Regression: Week 6, Tuesday

Niels Richard Hansen
May 26, 2013

Contents

1 Introduction **1**

2 Non-parametric estimators **2**

3 Hazards **3**

1 Introduction

Survival analysis

– or reliability analysis.

The topic has its own development with focus on aspects of models and distributions that differ from many other applications of statistics. This is primarily due to the following two issues:

- Survival distributions are skewed distributions on the positive half line. It is the *shape* of the distribution rather than the location of the distribution that is of interest.
- There is almost always a *censoring mechanism*, and certain aspects of the data are consequently missing. We need to deal with this in the modeling.

Example I

In medicine we want to test whether a new, promising drug can prolong the life of humans.

We set up a controlled, double-blinded experiment with 1000 individuals of age 55 given this drug and a control group of 1000 individuals of age 55 given a placebo drug (disregarding any ethical considerations at this point).

The test runs for 10 years, and those that survive for 10 years are all censored at that time. That is less problematic than if 10%, say, of the participants are censored because they abandon prematurely the experiment without dying.

Example II

In engineering we want to estimate the life time of an electrical component. We record whenever a component is put to work and whenever it fails. At a given time, all working components that have not yet failed are censored.

To estimate the life time based on the observed life times for the components that have failed up to this time will give a too pessimistic, biased result.

Example III

A "real" survival application.

Patients are enrolled in a study whenever they are diagnosed with a given (serious, life threatening) disease. Data on the subjects are collected – and may be collected regularly.

At a planned calendar time the statistical analysis is done, and patients alive at this time are censored.

Many questions are of interest, e.g. how different covariates are associated with the survival for this particular disease. One issue may be to compare the survival distributions for two or more treatments.

2 Non-parametric estimators

The setup

We consider n individuals, $T_1^*, \ldots, T_n^*$ independent, positive random variables (survival times). We observe

$$T_i = \min\{T_i^*, C_i\}$$

with *censoring times* $C_1, \ldots, C_n$. With

$$e_i = 1(T_i^* \leq C_i)$$

we observe the pairs

$$(T_1, e_1), \ldots, (T_n, e_n).$$

Survival function

The distribution function is $F(t) = P(T_1^* \leq t)$ and the *survival function* is

$$S(t) = 1 - F(t) = P(T_1^* > t).$$

Without censoring,

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1(T_i^* \leq t)$$

is the empirical distribution function and

$$\hat{S}(t) = 1 - \hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1(T_i^* > t).$$

the empirical survival function.

Introduce the process of *individuals at risk*

$$Y(t) = \sum_{i=1}^{n} 1(t \leq T_i).$$

The Kaplan-Meier estimator

Based on the censored survival observations (T_i, e_i), the process, $Y(s)$, of *at risk* individuals, and the (ordered) observed *survival* times t_i for $i = 1, \ldots, k$ up to time t the *Kaplan-Meier* estimator is

$$\begin{aligned}\hat{S}(t) &= \left(1 - \frac{1}{Y(t_1)}\right)\left(1 - \frac{1}{Y(t_2)}\right)\cdots\left(1 - \frac{1}{Y(t_k)}\right) \\ &= \prod_{i:t_i<t}\left(1 - \frac{1}{Y(s)}\right).\end{aligned}$$

This estimator is the survival analysis version of the empirical distribution function.

The version of the Kaplan-Meier estimator above assumes that only one death occurs at each time. An alternative formulation is as follows. With the *counting process* of deaths (non censored events)

$$N(t) = \sum_{i=1}^{n} 1(T_i \leq t, e_i = 1)$$

and the jumps for the counting process given as

$$\Delta N(t) = N(t) - N(t-) = N(t) - \lim_{\varepsilon \to 0+} N(t - \varepsilon),$$

the Kaplan-Meier estimator is

$$\hat{S}(t) = \prod_{s \leq t}\left(1 - \frac{\Delta N(s)}{Y(s)}\right).$$

The factors are equal to 1 for all s where $\Delta N(s) = 0$. In the notation above, $k = N(t)$. This version allows for $\Delta N(s) > 1$ to accommodate multiple deaths at the same time. This is in practice only a question of resolution. Survival times are typically in days, thus we can encounter multiple deaths at the same day.

