Discrete Time Survival Models

\[ \lambda_j = P(T_i = j \mid T_i \geq j, x_i) = h(\alpha_j + x_i'\beta), \]

where \( \lambda_j \) is the discrete hazard,

\[ \alpha = (\alpha_1, \ldots, \alpha_k)' \] are parameters characterizing the baseline hazard

\( x_i \) are time-independent covariates

\( \beta \) are regression coefficients
Proportional Hazards in Discrete Time

Assuming $\lambda(t) = \lambda_0(t) \exp(x_i'\beta)$, and let $S_i$ denote the continuous event time.

Suppose that $T_i = j$ if $S_i \in (a_{j-1}, a_j]$, for $j = 1, \ldots, k$.

Then, the discrete hazard is as follows:

$$
\Pr(T_i = j \mid T_i \geq j, x_i) = 1 - \frac{S(a_j)}{S(a_{j-1})} = 1 - \frac{\exp(-\Lambda_i(a_j))}{\exp(-\Lambda_i(a_{j-1}))}
$$

$$
= 1 - \exp \left( - \int_{a_{j-1}}^{a_j} \lambda_0(t) \exp(x_i'\beta) \, dt \right)
$$

$$
= 1 - \exp \left\{ - \exp(x_i'\beta)(\Lambda_0(a_j) - \Lambda_0(a_{j-1})) \right\}
$$

$$
= 1 - \exp \left\{ - \exp(\alpha_j + x_i'\beta) \right\},
$$

where $\alpha_j = \log(\Lambda_0(a_j) - \Lambda_0(a_{j-1}))$ and $\Lambda_0(t) = \int_0^t \lambda_0(s) \, ds$. 

Thus, a Cox proportional hazards model can be fit using a discrete-time approximation by using a binary response GLM with a complementary log-log link.

In doing this, the discrete event time $T_i$ must be coded as a $T_i \times 1$ vector of binary responses, $y_i = (0, 0, \delta_i)'$

The corresponding design matrix is then,

$$
X_i = (x_{i1}, \ldots, x_{iT_i})',
$$

where $x_{ij}$ is a $(k + p) \times 1$ vector consisting of 0s in each of the first $k$ positions except for the $j$th which has a 1. The last $p$ elements are fixed at $x_i$. 

Time-Varying Covariates

Often, in applications one or more of the predictors may vary over time.

For example, suppose that we are interested in assessing the effect of air pollution levels on mortality.

The air pollution levels vary from day to day.
Reasonable model for discrete hazard of death in age group $j$?

$$\Pr(T_i = j \mid T_i = j, x_i, z_{ij}) = h(\alpha_j + x_i'\beta + z_{ij}\psi),$$

where $z_{ij}$ is the level of population for individual $i$ at age $j$

This model accommodates the time-varying covariate

Are we making a restrictive assumption?
In the previous model, we assumed that the effect of air pollution was constant at different ages.

In fact, infants and the elderly are more susceptible to pollution-induced mortality.

How can we generalize the model, to account for this age-dependent susceptibility?
Time-Varying Coefficients

\[ \Pr(T_i = j \mid T_i = j, x_i, z_{ij}) = h(\alpha_j + x_i'\beta + z_{ij}\psi_j), \]

where we have now added a \( j \) subscript to the parameter \( \psi \) characterizing the air pollution effect.

Potentially, the effect of the time-independent predictors can also vary with time by allowing different \( \beta \)s for the different age intervals.

Dimensionality rapidly become problematic - Order Restrictions?
What about computation & inference from these models?

Well, if we’re frequentist, we can just fit the binary response GLM and proceed as before (maximum likelihood estimation, analysis of deviance, etc)

If we’re Bayesian, we can potentially also proceed as in binary response GLMs - either using adaptive rejection sampling or (if probit) the Albert and Chib approach
ContinuationRatioProbitModels

\[ \Pr(T_i = j \mid T_i = j, x_{ij}) = \Phi(x_{ij}' \beta), \]

where we can potentially parameterize \( x_{ij} \) to allow a nonparametric baseline and time-varying coefficients.

Note that \( T_i \in \{1, \ldots, k\} \), with \( k \) potentially large.

Thus, we have a potentially large number of parameters, including the time-varying coefficients.
By choosing a probit model, we can update the high dimensional vector jointly after augmenting the data with latent normal variables.

Now, we have \( y_i = (0, \ldots, 0, \delta_i) \) as a \( T_i \times 1 \) outcome vector for subject \( i \).

We introduce a \( z_i = (z_{i1}, \ldots, z_{iT_i})' \) vector of independent normal variables underlying \( y_i \).

\[
y_{ij} = 1(z_{ij} > 0) \quad \text{and} \quad z_{ij} \sim N(x_{ij}' \beta, 1), \quad \text{for} \quad j = 1, \ldots, T_i.
\]

Gibbs sampler proceeds as before.
What about the prior specification?

We have a potentially high-dimensional vector of time-varying baseline parameters and coefficients.

Potentially, no individuals with the event in certain intervals.

What type of information prior is reasonable?
Focusing initially on the model with no time-varying coefficients, we may want to do some smoothing.

Values of $\alpha_j$ and $\alpha_{j'}$ are likely to be similar if $j$ is close to $j'$

Autoregressive - Gaussian random walk prior:

$$\alpha_j \sim N(\alpha_{j-1}, \tau^{-1}),$$

where $\tau$ is a precision parameter controlling the degree of smoothing.
Penalized Likelihood

The autoregressive prior essentially penalizes values of $\alpha_j$ that are far from the neighboring values.

From a frequentist perspective, we can use a similar idea by including a penalty term in the likelihood and then maximizing the resulting penalized likelihood.

The penalty term can follow many forms, including an autoregressive normal density for the $\alpha$s.
Often, the degree of smoothing (controlled by the $\tau$ parameter) may be quite subjective.

Potentially, smoothing may obscure real features of the baseline hazard or time-varying coefficient function.

Order-restrictions are a powerful alternative.
Some Types of Parameter Restrictions

- **Simple Order**: $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_k$.

- **Umbrella Order**: $\alpha_1 \leq \ldots \leq \alpha_{j^*} \geq \ldots \geq \alpha_k$, where $j^* \in \{1, \ldots, k\}$ is an unknown changepoint.

- **Bathtub Order**: $\alpha_1 \geq \ldots \geq \alpha_{j^*} \leq \ldots \leq \alpha_k$.

Let $\Omega \subset \mathbb{R}^k$ denote the order-restricted space
Truncated Conjugate Priors for Parameter Constraints

Gelfand, Smith and Lee (1992)

Choose a prior ignoring the parameter constraint

Proceed with the Gibbs sampler as in the unconstrained case

Discard draws inconsistent with the order constraint
Homework Exercise: Describe a prior specification and Gibbs sampling algorithm for posterior computation of the model,

\[ \lambda_j = \Phi(\alpha_j + x_i'\beta), \]

under the constraint that \( \alpha_1 < \alpha_2 < \ldots < \alpha_k \). Let \( T_i \) denote the discrete event time and let \( \delta_i = 1 \) is the individual is uncensored and let \( \delta_i = 0 \) if the individual is censored, for \( i = 1, \ldots, n \).