Empirical Likelihood

F. P. Treasure

12-May-05
3010a

Objective:

to obtain point- and interval- estimates of time-to-event probabilities using a non-parametric, ‘empirical’, maximum likelihood estimate of the survivor function.

The survivor function

A time-to-event random variable T has survivor function F defined by: $F(t) := P\{T > t\}$.

The maximum likelihood estimator of F is $\hat{F}$.

Maximizing likelihood:

The maximum empirical likelihood estimator of the survivor function is the F which maximizes the likelihood subject to the conditions:

(1) $F(u) \geqslant F(v)$ when $u < v$;

(2) $F(t) \geqslant 0$;

and

(3) $F(0) = 1$.

Condition (3) – which implies no events at $t = 0$ – is not essential but aids exposition.

Constructing Likelihood

The likelihood is the product of the individual contributions to the likelihood. The form of the individual contribution depends on whether an event was observed or not for that individual, and – if not – what kind of censoring was involved.

Individual contributions

Here are the four most common kinds of contribution:

(1) ith individual *right* censored at x_i – we know $T > x_i$ so the contribution from that individual to the likelihood is

$$P\{T > x_i\} = F(x_i)$$

(2) ith individual with *event* at x_i – we know $T = x_i$ so the contribution from that individual to the likelihood is $P\{T = x_i\}$, equivalently:

$$P\{T \geqslant x_i\} - P\{T > x_i\} = \lim_{\Delta \downarrow 0} F(x_i - \Delta) - F(x_i) = F(x_i^-) - F(x_i)$$

(3) ith individual *left* censored at x_i – we know $T \leqslant x_i$ so the contribution from that individual to the likelihood is:

$$P\{T \leqslant x_i\} = 1 - F(x_i)$$

(4) ith individual *interval* censored in $\left[x_i^L, x_i^U\right[$ – we know $x_i^L < T \leqslant x_i^U$ so the contribution from that individual to the likelihood is:

$$P\{x_i^L < T \leqslant x_i^U\} = F(x_i^U) - F(x_i^L)$$

Empirical Likelihood Function

The empirical likelihood function is obtained by multiplying together all the individual contributions. If the dataset is large with a mixture of types of censoring then maximizing the likelihood is either only just feasible, or – more likely – infeasible.

There are useful tricks for simplifying the likelihood. Using these tricks we can maximize for

(1) small datasets;

(2) special cases – as an example, we will re-derive the Kaplan-Meier estimator.

Simplifying the Likelihood

The generic factor in the empirical likelihood has form $[F(b) - F(a)]$ and when maximizing we would like to make this factor as large as possible. If there are no factors $[F(c) - F(b)]$ in the likelihood then we should make $F(b)$ as large as possible, subject to the constraint $F(u) \geqslant F(b)$ when $u < b$. An immediate consequence is that $\hat{F}(0) = 1$.

Re-deriving Kaplan-Meier

The data comprises only of event times and right censoring times. The factors in the likelihood are therefore of form $F(x_i^-) - F(x_i)$ for an event and $F(x_i)$ for a censored observation.

If at time x_i there are no events but there is a censored observation then we can increase $F(x_i)$ indefinitely subject to decreasing F, that is:

$$\hat{F}(x_i) = \hat{F}\,(\text{largest preceding time at which there was an event})$$

Also, $\hat{F}(x_i^-)$ will never have a minus sign in front of it so:

$$\hat{F}(x_i^-) = \hat{F}\,(\text{largest preceding time at which there was an event})\,.$$

The implications are, firstly, that we need only work with times at which an event occurred and, secondly, $\hat{F}$ is constant except at an event time. So $\hat{F}$ is a decreasing piecewise constant function.

Notation for event times

The times at which there are at least one event are $a_1, \ldots, a_g$ with $a_0 := 0$ and $a_{g+1} := \infty$.

There are d_j events ($d_j > 0$) at a_j and c_j censored observations in $[a_j, a_{j+1}[$.

There are r_j individuals at risk at a_j with $r_g = d_g + c_g$ and, for $0 < j < g$, $r_j = d_j + c_j + r_{j+1}$.

Setting up and differentiating likelihood.

Using the simplification $\hat{F}(a_j^-) = \hat{F}\,(a_{j-1})$ we can write down the likelihood:

$$L = \prod_{j=1}^{g} \left[F(a_{j-1}) - F(a_j)\right]^{d_j} \left[F(a_j)\right]^{c_j} . \tag{1}$$

We maximize L to obtain the $\hat{F}\,(a_j)$. Then $\hat{F}\,(t)$ is given by:

$$\hat{F}\,(t) = \hat{F}\left(\max_{j:a_j \le t} a_j\right)$$

Simplifying the notation in (1) by defining $F_j := F\,(a_j)$ and taking logs:

$$\log L = S = \sum_{j=1}^{g} d_j \log(F_{j-1} - F_j) + \sum_{j=1}^{g} c_j \log(F_j) \ , \tag{2}$$

Partial differentiation by F_j gives:

$$\frac{\partial S}{\partial F_j} = \frac{-d_j}{F_{j-1} - F_j} + \frac{c_j}{F_j} + \frac{d_{j+1}}{F_j - F_{j+1}}$$

and we look for a maximum when $\frac{\partial S}{\partial F_j} = 0$, i.e. when

$$0 = \frac{-d_j}{\hat{F}_{j-1} - \hat{F}_j} + \frac{c_j}{\hat{F}_j} + \frac{d_{j+1}}{\hat{F}_j - \hat{F}_{j+1}} \tag{3}$$

for $j = 1, \ldots, g$.