The intuition behind the estimator is that each factor is an estimator of the conditional probability of surviving the time interval $(t_i, t_{i+1}]$ given survival beyond time t_i. Precisely,

$$\begin{aligned} S(t) = P(T_1^* > t) &= P(T_1^* > t \mid T_1^* > t_k) \times P(T_1^* > t_k \mid T_1^* > t_{k-1}) \\ &\quad \ldots \times P(T_1^* > t_2 \mid T_1^* > t_1) \times P(T_1^* > t_1). \end{aligned}$$

There is one individual out of $Y(t_i)$ that dies in the interval $(t_i, t_{i+1}]$, whence conditionally on having survived beyond t_i the probability of dying is estimated as $1/Y(t_i)$ and the probability of surviving beyond t_{i+1} is thus $1 - 1/Y(t_i)$. This argument is again under the assumption that multiple deaths do not occur at the same time.

3 Hazards

The hazard rate

If F is continuously differentiable with derivative f (the density for the survival distribution), we introduce the *hazard rate*

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

Observe that

$$\begin{aligned} \lambda(t) &= \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \frac{F(t+\varepsilon) - F(t)}{S(t)} \\ &= \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} P(T_i^* \in (t, t+\varepsilon] \,|\, T_i^* > t). \end{aligned}$$

Thus $\lambda(t)$ is the instantaneous rate of death at time t.

Examples

The *exponential distribution* has hazard rate

$$\lambda(t) = \lambda,$$

which is constant.

The *Weibull distribution* has hazard rate

$$\lambda(t) = \alpha\gamma t^{\gamma-1}$$

for $\alpha, \gamma > 0$.

The Weibull distribution can, in particular, capture increasing hazard rates over time by taking $\gamma > 1$.

The cumulative hazard function

Note that

$$\lambda(t) = -(\log S(t))'$$

hence

$$\Lambda(t) := \int_0^t \lambda(s)\mathrm{d}s = -\log S(t),$$

which is called the *cumulative hazard function*.

Observe that

$$S(t) = \exp\left(-\Lambda(t)\right).$$

The Nelson-Aalen estimator

With $\hat{S}$ the Kaplan-Meier estimator of the survival function the cumulative hazard can be estimated as $-\log \hat{S}$.

Alternatively, the *Nelson-Aalen* estimator is a direct estimator given as

$$\hat{\Lambda}(t) = \sum_{i:t_i<t} \frac{1}{Y(t_i)}$$

The corresponding estimator of the survival function is

$$\exp(-\hat{\Lambda}(t)) = \prod_{i:t_i<t} \exp\left(-\frac{1}{Y(t_i)}\right)$$

with

$$\exp\left(-\frac{1}{Y(t_i)}\right) \simeq \left(1 - \frac{1}{Y(t_i)}\right)$$

for large $Y(t_i)$ using the Taylor expansion $\exp(x) = 1 + x + o(x^2)$.

The intuition behind the Nelson-Aalen estimator is as follows. The probability that an individual that survived beyond time t_i dies in (a small) interval $(t_i, t_{i+1}]$ is approximately $\lambda(t_i)(t_{i+1} - t_i)$ by the definition of the hazard rate. The cumulative hazard function $\Lambda(t)$ is approximately the sum of these quantities for $t_i < t$. The probability of a death in $(t_i, t_{i+1}]$ is estimated as $1/Y(t_i)$ and aggregated by summation we get the Nelson-Aalen estimator of Λ. A technical justification that provides a deeper insight requires the framework of counting processes and martingales, which reveal that the estimator is unbiased.

Regression: Week 6, Thursday

Niels Richard Hansen
May 26, 2013

Contents

1The survival likelihood101 2Accelerated failure time models502 3Proportional hazards models903

1 The survival likelihood

Setup

Assume that T^* is a positive random variable with density f and survival function S, C is a positive random variable with density g and survival function H.

