Exercises and Lecture Notes, STK 4080, Autumn 2018

Version 0.17, 29-viii-2018

Nils Lid Hjort

Department of Mathematics, University of Oslo

Abstract

Exercises and Lecture Notes collected here are indeed for the Survival and Event History Analysis course STK 4080 / 9080, autumn semester 2018. The exercises will complement those given in the course book Aalen, Borgan, Gjessing, *Survival and Event History Analysis: A Process Point of View*, Springer, 2008.

1. Ancient Egyptian lifelengths

How long is a life? A unique set of lifelengths in Roman Egypt was collected by W. Spiegelberg in 1901 *(Ägyptische und griechische Eigennamen aus Mumienetiketten der römischen Kaiserzeit)* and analysed by (the very famous) Karl Pearson (1902) in the very first volume of (the very famous) Biometrika. The data set contains the age at death for 141 Egyptian mummies in the Roman period, 82 men and 59 women, dating from the last century b.C. The lifelengths vary from 1 to 96 years, and Pearson argued that these can be considered a random sample from one of the better-living classes in that society, at a time when a fairly stable and civil government was in existence (as we recall, the violent 'tax revolt' with ensuing long-lasting complications took place under Antoninus Pius later, in 139 AD). To access the data, go to `egypt-data` at the course website, reading them into your computer via

```
tt <- scan("egypt-data",skip=5)
```

Pearson did not attempt to fit any parametric models for these data, but discussed differences between the Egyptian age distribution and that of England 2000 years later. The purpose of the present exercise is to analyse aspects of the data by comparing the nonparametric survival curve (here a simplified version of the Kaplan–Meier curves, since there is no censoring; all the old Egyptians are dead) with a couple of parametric curves, in particular the Weibull.

(a) We start with the natural nonparametric estimate of the survival curve $S(t) = \Pr\{T \geq t\}$. Let the data be $t_1, \ldots, t_n$ (either the full set, or the subset for men, or that of the women). Since this is just a binomial probability, for each fixed t, we may put up the empirical survival function

$$S_{\rm emp}(t) = (1/n) \sum_{i=1}^{n} I\{t_i \geq t\} \quad \text{for } t > 0.$$

Show that $\mathrm{E}\, S_{\rm emp}(t) = S(t)$ and that $\mathrm{Var}\, S_{\rm emp}(t) = (1/n)S(t)\{1 - S(t)\}$.

(b) Compute the empirical survival curves, for men and for women, say $S_{m,\mathrm{emp}}(t)$ and $S_{w,\mathrm{emp}}(t)$, and display them in the same diagram, cf. Figure 0.1 below.

(c) Then consider the two-parameter Weibull model [note the Swedish pronunciation], which has a cumulative distribution of the form

$$F(t,a,b) = 1 - \exp\{-(at)^b\} \quad \text{for } t > 0,$$

with a and b positive parameters (typically unknown). (i) Find a formula for the median of the distribution. (ii) Show that the probability of surviving age t, given that one has survived up to t_0, is $\exp[-\{(at)^b - (at_0)^b\}]$, for $t > t_0$. (iii) Show that the density can be expressed as

$$f(t,a,b) = \exp\{-(at)^b\}a^b b t^{b-1} \quad \text{for } t > 0.$$

(d) Find formulae for the 0.20- and 0.80-quantiles, and set these equal to the observed 0.20- and 0.80-quantiles for the data. This yields two equations with two unknowns, which you can solve. In this fashion, find estimates $(\widetilde{a},\widetilde{b})$ for the men and for the women.

(e) While quantile fitting is a perfectly sensible estimation method, a more generally versatile method is that of maximum likelihood (ML), which will also be used later on in the course. By definition, the ML estimates $(\widehat{a},\widehat{b})$ are the parameter values maximising the log-likelihood function

$$\ell_n(a,b) = \sum_{i=1}^n \log f(t_i,a,b) = \sum_{i=1}^n \{-(at_i)^b + b\log a + \log b + (b-1)\log t_i\}.$$

This can be maximised numerically, as soon as you can programme the log-likelihood function. With data stored in your computer, called `tt`, try this, using R's powerful non-linear minimiser `nlm`:

```
logL <- function(para)
{
a <- para[1]
b <- para[2]
hei <- -(a*tt)^b + b*log(a) + log(b) + (b-1)*log(tt)
sum(hei)
}
# then:
minuslogL <- function(para)
{-logL(para)}
# then:
nils <- nlm(minuslogL2,c(0.20,1.00),hessian=T)
ML <- nils$estimate
```

It gives you the required ML estimates $(\widehat{a},\widehat{b})$. Carry out this estimation scheme, for the men and the women separately.

