Comparison of Treatments and Competing Risks in Clinical Trials

Ian McKeague
Department of Statistics
Florida State University

Outline

- Introduction: Parametric/nonparametric likelihood ratio methods
- Empirical likelihood (EL): advantages and disadvantages
- EL confidence bands for survival functions
- Comparison of survival functions in two-sample problems
- Comparison of independent competing risks
- Comparison of dependent competing risks (mark specific hazards)

Classical likelihood ratio method

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

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

Likelihood ratio statistic:

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

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:

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

Improvement over Wald-type confidence regions.

Empirical likelihood

- Introduced by Thomas and Grunkemeier (1975) in the context of survival function estimation. Owen (1988, 1990, . . . , 2001).

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

Likelihood ratio statistic:

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

where

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

Wilks theorem extends.

Empirical likelihood

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	

Example

$X_1, \ldots, X_n$ iid with cdf F

$\theta = \theta(F) = F(t)$

Likelihood ratio statistic 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 $\log(1+x) = x - x^2/2 + \ldots$ we have

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

under F_0.

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

$$\begin{aligned} -2\log R(t) \;&\xrightarrow{\mathcal{D}}\; \left(\frac{W^o(F_0(t))}{\sqrt{F_0(t)(1 - F_0(t))}}\right)^2 \\ &\stackrel{\mathcal{D}}{=}\; \left(\frac{W(\sigma^2(t))}{\sigma(t)}\right)^2, \end{aligned}$$

where W^o a standard tied-down Wiener process (Brownian bridge), W is a standard Wiener process, and

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

Test for $F = F_0$

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

Einmahl and McKeague (2001) showed that this approach extends to tests for symmetry, exponentiality, independence and changes in distribution (without censoring).

Confidence band for F

Invert to obtain a simultaneous confidence band for F over a given interval $[a, b]$:

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

where C_α is the upper α-quantile of

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

Hollander, McKeague, Yang (1997, JASA) developed this approach for right-censored data as an alternative to the standard (Wald-type) Hall–Wellner or Kolmogorov–Smirnov bands.

Observe n iid pairs (T_i, δ_i)
$T_i = \min(X_i, Y_i)$, $\delta_i = I\{X_i \leq Y_i\}$, X_i and Y_i independent.
F : cdf of X_i
$S = 1 - F$: survival function
$A = -\log S$: cumulative hazard function
Likelihood :

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

Two-sample problem with right-censoring

Assume $n_j/n \to p_j > 0$
Total sample size $n = n_1 + n_2$
Likelihood: $L(S_1, S_2) = L_1(S_1)L_2(S_2)$.

Methods for comparing two treatments:

- 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.

  Simulation method: 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 a simulation 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 $t' = F_2^{-1}(F_1(t))$.

- Multi-Q plot for k-sample comparisons.

- Relative survival $\theta(t) = S_1(t)/S_2(t)$. More relevant to medical practice and easier to interpret. McKeague and Zhao (2001) construct an EL confidence band:

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

  where C_α uses

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

- Ratios/differences of cumulative hazards:

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

  EL works, McKeague and Zhao (2001). Simulation needed for the ratio.

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

- Scaled pointwise confidence band: c_α replace by γc_α, where $\gamma > 1$ is selected by simulation to give simultaneous coverage over $[a, b]$.

- Test for equal hazard rates:

$$H_0 : \lambda_1(t) = \lambda_2(t),\ t \in [a, b].$$

  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 $\widehat{\theta}$ in the LR in place of $\theta(t) = S_1(t)/S_2(t)$:

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

Simulation needed.

- EL works in a similar way for testing proportional hazards:

$$A_1(t) = \beta A_2(t), \ t \in [a, b]$$

for some (unknown) constant β, provided $a > 0$. Simulation needed.