We define

$$T = \min\{T^*, C\} \quad \text{and} \quad e = 1(T^* \leq C).$$

Theorem 1. *If T and C are independent the joint distribution of (T, e) has density*

$$f(t)^e S(t)^{1-e} g(t)^{1-e} H(t)^e$$

w.r.t. the product measure $m \otimes \tau$ (the Lebesgue measure times the counting measure).

Proof. The derivation is as follows

$$\begin{aligned} P(T \leq t, e = 1) &= P(T^* \leq t, e = 1) \\ &= P(T^* \leq t, C \geq T^*) \\ &= \int_0^t f(s) \int_s^\infty g(u) \mathrm{d}u \mathrm{d}s \\ &= \int_0^t f(s) H(s) \mathrm{d}s. \end{aligned}$$

Likewise,

$$P(T \leq t, e = 0) = \int_0^t g(s) S(s) \mathrm{d}s,$$

and we conclude that the density is

$$h(t, e) = \begin{cases} f(t)H(t) & \text{if } e = 1 \\ g(t)S(t) & \text{if } e = 0. \end{cases}$$

□

The full likelihood

With $(T_1, e_1), \ldots, (T_n, e_n)$ i.i.d. with the same distribution as (T, e) the full likelihood is

$$L = \prod_{i=1}^{n} f(T_i)^{e_i} S(T_i)^{1-e_i} g(T_i)^{1-e_i} H(T_i)^{e_i}.$$

We assume that $f = f_\theta$ is parametrized by θ and that the distribution, given by g, of the censoring mechanism holds no information about θ. This implies that

$$L(\theta) = \prod_{i=1}^{n} f_\theta(T_i)^{e_i} S_\theta(T_i)^{1-e_i} K_i$$

with K_i depending on the observations but not the parameter θ.

Note that Harrell (and many others) "derive" the likelihood in a relatively informal way, which leaves some room for wondering exactly what the dominating measure is. This is explicit above. The derivation also makes it clear how the distribution of the censoring mechanism enters, and why it can be ignored if it does not depend on the unknown parameter θ. This has nothing to do with independence, which is important for the derivation of the full likelihood, but for the ignorability part, the assumption is that g does not depend upon θ.

It is possible to take a slightly different point of view and condition on the censoring variables instead. One arrives at essentially the same likelihood, but this time the dominating measure for the i'th observation becomes $m + \delta_{C_i}$ (δ_{C_i} is the Dirac measure in C_i).

The likelihood

From hereon the likelihood to consider is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f_\theta(T_i)^{e_i} S_\theta(T_i)^{1-e_i} \\ &= \prod_{i=1}^{n} \lambda_\theta(T_i)^{e_i} S_\theta(T_i) \end{aligned}$$

recalling the definition of the *hazard rate*

$$\lambda_\theta(t) = \frac{f_\theta(t)}{S_\theta(t)}.$$

The log-likelihood is

$$l(\theta) = \sum_{i=1}^{n} e_i \log \lambda_\theta(T_i) - \sum_{i=1}^{n} \Lambda_\theta(T_i)$$

recalling that the *cumulative hazard function* is

$$\Lambda_\theta(t) = -\log S_\theta(t).$$

MLE for the censored exponential

If the survival distribution is the exponential with parameter λ being the rate, the MLE is

$$\hat{\lambda} = \frac{n_u}{\sum_{i=1}^{n} T_i}$$

with n_u the number of failures/deaths (the number of uncensored observations).

- If we ignore censoring the MLE is

  $$\frac{n}{\sum_{i=1}^{n} T_i} > \hat{\lambda},$$

  which will overestimate the rate.

- If we discard censored observations the MLE is

  $$\frac{n_u}{\sum_{i=1}^{n} e_i T_i} > \hat{\lambda},$$

  which will overestimate the rate.

The derivation of the MLE is found in Harrell's book page 415.

