Année 2019-2020

Survival and longitudinal data analysis

Exercice 1

Fitting a parametric distribution: `Cabinet duration` data

This data set contains survival data on government coalitions in parliamentary democracies for the period 1945-1987 (see here for more details).

1. Compute and draw the empirical c.d.f. for the variable "duration".
2. Compute maximum likelihood estimator for the variable "duration" in the
   - exponential family
   - Weibull family
   - log-normal family
   - gamma family

   via the `fitdistr` of package `MASS`.
3. Draw the estimated c.d.f. in each family on the same plot as the empirical c.d.f.
4. Based on the plot, which distribution would you choose for the variable "duration" ?

Exercice 2

Your own code for the Kaplan-Meier estimator

- Develop a function to compute the Kaplan-Meier estimator that takes as inputs

  $$(t_1^C, \delta_1), (t_2^C, \delta_2), \ldots, (t_n^C, \delta_n),$$

  where the t_i^C are not necessarily in increasing order (nor distinct !).
- Consider the `pharmocoSmoking` data (available in package `asaur`), compare the results of your code to the one of the function `survfit` of package `survival`.
- Compute the Greenwood estimator of the variance of the Kaplan-Meier estimator.
- Is there a difference according to the treatment ?

Hint: you will need the following R functions `order`, `unique`

Exercice 3

Left-truncated and right-censored data

1. Load the `channing` dataset of the package `KMsurv`. From which problem(s) of observation do these data suffer ?
2. At age 901 how many residents are under observation and still alive ? In other words, how many patients are in the risk set at time 901 ?
3. They are 4 residents with ageentry = age. What happened to them ? Add 0.5 to the variable `age`.
4. Look at the option of the function `Surv` and estimate of the survival function via the `survfit` function.
5. Try to reproduce the figure below.

Exercice 4

You will need R packages `MASS`, `survival`, `asaur`, `KMsurv`.

Cnstruction and interpretation of a Cox model for the `pharmocoSmoking` dataset

1. How many covariates does the dataset contain ?
2. Fit a first Cox model with all the covariates you found in question 1. What is the problem ?
3. Fit a new Cox model with a subset of covariates, that solves the previous problem.
4. Do a backward procedure of variable selection based on Wald tests.
5. Interpret the coefficients in the final model.

Figure 4.11 shows the estimated probability of surviving beyond age t, given survival to 68 or 80 years for both males and females.

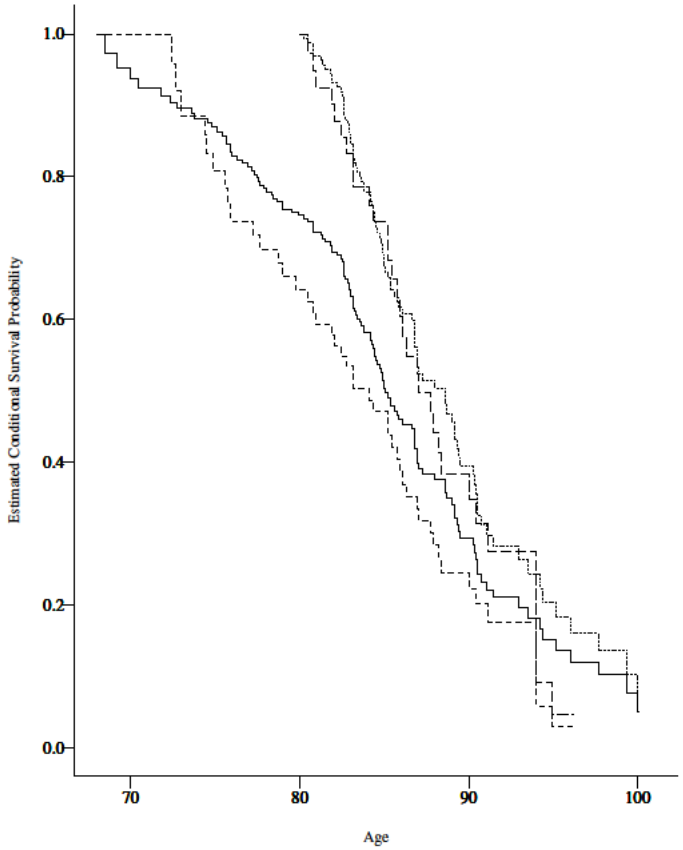

Figure 4.11 *Estimated conditional survival functions for Channing house residents. 68 year old females (————); 80 year old females (------); 68 year old males (— — —); 80 year old males (—— ——).*

Figure 1: Figure from page 125 of Klein and Moeschberger 2005