Survival Analysis

Special Features:

• Survival times are generally not symmetrically distributed - typically, positively skewed & normality not appropriate.

• Survival times are strictly positive.

• Survival times are frequently *censored*. That is, the *event* is unknown for an individual.
Types of Censoring

- **Right censoring**: Patient enters the study at $t_0$ and leaves the study at $t_0 + c$ without dying. Censoring occurs after (or to the right of) the last known survival time.

- **Left censoring**: Actual survival time is less than the observed time. For example, you might screen an individual at some age $t$ and note that they currently have a tumor, but the exact age of onset is only known to be less than the current age.

- **Interval censoring**: Individuals are known to have experienced a failure within an interval of time. For example, at exam one Bob doesn’t have a tumor, but at a later exam he does have a tumor - thus, tumor onset occurred between the exams.
Survivor function and hazard function

$t = \text{actual survival time of an individual}$

$T = \text{random variable associated with the survival time}$

$f(t) = \text{probability density function for the survival time}$

Distribution function of $T$:

$$F(t) = P(T < t) = \int_0^t f(s) \, ds.$$  

Survival Function:

$$S(t) = P(T \geq t) = 1 - F(t).$$
Hazard Function: probability that an individual dies at time $t$, conditional on having survived to that time.

Hazard function represents the instantaneous death rate for an individual surviving to time $t$:

$$\lambda(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t} \right\}.$$
It follows from the definition of $\lambda(t)$ that

$$\lambda(t) = \frac{P(t \leq T < t + \delta t)}{P(T \geq t)} \times \frac{1}{\delta t} = \frac{F(t + \delta t) - F(t)}{\delta t} \times \frac{1}{S(t)} = \frac{f(t)}{S(t)}.$$ 

So the hazard function is just the density function divided by the survival function.

It follows that

$$\lambda(t) = -\frac{d}{dt} \log S(t),$$

and we have

$$S(t) = \exp \{ -\Lambda(t) \},$$

where $\Lambda(t) = \int_0^t \lambda(s) \, ds$ is the cumulative hazard function.
Likelihood Function: Right Censored Data

Let \( t_i, i = 1, \ldots, n, \) denote the event (or censoring) time

Let \( \delta_i = 0 \) if individual \( i \) is censored and \( \delta_i = 1 \) otherwise

Likelihood:

\[
\prod_{i=1}^{n} \exp \{ - \Lambda(t_i) \} \lambda(t_i)^{\delta_i}.
\]
Proportional Hazards Model

\[ \lambda_i(t) = \lambda_0(t) \exp(x_i'\beta), \]

where \( \lambda_i(t) \) is the hazard function for the \( i \)th individual

\( \lambda_0(t) \) is the baseline hazard function

\( x_i \) is a vector of predictors

\( \beta \) are unknown regression coefficients

**Alternative Form:**

\[ \log \left\{ \frac{\lambda_i(t)}{\lambda_0(t)} \right\} = \beta_1 x_{i1} + \ldots + \beta_p x_{ip}. \]
Proportional Hazards Likelihood Function

Survival Probability:

\[ \exp \{ - \Lambda_i(t_i) \} = \exp \left\{ - \int_0^{t_i} \lambda_i(t) \, dt \right\} = \exp \left\{ - \Lambda_0(t_i) \exp(x_i'\beta) \right\}, \]

where \( \Lambda_0(t) = \int_0^t \lambda_0(s) \, ds. \)

Cox’s Partial Likelihood:

\[
L(\beta) = \prod_{i=1}^n \exp \left\{ - \Lambda_0(t_i) \exp(x_i'\beta) \right\} \{ \lambda_0(t_i) \exp(x_i'\beta) \}^{\delta_i}.
\]

\[\propto \prod_{i=1}^n \left[ \frac{\exp(x_i'\beta)}{\sum_{l \in R(t_i)} \exp(x_l'\beta)} \right]^{\delta_i}, \]

where the denominator is a sum over the individuals at risk at \( t_i \) (i.e., in risk set \( R(t_i) \)).
Note that a major advantage of the Cox’s proportional hazards model is that the *nuisance* parameters characterizing the baseline hazards function factor out of the likelihood for $\beta$.

Intervals between successive death times convey no information about the effect of explanatory variables on the hazard.

Hence, maximum likelihood estimation can proceed via a simple Newton-Raphson procedure implemented in any standard statistical package.

Formulation assumes no ties in the survival times.
Discrete-Time Survival Analysis

In many applications, \( T \in \{1, 2, \ldots, M\} \) and the survival time is said to be discrete.

Discrete-time survival models are also very useful as approximations to continuous time processes.

For example, one could partition the positive real time into intervals,

\[ (a_0, a_1], (a_1, a_2], \ldots, (a_{M-1}, a_M], \]

and then model the probability of falling into each of the bins.
Discrete Hazard: $\Pr(T = j \mid T \geq j)$

If the true survival time ($W$) is continuous, then we could say

$$\Pr(T = j \mid T \geq j) = \Pr(W \in (a_{j-1}, a_j] \mid W \geq a_{j-1}) = \frac{F(a_j) - F(a_{j-1})}{S(a_{j-1})} = \frac{S(a_{j-1}) - S(a_j)}{S(a_{j-1})}.$$ 

The advantage of the discrete-time formulation is that we can write the likelihood function as a simple binary-response glm