This is a third-order recurrence relationship. It simplifies considerably if we start with $\hat{F}_g$.

Starting the recurrence relationship

If we look at the rightmost factors of the likelihood: $\ldots [F_{g-1} - F_g]^{d_g} [F_g]^{c_g}$ we see that if $c_g = 0$ then:

$$\hat{F}_g = 0 \tag{4}$$

but if $c_g > 0$ then equation (3) simplifies to:

$$0 = \frac{-d_g}{\hat{F}_{g-1} - \hat{F}_g} + \frac{c_g}{\hat{F}_g}$$

which becomes:

$$\hat{F}_g = \frac{c_g}{c_g + d_g} \hat{F}_{g-1} \ . \tag{5}$$

Note that equation (5) includes equation (4) so we don't have to treat $c_g = 0$ as a special case.

Towards Kaplan-Meier

Re-writing (5) using r_g , the number of individuals in the risk set at a_g , we obtain:

$$\hat{F}_g = \left(1 - \frac{d_g}{r_g}\right) \hat{F}_{g-1}$$

which is immediately recognisable as a Kaplan-Meier recurrence.

Exercise: Prove by induction using equation (3) that:

$$\hat{F}_j = \left(1 - \frac{d_j}{r_j}\right) \hat{F}_{j-1} \ . \tag{6}$$

As we have assumed that there are no events at $t = a_o = 0$, we have:

$$\hat{F}_o = 1 \tag{7}$$

and, together, (6) and (7) are the familiar Kaplan-Meier estimator:

$$\hat{F}(t) = \prod_{j:a_j \leq t} \left(1 - \frac{d_j}{r_j}\right) \ .$$

Interval Estimates

The great advantage of deriving Kaplan-Meier in this way is that it provides a satisfactory method for deriving interval estimates. We have obtained the maximum likelihood estimate $\hat{F}(t)$ of $F(t)$. To obtain an interval estimate for $F(t)$ we need to calculate the likelihood where it is not at a maximum so we can apply Wilks's lemma. Specifically, we need $\tilde{L}(z,t)$ – the maximum likelihood constrained such that $F(t) = z$:

$$\tilde{L}(z,t) = \max_{F(t)=z} L(F_1, \ldots, F_g) \ .$$

Then a $1-\alpha$ likelihood ratio based confidence interval for $F(t)$ is given by:

$$\left\{ z : 2\log\left(\frac{\hat{L}}{\tilde{L}(z,t)}\right) \leqslant C_{1,1-\alpha} \right\} \tag{8}$$

where $\hat{L}$ is the unconstrained maximum likelihood $L(\hat{F}_1, \ldots, \hat{F}_g)$ and $C_{m,q}$ is the qth quantile of a chisquare(m) distribution.

Exercise: obtain an expression similar to (8) for a $1-\alpha$ confidence interval for the *median* of a time-to-event distribution.

Constrained Maximization

We need to maximize subject to the constraint $F(t) = z$ or, equivalently, subject to the constraint:

$$\log F_k - \log z = 0$$

where k is such that $a_k \leqslant t < a_{k+1}$. We use the method of Lagrange multipliers. If we can find a value λ such that the Lagrangian $S(\lambda)$:

$$S(\lambda) = \sum_{j=1}^{g} d_j \log(F_{j-1} - F_j) + \sum_{j=1}^{g} c_j \log(F_j) + \lambda(\log F_k - \log z) \tag{9}$$

has a maximum *and* the constraint is satisfied then we have found the constrained maximum.

The multiplier λ plays an interesting role in (9): it makes the RHS look like there are λ extra censored observations at time a_k. $S(\lambda)$ can be interpreted as what the log-likelihood would be if there were λ extra individuals in the dataset, all of whom were censored at a_k . (We are happy to let λ be non-integral or negative as these are mathematical individuals.) The recurrence relationship (6) can immediately be adapted:

$$\text{for } j > k \quad : \quad \tilde{F}_j = \left(1 - \frac{d_j}{r_j}\right)\tilde{F}_{j-1} \tag{10}$$

$$\text{for } j \leqslant k \quad : \quad \tilde{F}_j = \left(1 - \frac{d_j}{r_j(\lambda)}\right)\tilde{F}_{j-1} \tag{11}$$

where $r_j(\lambda) = r_j + \lambda$ is the *augmented* risk set.

We obtain $\tilde{F}_j$ for $j > k$ by starting (10) with $\tilde{F}_k = z$. We obtain $\tilde{F}_j$ for $j \leqslant k$ by starting (11) with $\tilde{F}_0 = 1$ and choosing λ such that $\tilde{F}_k = z$.

Exercise: show that λ can be so chosen.

The Interval Estimate

The unconstrained maximum log likelihood $\hat{S}$ is obtained by substituting the unconstrained estimates $\hat{F}_j$ – obtained from recurrence relation (6) – into the expression (2) for the log-likelihood.

The maximum log likelihood $\tilde{S}\,(z,t)$ constrained to $F(t) = z$ is obtained by substituting the unconstrained estimates $\tilde{F}_j$ – obtained from recurrence relations (10) and (11) – into expression (2).

The $1 - \alpha$ likelihood ratio based confidence interval for $F(t)$ is then the set:

$$\left\{ z : 2(\hat{S} - \tilde{S}\,(z,t)) \leqslant C_{1,1-\alpha} \right\}$$

(Computational note: as it is much easier to calculate $\tilde{F}_k$ from λ than vice-versa, the lower- and upper- limits are best found by varying λ rather than z.)