SOLUTIONS TO EXERCISES WEEK 47

Additional exercise 3

We assume that Z is gamma distributed with mean 1 and variance δ. Conditional on frailty, T_1 and T_2 are assumed independent and identically distributed with hazard

$$\alpha(t \mid Z) = Z\,\alpha(t)$$

We let $A(t) = \int_0^t \alpha(u)\,du$.

a) The joint survival function is given as

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$$

We know that Z has Laplace

transform $\mathcal{L}(c) = E(e^{-cZ}) = (1+\delta c)^{-1/\delta}$.

Therefore we have that

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$$

$$= E\{I(T_1 > t_1, T_2 > t_2)\}$$

$$= E\left[E\{I(T_1 > t_1, T_2 > t_2) \mid Z\}\right]$$

$$= E\{P(T_1 > t_1, T_2 > t_2 / Z)\}$$

$$= E\left(e^{-ZA(t_1)} e^{-ZA(t_2)}\right)$$

$$= E\left(e^{-(A(t_1)+A(t_2))Z}\right)$$

$$= \mathcal{L}(A(t_1) + A(t_2))$$

$$= \left(1 + \delta(A(t_1) + A(t_2))\right)^{-1/\delta} \qquad (1)$$

The conditional hazard rate for T_2 given $T_1 > t_1$, is given by

$$\mu(t_2 | t_1) = \frac{-S'(t_2 | t_1)}{S(t_2 | t_1)} \qquad (2)$$

where $S(t_2 | t_1) = P(T_2 > t_2 \mid T_1 > t_1)$ and differentiation in (2) is with respect to t_2.

b) By formula (1) we have that

$$S(t_2 | t_1) = P(T_2 > t_2 \mid T_1 > t_1)$$

$$= \frac{P(T_2 > t_2, T_1 > t_1)}{P(T_1 > t_1)}$$

$$= \frac{P(T_1 > t_1, T_2 > t_2)}{P(T_1 > t_1, T_2 > 0)}$$

$$= \frac{S(t_1, t_2)}{S(t_1, 0)}$$

$$= \frac{\left(1 + \delta\left(A(t_1) + A(t_2)\right)\right)^{-1/\delta}}{\left(1 + \delta\left(A(t_1) + A(0)\right)\right)^{-1/\delta}}$$

$$= \left(\frac{1 + \delta\left(A(t_1) + A(t_2)\right)}{1 + \delta A(t_1)}\right)^{-1/\delta}$$

$$= \left(1 + \frac{\delta A(t_2)}{1 + \delta A(t_1)}\right)^{-1/\delta}$$

By differentiation with respect to t_2 we obtain

$$S'(t_2 | t_1) = -\frac{1}{\delta}\left(1 + \frac{\delta A(t_2)}{1 + \delta A(t_1)}\right)^{-\frac{1}{\delta} - 1} \frac{\delta \alpha(t_2)}{1 + \delta A(t_1)}$$

$$= -\frac{\alpha(t_2)}{1 + \delta A(t_1)}\left(1 + \frac{\delta A(t_2)}{1 + \delta A(t_1)}\right)^{-\frac{1}{\delta} - 1}$$

c) The conditional hazard rate of T_2 given $T_1 > t_1$ becomes

$$\mu(t_2 \mid t_1) = \frac{-S'(t_2 \mid t_1)}{S(t_2 \mid t_1)}$$

$$= \frac{\dfrac{\alpha(t_2)}{1+\delta A(t_1)}\left(1 + \dfrac{\delta A(t_2)}{1+\delta A(t_1)}\right)^{-\frac{1}{\delta}-1}}{\left(1 + \dfrac{\delta A(t_2)}{1+\delta A(t_1)}\right)^{-1/\delta}}$$

$$= \frac{\dfrac{\alpha(t_2)}{1+\delta A(t_1)}}{1 + \dfrac{\delta A(t_2)}{1+\delta A(t_1)}}$$

$$= \frac{\alpha(t_2)}{1+\delta\,(A(t_1) + A(t_2))} \qquad (3)$$

By (6.8) in the ABG-book we have

$$\mu(t_2) = \frac{\alpha(t_2)}{1 + \delta A(t_2)} \qquad (6.8)$$

By comparing (6.8) and (3), we see that knowing that $T_1 > t_1$, reduces the hazard rate for T_2

