SOLUTION TO EXAM IN STK 4080/9080
9 DECEMBER 2014

Problem 1

a) Se pages 18-19, 22-23 and 114-115 in the ABG-book.

b) Se pages 77 and 115 in the ABG-book.

c) The Nelson-Aalen estimate for the cumulative hazard of complications is very steep for the first year or so, corresponding to a very high hazard for complications the first year after the transplantation. Then the estimate is not so steep, which means that the hazard is much lower. And after about four years the Nelson-Aalen estimate is almost horizontal which means that the hazard is close to zero.

The Nelson-Aalen estimate for other causes has a similar shape as the one for complications, but it less steep already after half a year or so. So the hazard for other causes is very high the first half year or so after the transplantation. But then it becomes lower, and after about two years it is very small.

The Nelson-Aalen estimate for death from relapse is less steep than the other two early on, corresponding to a smaller hazard than for the other two causes. It becomes less steep as time goes by, but not to a similar extent as for the other two causes. So even though the hazard of death due to relapse becomes smaller after about two years, it is clearly larger than the hazard for the other two causes.

d) The left-hand plot shows that a patient has almost 15% probability of dying from complications and 10% probability of dying from other causes the first year after the transplantation. And the probability of death from these two causes do only increase a few percentage points after that. However, as seen from the right-hand plot, a patient who survives the first half year after the transplantation, has a much lower probability of dying from these two causes. But the probability of dying from relapse is about the same just after transplantation and after half a year. In both cases the probability of dying of relapse within 10 years is about 15%.

Problem 2

The individual hazards, given frailty, are $\alpha(t|Z_i) = Z_i\,\alpha$.

a) When the frailties are given, we have a standard situation for censored survival data. Then we know that

$$P(H_i \mid Z_i) = \alpha(\tilde{T}_i \mid Z_i)^{D_i} e^{-\int_0^{\tilde{T}_i} \alpha(u \mid Z_i)\,du}$$

which for $\alpha(t \mid Z_i) = Z_i \alpha$ gives

$$P(H_i \mid Z_i) = (Z_i \alpha)^{D_i} e^{-Z_i \alpha \tilde{T}_i}$$

b) The Laplace transform of the frailty distribution is given by $\mathcal{L}(c) = E(e^{-cZ_i})$. It follows that $\mathcal{L}'(c) = -E(Z_i e^{-cZ_i})$. Thus we have that $E(Z_i^k e^{-cZ_i}) = (-1)^k \mathcal{L}^{(k)}(c)$ for $k = 0, 1$. Therefore

$$P(H_i) = E\{P(H_i \mid Z_i)\}$$

$$= E\{\alpha^{D_i} Z_i^{D_i} e^{-Z_i \alpha \tilde{T}_i}\}$$

$$= \alpha^{D_i} \, E\{Z_i^{D_i} e^{-Z_i \alpha \tilde{T}_i}\}$$

$$= \alpha^{D_i} (-1)^{D_i} \mathcal{L}^{(D_i)}(\alpha \tilde{T}_i)$$

c) For gamma distributed frailties with mean 1 and variance δ, we have

$$\mathcal{L}^{(0)}(c) = \mathcal{L}(c) = \{1+\delta c\}^{-1/\delta}$$

$$\mathcal{L}^{(1)}(c) = \mathcal{L}'(c) = -\frac{1}{\delta}\{1+\delta c\}^{-\frac{1}{\delta}-1}\delta$$

$$= -\{1+\delta c\}^{-(\frac{1}{\delta}+1)}$$

Thus we have for $k=0,1$:

$$\mathcal{L}^{(k)}(c) = (-1)^k \{1+\delta c\}^{-(\frac{1}{\delta}+k)}$$

Hence we have

$$P(H_i) = \alpha^{D_i} (-1)^{D_i} (-1)^{D_i} \{1+\delta\alpha\tilde{T}_i\}^{-(\frac{1}{\delta}+D_i)}$$

$$= \alpha^{D_i} \{1+\delta\alpha \tilde{T}_i\}^{-(\frac{1}{\delta}+D_i)}$$

and the logarithm of the marginal likelihood becomes

$$l(\alpha, \delta) = \log L_{marg}$$

$$= \sum_{i=1}^{n} \log P(H_i)$$

$$= \sum_{i=1}^{n} \left\{ D_i \log \alpha - \left(\frac{1}{\delta} + D_i\right) \log\left(1 + \delta\alpha\tilde{T}_i\right) \right\}$$

