bivariate 95% confidence regions calibrated by χ^2 and by the bootstrap (larger region); from Owen (2001).

Extensions

- Smooth functions of means: $\theta = h(\mu)$
- Linear functionals of F: $\theta = E(h(X)) = \int h(x)\, dF(x)$.
- Implicitly defined parameters: $E(m(X,\theta)) = 0$ where $m(X,\theta)$ is the estimating function; e.g., median, $m(X,\theta) = 1\{X \le \theta\} - .5$.

$$R(\theta) = \max\left\{ \prod_{i=1}^{n} np_i \colon \sum_{i=1}^{n} p_i m(X_i,\theta) = 0, p_i \ge 0, \sum_{i=1}^{n} p_i = 1 \right\}$$

Theorem Let $X_1, \ldots, X_n$ be iid, and suppose $m(X, \theta_0)$ has finite covariance matrix of rank $q > 0$. If $E(m(X, \theta_0)) = 0$, then

$$-2 \log R(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_q.$$

Proof Immediate from ELT upon some changes in notation: X is replaced by $m(X, \theta)$, which has mean zero when $\theta = \theta_0$.

Notice the basic ingredients:

- Taylor expansion of $g(\lambda)$ about 0 gives an explicit approximation to the Lagrange multiplier λ.
- Taylor expansion of $\log(np_i) = \log(1 + \lambda(X_i - \mu_0))$, then CLT.

EL simultaneous band for F

Local EL function at $\theta_0 = F_0(t)$:

$$
\begin{aligned}
R(t) &= \frac{\sup\{L(F) : F(t) = F_0(t)\}}{\sup\{L(F)\}} \\
&= \frac{\left(\frac{F_0(t)}{nF_n(t)}\right)^{nF_n(t)} \left(\frac{1-F_0(t)}{n(1-F_n(t))}\right)^{n(1-F_n(t))}}{\left(\frac{1}{n}\right)^n} \\
&= \left(\frac{F_0(t)}{F_n(t)}\right)^{nF_n(t)} \left(\frac{1-F_0(t)}{1-F_n(t)}\right)^{n(1-F_n(t))} .
\end{aligned}
$$

Hence

$$
\begin{aligned}
-2\log R(t) &= -2nF_n(t)\log\frac{F_0(t)}{F_n(t)} \\
&\quad -2n\left(1-F_n(t)\right)\log\frac{1-F_0(t)}{1-F_n(t)}.
\end{aligned}
$$

Taylor expanding $\log(1+x) = x - x^2/2 + O(x^3)$ we have

$$
-2\log R(t) = \left(\frac{\sqrt{n}(F_n(t)-F_0(t))}{\sqrt{F_0(t)(1-F_0(t))}}\right)^2 + o_P(1)
$$

As a *process* in $t \in [a,b]$:

$$
-2\log R(t) \xrightarrow{\mathcal{D}} \left(\frac{W^o(F_0(t))}{\sqrt{F_0(t)(1-F_0(t))}}\right)^2
$$

$$\stackrel{\mathcal{D}}{=} \quad \left(\frac{W(\sigma^2(t))}{\sigma(t)} \right)^2,$$

W^o standard tied-down Wiener process (Brownian bridge)
W standard Wiener process

$$\sigma^2(t) = \frac{F_0(t)}{1 - F_0(t)}.$$

Simultaneous confidence band for F over an interval $[a, b]$:

$$\{(t, F_0(t)) : -2 \log R(t) \leq C_\alpha, t \in [a, b]\}$$

C_α the upper α-quantile of

$$\sup_{t \in [\hat{\sigma}^2(a), \hat{\sigma}^2(b)]} \frac{W^2(t)}{t}.$$

Equal precision LR band. Narrower in tail than Hollander, McKeague, Yang (1997) band.

