Extra exercises in STK4080 2019

Solution to Exercise E1.1

Let $N(t)$ be a nonhomogeneous Poisson process (NHPP) with intensity function $\alpha(t)$, which means that, when defining $A(t) = \int_0^t \alpha(s)ds$,

- $N(t) - N(s) \sim$ Poisson$(A(t) - A(s))$ when $s < t$
- $N(t) - N(s)$ is independent of $\mathcal{F}_s$ when $s < t$

Suppose n NHPP-processes $N_1(t), \dots, N_n(t)$ as above have been observed, on time intervals, respectively, $[0, \tau_i]$ for $i = 1, \dots, n$, where the processes share the same $\alpha(t)$.

a) What is the 'at risk' indicator $Y_i(t)$ for the ith process?

Write down the intensity function $\lambda_i(t)$ for the ith process.

Solution: $Y_i(t) = I(\tau_i \geq t)$; $\lambda_i(t) = \alpha(t)Y_i(t)$

b) Show that a multiplicative intensity model for $\alpha(t)$ results from this.

Write down an expression for the Nelson-Aalen estimator for $A(t)$.

Solution: Assuming the processes do not jump at the same time, we observe the counting process

$$N(t) = \sum_{i=1}^{n} N_i(t)$$

which has intensity

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t) = \alpha(t) \sum_{i=1}^{n} Y_i(t) \equiv \alpha(t)Y(t)$$

The Nelson-Aalen estimator is hence

$$\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)} = \sum_{T_j \leq t} \frac{1}{Y(T_j)}$$

c) Suppose $n = 3$, $\tau_1 = 20, \tau_2 = 30, \tau_3 = 10$, and that the observed events for the three processes are given as

i=1: 5, 12, 17

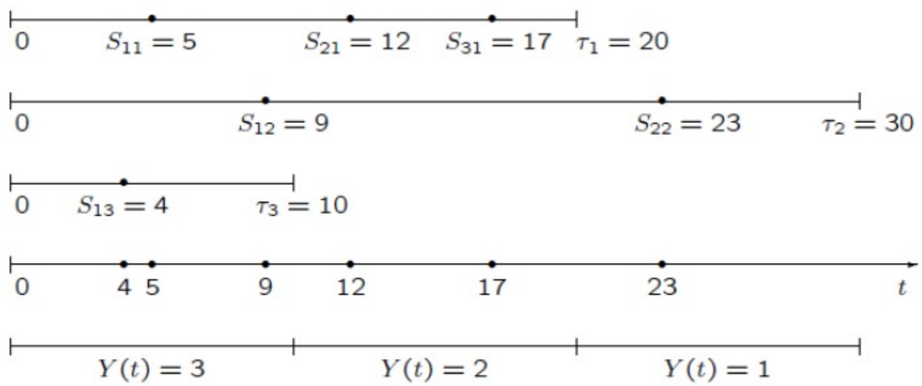
i=2: 9, 23

i=3: 4

Calculate the Nelson-Aalen estimator for $A(t)$ and draw a graph (by hand!).

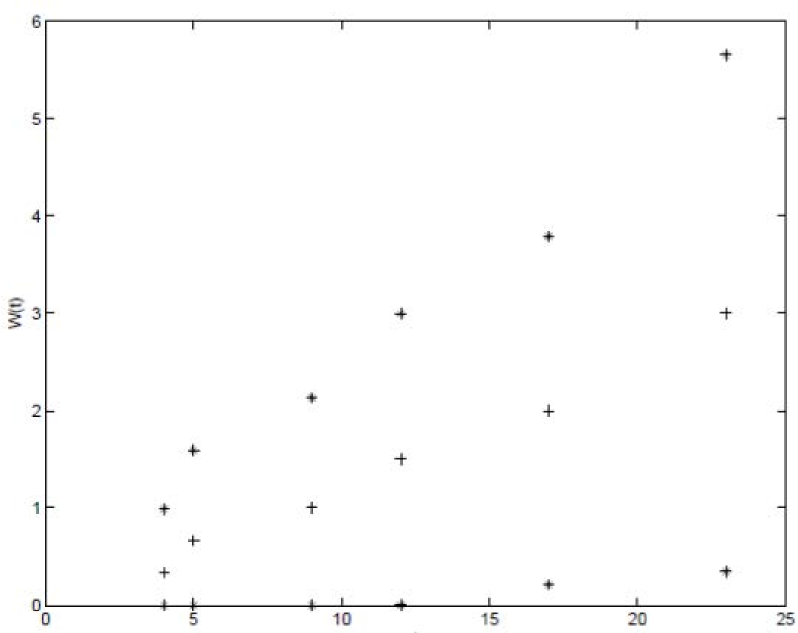
Calculate also the variance of the estimator and find pointwise 95% confidence bounds for the Nelson-Aalen curve.