(f) I find $(0.0270, 1.3617)$ for the men and $(0.0347, 1.5457)$ for the women. Display the two estimated Weibull survival curves, perhaps along with the two nonparametric ones, as in my Figure 0.1 here. Compute the estimatead median lifelengths, for men and for women, and comment.

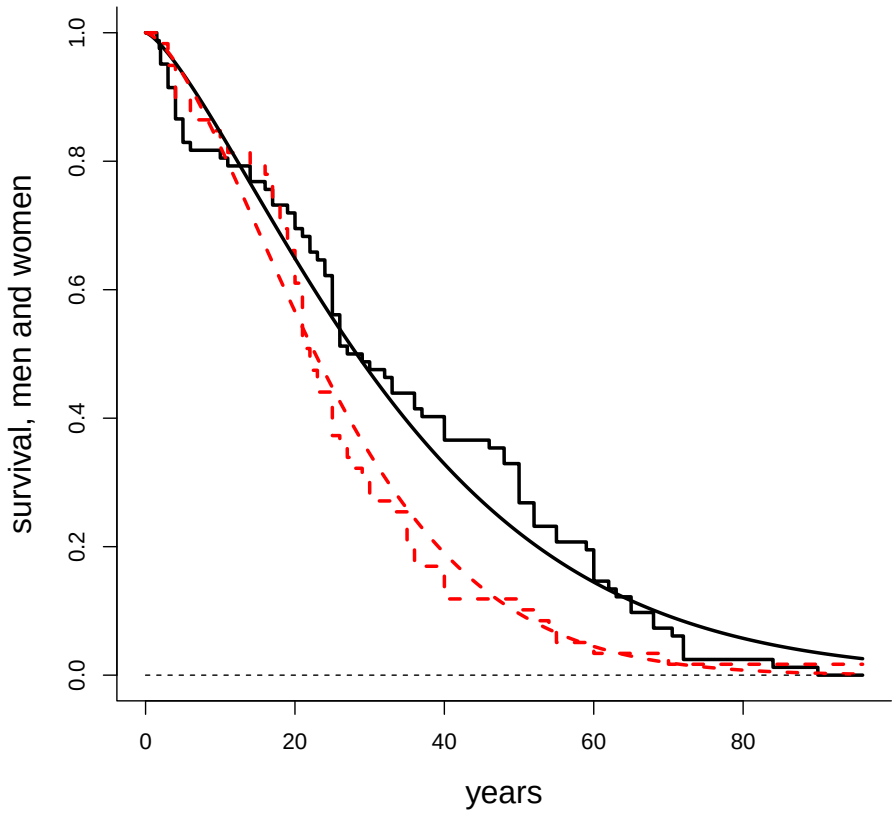

Figure 0.1: Survival curves from Roman era Egypt, two for men (black) and two for women (red). The step-functions are the empirical survival curves, type Kaplan–Meier; the continuous curves are the fitted Weibull curves.

(g) Compute and display also the estimated Weibull hazard rates, for men and for women. Comment on what you find.

(h) Considerations above invite statistical testing of the hypothesis H_0 that men and women of Roman era Egypt had the same lifelength distributions. Compute and display the 90% confidence bands

$$S_{m,\mathrm{emp}}(t) \pm 1.645\,\widehat{\tau}_m(t), \quad S_{w,\mathrm{emp}}(t) \pm 1.645\,\widehat{\tau}_w(t),$$

where

$$\begin{aligned}\widehat{\tau}_m(t)^2 &= (1/n_m)S_{m,\mathrm{emp}}(t)\{1-S_{m,\mathrm{emp}}(t)\},\\ \widehat{\tau}_w(t)^2 &= (1/n_w)S_{w,\mathrm{emp}}(t)\{1-S_{w,\mathrm{emp}}(t)\},\end{aligned}$$

the estimated variances. (We shall learn formal tests along such lines in the course.)

(i) Above I've forced you through the loops of things for one particular parametric model, namely the Weibull. Now do all these things for the Gamma(a,b) model too, with density $\{b^a/\Gamma(a)\}t^{a-1}\exp(-bt)$. Part of the point here is that this does not imply a doubling of your work efforts; you may edit your computer programmes, at low work cost, to accommodate

other parametric models, once you've been through one of them. The Weibull does a slightly better job than the Gamma, it turns out.

"Either man is constitutionally fitter to survive to-day [than two thousand years ago], or he is mentally fitter, i.e. better able to organise his civic surroundings. Both conclusions point perfectly definitely to an evolutionary progress." – *Karl Pearson*, 1902.

2. Did men live longer than women in Ancient Egypt?

As a follow-up to the Ancient Egypt analysis of Exercise 1, consider the following attempt to quantify more accurately the extent to which men and women had different lifelengths then.