Exam STK4080 December 2008 Problem 1

a) The survival function is given by

$$S(t) = P(T > t)$$

and the hazard rate is

given by

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \le T < t + \Delta t \mid T \geqslant t)$$

Now we have

$$\frac{1}{\Delta t} P(t \le T < t + \Delta t \mid T \geqslant t)$$

$$= \frac{P(t \le T \le t + \Delta t)}{\Delta t} \cdot \frac{1}{P(T \geqslant t)}$$

$$= \frac{S(t) - S(t + \Delta t)}{\Delta t} \cdot \frac{1}{S(t)}$$

$$\doteq - \frac{S(t + \Delta t) - S(t)}{\Delta t} \cdot \frac{1}{S(t)}$$

$$\to - S'(t) \frac{1}{S(t)} \quad \text{as } \Delta t \to 0$$

Thus we have

$$\alpha(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt}\log S(t)$$

By integrating and using that $S(0) = 1$, we obtain

$$A(t) = \int_0^t \alpha(u)\,du = -\log S(t)$$

This gives

$$S(t) = e^{-A(t)} = e^{-\int_0^t \alpha(u)\,du}$$

We now let $T_1, T_2, \ldots, T_n$ be iid with survival function $S(t)$ and hazard rate $\alpha(t)$. We only observe the right censored survival times $\tilde{T}_1, \dots, \tilde{T}_n$ and the censoring indicators $D_i = I\{\tilde{T}_i = T_i\}$.

b) We have independent censoring if an individual who is at risk at time t (i.e. $\tilde{T}_i \geq t$) has the same probability of experiencing the event in $[t, t+dt)$ as would have been the case in the situation without censoring.

A bit more formally:

$$P(t \leq \tilde{T}_i < t+dt, D_i = 1 \mid \tilde{T}_i \geq t, \text{past})$$

$$= P(t \leq T_i < t+dt \mid T_i \geq t)$$

cf ABG page 30

c) We may estimate $S(t)$ by the Kaplan-Meier estimator

$$\hat{S}(t) = \prod_{\{i : \tilde{T}_i \le t, D_i = 1\}} \left(1 - \frac{1}{Y(\tilde{T}_i)}\right)$$

where $Y(t) = \#\{i : \tilde{T}_i \geq t\}$ is the number at risk at time t.

The pth fractile ξ_p of the survival distribution is given by $S(\xi_p) = 1 - p$, cf page 95 in the ABG-book. The fractile is estimated by

