Exercise 3.1 ABK

Two groups, Placebo and 6-MB, time to remission. Assume continuous model with remission occurring before censorings.

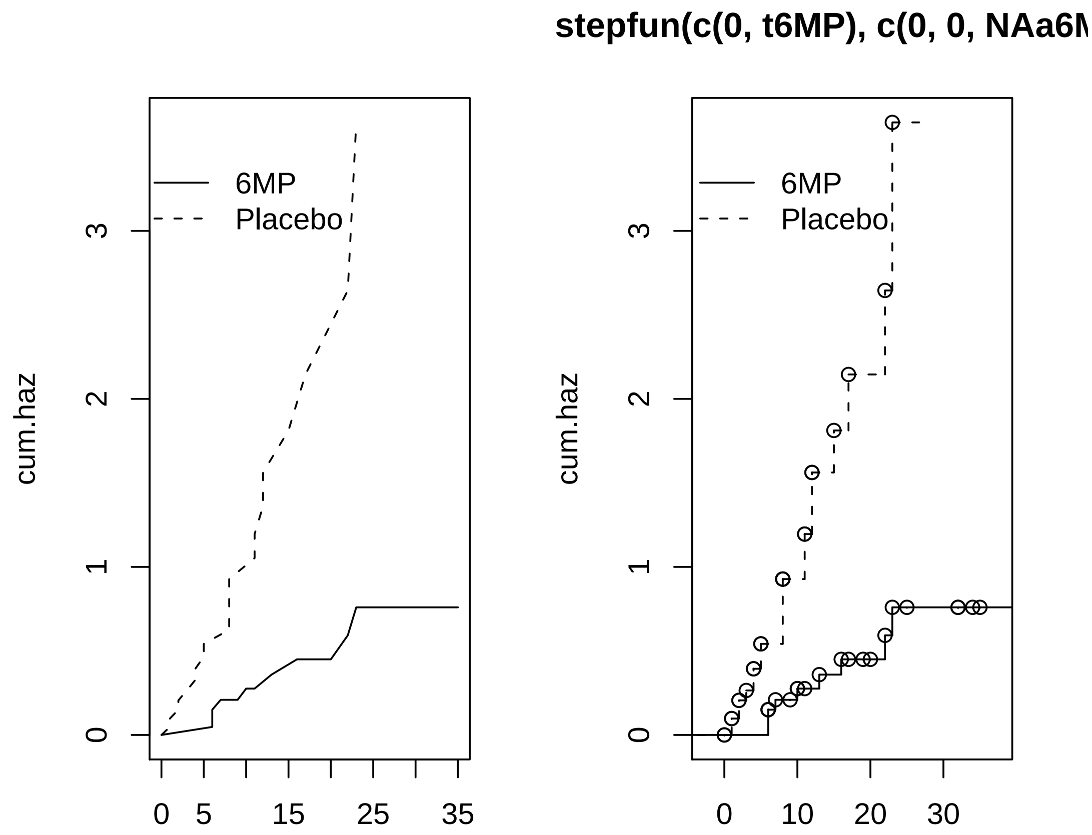
R-code, programming Nelson-Aalen with `cumsum`

```
tpl<-c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)
dpl<-rep(1,length(tpl))

t6MP<-c(6,6,6,6,7,9,10,10,11,13,16,17,19,20,22,23,25,32,32,34,35)
d6MP<-c(1,1,1,0,1,0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0)

ypl<-length(tpl):1
y6MP<-length(t6MP):1

NAapl<-cumsum(dpl/ypl)
NAa6MP<-cumsum(d6MP/y6MP)

plot(c(0,t6MP),c(0,NAa6MP),type="l",ylim=c(0,max(c(NAapl,NAa6MP))),
xlab="time",ylab="cum.haz")
lines(c(0,tpl),c(0,NAapl),lty=2)
legend(0,3.5,c("6MP","Placebo"),lty=1:2,bty="n")
```

Exercise 3.1 ABK

Alternatively, plotting with `stepfun`:

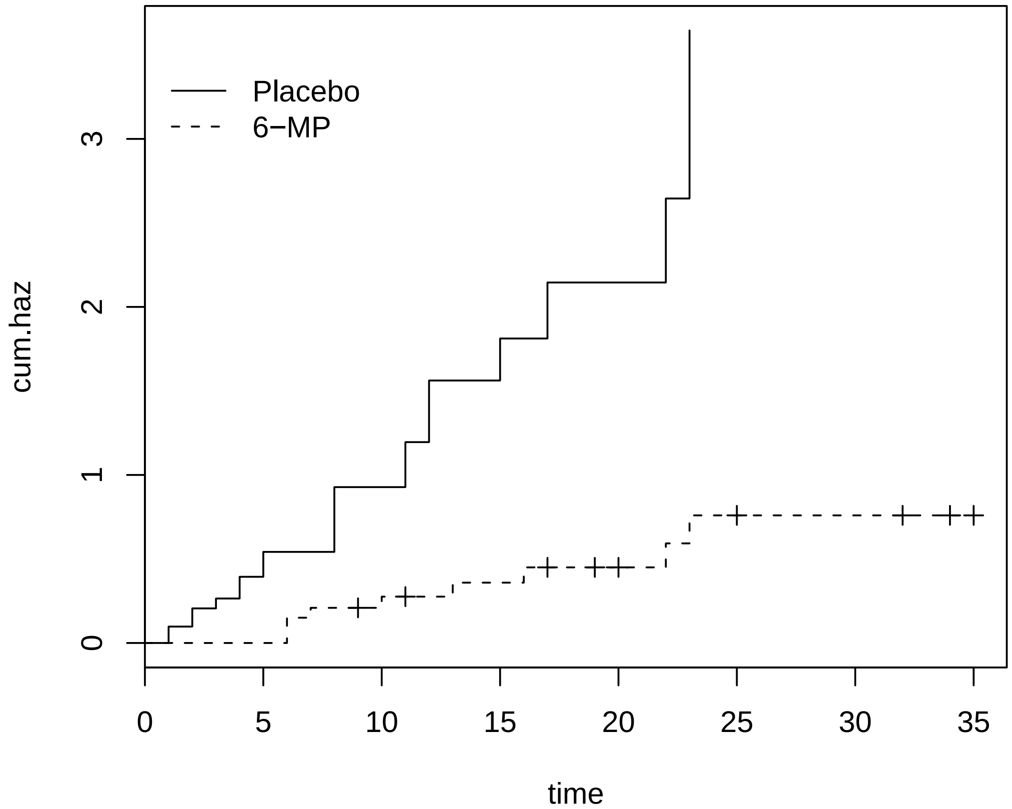
```
plot(stepfun(c(0,t6MP),c(0,0,NAa6MP)),ylim=c(0,max(c(NAapl,NAa6MP))),
xlab="time",ylab="cum.haz")
lines(stepfun(c(0,tpl),c(0,0,NAapl)),lty=2)
legend(-7,3.5,c("6MP","Placebo"),lty=1:2,bty="n")
```



Exercise 3.1 ABK

Or better, using the `survival` library:

```
times<-c(tpl,t6MP)
remis<-c(dpl,d6MP)
group<-c(rep(1,length(tpl)),rep(2,length(t6MP)))
library(survival)
fit<-survfit(Surv(times,remis)~group,type="fh2")
plot(fit,fun="cumhaz")
```



Exercise 3.2 ABK

With proportional hazards $\alpha_2(t) = R\alpha_1(t)$ (for hazard ration R):