(a) Plot the difference in survival function $D(t) = S_{m,\text{emp}}(t) - S_{w,\text{emp}}(t)$, and also the ratio funcion $S_{m,\text{emp}}(t)/S_{w,\text{emp}}(t)$. Comment on what these plots indicate.

(b) Find an expression for the variance $\kappa(t)^2$ of $D(t)$. Then construct and compute an empirical estimate, say $\widehat{\kappa}(t)$.

(c) Plot both $D(t)$ and the band $D(t) \pm 1.645\,\widehat{\kappa}(t)$. What is the interpretation of this band? What are your conclusions, regarding lifelengths in ancient Egypt? What are the likely reasons for differences you spot?

3. Survival functions and hazard rates

Consider a lifetime variable T with density f and cumulative distribution function F on the halfline (so, in particular, the distribution is continuous). Define the hazard rate function α as

$$\alpha(t)\,\mathrm{d}t = \Pr\{T \in [t, t+\mathrm{d}t] \,|\, T \geq t\},$$

for a small time window $[t, t+\mathrm{d}t]$; more formally,

$$\alpha(t) = \lim_{\varepsilon \to 0} (1/\varepsilon) \Pr\{T \in [t, t+\varepsilon] \,|\, T \geq t\}.$$

(a) First define the survival function as

$$S(t) = \Pr\{T \geq t\} = 1 - F(t).$$

What are its basic properties?

(b) Show that in fact $\alpha(t) = f(t)/S(t)$. So from knowledge of f we can always find the hazard rate from $\alpha = f/(1-F)$.

(c) Define also the cumulative hazard rate function as $A(t) = \int_0^t \alpha(s)\,\mathrm{d}s$. Show that

$$F(t) = 1 - \exp\{-A(t)\} \quad \text{and} \quad f(t) = \alpha(t) \exp\{-A(t)\}.$$

(d) Let T have the exponential distribution with density $f(t, \theta) = \theta \exp(-\theta t)$. Find its survival function and hazard rate.

(e) For the Weibull distribution, with $F(t) = 1 - \exp\{-(at)^b\}$, with the hazard rate function, and display it in a plot, for $a = 3.33$ and b equal to 0.9, 1.0, 1.1.

(f) Consider the Gamma distribution with parameters (a, b), which has the density

$$f(t, a, b) = \frac{b^a}{\Gamma(a)} t^{a-1} \exp(-bt) \quad \text{for } t > 0.$$

Show that the mean and variance are a/b and a/b^2. Take $b = 2.22$, compute the hazard rates for a equal to 0.8, 1.0, 1.2, and display these in a diagram. Give explicit formulae for the survival function and hazard rate for the case of $a = 2$.

(g) Consider a lifetime distribution with hazard rate $\alpha(t) = 1/(1+t)$. Find its survival function and density.

4. Maximum likelihood estimation with censored data

If we have observed independent lifetime data $t_1, \ldots, t_n$, from a suitable parametric density $f(t, \theta)$, the ML estimator is found by maximising the log-likelihood function $\sum_{i=1}^n \log f(t_i, \theta)$. This exercise looks into the required amendments in the case of censored data, say $(t_1, \delta_1), \ldots, (t_n, \delta_n)$, with

$$\delta_i = I\{t_i \text{ is the observed lifetime}\} = \begin{cases} 1 & \text{if } t_i \text{ is the observed lifelength,} \\ 0 & \text{if } t_i \text{ is the censored value.} \end{cases}$$

So, in the case of $\delta_i = 0$, this means that the real lifetime, say t_i^0, is at least as large as t_i, but we do not know more than that.

(a) Assume that the parametric model is given and perhaps primarily thought about via its hazard rate function, say $\alpha(t, \theta)$. Assume first that all t_i correspond to genuinely observed lifetimes, i.e. that there is no censoring. Show that the log-likelihood function above can be expressed as

$$\ell_n(\theta) = \sum_{i=1}^n \{\log \alpha(t_i, \theta) - A(t_i, \theta)\},$$

with $A(t, \theta)$ the cumulative hazard function for the model.

(b) For the very simple exponential model, with $\alpha(t, \theta) = \theta$, write up the log-likelihood function from the exression under (a), and show that the ML estimator is $\widehat{\theta} = n/\sum_{i=1}^n t_i = 1/\bar{t}$.

(c) Then consider the general case with censoring, i.e. some of the δ_i are equal to zero. Show that the log-likelihood function can be written

$$\ell_n(\theta) = \sum_{\delta_i=1} \log f(t_i, \theta) + \sum_{\delta_i=0} \log S(t_i, \theta) = \sum_{i=1}^n \{\delta_i \log \alpha(t_i, \theta) - A(t_i, \theta)\}.$$

Sometimes the first expression is more practical to work with, sometimes the second; also, as will be seen later, the second expression lends itself more easily to general counting process models.

