Statistics for Biology and Health
Chapter 5 Topics in Univariate Estimation

Qi Guo

July 21,2019

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

Contents

Introduction

Introduction

- Now we know KM estimator provides an estimate of the survival function, and NA estimator provides an estimate of the cumulative hazard rate, and the slope of the NA estimator provides a crude estimate of the hazard rate, but it usually hard to interpret.
- But these two estimator provide limited information about the mechanism of the process under study, as summarized by the hazard rate, so use kernel-smoothing technique to provide a better estimator of the hazard rate, which based on the NA estimator $\tilde{H}(t)$ and its variance $\hat{V}[\tilde{H}(t)]$.

Introduction

- Estimation techniques for both the additive and multiplicative models for excess mortality about the experience of the experimental subjects differs.
- In the end, show the problem of estimation of the survival function for right censored data is considered from a Bayesian perspective.

Estimating the Hazard Function

Introduction

- Now we know $\tilde{H}(t)$ is a step function with jumps at the event times, $0 = t_0 < t_1 < t_2 < \cdots < t_D$, let $\Delta\tilde{H}(t_i) = \tilde{H}(t_i) - \tilde{H}(t_{i-1})$ and $\Delta\hat{V}(\tilde{H}(t_i)) = \hat{V}(\tilde{H}(t_i) - \hat{V}(\tilde{H}(t_{i-1})$
- Note that $\Delta\hat{V}(\tilde{H}(t_i))$ provides a crude estimator of $h(t)$ at the death times. The kernel-smoothed estimator of $h(t)$ is a weighted average of these crude estimates over event times close to t.
- Closeness is determined by a bandwidth b, so that event times in the range $t - b$ to $t + b$ are included in the weighted average which estimates $h(t)$.
- The bandwidth is chosen either to minimize some measure of the mean-squared error or to give a desired degree of smoothness

The kernel function

- The weights are controlled by the choice of a kernel function, $K()$,defined on the interval $[-1, +1]$, which determines how much weight is given to points at a distance from t.
- Common choices for the kernel are the uniform kernel with

$$K(x) = 1/2 \ \ for \ -1 \leq x \leq +1$$

- The Epanechnikov kernel with

$$K(x) = 0.75(1 - x^2) \ \ for \ -1 \leq x \leq +1$$

- The biweight kernel with

$$K(x) = 0.75(1 - x^2) \ \ for \ -1 \leq x \leq +1$$

- The uniform kernel gives equal weight to all deaths in the interval $t - b$ to $t + b$, whereas the other two kernels give progressively heavier weight to points close to t.

The kernel-smoothed hazard rate estimator

- For any $t > 0$, for which $b \leq t \leq t_D - b$, the kernel smoothed estimator of $h(t)$ based on the kernel $K()$ is given by:

$$\hat{h}(t) = b^{-1} \sum_{i=1}^{D} K(\frac{t - t_i}{b}) \Delta \tilde{H}(t_i)$$

- The variance of $\hat{h}(t)$ is estimated by:

$$\delta^2[\hat{h}(t)] = b^{-2} \sum_{i=1}^{D} K(\frac{t - t_i}{b})^2 \Delta \hat{V}(\tilde{H}(t_i))$$

- When $t < b$, there are no events times less than 0 are observable, the use of an asymmetric kernel is suggested.

Gasser and Muller modified kernels

- Let $q = t/b$, define a modified kernel which accounts for the restricted range of the data, for the uniform kernel are expressed by:

$$K_q(x) = \frac{4(1+q^3)}{(1+q^4)} + \frac{6(1-q)}{(1+q)^3}x \quad for \ -1 \leq x \leq q$$

- And we also have versions for the Epanechnikov kernel and the biweight kernel.
- The confidence intervals or confidence bands for the hazard rate based on the smoothed hazard rate estimate is :

$$\hat{h}(t) exp\left[\pm \frac{Z_{1-\alpha/2}\sigma(\hat{h}(t))}{\hat{h}(t)} \right]$$

for a $(1-\alpha) \times 100\%$ pointwise CI based on a log transformation.

Example

- `Figure 6.3` shows the estimated hazard rate based on a bandwidth of 1 year for the uniform, Epanechnikov, and biweight kernels.