$$D \log \alpha - \sum_{i=1}^{n} \left(\frac{1}{\delta} + D_i\right) \log\left(1 + \delta\alpha\tilde{T}_i\right)$$

where $D = \sum_{i=1}^{n} D_i$

Note that when $\delta \to 0$, we have that $l(\alpha, \delta) \to l(\alpha)$, where

$$l(\alpha) = D \log \alpha - \sum_{i=1}^{n} \alpha T_i = D \log \alpha - \alpha T$$

with $T = \sum_{i=1}^{n} T_i$ (use L'Hôpital's rule).

This is the usual log-likelihood for exponentially distributed data.

d) To test $H_0: \delta = 0$ versus $H_A: \delta > 0$, we may use a one-sided likelihood ratio test. Let then $\hat{\alpha}, \hat{\delta}$ be the values of α and δ that maximize $l(\alpha, \delta)$, and let α^* be the value of α that maximizes $l(\alpha)$. We then compute the statistic

$$Z = 2\{l(\hat{\alpha}, \hat{\delta}) - l(\alpha^*)\}$$

Since the null hypothesis is on the boundary of the parameter space, ordinary likelihood theory does not apply. But one may show that

$$Z \sim \tfrac{1}{2}\chi^2_0 + \tfrac{1}{2}\chi^2_1$$

under H_0. Therefore one obtains the P-value by halving the P-value one would have obtained if Z was χ^2_1-distributed.

Problem 3

a) We have that

$$dN_i(t) = \lambda_i(t)dt + dM_i(t) \quad ; \quad i=1,\dots,n \qquad (\#)$$

where $M_i(t) = N_i(t) - \int_0^t \lambda_i(u)du$ is a martingale.

We introduce $M(t) = (M_1(t), \dots, M_n(t))^T$.

By [3] in the problem set we may then write (#) as

$$dN(t) = X(t)\, dB(t) + dM(t)$$

From (4) we then obtain

$$\hat{B}(t) = \int_0^t J(u)\{X(u)^T X(u)\}^{-1} X(u)^T \{X(u)dB(u) + dM(u)\}$$

$$= \int_0^t J(u)\, dB(u) + \int_0^t J(u)\{X(u)^T X(u)\}^{-1} X(u)^T dM(u)$$

$$= B^*(t) + \int_0^t J(u)\{X(u)^T X(u)\}^{-1} X(u)^T dM(u)$$

Thus we have

$$\hat{B}(t) - B^*(t) = \int_0^t J(u)\{X(u)^T X(u)\}^{-1} X(u)^T dM(u) \qquad (\#\#)$$

(##) shows that $\hat{B}(t) - B^{\circ}(t)$ is a vector-valued stochastic integral.

From this it follows that $\hat{B}(t) - B^{\circ}(t)$ is a mean zero martingale.

Hence

$$E\{\hat{B}(t) - B^{\circ}(t)\} = 0$$

and it follows that

$$E(\hat{B}(t)) = E(B^{\circ}(t)) = \int_0^t E\{J(u)\} \, dB(u)$$

$$= \int_0^t P\{X(u) \text{ has full rank}\} \, dB(u)$$

$$\approx B(t)$$

when $P(X(u) \text{ has full rank}) \approx 1$ for $u \in [0, t]$.

b) We know from question a) that $\hat{B}_1(t) - B_1^{\circ}(t)$ is a mean zero martingale. When H_0 holds true we have that

$B_1^{\circ}(t) = \int_0^t J(u) \beta_1(u) \, du = 0$, and hence

$\hat{B}_1(t)$ is a mean zero martingale under H_0. Now

$$Z_1(t_0) = \int_0^{t_0} L(u)\, d\hat{B}_1(u)$$

is a stochastic integral with respect to the martingale $\hat{B}_1(t)$, so $Z_1(t_0)$ is itself a mean zero martingale when H_0 holds true (when considered as a process in t_0). In particular $E\{Z_1(t_0)\} = 0$ when H_0 holds true.