EL test for $F = F_0$

$$\begin{aligned} T_n &= -2 \int_{-\infty}^{\infty} \log R(t)\, dF_n(t) \\ &\xrightarrow{\mathcal{D}} \int_0^1 \left(\frac{W^o(t)}{\sqrt{t(1-t)}} \right)^2 dt. \end{aligned}$$

EL test for symmetry

Einmahl and McKeague (2001): EL tests for symmetry, exponentiality, independence and changes in distribution.

$$H_0 : F(-x) = 1 - F(x-), \quad \text{for all } x > 0.$$

Local EL function:

$$R(x) = \frac{\sup\{L(\widetilde{F}) : \widetilde{F}(-x) = 1 - \widetilde{F}(x-)\}}{\sup\{L(\widetilde{F})\}}, \quad x > 0.$$

Treat $\widetilde{F}$ as a function of $0 \leq p \leq 1$, where $\widetilde{F}$ puts mass

- $p/2$ on $(-\infty, -x]$, and on $[x, \infty)$
- $1 - p$ on $(-x, x)$

Point masses on observations in the respective intervals:

$$\frac{p/2}{n\hat{p}_1}, \frac{p/2}{n\hat{p}_2}, \frac{1-p}{n(1-\hat{p})},$$

$\hat{p} = \hat{p}_1 + \hat{p}_2$, $\hat{p}_1 = F_n(-x)$, $\hat{p}_2 = 1 - F_n(x-)$. Maximum of

$$\left(\frac{p/2}{n\hat{p}_1}\right)^{n\hat{p}_1} \left(\frac{p/2}{n\hat{p}_2}\right)^{n\hat{p}_2} \left(\frac{1-p}{n(1-\hat{p})}\right)^{n(1-\hat{p})},$$

attained at $p = \hat{p}$.

$$\begin{aligned}
\log R(x) &= n\hat{p}_1 \log \frac{\hat{p}}{2\hat{p}_1} + n\hat{p}_2 \log \frac{\hat{p}}{2\hat{p}_2} \\
&= nF_n(-x) \log \frac{F_n(-x) + 1 - F_n(x-)}{2F_n(-x)} \\
&\quad + n(1 - F_n(x-)) \log \frac{F_n(-x) + 1 - F_n(x-)}{2(1 - F_n(x-))}
\end{aligned}$$

Test statistic:

$$T_n = -2 \int_0^\infty \log R(x)\, dG_n(x),$$

G_n is the empirical cdf of the $|X_i|$.

Theorem Let F be continuous. Then, under H_0

$$T_n \xrightarrow{\mathcal{D}} \int_0^1 \frac{W^2(t)}{t}\,dt$$

Survival analysis

Right-censored lifetime data

Observe n iid pairs (Z_i, δ_i)
$Z_i = \min(X_i, Y_i)$, $\delta_i = I\{X_i \leq Y_i\}$, X_i and Y_i independent.
F : cdf of X_i
G : cdf of Y_i
$S = 1 - F$: survival function, $S(0) = 1$
$\Delta F(t) = F(t) - F(t-)$: jump at t
A: cumulative hazard function (chf)

$$A(t) = \int_{(0,t]} \frac{dF(s)}{1 - F(s-)}$$

Review of some basics

There is a 1-1 correspondence between survival functions and cumulative hazards. If F is continuous: $A = -\log(S)$, $S = \exp(-A)$.

Lemma If F is a discrete cdf, the corresponding cumulative hazard function is

$$A(t) = \sum_{s \le t} \frac{\Delta F(s)}{1 - F(s-)}.$$

Conversely, if A is a discrete chf, the corresponding survival function is

$$S(t) = \prod_{s \le t} (1 - \Delta A(t))$$

Proof Given a discrete chf A, write $S(t) = \prod_{s \leq t}(1 - \Delta A(t))$. Then S has chf A, because $S(t-) = S(t)/(1 - \Delta A(t))$ and