$$A_2(t) = RA_1(t)$$

Might plot

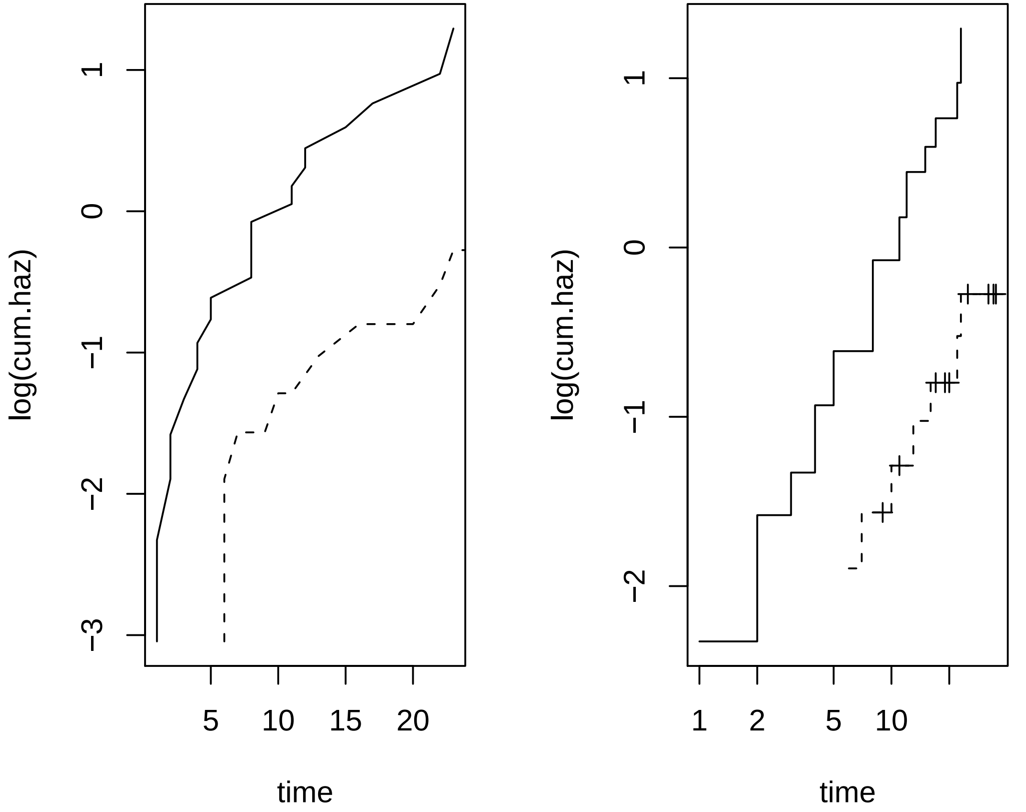
- $\frac{\hat{A}_2(t)}{\hat{A}_1(t)}$ over t; should be approximately horizontal line $y = R$
- $(\hat{A}_1(t), \hat{A}_2(t))$; should be straight line $y = Rx$
- $\log(\hat{A}_1(t))$ and $\log(\hat{A}_2(t))$ over t, should be two roughly paralell lines since $\log(A_2(t)) = \log(R) + \log(A_1(t))$

Last option is the standard plotting technique for checking proportional hazards graphically.

However, the method is rather coarse, and better tests have been developed and implemented.

Exercise 3.2 ABK

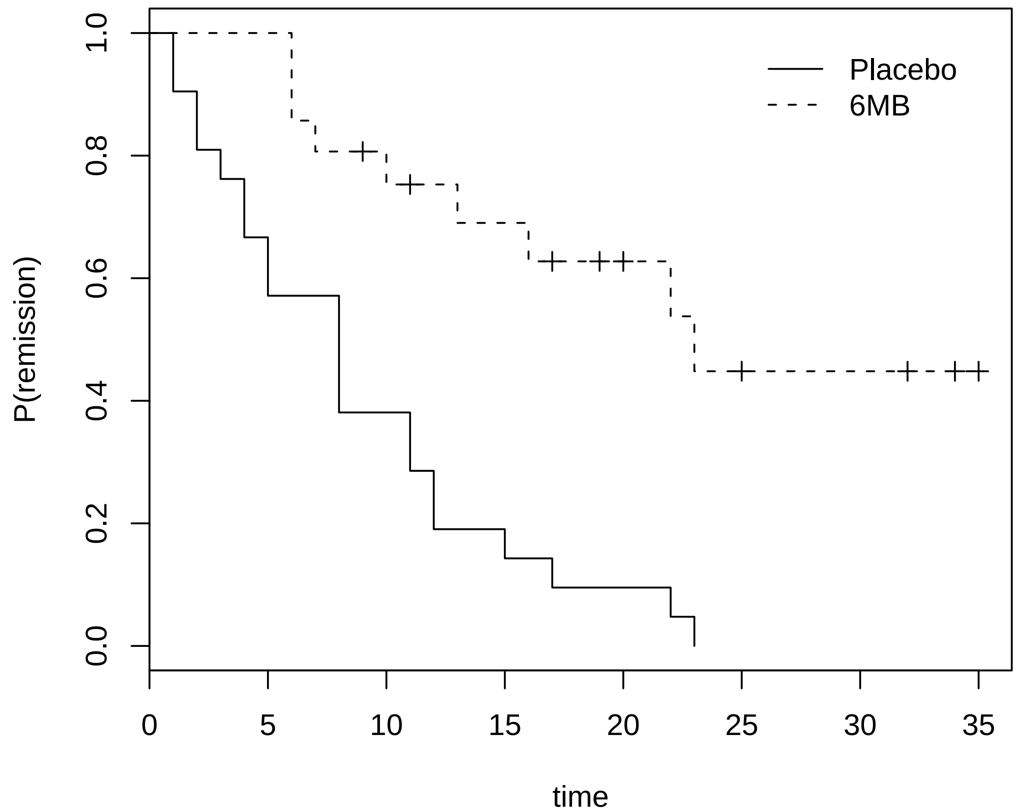
```
plot(tpl,log(NAapl),type="l",
ylim=c(min(log(c(NAapl,NAa6MP))),max(log(c(NAapl,NAa6MP)))),
xlab="time",ylab="log(cum.haz)")
lines(t6MP,log(NAa6MP),lty=2)

plot(fit,fun="cloglog",lty=1:2,xlab="time",ylab="log(cum.haz)")
legend(1,3.5,c("Placebo","6-MP"),lty=1:2,bty="n")
```



