Statistics for Biology and Health
Chapter 4 Estimation of Basic Quantities for Other Sampling Schemes

Qi Guo

July 20,2019

THE UNIVERSITY OF TEXAS AT DALLAS
School of Natural Sciences and Mathematics

Contents

Introduction

Introduction

- Last Chapter we focus on estimate of the right-censored and left-truncated data, which is most common.
- In this chapter, we examine about left censoring, censored individuals provide information indicating only that the event has occurred prior to entry into the study.
- Right-truncated data, which samples arise when one samples individuals from event records and, retrospectively determines the time to event.
- Estimation techniques for grouped data.

Estimation of the Survival Function for Left, Double, and Interval Censoring

Estimation of Left Censoring Data

- Examples of pure left censoring are rare. More common are samples which include both left and right censoring.
- And in this case we use a modified KM estimator suggested by Turnbull, is based on an iterative procedure which extends the notion of a self-consistent estimator.
- To construct the estimator we assume that there is a grid of time points $0 = t_0 < t_1 < t_2 < \cdots < t_m$ at which subjects are observed.
- Let d_i be the number of deaths at time t_i, and c_i,r_i are the number of individuals left-censored and right-censored at time t_i.

Algorithm

- The only information the left-censored observations at t_i give us is that the event of interest has occurred at some $t_j \leq t_i$
- The self-consistent estimator estimates the probability that this event occurred at each possible t_j less than t_i based on an initial estimate of the survival function.
- Using this estimate, we compute an expected number of deaths at t_j, which is then used to update the estimate of the survival function and the procedure is repeated until the estimated survival function stabilizes.

Algorithm

- `Step 0`: Produce an initial estimate of the survival function at each t_j,$S_0(t_j)$. Note any legitimate estimate will work. Turnbull?s suggestion is to use the KM estimate obtained by ignoring the left-censored observations.
- `Step` $(K+1)1$: Using the current estimate of S, estimate $p_{ij} = p[t_{j-1} < x \leq t_i | x \leq t_i]$ by $\frac{S_K(t_{j-1}) - S_K(t_j)}{1 - S_K(t_i)}$ for $j \leq i$.
- `Step` $(K+1)2$: Using the results of the previous step, estimate the number of events at time t_i by $\hat{d}_i = d_i + \sum_{i=j}^{m} c_i p_{ij}$.
- `Step` $(K+1)3$: Compute the KM estimator based on the estimated right-censored data with $\hat{d}_i$ events and r_i right-censored observations at t_i, ignoring the left-censored data. If this estimate, $S_{K+1}(t)$, is close to $S_K(t)$ for all t_i, stop the procedure; if not, go to step 1.

Estimation of the interval-censored data

- Sometimes data may be interval-censored. Here the only information we have for each individual is that their event time falls in an interval $(L_i, R_i], i = 1, ..., n$, but the exact time is unknown.
- And we still have a iteration like the estimarion of left-censored data in page 6.

Estimation of the Survival Function for Right-Truncated Data

Right-Truncated Data

- For right-truncated data, only individuals for which the event has occurred by a given date are included in the study. Right truncation arises commonly in the study of infectious diseases.
- Let T_i denote the chronological time at which the ith individual is infected and X_i the time between infection and the onset of disease.
- Sampling consists of observing (T_i, X_i) for patients over the period $(0\ to\ \tau)$. Note that only patients who have the disease prior to τ are included in the study.
- Estimation for this type of data proceeds by reversing the time axis. Let $R_i = \tau - X_i$. The R_i's are now left-truncated in that only individuals with value of $T_i \leq R_i$ are included in the sample.

Estimation of Survival in the Cohort Life Table

Estimation of Survival in the Cohort Life Table

- A“cohort” is a group of individuals who have some common origin from which the event time will be calculated
- They are followed over time and their event time or censoring time is recorded to fall in one of $k+1$ adjacent, nonoverlapping intervals, $(a_{j-1}, a_j]$, $j = 1, ..., k+1$.
- A traditional cohort life table presents the actual mortality experience of the cohort from the birth of each individual to the death of the last surviving member of the cohort.
- The basic construction of the cohort life table is a little complicated, and more details in columns in P152.