$$\Delta A(t) = 1 - \frac{S(t)}{S(t-)} = \frac{\Delta F(s)}{1 - F(s-)}.$$

Conversely, given a discrete survival function S, then

$$\begin{aligned} S(t) &= \prod_{u \leq t} \frac{S(u)}{S(u-)} = \prod_{u \leq t} \left(1 + \frac{\Delta S(u)}{S(u-)}\right) \\ &= \prod_{u \leq t} (1 - \Delta A(u)) \end{aligned}$$

where A is the chf. $\square$

Hazard functions

If F has density f, define the hazard function

$$\alpha(t) = f(t)/S(t) \approx P(X \in [t, t + dt)|X \geq t)/dt$$

Thus

$$P(X \in [t, t + dt)|X \geq t) \approx \alpha(t)\, dt$$

Cox proportional hazards model

$$\alpha(t|z) = \alpha_0(t) \exp(\beta' z)$$

adjusts for a (multi-dimensional) covariate z.

Counting process approach

$$N(t) = 1\{Z \le t, \delta = 1\}$$

At risk indicator: $Y(t) = 1\{Z \ge t\}$

Basic martingale: $M(t) = N(t) - \int_0^t Y(s)\alpha(s)\,ds$

$dN(t) \sim$ Bernoulli($Y(t)\alpha(t)\,dt$) given the past, so

$$E(dM(t)|\text{past}) = E(dN(t) - Y(t)\alpha(t)\,dt|\text{past}) = 0$$

Nonparametric likelihood

$$L(S) = L(F) = \prod_{i=1}^{n} (F(Z_i) - F(Z_i-))^{\delta_i} (1 - F(Z_i))^{1-\delta_i}.$$

To maximize $L(F)$, we only need consider F supported on the uncensored lifetimes.

Notation

Ordered uncensored lifetimes: $0 < T_1 \leq \ldots \leq T_k$, $T_0 = 0$

$h_j = \Delta A(T_j) = 1 - S(T_j)/S(T_{j-1})$ jump in chf at T_j

$r_j = \sum_{i=1}^n 1\{Z_i \geq T_j\}$ size of the risk set at T_j-, with $r_{k+1} = 0$.

$d_j \geq 1$ denotes the number of uncensored failures at T_j.

Lemma If F is supported on the uncensored lifetimes, then

$$L(S) = \prod_{j=1}^{k} h_j^{d_j}(1 - h_j)^{r_j - d_j}$$

Proof Note that the number of censored lifetimes in $[T_j, T_{j+1})$ is $r_j - d_j - r_{j+1}$, so

$$\begin{aligned}
L(S) &= \prod_{i=1}^{n} (S(Z_i-) - S(Z_i))^{\delta_i} (S(Z_i))^{1-\delta_i} \\
&= \left\{ \prod_{j=1}^{k} (S(T_j-) - S(T_j))^{d_j} \right\} \left\{ \prod_{j=1}^{k} S(T_j)^{r_j - d_j - r_{j+1}} \right\} \\
&= \left\{ \prod_{j=1}^{k} h_j^{d_j} S(T_{j-1})^{d_j} \right\} \left\{ \prod_{j=1}^{k} \frac{S(T_j)^{r_j - d_j}}{S(T_{j-1})^{r_j}} \right\} \\
&= \prod_{j=1}^{k} h_j^{d_j} (1 - h_j)^{r_j - d_j}
\end{aligned}$$

$\square$

Nonparametric MLEs

$L(S)$ is maximized when $h_j = d_j/r_j$, giving the Nelson–Aalen estimator:

$$A_n(t) = \sum_{j:T_j \le t} \frac{d_j}{r_j}$$

Kaplan–Meier estimator:

$$S_n(t) = \prod_{j:T_j \le t} \left(1 - \frac{d_j}{r_j}\right)$$

and $F_n = 1 - S_n$.