$$\hat{\xi}_p = \inf\{t : \hat{S}(t) \le 1 - p\}$$

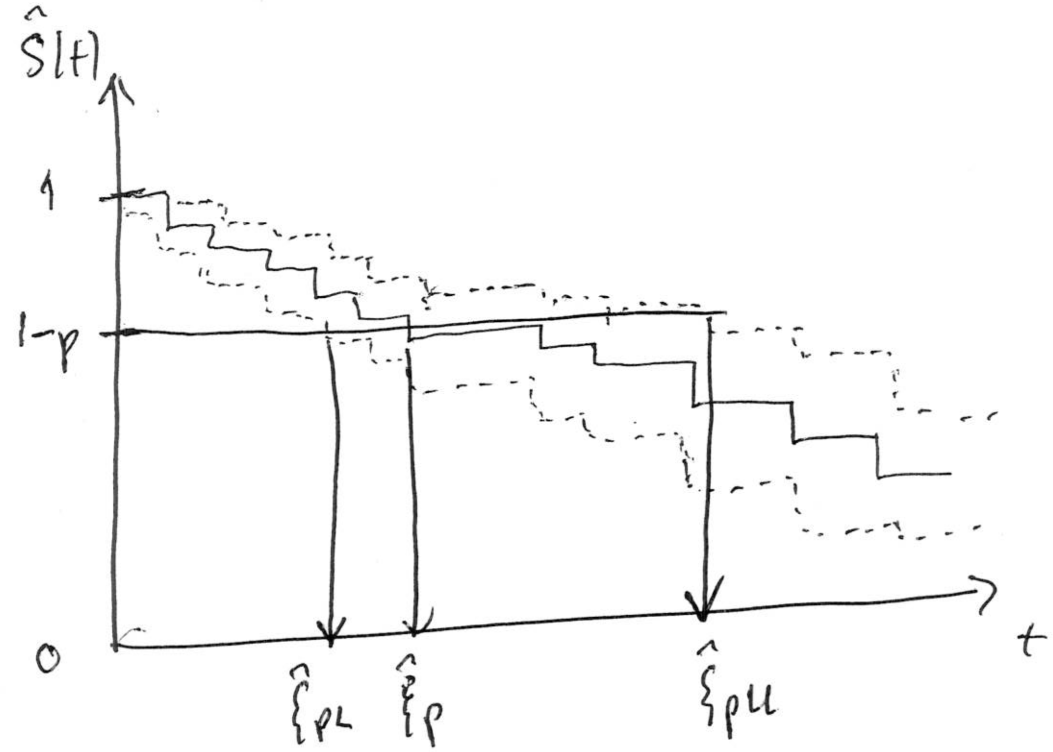
The quartiles are obtained for $p = 0.25$, $p = 0.50$ and $p = 0.75$

The lower and upper confidence limits, $\hat{\xi}_{pL}$ and $\hat{\xi}_{pU}$, are obtained in a similar manner from the confidence limits of the survival function. If we use the standard confidence limits for $S(t)$, i.e. $\hat{S}(t) \pm 1.96\,\hat{\tau}(t)$, we have

$$\hat{\xi}_{pL} = \inf\{t : \hat{S}(t) - 1.96\,\hat{\tau}(t) \leq 1-p\}$$

$$\hat{\xi}_{pU} = \inf\{t : \hat{S}(t) + 1.96\,\hat{\tau}(t) \leq 1-p\}$$

For further details, see section 3.2.3 and exercise 3.8 in the ABG-book. The sketch below illustrates the procedure



d) Output from R:

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
9	32	1	0.969	0.0308	0.798	0.996
11	31	1	0.938	0.0428	0.773	0.984
12	30	1	0.906	0.0515	0.737	0.969
20	29	2	0.844	0.0642	0.665	0.932
22	27	1	0.813	0.0690	0.629	0.911
25	26	2	0.750	0.0765	0.562	0.866
28	23	2	0.685	0.0826	0.493	0.816
31	21	1	0.652	0.0849	0.460	0.790
35	20	2	0.587	0.0880	0.397	0.736
46	18	1	0.554	0.0890	0.366	0.707
49	17	1	0.522	0.0895	0.336	0.678

We find $\hat{\xi}_{0.25} = 25$ days

$\hat{\xi}_{0.75,L} = 12$ days

$\hat{\xi}_{0.75,U} = 35$ days

Exam STK4080 December 2008 Problem 2

$N_1(t)$ and $N_2(t)$ are counting processes with intensity processes of the multiplicative form $\lambda_1(t) = \alpha_1(t) Y_1(t)$ and $\lambda_2(t) = \alpha_2(t) Y_2(t)$.

