Exam 2020/21

Exam nr. 30

21 januar 2021

Exercise 1

Let the setup be as stated in the exam.

a)

From the independence between X and Z we have

$$P(X = i, Z = j) = \begin{cases} \pi_x \pi_z, & (i,j) = (1,1) \\ \pi_x(1-\pi_z), & (i,j) = (1,0) \\ (1-\pi_x)\pi_z, & (i,j) = (0,1) \\ (1-\pi_x)(1-\pi_z), & (i,j) = (0,0) \end{cases}$$

Let

$$A_0(t) = \int_0^t \alpha_0(s)ds$$

be the cumulative baseline hazard function which is unspecified and can be seen as an infinite dimensional, non-parametric, unknown of the cox model. From the cox form we therefore have the cumulative hazard:

$$\begin{aligned} A(t) &= \int_0^t \alpha_0(s) \exp(X\beta + Z\gamma)ds \\ &= A_0(t) \exp(X\beta + Z\gamma) \end{aligned}$$

From which the survival function takes the form:

$$P(T > t) = \exp(-A(t))$$

one could also integrate out the statespace of X and Z to obtain:

$$\begin{aligned} P(T > t) &= \sum_{i,j} P(T > t, X = i, Z = j) \\ &= \sum_{i,j} P(T > t | X = i, Z = j) P(X = i, Z = j) \\ &= \sum_{i,j} \exp\left(-A_0(t) \exp(i\beta + j\gamma)\right) P(X = i, Z = j) \end{aligned}$$

The covariate distribution among surviviors, is then given by:

$$
\begin{aligned}
P(X=i,Z=j|T>t) &= \frac{P(T>t|X=i,Z=j)P(X=i,Z=j)}{P(T>t)} \\
&= \frac{\exp\left(-A_0(t)\exp(i\beta+j\gamma)\right)P(X=i,Z=j)}{\sum_{i,j}\exp\left(-A_0(t)\exp(i\beta+j\gamma)\right)P(X=i,Z=j)} \\
&= \frac{P(X=i,Z=j)}{\sum_{i',j'}\exp\left(A_0(t)\left[\exp(i\beta+j\gamma)-\exp(i'\beta+j'\gamma)\right]\right)P(X=i',Z=j')}
\end{aligned}
$$

From this we see that the limiting distribution of the covariates given survival depends on β and γ in the following way:

if $\beta<0,\gamma<0$:

$$\lim_{t\to\infty}P(X=i,Z=j|T>t)=\begin{cases}1, & i=1,j=1\\ 0, & \text{otherwise}\end{cases}$$

if $\beta>0,\gamma>0$:

$$\lim_{t\to\infty}P(X=i,Z=j|T>t)=\begin{cases}1, & i=0,j=0\\ 0, & \text{otherwise}\end{cases}$$

if $\beta<0,\gamma>0$:

$$\lim_{t\to\infty}P(X=i,Z=j|T>t)=\begin{cases}1, & i=1,j=0\\ 0, & \text{otherwise}\end{cases}$$

if $\beta<0,\gamma=0$:

$$\lim_{t\to\infty}P(X=i,Z=j|T>t)=\begin{cases}\pi_z, & i=1,j=1\\ 1-\pi_z, & i=1,j=0\\ 0, & \text{otherwise}\end{cases}$$

if $\beta=0,\gamma=0$:

$$\lim_{t\to\infty}P(X=i,Z=j|T>t)=P(X=i,Z=j)$$

Where the omitted cases are defined symetrically to the corresponding case above.

In general the covariates will not be independent for the surviors. This is due to the initial independence being distorted by a favoring of combinations of X and Z rendering a lower hazard, regardless of the initial joint distribution.

To have conditional independence we need to get:

$$P(X=i,Z=j|T>t)=P(X=i|T>t)P(Z=j|T>t)$$

for all $i,j\in\{0,1\},t>0$.

Lets look at the marginal conditional distribution for X first, following the same line of argument as above we get:

$$
\begin{aligned}
P(X=i|T>t) &= \frac{P(T>t|X=i)P(X=i)}{P(T>t)} \\
&= \frac{P(T>t|X=i,Z=0)P(Z=0|X=i)P(X=i)+P(T>t|X=i,Z=1)P(Z=1|X=i)P(X=i)}{P(T>t)} \\
&= \frac{P(T>t|X=i,Z=0)P(Z=0,X=i)+P(T>t|X=i,Z=1)P(Z=1,X=i)}{P(T>t)} \\
&= \frac{\exp\left(-A_0(t)\exp(i\beta)\right)(1-\pi_z)P(X=i)+\exp\left(-A_0(t)\exp(i\beta+\gamma)\right)\pi_z P(X=i)}{P(T>t)} \\
&= \frac{\exp\left(-A_0(t)\exp(i\beta)\right)(1-\pi_z)P(X=i)+\exp\left(-A_0(t)\exp(i\beta+\gamma)\right)\pi_z P(X=i)}{\sum_{i,j}\exp\left(-A_0(t)\exp(i\beta+j\gamma)\right)P(X=i,Z=j)}
\end{aligned}
$$

Similarly for $P(Z = j|T > t)$.

