3003
Cure

F. P. Treasure

November 8, 2000

Cure

A Simple Cure Model

The Concept of 'Cure'

A *cure* model is a survival model where a fraction of the population is not exposed to the hazard of interest.

Example: a treatment for a fatal disease prevents death (from that disease) in a proportion of those treated (the *cured* fraction). It may or may not delay death in the remaining individuals (the *vulnerable* fraction).

This is a special case of a *frailty* model. Whether or not a patient is cured is an 'unobserved explanatory variable'. Mixture distributions can be used. Great care needs to be taken with interpretation.

Notation

The probability of cure is π and we will usually be interested in its dependence, through a parameter vector β, on a vector of explanatory variables z.

The vulnerable fraction of the population is exposed to hazard h_D (D denoting 'diseased') with H_D, f_D and F_D for integrated hazard, density and survivor respectively.

We cannot directly observe h_D, H_D, f_D and F_D as we cannot select only the vulnerable fraction for analysis. What we can observe are the functions for the 'total' population: h_T, H_T, f_T and F_T.

The actual observations consist of survival times x_i, with $i = 1, \ldots, n$ and visibility indicators v_i with $v_i = \mathcal{I}\{x_i \text{ is an observed failure}\}$.

Derivation of Likelihood

The log-likelihood is given by the usual expression:

$$S = \sum_i \{v_i \log f_T^{(i)}(x_i) + (1 - v_i) \log F_T^{(i)}(x_i)\}$$

where we permit f_T and F_T to depend through parameters on an individual's explanatory variables.

We know however that

$$F_T^{(i)}(t) = \pi^{(i)} + (1 - \pi^{(i)}) F_D^{(i)}(t)$$

and

$$f_T^{(i)}(t) = (1 - \pi^{(i)}) f_D^{(i)}(t)$$

with π also depending through parameters on the individual.

The log-likelihood is therefore

$$S = \sum_i \left\{ v_i \log \left[(1 - \pi^{(i)}) f_D^{(i)}(x_i) \right] + (1 - v_i) \log \left[\pi^{(i)} + (1 - \pi^{(i)}) F_D^{(i)}(x_i) \right] \right\}. \quad (1)$$

Parametric Modelling

We let $\pi^{(i)}$ depend on explanatory variables $y^{(i)}$ linearly through parameters β. For convenience, we write $\pi^{(i)} = \pi(\beta^{\mathsf{T}} y^{(i)})$. A logistic link is usually used:

$$\pi(\eta) = \frac{1}{1 + e^{-\eta}}.$$

The survivor distribution depends on explanatory variables $z^{(i)}$ linearly through parameters γ. Typically, a Weibull distribution is is used with

$$H_D^{(i)}(t) = \left[e^{\gamma^{\mathsf{T}} z^{(i)}} t\right]^k$$

where k is the index of the distribution.

Maximum likelihood estimators of β, γ and k can be found using a numerical (not statistical) software package. (Note that the vectors y and z need not necessarily contain the same explanatory variables.)

Exposure to Other Hazards

In this section we extend the simple model to cope with individuals dying for other reasons than the disease we are interested in.

Background Mortality

The simple model is useful if the disease is rapidly fatal so that we can assume the individuals concerned are relatively unlikely to die of anything else. If we cannot make that assumption then we have to account for the background mortality.

We do this by considering every individual to be exposed to a 'background' hazard h_B (with corresponding H_B, f_B and F_B.) The hazard h_D affecting an individual vulnerable to the disease is now the sum of h_B and an 'excess' hazard h_E due directly to the disease. Individuals who are cured of the disease are exposed to h_B only.

Relative Survival

The integrated hazards H_D, H_B and H_E are related by:

$$H_D(t) = H_B(t) + H_E(t).$$

Using $H = -\log F$ we obtain the corresponding relationship between the survivor functions:

$$F_D(t) = F_B(t) F_E(t).$$

F_E is usually known as the *relative* survivor function.