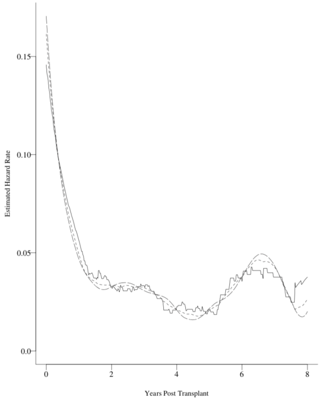

Figure 6.3 *Effects of changing the kernel on the smoothed hazard rate estimates for kidney transplant patients using a bandwidth of 1 year. Uniform kernel (———); Epanechnikov kernel (------) Biweight kernel (— — —)*

Determine the bandwidth

- Using kernel smoothing to obtain an estimate of the hazard rate is selection of the proper bandwidth.
- One way to pick a good bandwidth is to use a cross-validation technique for determining the bandwidth that minimizes some measure of how well the estimator performs.
- One such measure is the mean integrated squared error (*MISE*) of $\hat{h}$ over the range τ_L to τ_U defined by:

$$
\begin{aligned}
MISE(b) &= E\int_{\tau_L}^{\tau_U}[\hat{h}(u) - h(u)]^2 du \\
&= E[\int_{\tau_L}^{\tau_U}\hat{h}^2(u)du] - 2E[\int_{\tau_L}^{\tau_U}\hat{h}(u)h(u)du] + E[\int_{\tau_L}^{\tau_U}h^2(u)du]
\end{aligned}
$$

Determine the bandwidth

- The first term can be estimated by $\int_{\tau_L}^{\tau_U} \hat{h}^2(u)du$. If we evaluate $\hat{h}$ at a grid of points $\tau_L = u_1 < \cdots < u_M = \tau_U$, then, an approximation to this integral by the trapezoid rule is $\sum_{i=1}^{M-1} (\frac{u_{i+1}-u_i}{2})[\hat{h}^2(u_i) + \hat{h}^2(u_{i+1})]$.
- The second term can be estimated by a cross-validation estimate suggested by Ramlau-Hansen. This estimate is $b^{-1} \sum_{i \neq j} K(\frac{t_i - t_j}{b}) \Delta\tilde{H}(t_i) \Delta\tilde{H}(t_j)$, where the sum is over the event times between τ_L and τ_U.
- The last term depends on the unknown hazard rate, it is independent of the choice of the kernel and the bandwidth and can be ignored when finding the best value of b.

Determine the bandwidth

- Thus to find the best value of b which minimizes the *MISE* for a fixed kernel, find b which minimizes the function:

$$g(b) = \sum_{i=1}^{M-1} \left(\frac{u_{i+1} - u_i}{2}\right) [\hat{h}^2(u_i) + \hat{h}^2(u_{i+1})] - 2b^{-1} \sum_{i \neq j} K\left(\frac{t_i - t_j}{b}\right) \Delta\tilde{H}(t_i) \Delta\tilde{H}(t_j)$$

Example

- `Figure 6.4` shows the effects of changing the bandwidth on the estimate of $h(t)$. In this figure, based on the Epanechnikov kernel, we see that increasing the bandwidth provides smoother estimates of the hazard rate.

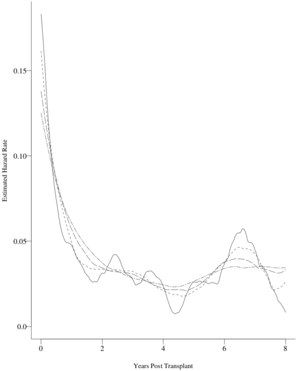

Figure 6.4 *Effects of changing the bandwidth on the smoothed hazard rate estimates for kidney transplant patients using the Epanechnikov kernel. bandwidth = 0.5 years (———) bandwidth = 1.0 years (-------) bandwidth = 1.5 years (— — —) bandwidth = 2.0 years (− · − · −)*

Estimation of Excess Mortality

Introduction

- Sometimes it's of interest to compare the mortality experience of a group of individuals to a known standard survival curve.
- Two simple models have been proposed to provide an inference on how the study population's mortality differs from that in the reference population.
- Suppose we have data on n individuals. Let $\theta_j(t)$ be the reference hazard rate for the jth individual in the study, which typically depends on the characteristics of the jth patient, such as race,sex,age,etc.

The relative mortality model

- The first model for excess mortality, commonly known as the relative mortality model, assumes that the hazard rate at time t for the jth patient under study conditions, $h_j(t)$, is a multiple, $\beta(t)$, of the reference hazard rate for this individual:

$$h_j(t) = \beta(t)\theta_j(t) \quad j = 1, 2, \cdots, n$$

- If $\beta(t) > 1$, individuals in the study group are experiencing the event of interest at a faster rate than comparable individuals in the reference population
- Let $B(t) = \int_0^t \beta(u)du$ be the cumulative relative excess mortality.
- The data available for estimating $B(t)$, for each individual, consists of study times and death indicators.