To keep conditional independence among the survivors one needs to offset the effect from β and γ corresponding to the initial distribution.
We start by noting the trival case of setting $\beta = \gamma = 0$ in which case on gets, from the formula above:

$$\begin{aligned} P(X = i, Z = j|T > t) &= \frac{\exp(-A_0(t))\, P(X = i, Z = j)}{\exp(-A_0(t))\,\{(1-\pi_x)(1-\pi_z) + \pi_x(1-\pi_z) + (1-\pi_x)\pi_z + \pi_x\pi_z\}} \\ &= P(X = i, Z = j) \\ &= P(X = i)P(Z = j) \end{aligned}$$

It is not clear that this is doable in general.

b)

Now let the hazard have the additive form, with interaction between X and Z modelled by the parameter ρ, now the cumulative hazard takes form:

$$A(t) = \int_0^t \alpha_0(s) + X\beta + Z\gamma + XZ\rho\, ds = A_0(t) + t(X\beta + Z\gamma + XZ\rho)$$

Hence we get the conditional survivor covariate distribution:

$$\begin{aligned} P(X = i, Z = j|T > t) &= \frac{P(T > t|X = i, Z = j)P(X = i, Z = j)}{P(T > t)} \\ &= \frac{\exp(-A_0(t) - t(i\beta + j\gamma + ij\rho))\, P(X = i, Z = j)}{\sum_{i',j'} \exp(-A_0(t) - t(i'\beta + j'\gamma + i'j'\rho))\, P(X = i', Z = j\prime)} \\ &= \frac{P(X = i, Z = j)}{\sum_{i',j'} \exp(t\,[\beta(i - i') + \gamma(j - j') + \rho(ij - i'j')])\, P(X = i', Z = j\prime)} \end{aligned}$$

for $i, j \in \{0, 1\}, t > 0$.

The conditional marginal covariate distribution is now seen to be:

$$\begin{aligned} P(X = i, Z = j|T > t) &= \frac{P(T > t|X = i, Z = j)P(X = i, Z = j)}{P(T > t)} \\ &= \frac{\exp(-A_0(t) - t(i\beta + j\gamma + ij\rho))\, P(X = i, Z = j)}{\sum_{i',j'} \exp(-A_0(t) - t(i'\beta + j'\gamma + i'j'\rho))\, P(X = i', Z = j\prime)} \end{aligned}$$

Again it is not clear whether conditional independence can be acheived. One would suspect that the multiplicative structure of the cox-form would be essential in maintaining such a property.

Exercise 2

a)

Since the event $\delta = 1$ occurs when T is less than U, given $U = t$ we simply find that

$$P(\delta = 1|U = t) = P(T < t) = 1 - P(T > t) = 1 - e^{-\theta t}.$$

likewise one has that

$$P(\delta = 0|U = t) = P(T > t) = e^{-\theta t}.$$

Now since the counting processes $N_j(t)$, $j = 0, 1$ can be seen in the light of the standard counting process setup, the difference being that no censoring is done. Since $N_0(t)$ only jumps, in $t = U$ when $U \leq T$ we have that given information at time t:

$$\mathcal{F}_t = \sigma\{1(U \leq s, \delta = 0), 1(U \leq s, \delta = 1)\}$$

then $dN_0(t)$ is er bernoulli variable conditioned on $\mathcal{F}_{t-}$ which can only jump if $t \leq U$, which is measurable since we assume that U is always observed, and $U < T$ since otherwise $N_0(t)$ will not jump. From the infitesimal interpretation of the hazard $\gamma(x)dx$ as the probability of an event in $[x, x + dx)$ where therefore get:

$$\begin{aligned}\mathbb{E}(dN_0(t)|\mathcal{F}_{t-}) &= 1(t \leq U)\gamma(t)dt\,\mathbb{E}(1(T \geq t)) \\ &= 1(t \leq U)\gamma(t)dt\,P(T \geq t) \\ &= 1(t \leq U)\gamma(t)dt\{e^{-\theta t}\}.\end{aligned}$$

Likewise for $N_1(t)$ since a jump can only occur if $T < U$ and $t \leq U$ we get:

$$\begin{aligned}\mathbb{E}(dN_1(t)|\mathcal{F}_{t-}) &= 1(t \leq U)\gamma(t)dt\,\mathbb{E}(1(T < t)) \\ &= 1(t \leq U)\gamma(t)dt\,P(T < t) \\ &= 1(t \leq U)\gamma(t)dt\{1 - e^{-\theta t}\}\end{aligned}$$

.

(b)

Now, by definition we need to show that

$$\langle M_{i0}, M_{k1}\rangle = 0$$

For all $1 \leq i, k \leq n$ Note that since

$$[M_{i,0}, M_{k,1}](t) = \sum_{s \leq t} \Delta M_{i,0}(s)\Delta M_{i,1}(s) = 0$$

Since the jumps of $M_{i,j}$ comes purely from the counting process $N_{i,j}(t)$, for $i \neq k$ the probability of simultanious jumps is zero. For $i = k$ we note that only one of $M_{i,0}(s), M_{i,1}(s)$ will make a jump in U_i, if $U_i \leq T_i$ then $M_{i,0}$ will jump while if $U_i > T_i$ then $M_{i,1}$ will make the jump.
Since the zero process in particular is a square integrable process, we get that the predictable covariation process $\langle M_{i0}, M_{k1}\rangle$ is the compensator of $[M_{i,0}, M_{k,1}](t) = 0$ which by the uniqueness of the Doob-Meyer is simply the zero-process.

