Exercise 3.3

a) Let g be a strictly increasing continuously differentiable function. Then we have the Taylor series expansion:

$$g\big(\widehat{A}(t)\big) \approx g\big(A(t)\big) + g'\big(A(t)\big)\big(\widehat{A}(t) - A(t)\big)$$

It follows that the distribution of $g\big(\widehat{A}(t)\big)$ is approximately the same as the distribution of

$$g\big(A(t)\big) + g'\big(A(t)\big)\big(\widehat{A}(t) - A(t)\big) \tag{E.1}$$

If we use that $\widehat{A}(t)$ is approximately normally distributed with mean $A(t)$, we see that (E.1) is approximately normally distributed with mean $g\big(A(t)\big)$ and variance $\big\{g'\big(A(t)\big)\big\}^2\,\mathrm{Var}\widehat{A}(t)$. The latter is estimated by $\big\{g'\big(\widehat{A}(t)\big)\big\}^2\,\widehat{\sigma}^2(t)$. This shows the result.

b) By the result in question a we have that

$$\frac{g\big(\widehat{A}(t)\big) - g\big(A(t)\big)}{g'\big(\widehat{A}(t)\big)\widehat{\sigma}(t)}$$

is approximately standard normally distributed. It then follows by a standard argument that

$$g\big(\widehat{A}(t)\big) \pm z_{1-\alpha/2}\; g'\big(\widehat{A}(t)\big)\,\widehat{\sigma}(t)$$

is an approximate $100(1-\alpha)\,\%$ confidence interval for $g(A(t))$. Here $\pm$ means $+$ for the upper limit and $-$ for the lower limit of the confidence interval.

c) With $g(x) = \log x$, we have $g'(x) = 1/x$. Thus an approximate $100(1-\alpha)\,\%$ confidence interval for log $A(t)$ is

$$\log\,\widehat{A}(t) \pm z_{1-\alpha/2}\,\widehat{\sigma}(t)/\widehat{A}(t)$$

By exponentiating the lower and upper limits of this confidence interval, we get the following confidence interval for $A(t)$:

$$\exp\Big\{\log\,\widehat{A}(t) \pm z_{1-\alpha/2}\,\widehat{\sigma}(t)/\widehat{A}(t)\Big\}$$

i.e.

$$\widehat{A}(t)\exp\Big\{\pm z_{1-\alpha/2}\,\widehat{\sigma}(t)/\widehat{A}(t)\Big\}$$

This is (3.7) in the ABG-book.

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 $N_1(t), \ldots, N_n(t)$ as above have been observed, on time intervals, respectively, $[0, \tau_i]$ for $i = 1, \ldots, n$, 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 $n = 3$, $\tau_1 = 30, \tau_2 = 20, \tau_3 = 10$, and that the observed events for the three processes are given as