(d) For the simple exponential model, again, but now with cenoring, show that the ML estimator is $\widehat{\theta} = \sum_{i=1}^n \delta_i / \sum_{i=1}^n t_i$, generalising the non-censored case above.

5. Counting process, at-risk process, intensity process, martingale

The eternal golden braid, for modelling and analysing survival and event history data, is the quadruple (N, Y, λ, M)! The ingredients are the counting process N, the at-risk process Y, the intensity process λ, and the martingale M. These matters are of course tended to in the ABG book, in several chapters, with various setups, specialisations, and generalisations. This particular exercise gives a separate, brief, and partial introduction to these four items, in the context of survival data. These are of the form $(t_1, \delta_1), \ldots, (t_n, \delta_n)$, as for Exercise 4. The distribution is continuous, so there are no ties among the t_i.

(a) The counting process N counts the number of observed events, over time:

$$N(t) = \#\{\text{observed events over } [0,t]\} = \sum_{i=1}^{n} I\{t_i \le t, \delta_i = 1\}.$$

It starts out at zero, at time zero, and then increases with jump size 1 each time a new observation is recorded.

(b) The at-risk process counts those individuals who are still at tisk, for each given time point:

$$Y(t) = \#\{\text{individuals at risk just before time } t\} = \sum_{i=1}^{n} I\{t_i \ge t\}.$$

The 'just before' thing can be formalised, e.g. via left continuity. The point is that an individual belongint to the risk set at time t, with this definition, can have his or her event in the time window $[t, t+\varepsilon]$. Note that $Y(t)$ counts both those t_i with $\delta_i = 1$ and those with $\delta_i = 0$ (since we do not know yet when events occur, or when censoring might occur).

(c) The intensity process $\lambda(s)$ can be defined in severeal ways, cf. the ABG book, and with somewhat different, but related, motivations and interpretations. The simplest way, in this framework, might be

$$\lambda(s)\,\mathrm{d}s = \Pr\{N[s, s+\mathrm{d}s] = 1 \,|\, \mathcal{F}_{s-}\}.$$

First,

$$\mathrm{d}N(s) = N[s, s+\mathrm{d}s] = N(s+\mathrm{d}s) - N(s-)$$

is the numer of observed events inside the small time window $[s, s+\mathrm{d}s]$. Second, $\mathcal{F}_{s-}$ is the full history of everything that has been observed up to just before time s, i.e. over $[0, s)$. In the present setup of survival data (i.e. without complications of more complex event history constructions), the relevant information in all of $\mathcal{F}_{s-}$ is simply 'how many are still at risk', i.e. $Y(s)$.

(d) In this setup, show that

$$\mathrm{d}N(s) \,|\, \mathcal{F}_{s-} \sim \mathrm{Bin}(Y(s), \alpha(s)\,\mathrm{d}s),$$

a simple binomial situation with $Y(s)$ at risk and with a small probability $\alpha(s)\,\mathrm{d}s$. Show that

$$\begin{aligned}
\Pr\{\mathrm{d}N(s) = 0 \,|\, \mathcal{F}_{s-}\} &= 1 - Y(s)\alpha(s)\,\mathrm{d}s + O((\mathrm{d}s)^2), \\
\Pr\{\mathrm{d}N(s) = 1 \,|\, \mathcal{F}_{s-}\} &= Y(s)\alpha(s)\,\mathrm{d}s + O((\mathrm{d}s)^2), \\
\Pr\{\mathrm{d}N(s) \ge 2 \,|\, \mathcal{F}_{s-}\} &= O((\mathrm{d}s)^2),
\end{aligned}$$

with order notation $g(\varepsilon) = O(\varepsilon^2)$ meaning that $g(\varepsilon)$ is of order ε^2 (more precisely, is not of a bigger order), defined as $g(\varepsilon)/\varepsilon^2$ remaining bounded as $\varepsilon \to 0$. The above means that all

the action here is in 0 (high chance) and 1 (slim chance, but important, and sooner or later it will kick in). In particular, show from (c) that

$$\lambda(s) = Y(s)\alpha(s).$$

This is a special case of Aalen's multiplicative intensity model (stemming from his Berkeley PhD thesis 1975, then from his Annals of Statistics paper 1978, and further discussed and used and generalised in dozens of books and a few hundreds of journal articles, etc.).