The orthogonality of $M_{\cdot 0}$ and $M_{\cdot 1}$ now simply follows from the predictable covariation process being bilinear, hence:

$$\langle M_{\cdot 0}, M_{\cdot 1} \rangle = \left\langle \sum_{i=1}^{n} M_{i0}, \sum_{k=1}^{n} M_{k1} \right\rangle = \sum_{1 \leq i,k \leq n} \langle M_{i0}, M_{k1} \rangle = 0$$

(c)

Since $N_i(t) = N_{i0}(t) + N_{i1}(t)$ always jumps in $t = U_i$ no matter the value of T we can follow the argument of seeing $dN_i(t)$ as a bernoulli variable. Since the hazard in both cases is $\gamma(s)$ one with probability $e^{-\theta t}$ the other with $1 - e^{-\theta t}$ we get

$$\mathbb{E}(dN_i(t)) = Y_i(t) \left[\gamma(t)(1 - e^{-\theta t})dt + \gamma(t)e^{-\theta t}dt \right] = Y_i(t)\gamma(t)dt$$

Where $Y_i(t) = 1(t \leq U_i)$. Hence the compensator of $N_i(t)$ is

$$\Lambda_i(t) = \int_0^t Y_i(s)\gamma(s)ds$$

With $Y(t) = \sum_{i=1}^{n} Y_i(t)$ we then have that the compensator for $N(t)$ is given by

$$\Lambda(t) = \int_0^t Y(s)\gamma(s)ds$$

From this we can write the decomposition:

$$M(t) = N(t) - \Lambda(t)$$

which is a locally square integrable martingale with respect to $\mathcal{F}_t$ Further

$$N(t) = \int_0^t Y(s)\gamma(s)ds + M(t) \iff dN(t) = Y(t)d\Gamma(t) + dM(t)$$

Since $dM(t)$ is a zero-mean process this motivates the estimating equation of $\Gamma(t) = \int_0^t \gamma(s)ds$:

$$\hat{\Gamma}(t) = \int_0^t \frac{1}{Y(s)} dN(s)$$

(d)

We now assume that we observe in $[0, t]$. Let $t_1, t_2, ..., t_n$ be the corresponding realisations of T_i. Assume that the realized sample of u_i has k samples outside of $[0, t]$ hence we have the order $u_{(1)} < u_{(2)} < \cdots < u_{(n-k-1)} < u_{(n-k)} < u_{(n-k+1)} = u_{(n-k+2)} = \ldots u_{(n)} = t$. we have 3 cases:

1)

For samples of u_i that do not fall in the interval we only know that $u_i > t$ which happens with probability/has likelihood

$$P(U_i > t) = \exp\left(- \int_0^t \gamma(s)ds \right)$$

from the assumtions of the assignment. This factor will enter k times.

2)

For the $n-k$ samples within the interval, we can have $u_i \leq t_i \Leftrightarrow \delta_i = 0$ Then the sample (u_i, δ_i) has likelihood given by hazard times survival evaluated in u_i times the probability that $u_i \leq T_i$ which is $e^{-\theta u_i}$ hence we have likelihood of $(u_i, \delta_i = 0)$:

$$\gamma(u_i) \exp\left(-\int_0^{u_i} \gamma(s)ds\right) e^{-\theta u_i}$$

3)

On the other hand we can have $t_i < u_i \Leftrightarrow \delta_i = 1$ Then the sample $(u_i, \delta_i = 1)$ has likelihood given by hazard times survival evaluated in u_i times the probability that $T_i < u_i$ which is $1 - e^{-\theta u_i}$ hence we get the likelihood of $(u_i, \delta_i = 1)$:

$$\gamma(u_i) \exp\left(-\int_0^{u_i} \gamma(s)ds\right) \left\{1 - e^{-\theta u_i}\right\}$$

Now, a neat way of book keeping the survival terms, which depend on the realised u_i's is through the at risk indicator, $Y(s) = \sum_{i=1}^n 1(s \leq U_i)$ using the ordering of u_i's given above, with k u_i's larger than t, we have:

$$\sum_{i=1}^n \int_0^{u_i} \gamma(s)ds = \underbrace{n}_{=Y(u_{(1)})} \int_0^{u_{(1)}} \gamma(s)ds + \underbrace{(n-1)}_{=Y(u_{(2)})} \int_{u_{(1)}}^{u_{(2)}} \gamma(s)ds + \cdots + \underbrace{k}_{=Y(t)} \int_{u_{(n-k)}}^{t} \gamma(s)ds = \int_0^t Y(s)\gamma(s)ds$$

Hence all of the terms of the form $\exp\left(-\int_0^{u_i} \gamma(s)ds\right)$ can be moved outside and kept in the term $\exp\left(-\int_0^t Y(s)\gamma(s)ds\right)$. Case 2 and 3 only contributes if $u_i \leq t$ hence these are exponentiated with the indicatorfunction $1(u_i \leq t)$. Hence we get the likelihood, written with random variables instead of realised data:

$$L_t(\gamma(s), \theta; (U_i, \delta_i), i = 1, ..., n) = L_t = \exp\left(-\int_0^t Y(s)\gamma(s)ds\right) \prod_{i=1}^n \left\{\gamma(U_i)(1 - e^{-\theta U_i})^{\delta_i} e^{-\theta U_i(1-\delta_i)}\right\}^{1(U_i \leq t)}$$