Solution:



t	$1/Y(t)$	$1/Y(t)^2$	$\hat{A}(t)$	$\widehat{Var\hat{A}}(t)$	$\widehat{SD\hat{A}}(t)$
4	1/3	1/9	1/3	1/9	0.3333
5	1/3	1/9	2/3	2/9	0.4714
9	1/3	1/9	1	1/3	0.5774
12	1/2	1/4	3/2	7/12	0.7638
17	1/2	1/4	2	5/6	0.9129
23	1	1	3	11/6	1.3540

Estimated $A(t)$ with 95% confidence limits:



How can you do this in R? (??)

d) Formulate a result on the asymptotic behavior of the estimator as the number n of processes tends to infinity. Consider in particular the case when $\tau_i \equiv \tau$ for all i.

Solution: First, from Slides 7:

Assume that there is a deterministic positive function $y(t)$ such that $Y(t)/n \to y(t) > 0$ in probability. Then

$$\sqrt{n}(\hat{A}(t) - A^*(t)) \text{ and } \sqrt{n}(\hat{A}(t) - A(t))$$

converge in distribution to the mean zero Gaussian martingale $U(t) = W(V(t))$ with predictable variation process

$$V(t) = \int_0^t v(s)ds = \int_0^t \frac{\alpha(s)}{y(s)} ds \qquad (*)$$

In the present application this leads to assuming

$$\frac{Y(t)}{n} = \frac{\sum_{i=1}^n I(\tau_i \geq t)}{n} \to y(t)$$

In particular, if all τ_i are equal to τ, we have $y(t) = 1$ for all $t \leq \tau$. In that case it is seen from (*) that $V(t) = A(t)$, so

$$\sqrt{n}(\hat{A}(t) - A(t)) \to W(A(t))$$

as functions of t. Thus, for a fixed t,

$$\sqrt{n}(\hat{A}(t) - A(t)) \to N(0, A(t))$$

This result corresponds well to the following consideration:

When $\tau_i \equiv \tau$, since $Y_i(t) = n$ for all $t \leq \tau$, we have

$$\hat{A}(t) = \sum_{T_j \leq t} \frac{1}{Y(T_j)} = \frac{1}{n} N(t) \qquad (**)$$

Now $N(t) \sim Poisson(nA(t))$, so $E(N(t)) = nA(t)$, $Var(N(t)) = nA(t)$. Thus the $\hat{A}(t)$ in (**) is an unbiased estimator with variance exactly $A(t)/n$.

Exercise E1.2

This exercise is based on the book ASAUR. In addition to the R-package 'survival' you will here need the package 'asaur' which contains several data sets from this book.

Example 1.2 p. 6 in ASAUR introduces the data "gastricXelox".

a) For the above data, use R to determine how many patients, N, had the event (death or progression). Then compute the total observation time for all patients, R, and consider the ratio

$$\frac{N}{R} = \frac{\text{“occurrences”}}{\text{“exposure”}}$$

which in Exercise 1.2 p. 11 in ASAUR is called *the event rate per person-week*. What is the interpretation of this quantity?