(e) Then consider the random process

$$M(t) = N(t) - \int_0^t \lambda(s)\,\mathrm{d}s = N(t) - \int_0^t Y(s)\alpha(s)\,\mathrm{d}s.$$

Demonstrate that it has the magical martingale property,

$$\mathrm{E}\,\{\mathrm{d}M(s)\,|\,\mathcal{F}_{s-}\} = 0,$$

with $\mathrm{d}M(s) = M(s+\mathrm{d}s) - M(s)$ the martingale increment.

(f) Show that the process

$$K(t) = M(t)^2 - \int_0^t Y(s)\alpha(s)\,\mathrm{d}s$$

also is a martingale. We shall see later, in the course and in exercises, that various central properties flow from these martingales, including results on limiting normality for classes of estimators.

(g) Consider again the situation of Exercise 4, with log-likelihood functions for censored data. With the golden quadruple on board, show that the log-likelihood function also can be expressed as

$$\ell_n(\theta) = \sum_{i=1}^n \{\delta_i \log\alpha(t_i,\theta) - A(t_i,\theta)\} = \int_0^\tau \{\log\alpha(s,\theta)\,\mathrm{d}N(s) - Y(s)\alpha(s,\theta)\,\mathrm{d}s\}.$$

Here the integral of a function with respect to a counting process is defined, simply, as

$$\int_0^\tau g(s)\,\mathrm{d}N(s) = \sum_{i=1}^n g(t_i)\delta_i,$$

a sum of the function evaluated precisely at the observed timelengths.

6. A parametric step-function for the hazard rate

Consider independent lifetime data of the form $(t_1,\delta_1),\ldots,(t_n,\delta_n)$, as met with in Exercises 4 and 5, and assume they stem from a common distribution with hazard rate $\alpha(s)$. We wish to estimate the cumulative hazard rate $A(t) = \int_0^t \alpha(s)\,\mathrm{d}s$. This can famously be done using the Nelson–Aalen estimator, see the following exercise, but I start out working through a parametric version, via a step-function. When the windows become small, this parametric estimator will actually converge to the Nelson–Aalen.

(a) Consider the parametric model where the hazard rate is a step-function, i.e. constant over windows. Suppose $[0, \tau]$ is the full time window of relevance (i.e. with a sufficiently big end-point τ), with windows $W_j = [s_{j-1}, s_j)$ for $j = 1, \ldots, k$, and $0 = s_0 < s_1 < \cdots < s_k = \tau$. The model is then of the form

$$\alpha(s) = \alpha_j \text{ on time window } W_j \quad \text{for } j = 1, \ldots, k.$$

Using the log-likelihood expression of Exercise 5, show that the log-likelihood function can be written as

$$\ell_n(\alpha_1, \ldots, \alpha_k) = \sum_{j=1}^{k} \int_{W_j} \{\log \alpha(s)\, \mathrm{d}N(s) - Y(s)\alpha(s)\, \mathrm{d}s\} = \sum_{j=1}^{k} (\Delta N_j \log \alpha_j - R_j \alpha_j),$$

in which

$$\Delta N_j = N(W_j) = N(s_j) - N(s_{j-1}) \quad \text{and} \quad R_j = \int_{W_j} Y(s)\, \mathrm{d}s.$$

So ΔN_j is the number of observed events, and R_j is the 'total time at risk', over window W_j.

(b) Show the ML estimators for the local constants become

$$\widehat{\alpha}_j = \frac{\Delta N_j}{R_j} = \frac{\Delta N_j}{\int_{W_j} Y(s)\, \mathrm{d}s}$$

for $j = 1, \ldots, k$. The ML estimator of the full cumulative hazard rate is hence the integral of the estimated step-function, which becomes the piecewise linear

$$\widehat{A}(t) = \begin{cases} \widehat{\alpha}_1 t & \text{for } t \in W_1, \\ \widehat{\alpha}_1 s_1 + \widehat{\alpha}_2 (t - s_1) & \text{for } t \in W_2, \\ \widehat{\alpha}_1 s_1 + \widehat{\alpha}_2 (s_2 - s_1) + \widehat{\alpha}_3 (t - s_2) & \text{for } t \in W_3, \\ \text{etc.} \end{cases}$$

(c) Find the Hessian matrix

$$J(\alpha) = -\partial^2 \ell_n(\alpha_1, \ldots, \alpha_k) / \partial\alpha \partial\alpha^{\mathrm{t}},$$

and show that it is diagonal. I write α for the full parameter vector $(\alpha_1, \ldots, \alpha_k)$ where convenient. Find also 'the observed information', which is this minus the second order derivative matrix computed at the ML. Show indeed that

$$\widehat{J} = J(\widehat{\alpha}) = \mathrm{diag}(R_1^2/\Delta N_1, \ldots, R_k^2/\Delta N_k).$$