The relative mortality model

- For the jth individual, let $Y_j(t)$ be 1 if the individual is at risk at time t and 0, otherwise.Define the function $Q(t) = \sum_{j=1}^{n} \theta_j(t) Y_j(t)$, let $t_1 < t_2 < \cdots < t_D$ be the times at which the events occur and d_i the number of events observed at time t_i. The estimator of $B(t)$ and its variance are:

$$\hat{B}(t) = \sum_{t_i \leq t} \frac{d_i}{Q(t_i)}$$

and

$$\hat{V}[\hat{B}(t)] = \sum_{t_i \leq t} \frac{d_i}{Q(t_i)^2}$$

- And we can also find the CI or CB, and a crude estimator of the relative risk function $\beta(t)$ is given by the slope of the estimated cumulative risk mortality estimator, and it can also improved by kernel smoothing like before.

The additive mortality model

- A second model used for comparing the study population to a reference population is the excess or additive mortality model.
- Assume that the hazard rate at time t for the jth individual under study is a sum of the population mortality rate $\theta_j(t)$ and an excess mortality function $\alpha(t)$.
- The $\alpha(t)$ is same for all individuals in the study group, and it's positive means study patients are dying faster than those in the reference population, is expressed by:

$$h_j(t) = \alpha(t) + \theta_j(t) \quad j = 1, \cdots, n$$

- Introduce the cumulative excess mortality function for convenient to estimate the $\alpha()$:

$$A(t) = \int_0^t \alpha(u)du$$

The additive mortality model

- The estimator of $A(t)$ is constructed from the difference of the observed hazard rate, estimated by the ordinary NA estimator $\tilde{H}(t)$ and an "expected" cumulative hazard rate $\Theta(t)$ based on the reference hazard rates.

$$\Theta(t) = \sum_{j=1}^{n} \int_0^t \theta_j(u) \frac{Y_j(u)}{Y(u)} du$$

  where $Y(t) = \sum_{j=1}^{n} Y_j(t)$ is the number of at risk at time t.
- The estimated variance is

$$\hat{V}[\hat{A}(t)] = \sum_{t_i \le t} \frac{d_i}{Y(t)^2}$$

- The survival cure, $S^*(t) = \exp[-\Theta(t)]$, provides an estimate of the expected survival curve if the reference mortality model is the same as the study population.

Bayesian Nonparametric Methods

Introduction

- An alternative to the classical nonparametric approach to estimating the survival function is to use Bayesian nonparametric methods.
- An investigator's a priori belief in the shape of the survival function is combined with the data to provide an estimated survival function.
- The parameters of the model are treated as random variables selected from the prior distribution,which is a multivariate distribution on the parameters, is selected to reflect the investigator's prior belief in the values of the parameters.
- It's complicated and just introduce the crude steps of the Bayesian methods, more details in P187-P197.

Two classes of prior distributions

- To obtain an estimate of the survival function, specify a loss function on which to base the decision rule.Analogous to the simple parametric case, use the squared-error loss function:

$$L(S, \hat{S}) = \int_0^{\infty} [\hat{S}(t) - S(t)]^2 d\omega(t)$$

  where $\omega(t)$ is a weight function.
- For this loss function, the value of $\hat{S}$, which minimizes the posterior expected value of $L(S, \hat{S})$, is the posterior mean and the Bayes risk $E[L(S, \hat{S})|DATA]$ is the posterior variance.
- And we have two classes of prior distributions.

Two classes of prior distributions

- Both lead to closed form estimates of the survival function using the squared-error loss function.
- These priors are chosen because they are conjugate priors for either the survival function or the cumulative hazard function. For a conjugate prior, the prior and posterior distributions are in the same family.
- The first prior is for the survival function, for this prior, we assume that the survival function is sampled from a Dirichlet process with a parameter function α.
- A second prior is to provide a prior distribution for the cumulative hazard function $H(t) = -\ln[S(t)]$. Here, we shall use a beta process prior.

Monte Carlo Bayesian methods

- The Monte Carlo Bayesian methods or the Gibbs sampler is more flexible than the other two approaches.
- For right-censored data,the closed form estimates of the survival function are available. But for other censoring or truncation schemes, such simple estimates are not available.
- The Monte Carlo Bayesian methods provides a way of simulating the desired posterior distribution of the survival function.