i=1: 5, 12, 17

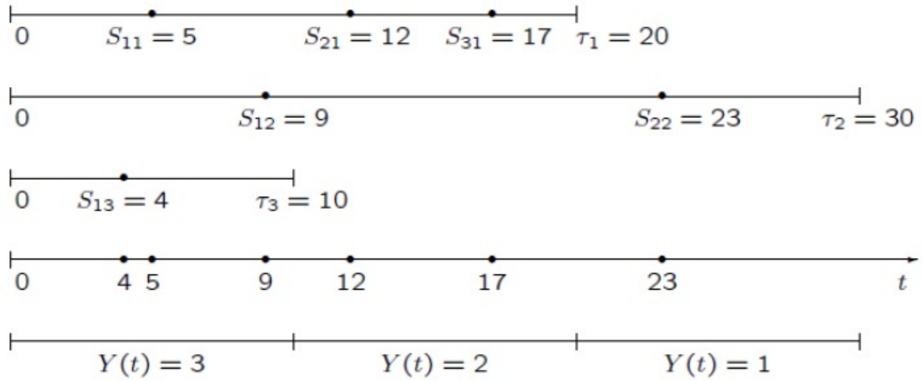
i=2: 9, 23

i=3: 4

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

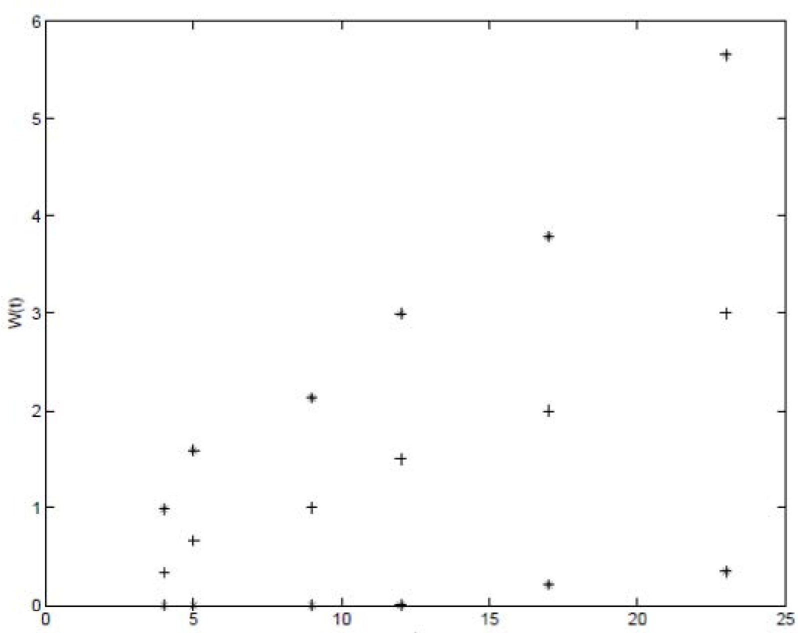
Calculate also the variance of the estimator and find pointwise 95% confidence bounds for the Nelson-Aalen curve.

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 $A(t)/n$.

```
# PRACTICAL EXERCISE 1
# ===================

# Read data:
path="http://www.uio.no/studier/emner/matnat/math/STK4080/h14/melanoma.txt"
melanoma=read.table(path,header=T)

# Load the survival library:
library(survival)


# Question a:
# -----------
# Nelson-Aalen plots for female patients:
fit.f=coxph(Surv(lifetime,status==1)~1,data=melanoma,subset=(sex==1))
surv.f=survfit(fit.f)
plot(surv.f,fun="cumhaz", mark.time=F,xlim=c(0,10),ylim=c(0,0.80),
    main="Females",xlab="Years since operation",ylab="Cumulative hazard")


# Question b:
# ----------
# Nelson-Aalen plots for male patients:
fit.m=coxph(Surv(lifetime,status==1)~1,data=melanoma,subset=(sex==2))
surv.m=survfit(fit.m)
plot(surv.m,fun="cumhaz", mark.time=F,xlim=c(0,10),ylim=c(0,0.80),
    main="Males",xlab="Years since operation",ylab="Cumulative hazard")

# Nelson-Aalen plots for both genders:
fit.b=coxph(Surv(lifetime,status==1)~strata(sex),data=melanoma)
surv.b=survfit(fit.b)
plot(surv.b,fun="cumhaz", mark.time=F,xlim=c(0,10),ylim=c(0,0.80),
    main="Both genders",xlab="Years since operation",ylab="Cumulative hazard",lty=1:2)
legend("topleft",c("females","males"),lty=1:2)

# If we include the option "conf.int=T" in the plot command,
# we obtain plots with confidence limits


# Question c:
# ----------
# Nelson-Aalen plots for patients according ulceration status:
fit.u=coxph(Surv(lifetime,status==1)~strata(ulcer),data=melanoma)
surv.u=survfit(fit.u)
plot(surv.u,fun="cumhaz", mark.time=F,xlim=c(0,10),ylim=c(0,1.0),
    main="Ulceration",xlab="Years since operation",ylab="Cumulative hazard",lty=1:2)
legend("topleft",c("with ulceration","without ulceration"),lty=1:2)


# Question d:
# ----------
# Nelson-Aalen plots for patients according to thickness group:
fit.t=coxph(Surv(lifetime,status==1)~strata(grthick),data=melanoma)
surv.t=survfit(fit.t)
plot(surv.t,fun="cumhaz", mark.time=F,xlim=c(0,10),ylim=c(0,1.0),
     main="Thickness",xlab="Years since operation",ylab="Cumulative hazard",lty=1:3)
legend("topleft",c("0-1 mm","2-5 mm","5+ mm"),lty=1:3)
```