Table 13.2 *Comparison of estimators for the leukaemia data*

Model (treatment of ties)	b_1	*SE*
Exponential	1.53	0.40
Weibull	1.73	0.41
Cox (Peto)	1.51	0.41
Cox (Cox)	1.63	0.43

methods. His results (after correcting observation 6 in sample 1, which was censored), together with those obtained above using parametric survival functions, are summarized in Table 13.2. The estimates all fall within a range of about half a standard error, and the increase in standard error from the Cox model as against the parametric survival functions is quite small. Efron (1977) and Oakes (1977) discuss this phenomenon from a theoretical viewpoint.

13.5 Cox's proportional-hazards model

Cox's (1972a) version of the proportional-hazards model is only partially parametric in the sense that the baseline hazard function $\lambda(t)$ is not modelled as a smooth function of t. Instead, $\lambda(t)$ is permitted to take arbitrary values and is irrelevant in the sense that it does not enter into the estimating equations derived from Cox's partial likelihood (Cox, 1975).

13.5.1 *Partial likelihood*

The argument used to derive the partial likelihood function is as follows. First observe that we need only consider times at which failures occur because, in principle at least, the hazard could be zero over intervals that are free of failures and no contribution to the likelihood would be made by these intervals. Let $t_1 < t_2 < \ldots$ be the distinct failure times and suppose for simplicity that there are no tied failure times. The risk set immediately prior to the jth failure, $R(t_j)$, is the set of individuals any of whom may be found to fail at time t_j. Thus, individuals who have previously failed or who have been censored are excluded from $R(t_j)$. Given that one failure is to occur in the interval $(t_j - \delta t, t_j)$, the relative probabilities of failure for the individuals in $R(t_j)$ are proportional

to the values of their hazard functions. Let $\mathbf{x}_j$ be the value of the covariate vector for the failed individual. The probability under the proportional-hazards model that the individual who fails at time t_j is the one actually observed is

$$\frac{\lambda(t)\exp(\boldsymbol{\beta}^T\mathbf{x}_j)}{\Sigma\lambda(t)\exp(\boldsymbol{\beta}^T\mathbf{x})} = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_j)}{\sum\exp(\boldsymbol{\beta}^T\mathbf{x})}, \tag{13.4}$$

where summation extends over the risk set $R(t_j)$.

This conditional probability is the probability of observing $\mathbf{x}_j$ in sampling from the finite population corresponding to the covariate vectors in $R(t_j)$, where the selection probabilities are proportional to $\exp(\boldsymbol{\beta}^T\mathbf{x})$. This is a generalization of the non-central hypergeometric distribution (section 7.3.2). This argument effectively reverses the roles of random failure times and fixed covariates to fixed failure times and covariates selected according to the probability distribution described above.

The partial likelihood for $\boldsymbol{\beta}$ is the product over the failure times of the conditional probabilities (13.4), and so independent of the baseline hazard function $\lambda(t)$. These conditional probabilities have the form of a linear exponential-family model so that $\boldsymbol{\beta}$ can be estimated by equating the vector sum of the covariates of the failed individuals to the sum of their conditional means. Note, however, that the conditioning event changes from one failure time to the next as individuals are removed from the risk set either through failure or through censoring.

13.5.2 *The treatment of ties*

The occurrence of ties among the failure times complicates the analysis, and several techniques have been proposed for dealing with this complication. One method due to Cox (1972a) is as follows. Suppose for definiteness that two failures occur at time t and that the vector sum of the covariates of these two failed individuals is $\mathbf{s}_j$. The factor corresponding to (13.4) is then defined to be

$$\exp(\boldsymbol{\beta}^T\mathbf{s}_j)/\sum\exp(\boldsymbol{\beta}^T\mathbf{s}), \tag{13.5}$$

where the sum in the denominator extends over all distinct pairs of individuals in $R(t_j)$. In other words we construct the finite

population consisting of sums of the covariate vectors for all distinct pairs of individuals in the risk set at time t_j. The probability under an exponentially weighted sampling scheme that the failures were those of the pair actually observed is given by (13.5), which again has the exponential-family form. Note however that the number of terms in the denominator of (13.5) quickly becomes exceedingly large for even a moderate number of ties at any failure time.

Any reasonable method for dealing with ties is likely to be satisfactory if the number of failed individuals constitutes only a small fraction of the risk set. In fact the likelihood contribution (13.5) is exact only if failures are thought of as occurring in discrete time. In practice, however, ties occur principally because of grouping. With grouped data the appropriate likelihood (Peto, 1972) involves the sum over all permutations of the failed individuals consistent with the ties observed. Suppose, for example, that two failures are tied and that the failed individuals have covariate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. The probability for the sequence in time $(\mathbf{x}_1, \mathbf{x}_2)$ or $(\mathbf{x}_2, \mathbf{x}_1)$, either of which is possible given the tie, is

$$\frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_1)}{\sum_R \exp(\boldsymbol{\beta}^T\mathbf{x})}\frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_2)}{\sum_{R_1} \exp(\boldsymbol{\beta}^T\mathbf{x})} + \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_2)}{\sum_R \exp(\boldsymbol{\beta}^T\mathbf{x})}\frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_1)}{\sum_{R_2} \exp(\boldsymbol{\beta}^T\mathbf{x})}, \tag{13.6}$$

where R_j is the risk set excluding $\mathbf{x}_j (j = 1, 2)$. Clearly the likelihood contribution becomes increasingly cumbersome as the number of ties becomes appreciable.