(e)

Now, denote the score process

$$U_t(\theta) = \frac{\partial}{\partial \theta} \log L_t$$

Starting by taking log of L_t we get:

$$\log L_t = -\int_0^t Y(s)\gamma(s)ds + \sum_{i=1}^n 1(U_i \leq t) \log\left[\gamma(U_i)(1 - e^{-U_i\theta})^{\delta_i} e^{-U_i\theta(1-\delta_i)}\right] \tag{1}$$

$$= -\int_0^t Y(s)\gamma(s)ds + \sum_{i=1}^n 1(U_i \leq t) \log \gamma(U_i) + \sum_{i=1}^n 1(U_i \leq t)\delta_i \log(1 - e^{-U_i\theta}) + \sum_{i=1}^n 1(U_i \leq t)(-U_i\theta(1-\delta_i)) \tag{2}$$

Differentiating with respect to θ the 2 first terms do not depend on θ, hence

$$\frac{\partial}{\partial \theta} \log L_t = \sum_{i=1}^n 1(U_i \leq t)\delta_i \frac{U_i e^{-U_i\theta}}{(1 - e^{-U_i\theta})} + \sum_{i=1}^n 1(U_i \leq t)(-U_i(1-\delta_i)) \tag{3}$$

Now, for the first sum notice that only terms with $\delta_i = 1$ are included, with $u_i \leq t$. Since the counting process $N_{\cdot 1}(s)$ only jumps if $T_i < U_i$ in the point $s = u_i$ we have that the sum can be written as an intergral wrt $N_{\cdot 1}(s)$ from 0 to t with integrand equal to sumnand, with points u_i replaces with variable s i.e.

$$\sum_{i=1}^{n} 1(U_i \leq t)\delta_i \frac{U_i e^{-U_i\theta})}{(1-e^{-U_i\theta})} = \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{\cdot 1}(s)$$

For the second sum the argument is the same, since only terms with $U_i \leq t$ is included, we need integrate from 0 to t, and by integrating wrt $N_{\cdot 0}(s)$ only terms with $\delta_i = 0 \Leftrightarrow U_i \leq T_i$ are included. Since the sum is over $-U_i$ the integrand is simply $-s$. i.e.

$$\sum_{i=1}^{n} 1(U_i \leq t)(-U_i(1-\delta_i)) = \int_0^t -s dN_{\cdot 0}(s)$$

and we therefore get that the score can be written as the sum of integrals integrated wrt the two counting processes.

$$\frac{\partial}{\partial\theta} \log L_t = \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{\cdot 1}(s) - \int_0^t s dN_{\cdot 0}(s)$$

The to find the compensator we argue using the intensity function. An argument, albeit heurestic, that the intensity function is zero and the compensatror therefore the zero-process, is to view the jumps of the score given $\mathcal{F}_{u-}$. we see that

$$\mathbb{E}\left[d\int_0^u \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{i1}(s)|\mathcal{F}_{u-}\right] = Y_i(u)(1-e^{-\theta u})\gamma(u)du \cdot u\frac{e^{-\theta u}}{1-e^{-\theta u}} = Y_i(u)\gamma(u)du \cdot ue^{-\theta u}$$

The term $Y_i(u)$ comes from the fact that the subject must be at risk for a jump to be able to occur, in that case a jump has probability $\gamma(u)dt$. The jump is only counted if $T_i < u$ which has probability $(1-e^{-\theta u})$. Finally the jump will be of size equal to the integrand.
Similairly,

$$\mathbb{E}\left[d\int_0^u s dN_{i0}(s)|\mathcal{F}_{u-}\right] = Y_i(u)e^{-\theta u}\gamma(u)du \cdot u$$

From this we see that the difference in zero, hence the increments of $\frac{\partial}{\partial\theta}\log L_t$ is a zero-mean process. Addin to this that the process is clearly adapted, due to $N_{\cdot 0}$ and $N_{\cdot 1}$ being adapted, and clearly integrable we get that it is in fact a martingale.
Loosely speaking the difference in probability of jumps with $\delta_i = 1, 2$ is being offset by the size of the jumps.

(f)

We start by looking at the quadratic covariation process of $U_t(\theta)$ and $\tilde{M}(t)$ i.e.

$$\left[U_t(\theta), \tilde{M}(t)\right] = \sum_{s\leq t} \Delta U_t(s)\Delta\tilde{M}(s)$$

the key observation is that jumps in $\tilde{M}(t)$ occur through the first term, given by jumps in $N(t)$, the jumps are of sizes $\frac{1}{Y(s)}$. Secondly $U_t(\theta)$ jumps as $N_{\cdot 1}(t)$ with sizes $t\frac{e^{-t\theta}}{1-e^{-t\theta}}$ and on the other hands with jumps as in $N_{\cdot 0}(t)$ with sizes $-t$. Combining these observations we see that

$$\left[U_t(\theta), \tilde{M}(t)\right] = \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}}\frac{1}{Y(s)} dN_{\cdot 1}(s) - \int_0^t s\frac{1}{Y(s)} dN_{\cdot 0}(s).$$

Arguing as in (c) we see that the compensator is the zero process. Since the predictive covariation process in this case is the compensator of the quadratic covariation process, i.e. zero, we have orthogonality between $U_t(\theta)$ and $\tilde{M}(t)$.

