Exercise Set 3
Duration Models

1. (Kaplan-Meier estimator of the survival function) Let $0 < a_1 < a_2 < \cdots < a_M$ be a positive, increasing set of constants, and let T be a non-negative random variable with $\Pr\{T > 0\} = 1$.

   (a) Show that, for any $m = 1, \ldots, M$

   $$\Pr\{T > a_m\} = \Pr\{T > a_m | T > a_{m-1}\} \Pr\{T > a_{m-1}\}$$

   (b) Hence show

   $$\overline{F}(a_r) = \Pr\{T > a_r\} = \prod_{r=1}^{n} \Pr\{T > a_r | T > a_{r-1}\}.$$

   Noting that a consistent estimator of $\Pr\{T > a_r | T > a_{r-1}\}$ is $\frac{N_r - E_r}{N_r}$, you have just derived the Kaplan-Meier estimator of the survival function

   $$\hat{S}(t) = \widehat{\overline{F}}(a_r) = \prod_{r=1}^{n} \frac{N_r - E_r}{N_r}$$

2. Consider the Weibull model with multiplicative heterogeneity,

   $$\lambda(t|x, v) = ve^{x\beta}\lambda_0(t)$$

   with $\lambda_0(t) = \alpha t^{\alpha-1}$. Assume that there are two types, with $\Pr\{v_i = \eta\} = \rho$ for A types, and $\Pr\{v_i = (1 - \rho\eta)/(1 - \rho)\} = 1 - \rho$ for B types (hence $E(v_i) = 1$) where $0 < \eta < 1$.

   (a) Find the distribution function of the duration variable T.

   (b) Find the log-likelihood function for observation i.

3. Assume that you have a flow sample, but all durations are censored.

   (a) Write down the log-likelihood function when all durations are censored.

   (b) Find the special case of the Weibull distribution in part (a).

   (c) Assume that covariates x do not affect the hazard, only on intercept term is to be included, $\exp(\beta)$. Show that the Weibull log likelihood cannot be maximized for any real numbers β and α. What do you conclude about estimating duration models from flow data when all durations are right censored ?

(d) Now assume that only some durations are censored, but you do not observe the individual durations: let c_i be individual's i censoring time, and $d_i = 1$ if the i's duration is censored. Hence you only observe this indicator variable. Show that the conditional likelihood function has the "binary response" form.

4. Consider the problem of estimation using a stock sample. Let the non-censoring indicator be $d_i = 1$(spell not censored), c_i is the individual censoring date, assumed to be uninformative. $T_i = \min\{T_i^*, c_i\}$. The spell starts at time τ_i before interview date b. Consider (τ_i, c_i, x_i, t_i), which is a random draw from the population of all spells starting in $[0, b]$. The individual is in the sample if the sample selection indicator $s_i = 1(T_i^* \geq b - \tau_i)$ equals one. Assume that $c_i > b - \tau_i$ for all i, so that we always observe part of each spell after the sampling date, b.

   (a) For $b - \tau_i < t < c_i$ show that

   $$\Pr\{T_i^* \leq t | \tau_i, c_i, x_i, s_i = 1\} = \frac{F(t|x_i) - F(b - \tau_i|x_i)}{1 - F(b - \tau_i|x_i)}.$$

   (b) Hence obtain the density of T_i^* conditional on $(\tau_i, c_i, x_i, s_i = 1)$.

   (c) Show that

   $$\Pr\{T_i = c_i | \tau_i, c_i, x_i, s_i = 1\} = \frac{1 - F(c_i|x_i)}{1 - F(b - \tau_i|x_i)}.$$

   (d) Hence show that the likelihood is therefore

   $$\sum_{i=1}^{n} d_i \log(f(t_i|x_i, \theta)) + (1 - d_i)\log(1 - F(t_i|x_i, \theta)) - \log(1 - F(b - \tau_i|x_i, \theta)).$$