Expressions (13.5) and (13.6) for the contribution to the likelihood can both be derived by arguments involving exponentially weighted sampling from a finite population without replacement. If the number of ties is small we may use the simpler expression

$$\frac{\exp(\boldsymbol{\beta}^T\mathbf{s})}{\{\sum_R \exp(\boldsymbol{\beta}^T\mathbf{x})\}^m}, \tag{13.7}$$

where $\mathbf{s}$ is the sum of the covariate vectors of the m tied individuals (Peto, 1972). This term corresponds to sampling with replacement.

13.5.3 *Numerical methods*

The likelihood formed by taking the product over failure times of the conditional probabilities (13.4) can, in principle, be maximized directly using the weighted least-squares method discussed in Chapters 2 and 8. Alternatively we can regard the covariate vector of the failed individuals as the response and condition on the set of covariates of all individuals in the risk set at each failure time, these being regarded as fixed. If we write $\mathbf{y}$ for the covariate vector of the failed individual the log likelihood for one failure time takes the form

$$\boldsymbol{\beta}^T\mathbf{y} - \log\{\textstyle\sum \exp(\boldsymbol{\beta}^T\mathbf{x})\},$$

with summation over the risk set. This has the form of an exponential family model with canonical parameter $\boldsymbol{\beta}$ and $b(\theta)$ (in the notation of section 2.2) equal to $\log\{\sum \exp(\boldsymbol{\beta}^T\mathbf{x})\}$. The (conditional) mean is then given by $b'(\theta)$ and the variance by $b''(\theta)$. However, this formulation is unhelpful computationally because there is no explicit expression for the quadratic weight (here equal to the variance function) as a function of the mean.

The computational difficulty can be avoided by a device similar to that used in section 13.4. Suppose that k_j individuals are at risk immediately prior to t_j and that just one individual is about to fail. If we regard the observation on the failed individual as a multinomial observation with k_j categories, taking the value 1 for the failed observation and 0 for the remainder, then the contribution to the likelihood is again of the form (13.4), but now interpreted as a log-linear model for the cell probabilities. Thus the numerical methods of Chapter 5 may be used provided that the algorithm allows variable numbers of categories for the multinomial observations.

Alternatively (Whitehead, 1980) a Poisson log likelihood may be used provided that a blocking factor associated with failure times is included. The idea here is that at each failure time each individual in the risk set contributes an artificial Poisson response of 1 for failure and 0 for survival. The mean of this response is $\exp(\alpha+\boldsymbol{\beta}^T\mathbf{x})$ for an individual whose covariate value is $\mathbf{x}$ and α represents the blocking factor associated with failure times. Because of the equivalence of the Poisson and multinomial likelihoods discussed in section 6.4, the estimate of $\boldsymbol{\beta}$ and the estimate of its precision are identical to those obtained from the multinomial likelihood and

hence to the partial likelihood.

The computations can be simplified if the number of distinct covariate vectors is small so that individuals in the risk set may be grouped into sets of constant hazard. The adjustment for ties is simple for the third method described above (often called Peto's method). In the multinomial log likelihood we set the multinomial total equal to the observed number of tied failures at that time. No adjustment to the algorithm is required. The corresponding Poisson log likelihood is equivalent to Peto's version of the partial likelihood.

Whitehead (1980) describes the adjustments to the Poisson likelihood required to maximize the likelihood corresponding to Cox's method for dealing with ties.

13.6 Bibliographic notes

The recent literature on the analysis of survival data includes books by Cox and Oakes (1984), Elandt-Johnson and Johnson (1980), Gross and Clark (1975), Lawless (1982), Lee (1980), Kalbfleisch and Prentice (1980) and Miller (1981).

Cox's model was proposed by Cox (1972a), and fitting via GLIM discussed by Whitehead (1980); the pseudo-Poisson model for parametric survival functions was proposed by Aitkin and Clayton (1980), who also discuss the definition of residuals and the necessary adaptation of standard graphical techniques (see also Crowley and Hu 1977). For a comparison of Cox and Weibull models, see Byar (1983).

13.7 Further results and exercises 13

13.1 In medical trials the recruitment of patients frequently continues over a prolonged period, spanning perhaps the entire trial. Consider such a trial to test a new drug that is claimed to benefit patients suffering from angina by reducing the incidence of coronary disease. The protocol specifies eligible patients to be those aged 55–75, showing symptoms of angina who have no previous record of heart attack and are taking no other medication. After being judged eligible and consent has been obtained, a patient

is randomized to one of two groups, either the new drug or the standard treatment.

Discuss how you might analyse the data that have accumulated after two years in such a trial. Consider in particular the following points.

1. What are appropriate definitions of failure:
   - deaths from all causes;
   - deaths from coronary disease only;
   - all heart attacks whether fatal or not.
2. Choice of origin for the time scale:
   - calendar time from the beginning of the study;
   - time from individual patient randomization;
   - time from first appearance of patient's angina symptoms.
3. Non-compliance because of non-fatal side-effects:
4. Who to include in the risk set:
   - all known survivors among those randomized;
   - all survivors excluding those no longer complying.

13.2 Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be an ordered sample of *i.i.d.* exponential random variables of unit mean. Define the normalized differences

$$Y_1 = nX_{(1)}, \qquad Y_i = (n-i+1)(X_{(i)} - X_{(i-1)}), \quad i = 2, \ldots, n.$$

Show that $Y_1, \ldots, Y_n$ are *i.i.d.* exponential random variables of unit mean.