Asymptotics

Assume now F is continuous. Then

$$\sqrt{n}(A_n(t) - A(t)) \xrightarrow{\mathcal{D}} W(\sigma^2(t))$$

$$\sqrt{n}(S_n(t) - S(t)) \xrightarrow{\mathcal{D}} S(t)W(\sigma^2(t))$$

where

$$\sigma^2(t) = \int_0^t \frac{dF(s)}{(1 - F(s))^2(1 - G(s-))}$$

Without censoring, simplifies to

$$\sigma^2(t) = \frac{F(t)}{1 - F(t)}.$$

EL function

$$R(\theta_0) \quad = \quad \frac{\sup\{L(S) : \theta(S) = \theta_0\}}{\sup\{L(S)\}}$$

EL suddenly becomes difficult because of the censoring!

Unless $\theta(S)$ has a particularly simple form, $R(\theta_0)$ may be intractable.

Known tractable forms of $\theta(S)$ or $\theta(A)$:

- $S(t_0)$
- $A(t_0)$
- quantiles
- linear functionals $\theta(F) = \int h(t)\, dF(t)$
- linear functionals $\theta(A) = \int h(t)\, dA(t)$

Thomas and Grunkemeier (1975), Li (1995), Murphy (1995), Pan and Zhou (2002)

EL for means

Linear functional

$$\theta(F) = E(h(X)) = \int h(x)\, dF(x)$$

(e.g., mean lifetime).

F_n is an inverse-probability-of-censoring weighted average:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1\{Z_i \leq t\}\delta_i}{1 - G_n(Z_i-)}$$

Robins and Rotnitzky (1992)

$E(h(X))$ can be estimated by

$$\theta(F_n) = \frac{1}{n}\sum_{i=1}^{n} \frac{h(Z_i)\delta_i}{1 - G_n(Z_i-)}$$

Lemma

$$E(h(X)) = E\left(\frac{h(Z)\delta}{1 - G(Z-)}\right)$$

Proof

$$E\left(\frac{h(\min(X,Y))1\{X \le Y\}}{1-G(\min(X,Y)-)}\right) = \iint_{x\le y} \frac{h(x)}{1-G(x-)}\,dF(x)dG(y)$$

$$= \int_0^\infty \frac{h(x)}{1-G(x-)}\int_x^\infty dG(y)dF(x)$$

Proof

$$
\begin{aligned}
E\left(\frac{h(\min(X,Y))1\{X \leq Y\}}{1-G(\min(X,Y)-)}\right) &= \iint_{x\leq y} \frac{h(x)}{1-G(x-)} dF(x)dG(y) \\
&= \int_0^\infty \frac{h(x)}{1-G(x-)} \int_x^\infty \, dG(y)dF(x) \\
&= \int_0^\infty \frac{h(x)}{1-G(x-)} (1-G(x-))\, dF(x) \\
&= E(h(X))
\end{aligned}
$$

$\square$

If censoring cdf G were known, standard EL for means could be used (everything inside the expectation is observable):

$$-2\log R(\theta_0, G) \xrightarrow{\mathcal{D}} \chi_1^2.$$

Wang and Jing (2001) replace G by its Kaplan–Meier estimator and show

$$-2\log R(\theta_0, G_n) \xrightarrow{\mathcal{D}} c\chi_1^2$$

where c is an estimable constant.

Murphy and van der Vaart (1997) established an ELT for $\theta(F) = E(h(X))$ (doubly censored data) but EL function may be difficult to compute (has it been tried?).

EL for the Cox model regression parameters

$$\alpha(t|z) = \alpha_0(t)\exp(\beta' z)$$

Estimating equation for β:

$$E(U(\beta_0)) = 0$$

where U is the partial likelihood score function.

Qin and Jing (2001): standard EL for this estimating equation.

Murphy and van der Vaart (1997): a profile EL for β for current status data.

EL for survival function at a fixed point

$p = S(t_0)$, with t_0 fixed, $0 < p < 1$.

Method of Lagrange multipliers is tractable.