Exercise 3.4 ABK Kaplan-Meier

Computing the Kaplan-Meier by `cumprod` and `survfit`

```
KMpl<-cumprod(1-dpl/ypl)
KM6MP<-cumprod(1-d6MP/y6MP)

fit<-survfit(Surv(times,remis)~group)
plot(fit,xlab="time",ylab="P(remission)",lty=1:2)
legend(25,1,lty=1:2,c("Placebo","6MB"),bty="n")
```



Exercise 3.5 Kaplan-Meier without censoring

With event time $t_1 < t_2 < \cdots t_n$ and no censorings we have the number at risk at t_j equal to $Y(t_j) = n - j + 1$. Thus, Add.Ex. 2,

$$\hat{S}(t) = \prod_{t_k \leq t_j} [1 - \frac{1}{n-k-1}] = \prod_{t_k \leq t_j} \frac{n-k-2}{n-k-1} = \frac{Y(t_j+)}{n} = 1 - \hat{F}(t_j)$$

where $\hat{F}(t) = \frac{\#(T_i \leq t)}{n}$ is the empirical cumulative distr. function. The Greenwood formula variance estimator of $\hat{S}(t)$ is given by

$$\hat{S}(t)^2 \int_0^t \frac{dN(s)}{Y(s)(Y(s)-1)} = \hat{S}(t)^2 \sum_{\tilde{T}_i \leq t} \frac{D_i}{Y(\tilde{T}_i)(Y(\tilde{T}_i)-1)}$$

where the term $Y(\tilde{T}_i) - 1$ differ from what is used for the Nelson-Aalen variance est.

Exercise 3.5 Kaplan-Meier without censoring

However, with no censoring and all $D_i = 1$

$$\sum_{\tilde{T}_i \le t} \frac{D_i}{Y(\tilde{T}_i)(Y(\tilde{T}_i)-1)} = \sum_{\tilde{T}_i \le t} \left[\frac{1}{Y(\tilde{T}_i)-1} - \frac{1}{Y(\tilde{T}_i)}\right]$$
$$= \sum_{\tilde{T}_i \le t} \left[\frac{1}{n-i} - \frac{1}{n-i+1}\right] = \frac{1}{n-k} - \frac{1}{n} = \frac{k}{n(n-k)}$$

when $\tilde{T}_{k-1} \le t < \tilde{T}_k$. But then Greenwoods formula becomes

$$\hat{S}(t)^2 \frac{\hat{F}(t)}{n\hat{S}(t)} = \frac{\hat{F}(t)(1-\hat{F}(t))}{n}$$

thus the usual binomial variance for $\hat{F}(t)$.