We want to test

$$H_0: \alpha_1(t) = \alpha_2(t) \quad \text{for all } t \in [0, t_0]$$

a) We will base the test on a statistic of the form

$$Z_1(t_0) = \int_0^{t_0} L(t) \left(d\hat{A}_1(t) - d\hat{A}_2(t) \right) \quad (*)$$

where $L(t) \geq 0$ is a weight process and $\hat{A}_1(t)$ and $\hat{A}_2(t)$ are the Nelson-Aalen estimators

If H_0 is true we will have $d\hat{A}_1(t)$ and $d\hat{A}_2(t)$ fairly equal (on the average), and $Z_1(t_0)$ will be close to zero (in fact, it will have mean zero as shown in 6).

If, however, $\alpha_1(t) > \alpha_2(t)$ [or $\alpha_1(t) < \alpha_2(t)$] then $d\hat{A}_1(t)$ will tend to be larger [smaller] than $d\hat{A}_2(t)$, and $Z_1(t_0)$ will be quite a bit larger [smaller] than zero.

Thus $Z_1(t_0)$ is a reasonable test statistic for "non-crossing" hazards alternatives

b) For $h = 1, 2$, the Nelson–Aalen estimators are given by

$$\hat{A}_h(t) = \int_0^t \frac{J_h(u)}{Y_h(u)} \, dN_h(u)$$

where $J_h(u) = I(Y_h(u) > 0)$.

It is assumed that the weight process $L(t) = 0$ whenever at least one of $Y_1(t)$ and $Y_2(t)$ is zero.

Then we may reformulate $Z_1(t_0)$ as [cf pp 105-106 in ABG]

$$Z_1(t_0) = \int_0^{t_0} L(t) \left\{ \frac{J_1(t)}{Y_1(t)} dN_1(t) - \frac{J_2(t)}{Y_2(t)} dN_2(t) \right\}$$

$$= \int_0^{t_0} \frac{L(t)}{Y_1(t)} dN_1(t) - \int_0^{t_0} \frac{L(t)}{Y_2(t)} dN_2(t)$$

We write $\alpha(t)$ for the common value of $\alpha_1(t)$ and $\alpha_2(t)$ under H_0.

If H_0 holds true we have the decomposition

$$dN_h(t) = \alpha(t) Y_h(t) dt + dM_h(t)$$

for $h = 1, 2$, where the $M_h(t)$'s are martingales.

So if H_0 holds true we have

$$Z_1(t_0) = \int_0^{t_0} \frac{L(t)}{Y_1(t)} \{\alpha(t) Y_1(t)\, dt + dM_1(t)\}$$

$$- \int_0^{t} \frac{L(t)}{Y_2(t)} \{\alpha(t) Y_2(t) dt + dM_2(t)\}$$

$$= \int_0^{t_0} \frac{L(t)}{Y_1(t)}\, dM_1(t) - \int_0^{t} \frac{L(t)}{Y_2(t)}\, dM_2(t)$$

Thus $Z_1(t_0)$, considered as a

process in t_0, is a difference of two stochastic integrals. Since a stochastic integral is itself a mean zero martingale, it follows that $Z_1(t_0)$ is a mean zero martingale.

To derive an estimator for the variance of $Z_1(t_0)$ under H_0, we consider the predictable variation process of $Z_1(t_0)$. This becomes

$$\langle Z_1 \rangle (t_0) = \int_0^{t_0} \left(\frac{L(t)}{Y_1(t)} \right)^2 \alpha(t)\, Y_1(t)\, dt$$

$$+ \int_0^{t} \left(\frac{L(t)}{Y_2(t)} \right)^2 \alpha(t)\, Y_2(t)\, dt$$

cf (7.48) in the ABG-book

Now $\langle Z_1 \rangle (t_0)$ simplifies to

$$\langle Z_1 \rangle (t_0) = \int_0^{t_0} \frac{L(t)^2}{Y_1(t)} \alpha(t)dt + \int_0^{t_0} \frac{L(t)^2}{Y_2(t)} \alpha(t)dt$$