Large-sample theory, dealt with later in the course and in ABG Ch. 5, says that

$$\widehat{\alpha} \approx_d \mathrm{N}_k(\alpha, \widehat{J}^{-1}).$$

Show that this translates to the $\widehat{\alpha}_j$ being approximately unbiased, normal, and independent, with

$$\mathrm{Var}\, \widehat{\alpha}_j \doteq \Delta N_j / R_j^2.$$

(d) Argue, perhaps heuristically, that when the time windows become carefully small, then the above $\widehat{A}(t)$ is in effect a nonparametric estimator of the cumulative hazard function, approximately unbiased and normal, and with variance estimable by

$$\widehat{\kappa}(t)^2 = \sum_{\text{windows left of } t} \frac{\Delta N_j}{(R_j/d_j)^2},$$

with $d_j = s_j - s_{j-1}$ the width of window W_j. Since $R_j/d_j = (1/d_j)\int_{W_j} Y(s)\,\mathrm{d}s$, this is close to $Y(s_j^*)$, with s_j^* the mid-point of W_j. A further approximation, which becomes correct in a fine-tuned large-sample setup with cells becoming small at the right rate, is then

$$\widehat{\kappa}(t)^2 = \int_0^t \frac{\mathrm{d}N(s)}{Y(s)^2}.$$

Via the step-function model, and some extra analysis, we have essentially reinvented the Nelson–Aalen estimator, along with its properties; see Hermansen and Hjort (2015).

7. The Nelson–Aalen estimator

Consider again independent lifetime data of the form $(t_1, \delta_1), \ldots, (t_n, \delta_n)$, as met with in Exercises 4, 5, 6, and assume they stem from a common distribution with hazard rate $\alpha(s)$. In the previous exercise I set up a step-function model for $\alpha(s)$, which led to an almost nonparametric estimator for the cumulative $A(t) = \int_0^t \alpha(s)\,\mathrm{d}s$. The canonical nonparametric estimator is indeed this fine-tuned limit, namely the Nelson–Aalen estimator.

(a) We start with the definition, using the counting process and at-risk process notation of Exercise 5 (and used in the book). The Nelson–Aalen estimator for $A(t)$ is

$$\widehat{A}(t) = \int_0^t \frac{\mathrm{d}N(s)}{Y(s)} = \sum_{t_i \le t} \frac{\delta_i}{Y(t_i)}.$$

Again, the integral with respect to a counting process is simply the finite sum over the appropriate integrand, over the observed event times. It is easy to make a programme computing $\widehat{A}(t)$. Do this, for a dataset of your choice, perhaps simulated. Often one is content to compute and plot $\widehat{A}(t)$ just at the observed values t_i, in which case a simpler programme than the one below can be put up, but in various contexts it is useful to compute, plot, compare for a full fine grid of values, say, as here. This little programme requires that the (t_i, δ_i) are predefined as `tt` and `delta`.

```
eps <- 0.001
tval <- seq(0,20,by=eps)
Yval <- 0*tval
DeltaNval <- 0*tval
# then:
for (j in 1:length(tval))
{
tnow <- tval[j]
Yval[j] <- sum(1*(tt >= tnow))
ok <- 1*(tt >= tnow)*(tt < tnow+eps)*delta
```

```
DeltaNval[j] <- sum(ok)
}
# then:
jumps <- DeltaNval/Yval
Ahat <- cumsum(jumps)
matplot(tval,Ahat,type="l",xlab="time",ylab="look at Nelson-Aalen")
```

(b) Then we ought to spend a few minutes thinking about why the Nelson–Aalen $\widehat{A}(t)$ is a natural estimator of $A(t)$. Using the martingale M of Exercise 5, we may write

$$\mathrm{d}N(s) = Y(s)\alpha(s)\,\mathrm{d}s + \mathrm{d}M(s) = \text{structure} + \text{random fluctuations},$$

which implies

$$\mathrm{d}N(s)/Y(s) = \alpha(s)\,\mathrm{d}s + \text{noise}.$$

Argue that this points to the Nelson–Aalen.