For $\langle U_t(\theta)\rangle$ we get:

$$\langle U_t(\theta)\rangle = \left\langle \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{\cdot 1}(s) - \int_0^t s dN_{\cdot 0}(s), \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{\cdot 1}(s) - \int_0^t s dN_{\cdot 0}(s) \right\rangle \tag{4}$$

$$= \left\langle \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{\cdot 1}(s) \right\rangle - 2\left\langle \int_0^t \frac{se^{-s\theta}}{1-e^{-s\theta}} dN_{\cdot 1}(s), \int_0^t s dN_{\cdot 0}(s) \right\rangle + \left\langle \int_0^t s dN_{\cdot 0}(s) \right\rangle \tag{5}$$

(g)

we have from the decomposition of $N = M + \Lambda$ that

$$\hat{\Gamma}(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s) + \int_0^t \frac{J(s)}{Y(s)} d\Lambda(s) = \int_0^t \frac{J(s)}{Y(s)} dM(s) + \int_0^t \frac{J(s)}{Y(s)} Y(s)\gamma(s) ds = \int_0^t \frac{J(s)}{Y(s)} dM(s) + \int_0^t J(s)\gamma(s) ds$$

hence

$$n^{1/2}\left(\hat{\Gamma}(t) - \Gamma^*(t)\right) = n^{1/2} \int_0^t \frac{J(s)}{Y(s)} dM(s) = \int_0^t \frac{J(s)}{n^{-1/2}Y(s)} dM(s)$$

now since M is a martingale and the integrand is clearly locally bounded and predictable, the above is a locally square integrable martingale, following from Theorem 2.2.2 in MS.
We wish to use the martingale CLT, Theorem 2.5.1 and will therefore give some heurestic arguments, leaving out rigid arguments for the sake of time.

First, lets look at the predictable variation process

$$\left\langle \int_0^t \frac{J(s)}{n^{-1/2}Y(s)} dM(s) \right\rangle = \int_0^t \frac{J(s)}{n^{-1}Y^2(s)} d\langle M\rangle(s)$$

where we have used one of the statements of Theorem 2.2.2.
Since $[M](t) = N(t)$, because M only jumps when N does, we get that $\langle M\rangle(t)$ is the compensator of $[M](t)$ which is $\Lambda(t) = \int_0^t Y(s) d\Gamma(s)$

Hence

$$\begin{aligned}
\int_0^t \frac{J(s)}{n^{-1}Y^2(s)} d\langle M\rangle(s) &= \int_0^t \frac{J(s)}{n^{-1}Y^2(s)} d\Lambda(s) \\
&= \int_0^t \frac{J(s)}{n^{-1}Y^2(s)} Y(s)\gamma(s) d(s) \\
&= \int_0^t \frac{J(s)}{n^{-1}Y(s)} \gamma(s) d(s)
\end{aligned}$$

Now, $J(s) = 1(Y(s) > 0)$ indicates whether there are still subjects at risk. If the hazard $\gamma(s)$ is positive in s it follows that $P(J(s) = 1) \to 1$ as $n \to \infty$.

as for $\frac{1}{n}Y(s)$, the iid property of $U_i, i = 1, ..., n$ makes us able to invoke the Law of Large numbers yielding:

$$\frac{1}{n}Y(s) = \frac{1}{n}\sum_{i=1}^n 1(s \le U_i) \xrightarrow{as} \mathbb{E}1(s \le U) = P(s \le U)$$

if we insert this we get that for n large the integral is

$$\int_0^t \frac{\gamma(s)}{P(s \le U)} ds = \int_0^t \frac{f_U(s)}{P(s \le U)^2} ds$$

where $f_U(s)$ denotes the density of U, where we have used the definition of the hazard as the ratio between the density and survival function. Now the antiderivative is seen by noting:

$$\frac{d}{ds}\left(\frac{1}{P(s \leq U)}\right) = \frac{-1}{P(s \leq U)^2}\frac{d}{ds}P(s \leq U) = \frac{f_U(s)}{P(s \leq U)^2}$$

hence the integral becomes

$$\begin{aligned}
\int_0^t \frac{f_U(s)}{P(s \leq U)^2} ds &= \left[\frac{1}{P(s \leq U)}\right]_{s=0}^{s=t} \\
&= \frac{1}{P(t \leq U)} - \frac{1}{P(0 \leq U)} \\
&= \frac{1}{P(t \leq U)} - 1 \\
&= \frac{1 - P(t \leq U)}{P(t \leq U)} \\
&= \frac{P(U < t)}{P(U \geq t)}
\end{aligned}$$

To invoke the Martingale CLT we still need to argue that the jumps sizes of the process becomes neglible, as n tends to infinity. This should be clear from the fact that we normalize with $n^{1/2}$. Hence, from the martingale CLT we have that

$$n^{1/2}\left(\hat{\Gamma}(t) - \Gamma^*(t)\right) \xrightarrow{D} U(t)$$

where U is a Gaussian Martingale with variance function $V(t) = \frac{P(U<t)}{P(U \geq t)}$. Which is consistently estimated by the quadratic variation process

$$\left[\int_0^t \frac{J(s)}{n^{-1/2}Y(s)} dM(s)\right] = \int_0^t \frac{J(s)}{n^{-1}Y^2(s)} d\left[M\right](s) = \int_0^t \frac{J(s)}{n^{-1}Y^2(s)} dN(s)$$