If $\beta_1(u) > 0$ for all $u \in [0, t_0]$, $d\hat{B}(u)$ will tend to be positive at the event times and hence $Z_1(t_0)$ will tend to be positive. And similarily, if $\beta_1(t) < 0$ $Z_1(t_0)$ will tend to be negative. This explains why $Z_1(t_0)$ is a reasonable test statistic for one-sided alternatives (i.e $\beta_1(t) > 0$ or $\beta_1(t) < 0$)

c) When H_0 holds true we have by (3) that $\lambda_i(t) = Y_i(t)\beta_0(t)$. It then follows by (#), (5) and (6) that

$$Z_1(t_0) = \int_0^{t_0} L(u)\, d\hat{B}_1(u)$$

$$= \sum_{i=1}^{n} \int_0^{t_0} L(u) J(u) \frac{Y_i(u)\{x_i - \bar{x}(u)\}}{S_{xx}(u)} dN_i(u)$$

$$= \sum_{i=1}^{n} \int_0^{t_0} L(u) \frac{Y_i(u)\{x_i - \bar{x}(u)\}}{S_{xx}(u)} \{\lambda_i(u)du + dM_i(u)\}$$

$$= \sum_{i=1}^{n} \int_0^{t_0} L(u) \frac{Y_i(u)\{x_i - \bar{x}(u)\}}{S_{xx}(u)} Y_i(u)\beta_0(u)du$$

$$+ \sum_{i=1}^{n} \int_0^{t_0} L(u) \frac{Y_i(u)\{x_i - \bar{x}(u)\}}{S_{xx}(u)} dM_i(u)$$

$$= \sum_{i=1}^{n} \int_0^{t_0} L(u) \frac{Y_i(u)\{x_i - \bar{x}(u)\}}{S_{xx}(u)} dM_i(u) \qquad (\#\#\#)$$

where the last equality follows since

$$\sum_{i=1}^{n} Y_i(u)\{x_i - \bar{x}(u)\} = 0.$$

By (###) and the relation in the hint, we have

$$\langle Z_1 \rangle (t_0) = \sum_{i=1}^{n} \int_0^{t_0} \left\{ L(u) \frac{Y_i(u)\{x_i - \bar{x}(u)\}}{S_{xx}(u)} \right\}^2 \lambda_i(u)du$$

$$= \sum_{i=1}^{n} \int_0^{t_0} L(u)^2 \frac{Y_i(u)^2\{x_i - \bar{x}(u)\}^2}{S_{xx}(u)^2} Y_i(u)\beta_0(u)du$$

$$= \int_0^{t_0} L(u)^2 \frac{\sum_{i=1}^{n} Y_i(u)\{x_i - \bar{x}(u)\}^2}{S_{xx}(u)^2} \beta_0(u)du$$

$$= \int_0^{t_0} \frac{L(u)^2}{S_{xx}(u)} \beta_0(u)du$$

As noted in the problem the result in question c) motivates the variance estimator

$$V_{11}(t_0) = \int_0^{t_0} \frac{L(u)^2}{S_{xx}(u)} \frac{dN_\cdot(u)}{Y_\cdot(u)}$$

and one may show that $Z_1(t_0)/\sqrt{V_{11}(t_0)}$ is approximately standard normal under H_0

d) The baseline hazard rate ($\beta_0(t)$) corresponds to the hazard rate for an individual with centred tumor thickness equal to zero, i.e. with tumor thickness equal to the mean thickness among all patients.
The estimate $\hat{B}_0(t)$ of the cumulative baseline hazard is fairly linear the first five years after operation with slope approximately equal to

$\frac{0.28}{5} = 0.056$. Thus a patient with mean tumor thickness has a hazard which is about 0.056 per year the first five years after operation.
Between 5 and 8 years after the operation, such a patient has a hazard which is about
$(0.40 - 0.28)/(8-3) = 0.040$ per year.

The estimate $\hat{B}_1(t)$ for the cumulative regression function for tumor thickness has a slope which is about $0.08/4 = 0.02$ the first four years after operation. So in this period the hazard rate increases with about 0.02 per year for each mm increase in tumor thickness. After about four years the estimate $\hat{B}_1(t)$ is fairly horizontal, corresponding to no effect of tumor thickness (i.e. $\beta_1(t) \approx 0$) after four years.

e) We have $Z_1(t_0) = 102.7$ and $V_{11}(t_0) = 364.1$ The standardized test statistic takes the value $Z_1(t_0)/\sqrt{V_{11}(t_0)} = 102.7/\sqrt{364.1} = 5.36$ which is highly significant when compared with a standard normal distribution. Thus the mortality of the patients increases significantly with tumor thickness