- Adjusting for a covariate effect (Z).
Li and Van Keilegom (2001), one sample:
$L(t|z) = \prod_{i=1}^{n} \{(F(T_i|z) - F(T_i - |z))^{\delta_i} (1 - F(T_i|z))^{1-\delta_i}\}^{W_i(z)}$

$$W_i(z) = \frac{K\left(\frac{z - Z_i}{b_n}\right)}{\sum_{l=1}^{n} K\left(\frac{z - Z_l}{b_n}\right)}$$

Comparison of independent competing risks

X, Y independent. Observe $T = \min(X, Y)$ and cause of failure $V \in \{1, 2\}$.

EL works (without censoring) for:

- Confidence band for ratio of survival functions $S_X(t)/S_Y(t)$. Uses same C_α with

$$\sigma^2(t) = \sigma_X^2(t) + \sigma_Y^2(t).$$

- Confidence band for ratio of cumulative hazards $A_X(t)/A_Y(t)$. Simulation needed.

- Testing hypothesis of equal hazards over $[a, b]$, $a > 0$. Simulation needed.

- Adjustment for covariate effects possible.

Comparison of dependent competing risks

Gilbert, McKeague and Sun (2001)

Example. Clinical trial involving an HIV drug regimen. Risk of treatment failure can increase with HIV genetic evolution away from a baseline virus sequence. Detecting such an association can help in designing treatments that overcome the problem of drug resistance.

Failure time T, continuous mark variable V with support $[0, 1]$, independent censoring C. Observe $(X, \delta, \delta V)$, where $X = \min(T, C)$ and $\delta = I(T \leq C)$.

Mark-specific hazard rate function:

$$\lambda(t, v) = \lim_{h_1, h_2 \to 0} P\{T \in [t, t+h_1), V \in [v, v+h_2) | T \geq t\}/h_1 h_2,$$

H_0: $\lambda(t, v)$ does not depend on v for $t \in [0, \tau]$,
H_a: $\lambda(t, v_1) \leq \lambda(t, v_2)$ for all $v_1 \leq v_2,\ t \in [0, \tau]$
with strict inequality for some t, v_1, v_2 in H_a.

Test procedure based on estimates of the doubly cumulative mark-specific hazard function

$$\Lambda(t, v) = \int_0^v \int_0^t \lambda(s, u)\, ds\, du.$$

Nonparametric estimator of $\Lambda(t, v)$ is provided by the Nelson–Aalen-type estimator

$$\hat{\Lambda}(t, v) = \int_0^t \frac{N(ds, v)}{Y(s)}, \ \ t \geq 0, \ \ v \in [0, 1], \qquad (1)$$

where $Y(t) = \sum_{i=1}^n I(X_i \geq t)$ is the size of the risk set at time t, and

$$N(t, v) = \sum_{i=1}^n I(X_i \leq t, \delta_i = 1, V_i \leq v)$$

is the marked counting process with jumps at the uncensored failure times X_i and associated marks V_i.

H_0 holds if and only if T and V are independent and V is uniformly distributed over $[0, 1]$, so under H_0 use $\bar{\Lambda}(t, v) = v\hat{\Lambda}(t, 1)$

Test process:

$$L_n(t, v) = \sqrt{n} \int_0^t H_n(s)\, (\hat{\Lambda} - \bar{\Lambda})(ds, v)$$

where $H_n(\cdot)$ is a suitable weight process.

Test statistic

$$T_n = \sup_{v_1 < v_2} \sup_{0 \le s \le t < \tau} (\Delta(t, v_1, v_2) - \Delta(s, v_1, v_2))$$

$$\Delta(t, v_1, v_2) = L_n(t, v_1) + L_n(t, v_2) - 2L_n(t, (v_1 + v_2)/2).$$

Bootstrap procedure

T_n^* formed by replacing the V_i by iid uniform(0,1).

Null distribution of T_n coincides in the limit with the conditional distribution of T_n^* given the observed data.

HIV clinical trial

No evidence of drug resistance: $T_n = 0.487$, p-value $= 0.64$

Conclusion

- EL leads to a number of useful methods for the comparison of treatments and competing risks

- EL currently has limited applicability to complex clinical trials

- For more . . . come to the IMS Invited paper session "A Decade of Empirical Likelihood" at the 2002 JSM in New York City.