Whole Population Density and Survivor

The survivor function for the whole population, cured and vulnerable, is now given by:

$$F_T(t) = \pi F_B(t) + (1-\pi)F_B(t)F_E(t)$$

or

$$F_T(t) = F_B(t)(\pi + (1-\pi)F_E(t)).$$

The whole population density is therefore given by

$$f_T(t) = f_B(t)\left[\pi + (1-\pi)F_E(t)\right] + F_B(t)(1-\pi)f_E(t)$$

which simplifies to

$$f_T(t) = F_B(t)\left[\pi h_B(t) + (1-\pi)F_E(t)(h_B(t) + h_E(t))\right].$$

Likelihood Accounting for Background Mortality

The log-likelihood becomes

$$\begin{aligned} S &= \sum_i \log F_B^{(i)}(x_i) \\ &\quad + \sum_i v_i \log\left[\pi^{(i)} h_B^{(i)}(x_i) + (1-\pi^{(i)})F_E^{(i)}(x_i)\left(h_B^{(i)}(x_i) + h_E^{(i)}(x_i)\right)\right] \\ &\quad + \sum_i (1-v_i)\left[\pi^{(i)} + (1-\pi^{(i)})F_E^{(i)}(x_i)\right]. \end{aligned}$$

where I have indicated by a superscript (i) that the background survival distribution will be different for each individual (usually depending on age and sex). Neither $F_B^{(i)}$ or $h_B^{(i)}$ are functions of the parameters of interest: $F_B^{(i)}$ can be ignored in the maximization and $h_B^{(i)}(x_i)$ can be obtained from published demographic tables. The excess survival distribution and $\pi^{(i)}$ depend, as in the simple model described previously, on parameter vectors β and γ: maximum likelihood estimates can be found as usual.

Example

De Angelis *et al* (1999) analysed the survival of Finish colon-cancer patients using a cure model adjusting for background mortality. They found that survival worsened with age but

1. middle-aged patients had a worse prognosis than young patients because a smaller proportion were cured;

2. old patients had a worse prognosis than middle aged patients because the vulnerable fraction did not survive for so long.

Why colon cancer? A reasonably common cancer with a reasonable proportion of patients apparently cured. Why Finland? The Finns have for a long time been very good at keeping medical records and using them for medical research (at the time of writing such use is illegal in the United Kingdom).

It is important that as much as possible of the survival curve is observed in order to give scope for the vulnerable fraction to fail. De Angelis *et al* tested this by artificially censoring the data early to see if this affected the parameter estimates. It is also important to check that the parameter correlation matrix is not nearly singular: if the same explanatory variables are in both y and z there may be a problem with identifiability.

Semi-Parametric Modelling of Vulnerable Fraction

In this section we extend the simple model by using a semi-parametric survival distribution for the vulnerable fraction. (Exercise: what difficulties would there be in using a non-parametric survival distribution for the vulnerable fraction?)

Proportional Hazards Models

The usual choice for a semi-parametric model would be a proportional hazards model. We model the hazard in the vulnerable fraction using a baseline hazard h_0 (with corresponding H_0, f_0 and F_0). The hazard experienced by the ith individual (were that individual to belong to the vulnerable fraction) is modelled by:

$$h_D^{(i)} = \exp(\gamma^{\mathsf{T}} z^{(i)}) h_0(t)$$

with survivor and density function given respectively by

$$F_D^{(i)} = [F_0(t)]^{\exp(\gamma^{\mathsf{T}} z^{(i)})}$$

and

$$f_D^{(i)} = \exp(\gamma^{\mathsf{T}} z^{(i)}) h_0(t) \, [F_0(t)]^{\exp(\gamma^{\mathsf{T}} z^{(i)})} .$$

There are difficulties. Hazards which are proportional in the vulnerable fraction are not necessarily proportional in the whole population. The argument which constructs a (partial) likelihood from the risk set at each failure time no longer holds as we do not know who is in the vulnerable fraction. We consequently cannot lose the baseline hazard by cancelling it out.

We must start by writing down the full likelihood.

Proportional Hazards Likelihood

Putting the expressions for f_D and F_D into the simple model likelihood (equation 1) gives

$$S = \sum_i \left\{ v_i \log \left[\left(1 - \pi(\beta^{\mathsf{T}} y^{(i)}) \right) \exp(\gamma^{\mathsf{T}} z^{(i)}) h_0(x_i) \, [F_0(x_i)]^{\exp(\gamma^{\mathsf{T}} z^{(i)})} \right] \right\}$$

$$+\sum_i \left\{(1-v_i)\log\left[\pi(\beta^{\mathsf{T}}y^{(i)}) + \left(1-\pi(\beta^{\mathsf{T}}y^{(i)})\right)\left[F_0(x_i)\right]^{\exp(\gamma^{\mathsf{T}}z^{(i)})}\right]\right\}.$$

This expression is impossible in practice to maximize because of the non-parametric nature of h_0 and F_0.