$$= \int_0^{t_0} \frac{L(t)^2 \,(Y_2(t) + Y_1(t))}{Y_1(t)\,Y_2(t)} \alpha(t)dt$$

$$= \int_0^{t} \frac{L(t)^2\, Y_\bullet(t)}{Y_1(t)\,Y_2(t)} \alpha(t)dt \qquad (\#)$$

where $Y_\bullet(t) = Y_1(t) + Y_2(t)$

To obtain an estimator for the variance of $Z_1(t_0)$, we replace $\alpha(t)dt$ in $(\#)$ by $d\hat{A}(t)$, where $\hat{A}(t) = \int_0^t \frac{J(u)}{Y_\bullet(u)}\, dN_\bullet(u)$ is the Nelson-Aalen estimator based on the aggregated process $N_\bullet(t) = N_1(t) + N_2(t)$

and $J(u) = I(Y_\bullet(u) > 0)$.

We then obtain the variance estimator

$$V_{11}(t_0) = \int_0^{t_0} \frac{L(t)^2\, Y_\bullet(t)}{Y_1(t) Y_2(t)} \frac{dN_\bullet(t)}{Y_\bullet(t)}$$

$$= \int_0^{t_0} \frac{L(t)^2}{Y_1(t) Y_2(t)} dN_\bullet(t)$$

One may show that this estimator is unbiased under H_0 (exercise 3.10)

c) The log-rank test corresponds to the weight process

$L(t) = Y_1(t)\, Y_2(t) \,/\, Y_\bullet(t)$

Then $Z_1(t_0)$ may be rewritten

as

$$Z_1(t_0) = \int_0^{t_0} \frac{L(t)}{Y_1(t)}\, dN_1(t) - \int_0^{t_0} \frac{L(t)}{Y_2(t)}\, dN_2(t)$$

$$= \int_0^{t_0} \frac{Y_2(t)}{Y_0(t)}\, dN_1(t) - \int_0^{t} \frac{Y_1(t)}{Y_0(t)}\, dN_2(t)$$

$$= \int_0^{t_0} \frac{Y_2(t)}{Y_0(t)}\, dN_1(t) - \int_0^{t} \frac{Y_1(t)}{Y_0(t)} \{dN_0(t) - dN_1(t)\}$$

$$= \int_0^{t_0} \left(\frac{Y_2(t)}{Y_0(t)} + \frac{Y_1(t)}{Y_0(t)} \right) dN_1(t) - \int_0^{t_0} \frac{Y_1(t)}{Y_0(t)}\, dN_0(t)$$

$$= \int_0^{t_0} dN_1(t) - \int_0^{t_0} \frac{Y_1(t)}{Y_0(t)}\, dN_0(t)$$

$$= N_1(t_0) - E_1(t_0)$$

where $E(t_0) = \int_0^{t_0} \frac{Y_1(t)}{Y_0(t)}\, dN_0(t)$

$E_1(t_0)$ may be interpreted as the expected number of events under H_0

For the data example we get the R output

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
treat=0	32	5	10.2	2.65	5.49
treat=1	32	15	9.8	2.75	5.49

Chisq= 5.5 on 1 degrees of freedom, p= 0.0192

Here "treat =0" corresponds to group 1 and "treat = 1" to group 2.

We see from the output that $N_1(t_0) = 5$ and $E_1(t_0) = 10.2$

The test statistic

$$X^2 = \frac{Z_1(t_0)^2}{V_{11}(t_0)} = \frac{(N_1(t_0) - E_1(t_0))^2}{V_{11}(t_0)}$$

takes the value 5.5, which should be compared with a chi-square distribution with 1 degree of freedom. This gives the P-value 1.9%, so the difference between the groups is significant

Since $N_1(t_0) = 5$ is smaller than $E_1(t_0) = 10.2$, the treatment with both MTX and CSP reduce the risk of the life threatening complication.