$\hat{S}$ maximizing $L(S)$ subject to the constraint $S(t_0) = p$ is

$$\hat{S}(t) = \prod_{j: T_j \le t} \left(1 - \frac{d_j}{r_j + \lambda}\right)$$

where the Lagrange multiplier λ is the solution to

$$\prod_{j: T_j \le t_0} \left(1 - \frac{d_j}{r_j + \lambda}\right) \;=\; p.$$

Equivalently,

$$g(\lambda) = \sum_{j:T_j \le t_0} \log\left(1 - \frac{d_j}{r_j + \lambda}\right) = \log p = -A(t_0)$$

Theorem If S is continuous, $0 < p = S(t_0) < 1$ and $G(t_0) < 1$, then

$$-2\log R(p) \xrightarrow{\mathcal{D}} \chi_1^2$$

Thomas and Grunkemeier (1975), Li (1995), Murphy (1995)

Proof Same technique as in the standard ELT, except instead of using the standard CLT, a martingale CLT is applied to the Nelson–Aalen estimator.

Taylor expansion of g leads to

$$\lambda = n(A(t_0) - A_n(t_0))/\hat{\sigma}^2 + O_P(1)$$

where $\hat{\sigma}^2$ is an estimate of $\sigma^2(t_0)$.

$$\begin{aligned}
-2\log R(p) &= -2(\log(L(\hat{S}) - \log(L(S_n)) \\
&= -2 \sum_{i:T_j \le t_0} \left\{ (r_j - d_j)\log\left(1 + \frac{\lambda}{r_j - d_j}\right) \right. \\
&\qquad \left. - r_j \log\left(1 + \frac{\lambda}{r_j}\right) \right\} \\
&= \lambda^2 \hat{\sigma}^2/n + o_P(1) \\
&= n(A_n(t_0) - A(t_0))^2/\hat{\sigma}^2 + o_P(1) \\
&\xrightarrow{\mathcal{D}} \chi_1^2 \qquad \square
\end{aligned}$$

EL simultaneous band for S

As a *process* in $t \in [a, b]$,

$$-2\log R(t) \xrightarrow{\mathcal{D}} \left(\frac{W(\sigma^2(t))}{\sigma(t)}\right)^2,$$

Simultaneous confidence band for S over an interval $[a, b]$:

$$\{(t, S(t)) : -2\log R(t) \le C_\alpha, t \in [a, b]\}$$

C_α the upper α-quantile of

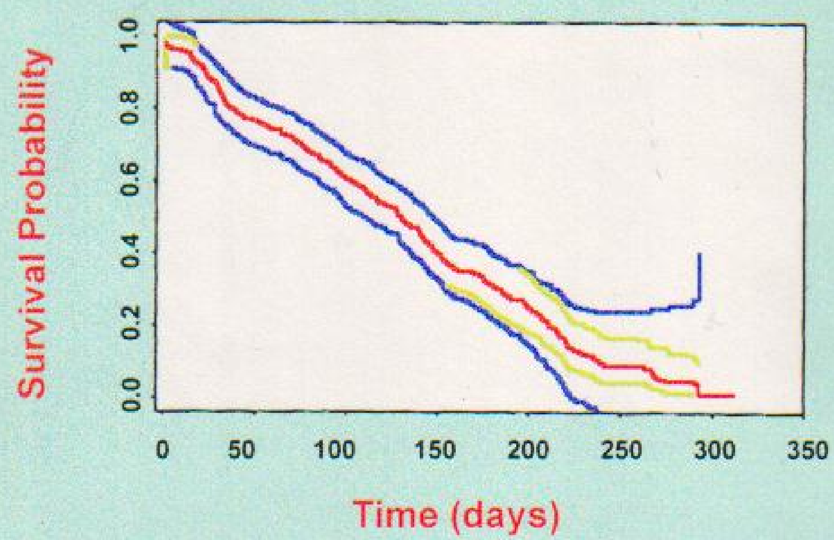
$$\sup_{t\in[\hat\sigma^2(a),\hat\sigma^2(b)]} \frac{W^2(t)}{t}$$

Equal precision LR band. Narrower in the tail than Hollander, McKeague, Yang (1997) band.

