7. Parametric models in survival analysis

§7.1. Univariate parametric survival analysis

• **Common distributions for survival analysis**
  - Exponential distribution
  - Weibull distribution
  - Gamma distribution
  - Log-normal distribution
  - Log-logistic distribution

• **Likelihood functions under censoring and truncation**

  Let \( f(t, \theta), F(t, \theta) \) and \( S(t, \theta) \) be the density function, cumulative distribution function and survival function respectively.

  The forms of likelihood with different types of censoring and truncations are as follows:

  i) Survival time \( T_1 \):

  \[ f(T_1, \theta). \]

  ii) Right censoring time \( T_2 \):

  \[ S(T_2, \theta). \]
iii) Left censoring time $T_3$:

\[ F(T_3, \theta). \]

iv) Interval censoring time $[T_{41}, T_{42}]$:

\[ S(T_{41}, \theta) - S(T_{42}, \theta). \]

v) Survival time $T_5$ with left truncation at $L_5$:

\[ \frac{f(T_5, \theta)}{S(L_5, \theta)}. \]

vi) Right censoring time $T_6$ with left truncation at $L_6$:

\[ \frac{S(T_6, \theta)}{S(L_6, \theta)}. \]

The joint likelihood of $T_2, T_2, T_3, (T_{41}, T_{42}), (T_5, L_5), (T_6, L_6)$ is the product of the individual likelihoods, i.e.,

\[
L(\theta) = f(T_1, \theta) S(T_2, \theta) F(T_3, \theta) [S(T_{41}, \theta) - S(T_{42}, \theta)] \\
\times \frac{f(T_5, \theta)}{S(L_5, \theta)} \frac{S(T_6, \theta)}{S(L_6, \theta)}.
\]

**Estimation and inference**

The unknown parameters are estimated by MLE.

The inference is based on the asymptotic distribution of the MLE:

\[ \hat{\theta} \sim N(\theta, I(\hat{\theta})^{-1}). \]
where
\[ I(\hat{\theta}) = -\frac{\partial^2 \ln L(\hat{\theta})}{\partial \theta \partial \theta^t}. \]

§7.2. Parametric regression models for survival analysis

- **Accelerated failure time models**

  \[ T = \exp(\beta_0 + \beta_1^t x)\epsilon, \]
  where \( \epsilon \sim F_0(\epsilon, \theta) \) for some specified distribution function \( F_0 \). The model is equivalent to
  \[ \ln T = \beta_0 + \beta_1^t x + \ln \epsilon. \]

  The survival function of the accelerated failure time is
  \[ S(t|x) = S_0(t\exp\{-\beta_0 + \beta_1^t x\}) \]
  \[ = S(t\exp\{-\beta_1^t x\}|x = 0). \]

  The survival function with covariate \( x \) can be considered as the survival function with covariate \( x = 0 \) accelerated by a factor \( \exp\{-\beta_1^t x\} \), hence the name of the model.

  The common distributions specified for \( \epsilon \) in the accelerated model are: *Exponential*, *Weibull*, *log-normal*, *log-logistic*. The corresponding distributions for \( \ln T \): *minimum extreme value*, *normal*, *logistic*. 

3
• Additive models:

\[ T = \beta_0 + \beta^t x + \epsilon, \]

where \( \epsilon \sim F_0(\epsilon, \theta) \) for some specified distribution function. The common specified distributions: Exponential, Weibull, Gamma, log-normal, log-logistic.

• Estimation and Inference

Consider right censoring only:

**Likelihood function for accelerated model**

\[
L(\beta, \theta) = \prod_{i=1}^{n} \left[ f(T_i e^{-x_i^t \beta}) e^{-x_i^t \beta} c_i \right] S(T_i e^{-x_i^t \beta})^{1-c_i}.
\]

**Likelihood function for additive model**

\[
L(\beta, \theta) = \prod_{i=1}^{n} \left[ f(T_i - x_i^t \beta) \right] c_i S(T_i - x_i^t \beta)^{1-c_i}.
\]

The parameters are estimated by MLE based on the likelihood function.

The inference is based on the theory of MLE.
• Splus functions for parametric regression models

survReg(formula, data=, dist="weibull", parm, init, scale=0, ...)

formula
a formula expression as for other regression models.

dist
assumed distribution for y variable. Allowed values include "weibull", "exponential", "gaussian", "logistic", "lognormal" and "loglogistic".

parm
a list of fixed parameters. For the t-distribution for instance this is the degrees of freedom.

init
optional vector of initial values for the parameters.

scale
optional fixed value for the scale. If set to <=0 then the scale is estimated.