Multivariate Survival Analysis

Previously we have assumed that either \((X_i, \delta_i)\) or \((X_i, \delta_i, Z_i)\), \(i = 1, \ldots, n\), are i.i.d.. This may not always be the case.

Multivariate survival data can arise in practice in different ways:

- **Clustered survival data.** This happens when failure times (often of the same type, e.g., death) from the same cluster are dependent, such as in
  - multicenter clinical trials: cluster = clinical center
  - familial data: cluster = family
  - matched pairs: cluster = pair

- **Recurrent event data.** This happens when an individual may experience events of the same type more than once, e.g., infections. It is actually a special case of clustered data, where the failure times are ordered.

What are the failure time random variables in each case?
Work has been done (and still being done) to estimate non-parametrically the joint survival distribution, without any predictors, eg. bivariate survival function. This turns out to be much harder than the one dimensional case (KM), mainly due to censoring.

In practice, we are often interested in relating certain covariates to the survival time (regression setting), while taking into consideration the dependence among the survival times.

Here we will introduce extensions of the Cox model to handle correlated survival data.

There are two types of approaches:

- the conditional approach (using random effects)
- the marginal (WLW) approach
The random effects approach

The dependence among failure times may be viewed as induced by unmeasured factors that are common to subjects within a cluster but may vary from cluster to cluster.

For example,

- EST 1582 is a phase III multi-institutional lung cancer trial to compare two treatments, CAV vs. CAV-HEM. It was shown (Gray 1995) that there is significant institutional variation in the treatment effects (see plot);

- for familial data, common genetic or environmental factors might have induce the dependence of survival times among the family members;

- in recurrent events, multiple events of the same individual are likely correlated due to some factor unique to that individual but different from individual to individual.
One may model these cluster-to-cluster variation as **fixed effects**. For example, in the E1582 data, if $Z = 1$ for CAV-HEM, and $Z = 0$ for CAV, instead of using a single $\beta$ for the treatment difference, we may use $\beta_1$ for the treatment effect in institution 1, $\beta_2$ for the treatment effect in institution 2, ... there are 31 institutions (579 observations). How do we write down such a fixed effects Cox model?

Obviously the above fixed effects model will introduce many parameters. In the above figure, that is NOT how we obtained the estimated effects.
Alternatively one may try to use a stratified Cox model. But this will still have problems with small strata. In fact, the number of patients per institution varied from 1 to 56.

One way to solve this problem is to turn the many (31) fixed treatment effects into random effects, i.e. put a (prior) distribution on them (in a Bayesian sense). This effectively shrinks the magnitude of the estimated effects towards zero.

Shrinkage estimators are known to be biased, but nonetheless have smaller MSE (Morris, 1983).
The univariate frailty model

The simplest random effects model for clustered survival data is the univariate frailty model

\[ \lambda_{ij}(t) = \lambda_0(t) \exp\{\beta'Z_{ij} + b_i\}. \]

Here \( i \) indicates the cluster, and \( j \) the individual from cluster \( i \) (\( i = 1, ..., n, \ j = 1, ..., n_i \)), and \( b_i \) is the ‘random cluster effect’. This is sometimes called a hierarchical model.

The assumption of the random effects model is that \( b_1, ..., b_n \) are i.i.d. random variables, from a distribution known up to a finite number of parameters. For example, \( b_i \sim N(0, \sigma^2) \). This can substantially reduce the number of parameters compared to a fixed effects model.

Such modeling is most suitable when there are many clusters and each of relative small size.

In general \( b_i \) induces correlation among the individuals in the same cluster \( i \).
We can equivalently write the above model as
\[ \lambda_{ij}(t) = \lambda_0(t) \exp\{\beta'Z_{ij}\}\omega_i. \]
so that \( b_i = \log \omega_i; \) \( \omega_i \) is often called the \textit{frailty} term. Early use of frailty models was mainly to count for ‘unobserved heterogeneity’, or unmeasured covariates, for i.i.d. data \((n_i = 1)\).

In the past, the commonly used distribution for the \( \omega_i \)'s is the gamma distribution with mean of one and unknown variance. This is mainly due to mathematical convenience (conjugate prior).

The mean of \( b_i \) (or \( \omega_i \)) needs to be fixed to avoid problems of identifiability (why?).

The univariate frailty model can be fitted using \texttt{coxph()} with option \texttt{frailty}. The distributions allowed for \( \omega_i \) are gamma and log-normal.
In terms of the data generating mechanism, we can think of the clusters as being sampled from a population.

\( b_i \) here is sometimes called the random intercept; it’s a random effect on the baseline hazard.

This model counts for the correlation among the failure times within a cluster, but not the covariate effects (like in E1582) that vary from cluster to cluster.
General random effects Cox model

A more general random effects Cox model is

$$\lambda_{ij}(t) = \lambda_0(t) \exp\{\beta'Z_{ij} + b_i'W_{ij}\},$$

Here $W_{ij}$ is a vector of covariates that have random effects, such as treatment in E1582 data; $b_i$ is the ‘covariate by cluster interaction’ (compare to the fixed effects model we might have written down earlier).

When $W_{ij} = 1$, the above is the univariate frailty model.

Often the covariates in $W$ are also included in $Z$, so that they have fixed effects (can be thought of as ‘main effects’), and we impose that $E(b_i) = 0$.

The above model is also called the proportional hazards mixed-effects model (PHMM). It parallels the linear mixed-effects model (LMM) and generalized mixed-effects model (GLMM).