Li and Van Keilegom (2001): adjustment for a covariate effect (continuous one-dimensional covariate).

Example

Data on 432 manuscripts submitted to JASA during 1994. Time to first review censored by the end of the year.



Estimated Probability (Red Line) that a First Review of a JASA Manuscript Takes Longer Than t Days With 95 % LR Band (Green Line) and HW Band (Blue Line).

Source : Hollander, McKeague and Yang (1997, Journal of the American Statistical Association)

Two-sample problem with censoring

Comparison of treatment and placebo groups.

Notation

Index sample by j

Assume $n_j/n \to p_j > 0$

Total sample size $n = n_1 + n_2$

Nonparametric likelihood: $L(S_1, S_2) = L_1(S_1)L_2(S_2)$.

- Standard method: logrank test for $S_1 = S_2$.
- Wald-type comparison of S_1 and S_2 using some smooth functional $\varphi(S_1, S_2)$ and the functional delta method typically leads to intractable limiting distributions. Simulation needed.

Gaussian multiplier simulation technique

Martingale increments $dM_i(t)$ replaced by $G_i dN_i(t)$, where $G_i \sim N(0,1)$. (Lin, Wei and Ying, 1993)

Parzen, Wei and Ying (1997) constructed a Wald-type confidence band for $S_1(t) - S_2(t)$ using this technique.

Q-Q plot

$\{(F_1^{-1}(p), F_2^{-1}(p)) : 0 < p < 1\}$

Einmahl and McKeague (1999) constructed an EL confidence band for the Q-Q plot:

$$\{(t_1, t_2)\colon -2\log R(t_1, t_2) \le C_\alpha, t_1 \in [a, b]\}$$

where C_α uses

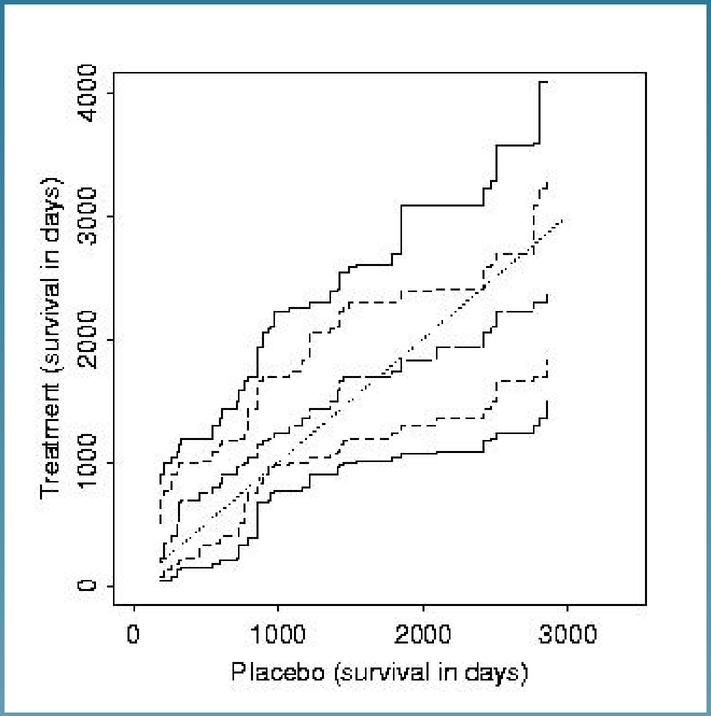
$$\sigma^2(t) = \sigma_1^2(t)/p_1 + \sigma_2^2(t')/p_2$$

and $t' = F_2^{-1}(F_1(t))$. Simulation not needed.