(h)

It was already estabilised that U_t was a martingale, the question now is whether we can normalize it in some way such that we will have convergence in distribution og U_t when n tends to infinity.

since

$$n^{-1/2}U_t(\theta) = \frac{1}{n^{1/2}}\int_0^t s\frac{e^{-\theta s}}{1 - e^{-\theta s}} dN_{\cdot 1}(s) - \frac{1}{n^{1/2}}\int_0^t s dN_{\cdot 0}(s)$$

we have that the left integrand is bounded from above by 1 (when $s = 0$) while the second integrand is bounded by t. Hence the jump sizes will tend to zero when we normalise by the factor $n^{-1/2}$.

inserting the middle point $\Gamma(t) = \int_0^t \gamma(s)ds$

$$\begin{aligned}
n^{1/2}\tilde{M}(t) &= n^{1/2}\left(\hat{\Gamma}(t) - \Gamma^*(t)\right) \\
&= n^{1/2}\left(\hat{\Gamma}(t) - \Gamma(t) + \Gamma(t) - \Gamma^*(t)\right)
\end{aligned}$$

(i)

(j)

Practical part

2.1

Setup

We start by simulating the data as given in the exam. T and U are standard exponential i.e.

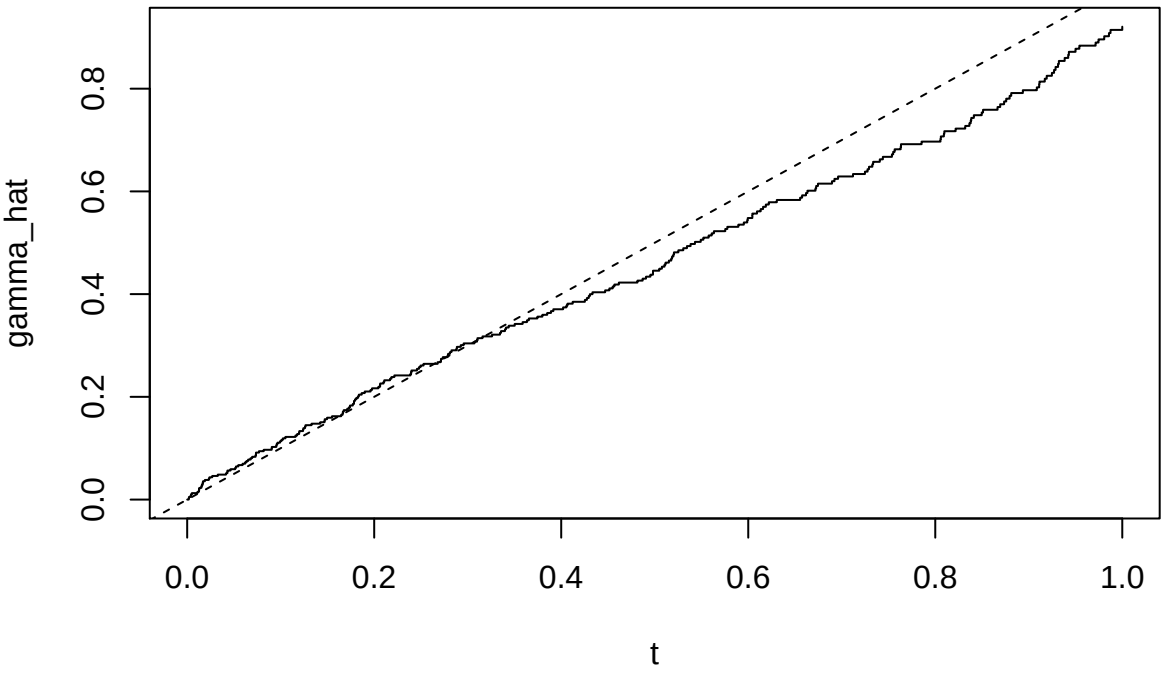
$$\theta = 1, \quad \gamma(s) = 1$$

hence $\Gamma(t) = \int_0^t \gamma(s)ds = t$

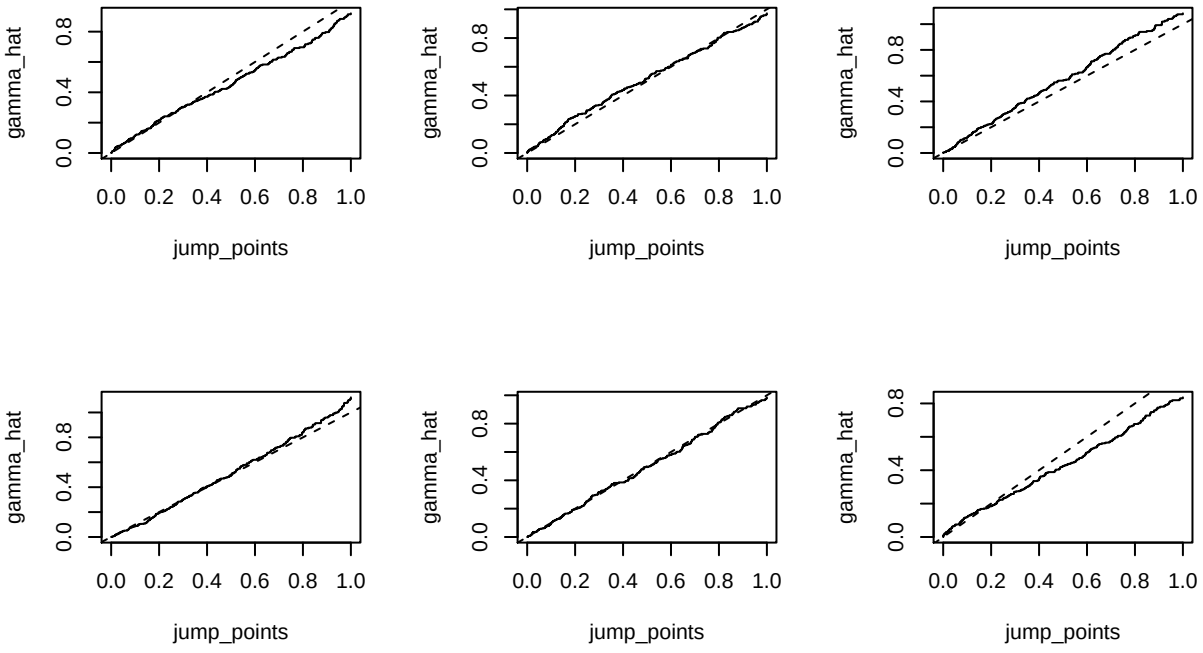
```
n <- 400
T <- rexp(n = n, rate = 1)
U <- rexp(n = n, rate = 1)
delta <- as.numeric(T < U)
tau <- 1

#Observed data:
U.obs <- U*as.numeric(U < tau) + tau*as.numeric(U >= tau)
#delta is set missing/999 is U is not observed
delta.obs <- delta*as.numeric(U.obs < tau) + 999*as.numeric(U >= tau)
#status as to wheter U was observed or not
status <- as.numeric(U.obs < tau)
```

We calculate $\hat{\Gamma}(t) = \int_0^t \frac{1}{Y(s)} dN(s)$ by noting that this is simply a sum adding one over the number of subjects at risk `sum(U.obs >= s)` at each timepoints of jumps of $N(t) = N_{\cdot 0}(t) + N_{\cdot 1}(t)$. Hence we calculate the vector of timepoints, and the cumulative reciprocal at risk set.

```
jump_points <- unique(sort(U.obs))

risk_inv <- 1/sapply(X = jump_points, FUN = function(s) sum(U.obs >= s))

#add 0 point with 0 cum hazard
jump_points <- c(0, jump_points)
gamma_hat <- c(0,cumsum(risk_inv))

plot(jump_points, gamma_hat, type = "s", xlab = "t")
abline(0,1, lty = 2)
```



since $\hat{\Gamma(t)}$ was our estimator of $\Gamma(t) = \int_0^t \gamma(s)ds$ which for U standard exponential gives $\Gamma(t) = \int_0^t \gamma(s)ds = \int_0^t ds = t$ the estimator should lie around the line $y = t$ i.e. slope 1 and intercept 0. Which it seems to do pretty well. Lets run just a couple of times to be sure.



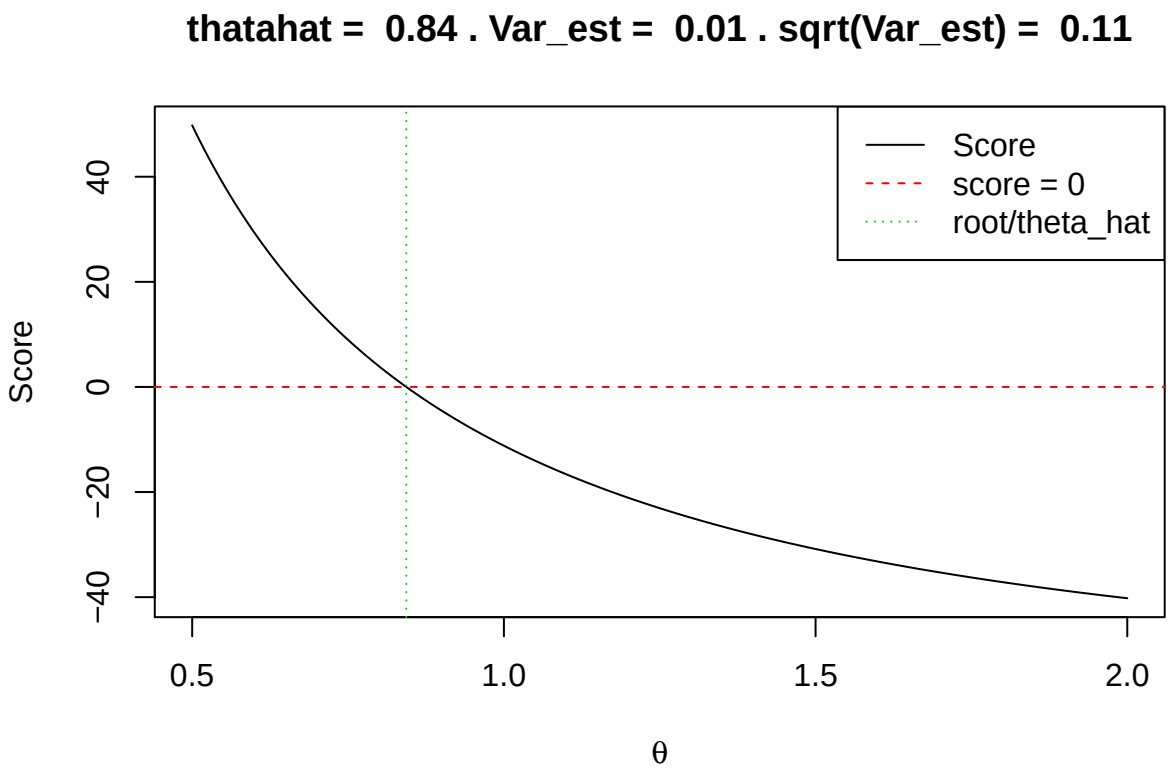
(b)