Solution:

```
> N=sum(gastricXelox$delta)
> N
[1] 32
> R=sum(gastricXelox$timeWeeks)
> R
[1] 2866
> N/R
[1] 0.01116539
```

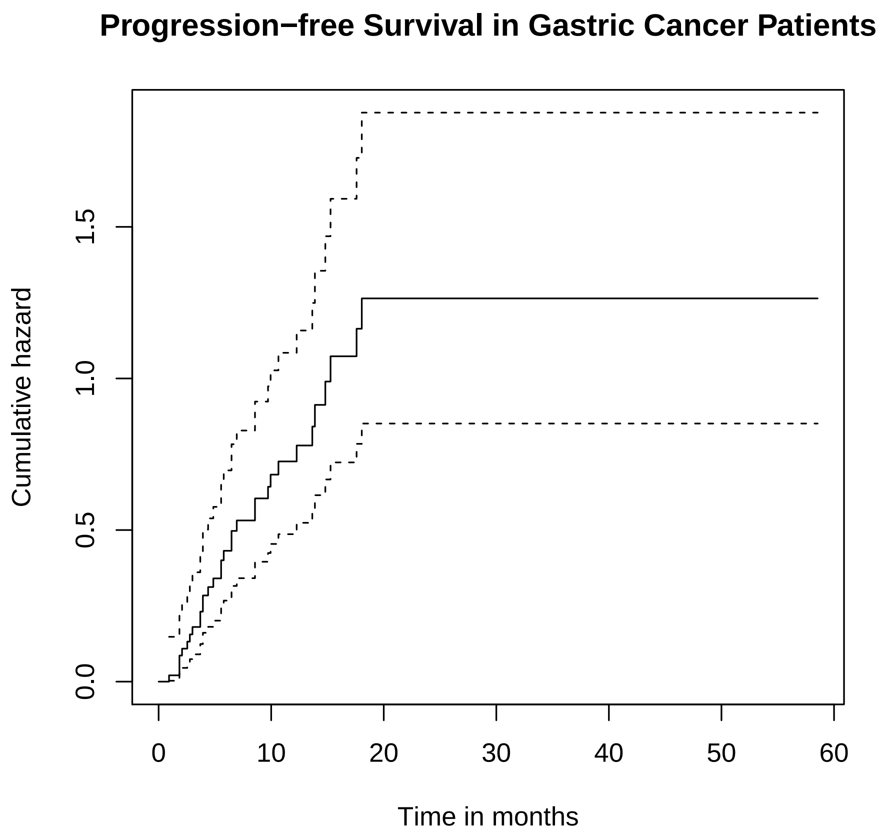
b) Run the R-code on page 30 in ASAUR where the above data are analyzed by the Kaplan-Meier estimator. Give an interpretation of what you see from the plot.

Solution: Plot is given on p. 30 in ASAUR. The survival curve decreases rather sharply in the beginning, but seems to flatten out after approximately 18 months. A possible explanation is that the patients that have reached say, 20 months, will survive for the nearest years (follow-up time here is approximately 5 years).

c) Do the corresponding analysis using the Nelson-Aalen estimator. Give an interpretation of what you see from the plot. What is the connection with the quantity calculated in question (a)?

Solution: The Nelson-Aalen plot is not used in ASAUR (it is used only for providing another estimator of $S(t)$ as an alternative to the Kaplan-Meier estimator.

```
> timeMonths <- gastricXelox$timeWeeks*7/30.25
> delta <- gastricXelox$delta
> result.na <- survfit(coxph(Surv(timeMonths, delta) ~ 1))
> plot(result.na, fun="cumhaz",xlab="Time in months", ylab="Cumulative hazard")
> title("Progression-free Survival in Gastric Cancer Patients")
```



The hazard rate (derivative of $A(t)$) seems to be fairly constant up to 18 months. The observed rate is then approximately $1.2/18 \approx 0.067$ per month, while in (a) above we got approximately 0.011 per week or 0.048 per month. The difference can be explained from the figure, since there are no events after 20 months.