(c) With a bit of heuristics, we have

$$\widehat{A}(t) - A(t) = \int_0^t \frac{\mathrm{d}N(s)}{Y(s)} - A(t) \doteq \int_0^t \frac{\mathrm{d}M(s)}{Y(s)} = \int_0^t \frac{1}{\widehat{y}(s)}\frac{\mathrm{d}M(s)}{n},$$

where $\widehat{y}(s) = Y(s)/n$ is a steadily more precise estimate of its limit in probability, say $y(s) = \Pr\{T \geq s, C \geq s\}$, with C the censoring mechanism. It follows that

$$\sqrt{n}\{\widehat{A}(t) - A(t)\} \doteq \int_0^t \frac{1}{\widehat{y}(s)}\frac{\mathrm{d}M(s)}{\sqrt{n}}.$$

But $M(t)/\sqrt{n} \to_d V(t)$, say, a zero-mean Gaussian martingale with incremental variances $\mathrm{Var}\,\mathrm{d}V(s) = y(s)\alpha(s)\,\mathrm{d}s$, by results in ABK (Chs. 4, 5). This, at least heuristically, is seen to imply

$$Z_n(t) \to_d Z(t) = \int_0^t \frac{1}{y(s)}\mathrm{d}V(s),$$

which is another zero-mean Gaussian martingale with incremental variances

$$\mathrm{Var}\,\mathrm{d}Z(s) = \mathrm{Var}\,\frac{\mathrm{d}V(s)}{y(s)} = \frac{\alpha(s)\,\mathrm{d}s}{y(s)}.$$

(d) So the Nelson–Aalen is for large samples approximately unbiased, approximately normal, and with variance

$$\sigma(t)^2 = \mathrm{Var}\,\widehat{A}(t) \doteq \frac{1}{n}\int_0^t \frac{\alpha(s)\,\mathrm{d}s}{y(s)}.$$

Give arguments supporting the estimator

$$\widehat{\sigma}(t)^2 = \int_0^t \frac{\mathrm{d}N(s)}{Y(s)^2}.$$

So a programme for the Nelson–Aalen just needs a few lines more to produce also $\widehat{\sigma}(t)$. In particular, confidence bands are now easy to construct, say

$$\widehat{A}(t) \pm 1.645\,\widehat{\sigma}(t) \quad \text{for } t \in [0, \tau],$$

where $[0, \tau]$ is a relevant time window for the data. Try to show that this band contains the true $A(t)$ with probability converging to 0.90, for each fixed t.

8. IUD expulsion

Data have been collected for IUD use for $n = 100$ women (I believe they stem from a Stanford PhD 1975, with data later on forwarded to and worked with by Aalen, then to Borgan and myself). The `iud-data` file has three columns: the index $i = 1, \ldots, n$; the time t_i to 'event', measured in days, from the first day of use; and an index for 'event', from 1 (she's pregnant!, which however does not happen here), to 2 (expulsion), to 3 and 4 (removal for pains, or bleeding, or other medical reasons), to yet other categories 5, 6, 7, 8, 9 of less interest here.

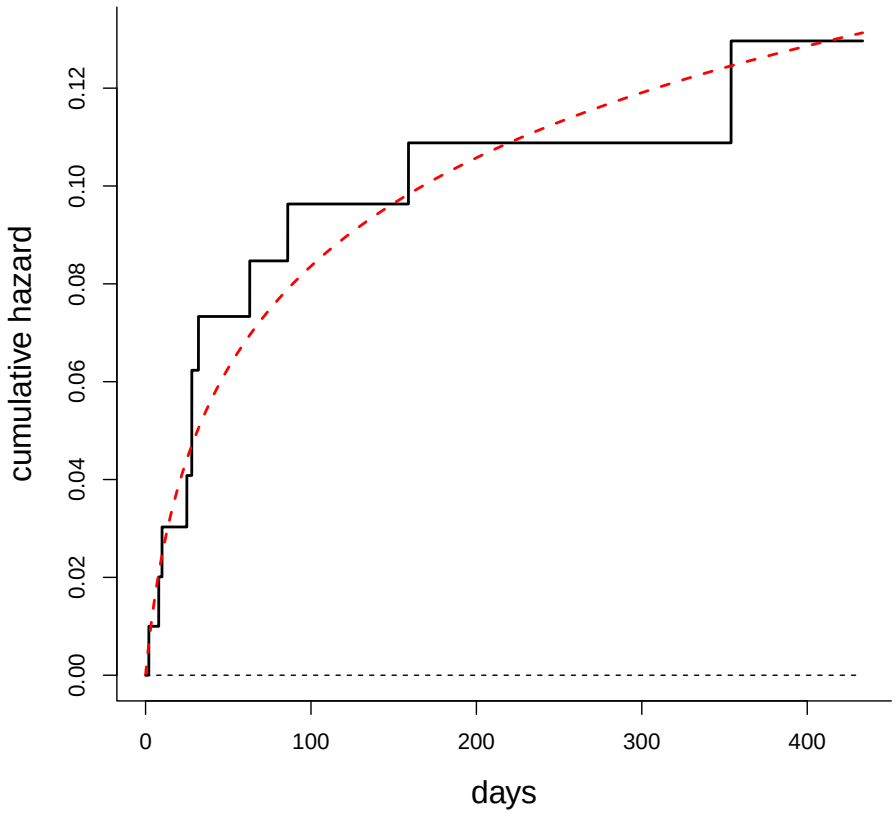

Figure 0.2: Estimated cumulative hazard rate for time to expulsion of IUD, via the nonparametric Nelson–Aalen estimator and the parametric frailty model.