The $b_i$’s are often assumed to be normal. Gamma distribution is no longer suitable, as it is not scale invariant.

This model was considered in Vaida and Xu (2000), and Ripatti and Palmgren (2000).
Figure 2: Top: cluster effects ($\beta Z + b_{i0}$); bottom: cluster x treatment interaction ($\beta Z + b_{i0} + b_{i1}Z$).
Inference

Inference under the random effects Cox model is carried out using the *full likelihood* (as opposed to the partial likelihood).

- The assumptions are: data from different clusters are independent; within the same cluster \( i \), the dependence is induced by \( b_i \); in other words, conditional on \( b_i \), the individuals in cluster \( i \) are independent.

- Conditional on \( b_i \), we may write down the full likelihood for cluster \( i \):

\[
L_i = \prod_{j=1}^{n_i} \lambda_{ij}(X_{ij})^{\delta_{ij}}S_{ij}(X_{ij})
\]

\[
= \prod_{j=1}^{n_i} \{\lambda_0(X_{ij})e^{\beta'Z_{ij}+b_i'W_{ij}}\}^{\delta_{ij}} \exp\{-\Lambda_0(X_{ij})e^{\beta'Z_{ij}+b_i'W_{ij}}\}
\]

- The whole likelihood condition on the \( b_i \)'s is

\[
\prod_{i=1}^{n} L_i.
\]
The likelihood for observed data is then

\[ L(\beta, \lambda_0, \sigma) = \prod_{i=1}^{n} \int L_i \cdot p(b_i; \sigma) db_i. \]

Here \( \sigma \) denotes the unknown parameter(s) in the distribution of the random effects, and \( p(b_i; \sigma) \) is the density of \( b_i \). The parameters to be estimated are \((\beta, \lambda_0, \sigma)\).

In order to solve for MLE, one needs to evaluate the above integrals. Often these integrals are not of closed forms. Iterative algorithms such as EM are often used.

An R package `phmm` is available for the above NPMLE.

Table 1: Estimates (standard errors) from E1582 lung cancer trial

<table>
<thead>
<tr>
<th></th>
<th>( d = 0 )</th>
<th>( d = 1 )</th>
<th>( d = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>( \beta )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>treatment</td>
<td>-0.254 (0.085)</td>
<td>-0.250 (0.104)</td>
<td>-0.247 (0.119)</td>
</tr>
<tr>
<td>bone</td>
<td>0.223 (0.093)</td>
<td>0.212 (0.095)</td>
<td>0.230 (0.144)</td>
</tr>
<tr>
<td>liver</td>
<td>0.429 (0.090)</td>
<td>0.423 (0.091)</td>
<td>0.393 (0.094)</td>
</tr>
<tr>
<td>ps</td>
<td>-0.602 (0.104)</td>
<td>-0.641 (0.109)</td>
<td>-0.649 (0.131)</td>
</tr>
<tr>
<td>wt loss</td>
<td>0.200 (0.087)</td>
<td>0.218 (0.089)</td>
<td>0.208 (0.092)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( \text{Var}(b) )</th>
<th>( \text{Var}(b) )</th>
<th>( \text{Var}(b) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>0.071 (0.069)</td>
<td>0.046 (0.184)</td>
<td></td>
</tr>
<tr>
<td>bone</td>
<td></td>
<td>0.129 (0.083)</td>
<td></td>
</tr>
<tr>
<td>liver</td>
<td></td>
<td></td>
<td>0.198 (0.083)</td>
</tr>
<tr>
<td>ps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt loss</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Notes

In a previous figure of estimated treatment effects by institution, those are the ‘estimated’ $b_i$’s under the $d = 1$ model. They are obtained as empirical Bayes estimate following the MLE; they are the posterior means (can also be mode) of the $b_i$’s given the observed data.

In addition,

- R package ‘phmm’ gives the NPMLE, as well as the empirical Bayes estimate of the random effects, and various testing, AIC etc.;

- Asymptotic properties for the NPMLE has been established, with also stable numerical properties (Gamst et al, 2009).

- `coxph()` with ‘frailty’ option gives a penalized partial likelihood (PPL) estimator;

- there is also an R package ‘coxme’ that gives the PPL under the general PHMM.

- No theoretical properties for the PPL has been established (inconsistent in general). It has numerical problems sometimes, and the estimated variances (hence CI’s) can be inaccurate.

- PPL is faster to compute.
Figure 3: Estimated random effects from univariate lognormal frailty model and 95% confidence intervals (CGD data). (a) without adjustment for covariates other than treatment; (b) with adjustment for covariates.

Notes on recurrent events

Both the marginal model and the random effects model can be used for recurrent event data. When the random effects model is used, the assumption is a common baseline hazard function for the events of the same individual, like in the CGD example.
• Other methods have been studied that are aimed specifically for recurrent event data. One popular such method is **PWP** (Prentice, Williams and Peterson 1981), which is a *conditional approach* that takes into account the ordering of the events, and was the recommended method by Oakes (1992, in *Survival Analysis, State of the Art*, ed. Klein and Goel).

• Another simple approach that is sometimes used is the **Anderson-Gill** model, or the **counting process** approach. This has the favor of the simple Cox model, and requires the relatively strong assumption that the times between successive events be conditionally independent given the history. The covariate ‘prior number of events’ is sometimes included in such a model.

• Therneau and Hamilton (1997, *Stat in Med*, p.2029-2047) compared the above methods using example datasets, showing how they are implemented using popular statistical software. Similar material can also be found in Therneau and Grambsch’s book.