- Problem 1
  - Survival analysis involves the modelling of time to event data, in this context, death or failure is considered an "event". Survival analysis attempts to answer questions such as: what is the fraction of a population which will survive past a certain time? Of those that survive, at what rate will they die or fail? How do particular circumstances or characteristics increase or decrease the odds of survival?
  - $$\lambda(t; x_i) = \lambda_0(t)e^{\beta x_i},$$

    where x_i is $p \times 1$ vector ofcovariates and β is a $1 \times p$ vector of parameters.

    $$\lambda(t; x_i(u), u \leq t) = \lambda_0(t)e^{\beta x_i(t)},$$

    where $x_i(t)$ are time-dependent covariates.
  - Independent censoring in a one sample setting means that the failure times are independent with censoring times, i.e., the value of C could not have implication on the value of T.

    Independent censoring in the PHM means that, conditional on covariates at each duration, the censored items are "representative" of those under observation at the same time. In other words, items may not be censored (withdrawn) because they have a higher or lower risk than the average, given the covariates.
  - In short, informative censoring means that the censoring distribution may depend on the unknown parameters in the model.

    In general, getting sicker is a common reason of informative censoring. For example, compared to placebo, drug treatment group would be selectively depleted of the sickest persons, which would make drug treatment group appear better
- Problem 2

~~If the largest observed time $y_{(k)}$ is uncensored, $\hat{S}(y_{(k)}) = 0$. Otherwise, the Kaplan-Meier estimate will not go down to 0 and is unreliable (undetermined) for $t > y_{(k)}$.~~

- ~~Problem 3~~

If all of the largest observed times are uncensored, the Kaplan-Meier-estimate will eventually reach the value 0. If one or more of the largest observed times are censored, then the K-M-estimate will not reach 0, and is undetermined for $t >$ largest uncensored time $Y_{(K)}$.

• Problem 3

–

$$\lambda_1(t) = \lambda_0(t)e^{\beta},$$

where $\lambda_1(t)$ is the hazard function of treatment group and $\lambda_0(t)$ is the hazard function of control group.

The partial likelihood is

$$L_p = \frac{1}{3+4e^{\beta}} \frac{e^{\beta}}{1+4e^{\beta}} \frac{e^{\beta}}{2e^{\beta}} \frac{e^{\beta}}{e^{\beta}}$$

$$= \frac{1}{3+4e^{\beta}} \frac{e^{\beta}}{1+4e^{\beta}} \frac{1}{2}$$

Obtain $\hat{\beta}$ by maximizing L_p.

– Small sample size results in large variance of estimates from (a).

– Kaplan-Meier estimate for the treatment data:

$$\hat{S}(7) = (1 - 1/4) = 3/4$$

$$\hat{S}(11) = 3/4(1 - 1/2) = 3/8$$

$$\hat{S}(11) = 3/8(1 - 1/1) = 0$$

- Problem 4

  See homework solution

- Problem 5

  See homework solution