(a) Fit first the simple model that takes the hazard rate to be a constant θ. Under this model, what is the estimated median time to expulsion, for women using IUD (supposing they do not quit on their own)? (I have no idea whether these 1975 IUD data would look very differently now.) Compute also $\ell_{n,0,\max} = \ell_{n,0}(\widehat{\theta})$, the attained log-likelihood maximum for that model.

(b) Then assume that each woman has an exponential IUD expulsion time, say θ, but that this parameter varies from woman to woman, according to a Gamma distribution (a, b). Show that the survival function in the population then becomes

$$S(t) = \Pr\{\text{IUD still in place at time } t\} = \frac{1}{(1 + t/b)^a} = \exp\{-a \log(1 + t/b)\}.$$

(c) Show that the ensuing hazard rate function becomes

$$\alpha(t) = \frac{a/b}{1+t/b} = \frac{\theta_0}{1+t/b},$$

writing for emphasis $\theta_0 = a/b$ for the mean value of the Gamma distribution of the women's random intensities. If b is large, the variance of the random θ is small, and we're back to the simpler model with a common θ_0 for all IUD users.

(d) Fit the expulsion data to this two-parameter model. Produce a version of Figure 0.2, with both the parametric and nonparametric Nelson–Aalen estimates. Does the model appear to fit? Under this two-parameter model, what is the estimated median time until expulsion (again, assuming the woman does not quit on her own)? Compute also $\ell_{n,\max} = \ell_n(\widehat{a}, \widehat{b})$, the attained log-likelihood maximum for this model, and compare to the corresponding number for the simpler model.

(e) In addition to producing a version of Figure 0.2, pertaining to cumulative hazard, make a simular figure for the estimated survival functions (parametric and nonparametric), i.e. the probability that the IUD is not yet expulsed.

References

Aalen, O.O. (1975). Statistical Inference for a Family of Counting Processes. PhD thesis, Department of Statistics, University of Berkeley, California.

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics* **6**, 701–726.

Aalen, O.O., Borgan, Ø., and Gjessing, H.K. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer.

Andersen, P.K., Borgan, Ø., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics* **11**, 1–16.

Claeskens, G. and Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.

Cunen, C., Hjort, N.L., and Nygård, H. (2018). Statistical sightings of better angels. Submitted for publication.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge.

Hermansen, G.H. and Hjort, N.L. (2015). Bernshteĭn–von Mises theorems for nonparametric function analysis via locally constant modelling: A unified approach. *Journal of Statistical Planning and Inference* **166**, 138–157.

Hjort, N.L. (1985). Discussion contribution to P.K. Andersen and Ø. Borgan's 'Counting process models for life history data: A review'. *Scandinavian Journal of Statistics* **12**, 141–151.

Hjort, N.L. (1985). An informative Bayesian bootstrap. Technical Report, Department of Statistics, Stanford University.

Hjort, N.L. (1986). Discussion contribution to P. Diaconis and D. Freedman's paper 'On the consistency of Bayes estimators'. *Annals of Statistics* **14**, 49–55.

Hjort, N.L. (1986). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scandinavian Journal of Statistics* **13**, 63–75.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Annals of Statistics* **18**, 1259–1294.

Hjort, N.L. (1991). Bayesian and empirical Bayesian bootstrapping. Statistical Research Report, Department of Mathematics, University of Oslo.

Hjort, N.L. (2003). Topics in nonparametric Bayesian statistics [with discussion]. In *Highly Structured Stochastic Systems* (eds. P.J. Green, N.L. Hjort, S. Richardson). Oxford University Press, Oxford.

Hjort, N.L. (2018). Towards a More Peaceful World [Insert '!' or '?' Here]. FocuStat Blog Post.

Hjort, N.L., Holmes, C.C., Müller, P., and Walker, S.G. (2010). *Bayesian Nonparametrics.* Cambridge University Press, Cambridge.

Hjort, N.L. and Petrone, S. Nonparametric quantile inference using Dirichlet processes. In *Festschrift for Kjell Doksum* (ed. V. Nair).

Pearson, K. (1902). On the change in expectation of life in man during a period of circa 2000 years. *Biometrika* **1**, 261–264.

Schweder, T. and Hjort, N.L. (2016). *Confidence, Likelihood, Probability.* Cambridge University Press, Cambridge.