2 Accelerated failure time models

The log-logistic distribution

If X has a logistic distribution the distribution function is

$$G(x) = \frac{e^{\lambda x}}{1 + e^{\lambda x}}$$

and $Y = e^X$ – the *log-logistic* distribution – has distribution function

$$F(y) = G(\log y) = \frac{y^\lambda}{1 + y^\lambda}$$

and density

$$f(y) = F'(y) = \frac{\lambda y^{\lambda - 1}}{(1 + y^\lambda)^2}.$$

We introduce a *scale* parameter as follows

$$f_\eta(y) = \frac{\lambda e^{-\lambda\eta} y^{\lambda - 1}}{(1 + e^{-\lambda\eta} y^\lambda)^2}.$$

Exercise

- Show that if Y has a log-logistic distribution with density $f = f_0$ then

$$e^{\eta} Y$$

  has density f_η.

- Show that if Y has a log-logistic distribution with density f_η then

$$\log Y - \eta$$

  has a logistic distribution.

Linear predictors

If $\eta = \sum_j x_j \beta_j$ is a *linear predictor* we assume that it affects the survival distribution as a *scale* transformation

$$e^{\eta} Y$$

of the baseline distribution of Y.

The λ parameter is a nuisance parameter that determines the shape of the baseline distribution.

The survival function is

$$S_\eta(t) = \frac{1}{1 + e^{\lambda\eta} y^\lambda}.$$

Accelerated failure time models

Definition 2. An AFT model has survival function given as

$$S_\eta(t) = \psi((\log t - \eta)/\sigma)$$

with η the linear predictor, ψ a survival function (on $\mathbb{R}$), and $\sigma > 0$ called the scale parameter.

For the log-logistic model the scale parameter is $\sigma = \lambda^{-1}$.

A unit change of x_j increases – or *accelerates* – the failure by a factor e^{β_j}.

3 Proportional hazards models

Proportional hazards models

Definition 3. The proportional hazards model has hazard rate

$$\lambda(t) = \lambda_0(t)e^{\eta}$$

with η the linear predictor and λ_0 the baseline hazard rate.

It follows that for the cumulative hazard function

$$\Lambda(t) = \Lambda_0(t)e^{\eta}$$

the proportionality holds too.

The factor e^{β_j} is the *hazard ratio* between two models corresponding to a unit change of x_j.

Weibull example

The Weibull baseline hazard rate and cumulative hazard function are

$$\lambda_0(t) = \gamma t^{\gamma-1} \qquad \Lambda_0(t) = t^{\gamma}.$$

The log-likelihood is

$$\begin{aligned} l &= \sum_{i=1}^{n} e_i \log(\gamma T_i^{\gamma-1} e^{\eta_i}) - T_i^{\gamma} e^{\eta_i} \\ &= \underbrace{\sum_{i=1}^{n} e_i \log(T_i^{\gamma} e^{\eta_i}) - T_i^{\gamma} e^{\eta_i}}_{\text{Poisson log-likelihood}} + \sum_{i=1}^{n} e_i \log(\gamma T_i^{-1}). \end{aligned}$$

This is (surprisingly) up to a constant the log-likelihood for a Poisson model of the e_i's with log link and mean value $T_i^{\gamma} e^{\eta_i}$ for fixed γ.

The glm-framework can be used to fit the model (for fixed γ) with the survival times entering as an offset term.

The α parameter has been dropped in the Weibull distribution above as it is captured by an intercept in the linear predictor.

Weibull example

To estimate γ we can use an iterative procedure or compute the profile likelihood using the glm-framework for the optimization for a range of γ-parameters.

For fixed linear predictor we find that

$$\partial_{\gamma} l = \sum_{i=1}^{n} (e_i - T_i^{\gamma} e^{\eta_i}) \log T_i + \frac{e_i}{\gamma}.$$

Thus γ solves the equation

$$\gamma = \frac{n_u}{\sum_{i=1}^{n}(T_i^{\gamma} e^{\eta_i} - e_i)\log T_i}$$

with n_u the number of uncensored observations.

There is no closed form solution to the equation. One idea is to use an iterative procedure and approximate the solution by

$$\gamma^{(k)} = \frac{n_u}{\sum_{i=1}^{n}(T_i^{\gamma^{(k-1)}} e^{\eta_i} - e_i)\log T_i}$$

and then reestimate the linear predictors before the next iteration.