Ex. 3.3 ABK

The delta method: Assume that $\sqrt{n}(\hat{\theta} - \theta) \to \mathrm{N}(0, \sigma^2)$ and let $\gamma = g(\theta)$ and $\hat{\gamma} = g(\hat{\theta})$ for a smooth function $g()$. Then

$$\sqrt{n}(\hat{\gamma} - \gamma) \to \mathrm{N}(0, g'(\theta)^2\sigma^2)$$

Proof: By a 1. order Taylor expansion
$\hat{\gamma} = g(\hat{\theta}) = g(\theta) + g'(\theta)(\hat{\theta} - \theta) +$ Rest term
and thus

$$\sqrt{n}(\hat{\gamma} - \gamma) = g'(\theta)\sqrt{n}(\hat{\theta} - \theta) + \text{ Rest term}'$$

and $g'(\theta)\sqrt{n}(\hat{\theta} - \theta) \to \mathrm{N}(0, g'(\theta)^2\sigma^2)$

Ex. 3.3 ABK

Confidence interval by the delta method: Let $g(\theta)$ be an increasing function. Let also $(\hat{\gamma}_L, \hat{\gamma}_U) = \hat{\theta} \pm 1.96 \frac{\tau^2}{\sqrt{n}}$ be an approximate 95% CI for γ. Then, with $g^{-1}()$ the inverse function of $g()$, $\hat{\theta}_L = g^{-1}(\hat{\gamma}_L)$ and $\hat{\theta}_U = g^{-1}(\hat{\gamma}_U)$

$$\begin{aligned} \mathrm{P}(\hat{\theta}_L < \theta < \hat{\theta}_U) &= \mathrm{P}(g(\hat{\theta}_L) < g(\theta) < g(\hat{\theta}_U)) \\ &= \mathrm{P}(\hat{\gamma}_L < \gamma < \hat{\gamma}_U) \approx 0.95 \end{aligned}$$

and $(\hat{\theta}_L, \hat{\theta}_U) = (g^{-1}(\hat{\gamma}_L), g^{-1}(\hat{\gamma}_U))$ is an approximate 95% CI of θ.

For a well chosen function $g()$ we may have a better normal approximation for $\hat{\gamma} = g(\hat{\theta})$ than for $\hat{\theta}$. Then first constructing a symmetric CI for γ and transforming back we get a better CI for θ.

Ex. 3.3 ABK

Application of the delta method to Nelson-Aalen:
$\hat{\theta} = \hat{A}(t) \sim \mathrm{N}(\theta, \hat{\sigma}^2(t))$ where $\theta = A(t)$ and $\hat{\sigma}^2(t) = \int \frac{dN(s)}{Y(s)^2}$.

Then $\hat{\gamma} = \log(\hat{\theta}) = \log(\hat{A}(t)) \sim \mathrm{N}(\gamma, \hat{\sigma}^2(t)/A(t)^2)$ for $\gamma = \log(A(t))$ and