Mayo Clinic trial Q-Q plot

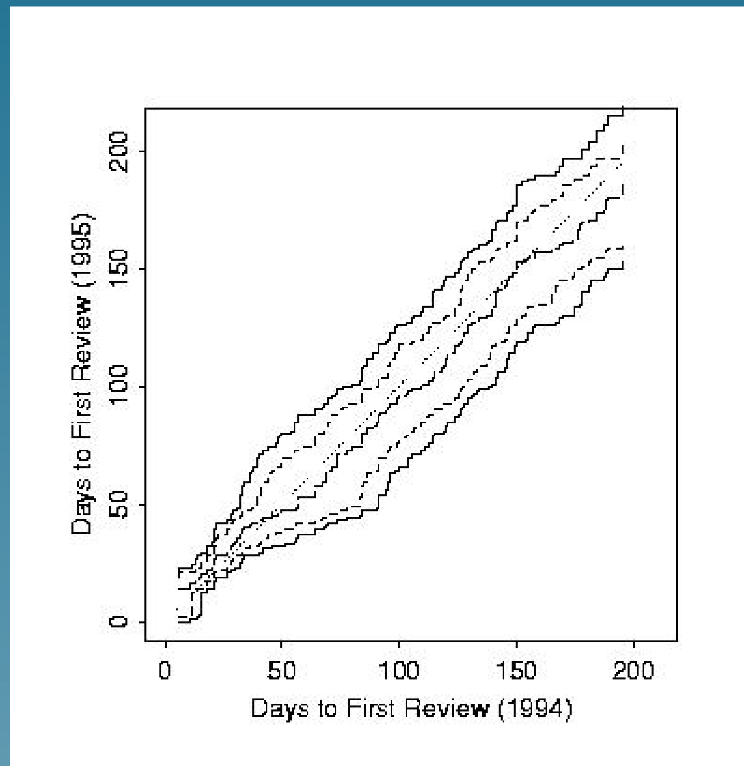
Randomized clinical trial for primary biliary cirrhosis of the liver.

158 patients in treatment group, and 154 in placebo group.



JASA time to first review Q-Q plot

JASA manuscripts data: 432 submitted in 1994, and 444 in 1995.



Relative survival

$\theta(t) = S_1(t)/S_2(t)$

More relevant than a Q-Q plot to medical practice and easier to interpret.

McKeague and Zhao (2002) construct an EL simultaneous band:

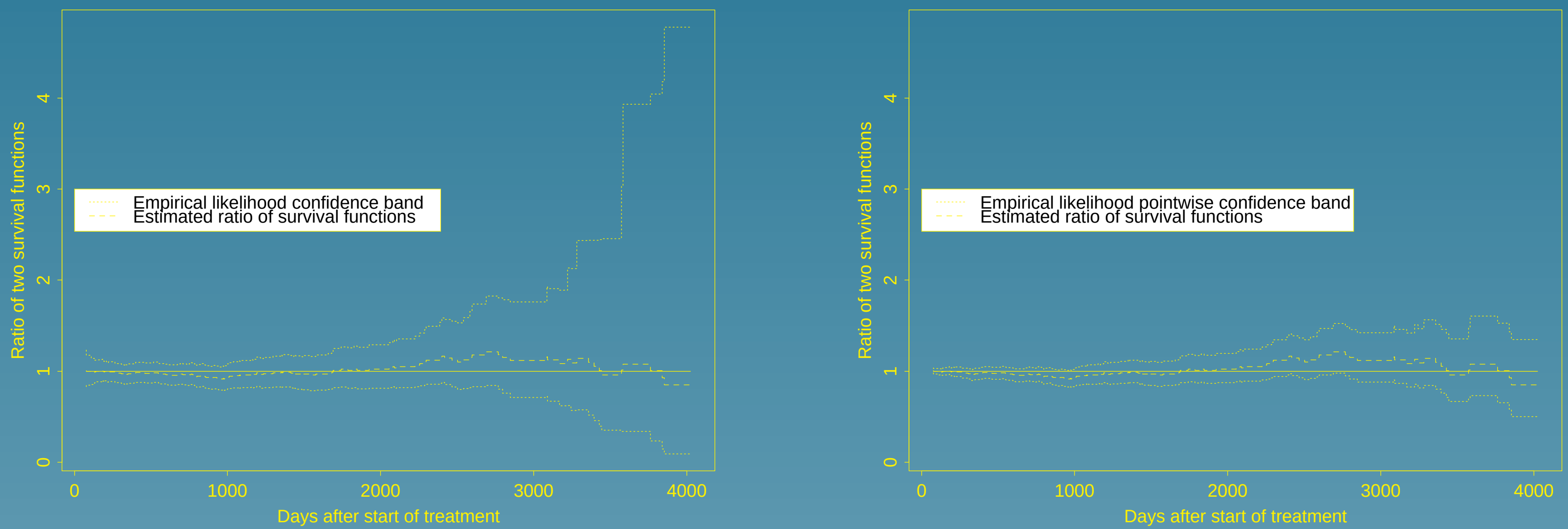
$$\{(t, \theta(t)) \colon -2 \log R(t) \leq C_\alpha, t \in [a, b]\}$$