Notice that there would not of course be a problem if there is no cured fraction because we can then use ordinary proportional hazards techniques. By extension, there would not be a problem if we knew for certain which fraction, cured or vulnerable, each individual belong to. This is a very common situation in applied statistics where the problem becomes much more tractable if we knew a little more data. Here we would like to know for each individual $q_i = \mathcal{I}\{i\text{th individual is vulnerable}\}$ but we only know that $v_i = 1$ implies $q_i = 1$.

The technique to use here is *expectation-maximization*.

Expectation-Maximization

Expectation-Maximization (*EM*) is a whole lecture course in itself. Here we give an indication of what it is about.

If we knew whether or not an individual was cured then the log-likelihood splits into two terms. The first term corresponds to membership of the cured or vulnerable fraction and is binomial with form

$$\sum_i \left\{(1-q_i)\log\pi(\beta^{\mathsf{T}}y^{(i)}) + q_i\log[1-\pi(\beta^{\mathsf{T}}y^{(i)})]\right\}. \tag{2}$$

The second term - which does not contain β - corresponds to the survival distribution in the vulnerable group. We do not need to consider the full likelihood for this survival distribution because we can use the normal proportional hazards techniques (only order of failure important, writing down conditional probabilities, baseline hazard cancels out etc.) while we are pretending we know which individuals are in the vulnerable group.

The essence of the EM technique is that we pretend we know the q_i, we maximize the likelihood, we update our knowledge of q_i, maximize the likelihood again, and so on until the parameter estimates converge (which they usually do).

In more detail:

1. Make an initial guess $\hat{q}_i$ for q_i: $\hat{q}_i = v_i$ will do.
2. Estimate β using the binomial likelihood (2).
3. Within the fraction $\hat{q}_i \neq 0$, estimate γ by proportional hazards.
4. Also within $\hat{q}_i \neq 0$, estimate the baseline hazard and survivor $\hat{h}_0(t)$ and $\hat{F}_0(t)$ (see end of lecture on proportional hazards). Note that it is usual to set $\hat{F}_0$(last failure time) equal to zero to avoid problems with identifiability.

5. Each individual observed to fail has probability 1 of being in the vulnerable fraction. An estimate of the probability of an individual censored at x_i belonging to the vulnerable fraction is

$$\frac{\left(1-\pi(\hat{\beta}^{\mathsf{T}}y^{(i)})\right)\left[F_0(x_i)\right]^{\exp(\hat{\gamma}^{\mathsf{T}}z^{(i)})}}{\pi(\hat{\beta}^{\mathsf{T}}y^{(i)})+\left(1-\pi(\hat{\beta}^{\mathsf{T}}y^{(i)})\right)\left[F_0(x_i)\right]^{\exp(\hat{\gamma}^{\mathsf{T}}z^{(i)})}}.$$

   We update the $\hat{q}_i$ to these probabilities. This is the *expectation* step.

6. We repeat the estimations in steps 2 to 4 with the new $\hat{q}_i$. The $\hat{q}_i$ will now be fractional, so the observations in the proportional hazards calculations should be weighted by $\hat{q}_i$. This is the *maximization* step.

7. We repeat steps 5 and 6 until the parameter estimates converge.

Examples

See Sy and Taylor (2000) and Peng and Dear (2000) for applications of this technique to tonsil cancer and throat cancer respectively.

References

De Angelis R., Capocaccia R., Hakulinen T., Soderman B. and Verdecchia A. (1999) Mixture models for cancer survival analysis: applications to population-based data with covariates *Statistics in Medicine* 18 (4) 441-454

Peng Y. and Dear K. B. G. (2000) A nonparametric mixture model for cure rate estimation *Biometrics* 56 (1) 237-243

Sy J. P. and Taylor J. M. G. (2000) Estimation in a Cox proportional hazards cure model *Biometrics* 56 (1) 227-236