recall one of the forms of the score, which will be more useful to us for programming.

$$U_t(\theta) = \frac{\partial}{\partial\theta} \log L_t = \sum_{i=1}^{n} 1(U_i \leq t)\delta_i \frac{U_i e^{-U_i\theta}}{(1 - e^{-U_i\theta})} - \sum_{i=1}^{n} 1(U_i \leq t)(U_i(1 - \delta_i))$$

The R code for calculating and solving the score equation, which we do by the base R function `uniroot` - and doing plots is given below, with further explanation in the comments:

```
#first pick out indices of terms going into left- resp. right sum. then do calculation
Utheta <- function(theta, t = tau, UU, DD){
  index_l <- which(UU <= t & DD == 1)
  index_r <- which(UU <= t & DD == 0)
  #return U(theta)
  sum(UU[index_l]*exp(-UU[index_l]*theta) / (1 - exp(-UU[index_l]*theta))) -
    sum(UU[index_r])
}

#find root
theta_hat <- uniroot(f =function(x) Utheta(theta = x, t = 1, UU = U.obs, DD = delta.obs),
                     interval = c(0.1, 4))$root

#calculate var estimate
#Need estimate of Gamma:
jump_points <- unique(sort(U.obs))
risk_inv <- 1/sapply(X = jump_points, FUN = function(s) sum(U.obs >= s))
gamma_hat <- cumsum(risk_inv)

#Var estimate
var <- 1/n*1/sum(
  jump_points^2*exp(-theta_hat*jump_points)/(1-exp(-theta_hat*jump_points))*
  exp(-gamma_hat)*risk_inv
)
#Standard error
sd <- sqrt(var)

#plot the score as function of theta, and the theta solveing U(theta) = 0
theta_seq <- seq(from = 0.5, to = 2, length.out = 1e3)
Ut_seq <- sapply(X = theta_seq,
                 FUN = function(x) Utheta(theta = x, UU = U.obs, DD = delta.obs))

plot(theta_seq, Ut_seq,
     type = "l", xlab = expression(theta), ylab = "Score",
     main = paste("thatahat = ", round(theta_hat,2),
                  ". Var_est = ", round(var,2),
                  ". sqrt(Var_est) = ", round(sd,2)))

legend("topright", c("Score", "score = 0", "root/theta_hat"), col = 1:3, lty = 1:3)
abline(h=0, lty =2, col = 2)
abline(v = theta_hat, lty = 3, col = 3)
```



(c)

We now do the above exercise repeatedly 2000 times, to simulate out the variability of $\hat{\theta}$ and to asses if it is central. The base-R function `replicate` is very useful for tasks like this.

```
hh <- replicate(n = 2000, expr = {
  #data sim
n <- 400
T <- rexp(n = n, rate = 1)
U <- rexp(n = n, rate = 1)
delta <- as.numeric(T < U)
tau <- 1

#Observed data:
U.obs <- U*as.numeric(U < tau) + tau*as.numeric(U >= tau)
#delta is set missing/999 is U is not observed
delta.obs <- delta*as.numeric(U.obs < tau) + 999*as.numeric(U >= tau)

#
jump_points <- unique(sort(U.obs))
risk_inv <- 1/sapply(X = jump_points, FUN = function(s) sum(U.obs >= s))
gamma_hat <- cumsum(risk_inv)

#Calculate Theta hat
theta_hat <- uniroot(f =function(x) Utheta(theta = x, t = 1, UU = U.obs, DD = delta.obs),
                     interval = c(0.1, 4))$root

#Var estimate
var <- 1/n*1/sum(
```

```
  jump_points^2*exp(-theta_hat*jump_points)/(1-exp(-theta_hat*jump_points))*
  exp(-gamma_hat)*risk_inv
)
#standard error
sd <- sqrt(var)

#return the estimate of theta and estimated SD given data
c(theta_hat, sd)

})

mean(hh[1,]) #mean of 2000 runs of theta hat
```

```
## [1] 1.005189
```

```
sd(hh[1,])   #sd of 2000 runs of theta hat
```

```
## [1] 0.1132163
```

```
mean(hh[2,]) #mean of 2000 calculations of SD of theta hat
```

```
## [1] 0.1139038
```

We see that the estimator fares quite well, on average being very close to the true value of $\theta = 1$ and the standard error of the estimator produces reliable results close to the sample standard error of the estimate which is seen to be around 0.11 - this would justify e.g. creating confidence intervals or simple wald-type hypothesis testing for $\hat{\theta}$ based on the estimator.