where C_α uses

$$\sigma^2(t) = \sigma_1^2(t)/p_1 + \sigma_2^2(t)/p_2.$$

Simulation not needed.

Mayo clinic trail: placebo/treatment relative survival



- ROC curve (P-P plot) $\{(F_1(x), F_2(x)) : x \in \mathbb{R}\}$.
  Claeskens, Jing, Peng and Zhou (2001): pointwise EL band using kernel smoothing; no censoring.

- Simultaneous band for differences in cumulative hazards:

$$A_1(t) - A_2(t) = -\log(S_1(t)/S_2(t))$$

  EL works without simulation, McKeague and Zhao (2002).

- Simultaneous band for relative cumulative risk

$$A_1(t)/A_2(t) = \log S_1(t)/\log S_2(t)$$

  EL works, McKeague and Zhao (2002). Gaussian multiplier simulation needed.

- Simultaneous band for vaccine efficacy: measured as 1 minus some measure of relative risk (RR) in the vaccinated group compared with the unvaccinated group (VE = 1 - RR):

$$\mathrm{VE}(t) = 1 - \frac{\alpha_{\mathrm{vaccine}}(t)}{\alpha_{\mathrm{placebo}}(t)}$$

$$\mathrm{VE}_c(t) = 1 - \frac{A_{\mathrm{vaccine}}(t)}{A_{\mathrm{placebo}}(t)}$$

Halloran, Struchiner and Longini (1997)

EL works, McKeague and Zhao (2002). Gaussian multiplier simulation needed.

- Ratios of cdfs: $F_1(t)/F_2(t)$, EL intractable?

EL test for equal hazard rates

$$H_0 : \alpha_1(t) = \alpha_2(t), \;\; t \in [a, b]$$

EL works if $a > 0$ as H_0 is then equivalent to constant relative survival:

$$S_1(t)/S_2(t) = \theta, \;\; t \in [a, b]$$

for some (unknown) constant θ.

Use a plug-in estimate $\hat{\theta}$ in the EL function in place of $\theta(t) = S_1(t)/S_2(t)$:

$$T_n = \sup_{t \in [a,b]} -2 \log R(t, \hat{\theta}).$$

Gaussian multiplier simulation needed.

McKeague and Zhao (2002)

Conclusion

- EL shows great promise for further development in more complex clinical trial settings.
- As we have seen, simulation is often needed to adequately calibrate EL for simultaneous inference in survival analysis.
- A commercial plug: come to the IMS Invited paper session *A Decade of Empirical Likelihood* at the August 2002 Joint Statistical Meetings in New York City!