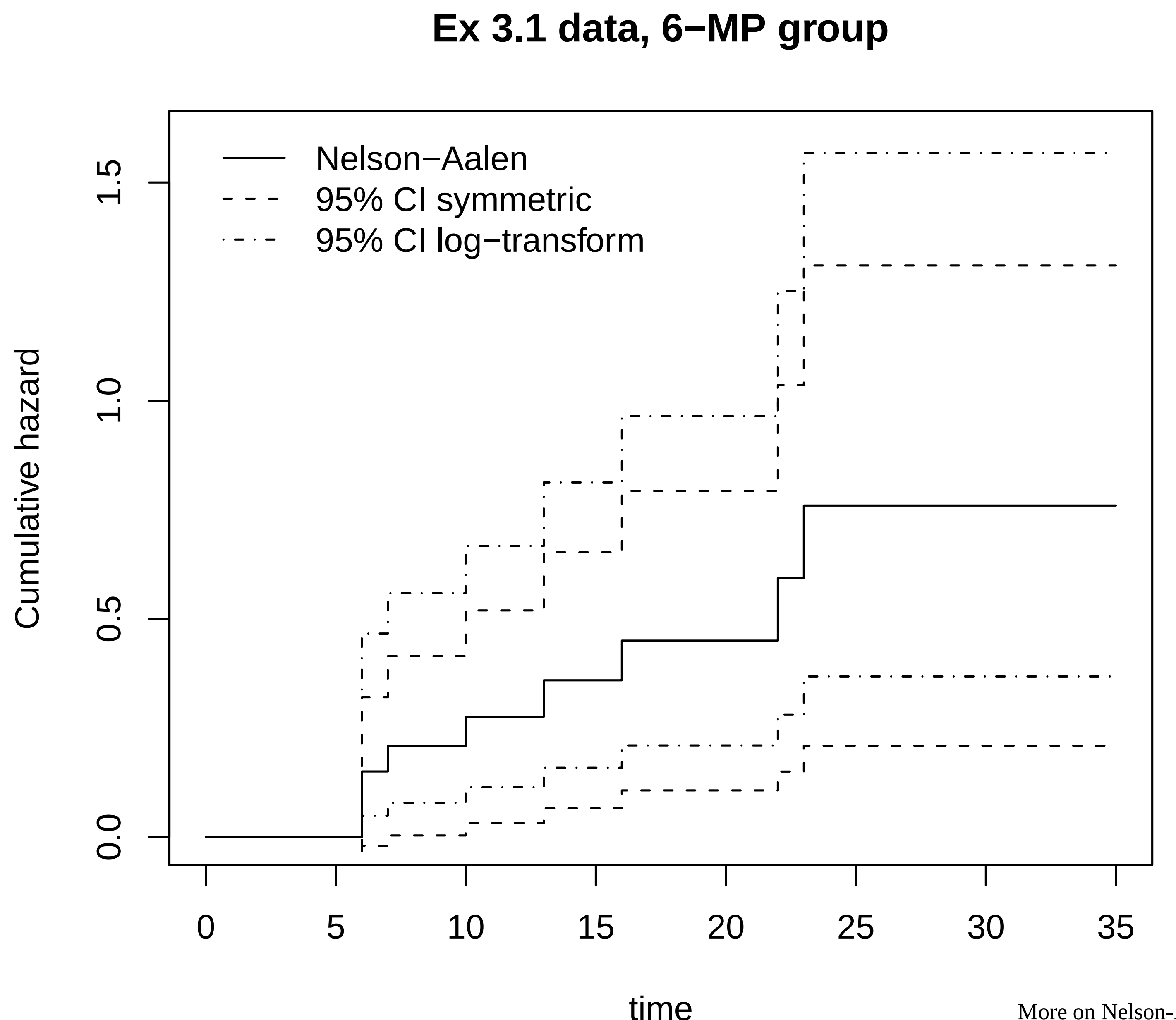
$$\log\hat{\theta} \pm 1.96\hat{\sigma}(t)/\hat{\theta} = \log\hat{A}(t) \pm 1.96\hat{\sigma}(t)/\hat{A}(t)$$

an approximate 95% CI for $\log(A(t))$. But transforming back with $\theta = g^{-1}(\gamma) = \exp(\gamma)$ this gives 95% CI for $\theta = A(t)$ by

$$\hat{A}(t)\exp(\pm 1.96\hat{\sigma}(t)/\hat{A}(t))$$

or eq. (3.7) in ABK.

Ex. 3.3 Illustration, Note that log-transformed CI > 0



Ex. 3.3 Illustration, R-code

```
stepfun2<-function(x,y){
n<-length(x)
x0<-sort(c(x[1:(n-1)],x[2:n]))
indeks<-numeric(0)
for (i in 1:(n-1)) indeks<-c(indeks,i,i)
y0<-y[indeks]
cbind(x0,y0)
}

V<-cumsum(d6MP/y6MP^2)
plot(stepfun2(c(0,t6MP),c(0,NAa6MP)),type="l",ylim=c(0,1.6),
xlab="time",ylab="Cumulative hazard")
lines(stepfun2(c(0,t6MP),c(0,NAa6MP)+sqrt(c(0,V))*1.96),lty=2)
lines(stepfun2(c(0,t6MP),c(0,NAa6MP)-sqrt(c(0,V))*1.96),lty=2)
lines(stepfun2(t6MP,NAa6MP*exp(sqrt(V)*1.96/NAa6MP)),lty=4)
lines(stepfun2(t6MP,NAa6MP*exp(-sqrt(V)*1.96/NAa6MP)),lty=4)
legend(-0.5,1.65,c("Nelson-Aalen","95% CI symmetric","95% CI log-transf
title("Ex 3.1 data, 6-MP group")
```