C.1 Modern Survival Problem sheet 1: Counting processes and martingales

(1) (a) We have a point process with intensity

$$\lambda(t) = \begin{cases} 1 & \text{if } 0 \le t < 5, \\ 2 & \text{if } 5 \le t < 10, \\ 0 & \text{otherwise.} \end{cases}$$

Let $(N(t))_{t\ge 0}$ be the counting process and let T_i be the time of the i-th event, or ∞ if there is no i-th event.

i. Find the distribution of the total number of events.

The cumulative intensity is

$$\Lambda(t) = \begin{cases} t & \text{if } t \le 5, \\ 2t - 5 & \text{if } 5 < t \le 10, \\ 15 & \text{if } t > 10. \end{cases}$$

The distribution of the total number of events is Poisson with parameter 15.

ii. Find the distribution of $T_2 - T_1$.

The inverse cumulative intensity is

$$\Lambda^{-1}(s) = \begin{cases} s & \text{if } s \le 5, \\ \frac{s}{2} + 2.5 & \text{if } 5 < s \le 15 \\ \infty & \text{if } s > 15. \end{cases}$$

We may think of $N(t)$ as $N'(\Lambda(t))$, where N' is a counting process of a Poisson process with unit intensity. Thus

$$\begin{aligned} \mathbb{P}\{T_2 - T_1 \ge t\} &= \mathbb{P}\{\Lambda^{-1}(T_2') - \Lambda^{-1}(T_1') \ge t\} \\ &= \int_0^{15} \mathbb{P}\{T_2' \ge \Lambda\big(t + \Lambda^{-1}(s)\big) \mid T_1' = s\} e^{-s} ds \\ &= \int_0^{15} e^{-\Lambda\left(t+\Lambda^{-1}(s)\right)+s} e^{-s} ds \\ &= \int_0^{5} e^{-\Lambda(t+s)} ds + \int_5^{15} e^{-\Lambda(t+s/2+2.5)} ds. \end{aligned}$$

For $t < 5$ this is

$$\begin{aligned}\int_t^5 e^{-s}ds + \int_5^{5+t} e^{-(2s-5)}ds + \int_5^{15-2t} e^{-2t-s)}ds + \int_{15-2t}^{15} e^{-15}ds \\ = \left(e^{-t} - e^{-5}\right) + \frac{1}{2}\left(e^{-5} - e^{-5-2t}\right) + \left(e^{-5-2t} - e^{-15}\right) + 2te^{-15} \\ = e^{-t} - \frac{1}{2}e^{-5} + \frac{1}{2}e^{-5-2t} + (2t-1)e^{-15}.\end{aligned}$$

For $10 \geq t \geq 5$ it is

$$\int_t^{10} e^{-2s+5} + (t+5)e^{-15} = \frac{1}{2}e^{-2t+5} + (t+4.5)e^{-15}$$

Finally, for $t \geq 10$ we have $\mathbb{P}\{T_2 - T_1 \geq t\} = 15e^{-15}$.
This may seem not quite right, since we might be inclined to say

$$\mathbb{P}\{T_2 = \infty\} = \mathbb{P}\{N(\infty) \leq 1\} = 16e^{-15}.$$

The difference is the event $\{T_1 = \infty\}$, which has probability e^{-15}. The calculation above integrated over values of T_1 on $(0, \infty)$, so implicitly excluded the event $\{T_1 = \infty\}$. In fact, it is not clear that $T_2 - T_1$ is defined on this event.

iii. Show that $T_i \to \infty$ in probability.
For any fixed positive real K,

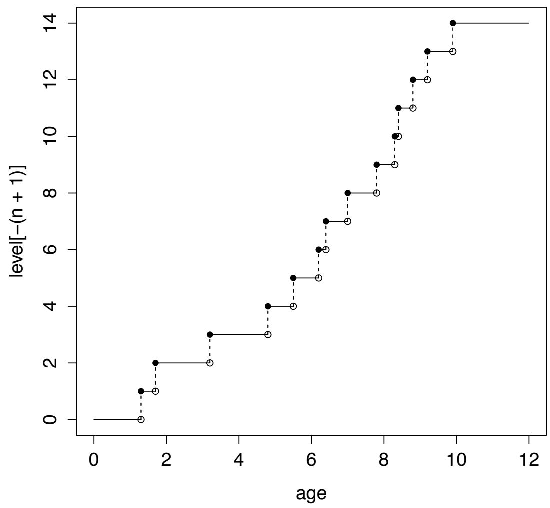
$$\mathbb{P}\{T_i \leq K\} = \mathbb{P}\{N(K) \geq i\} \leq \mathbb{P}\{N(\infty) \geq i\} \xrightarrow{i\to\infty} 0.$$

(b) i. NOTE: This question was a bit inconsistent. The numbers relate to a version of the question in which you were supposed to sketch this particular realisation. I changed it to sketching a "typical realisation?? but left the numbers in. Apologies if this was confusing.

ii. What is the compensator $A(t)$ for $N(t)$?
The compensator for the counting process is the cumulative hazard rate

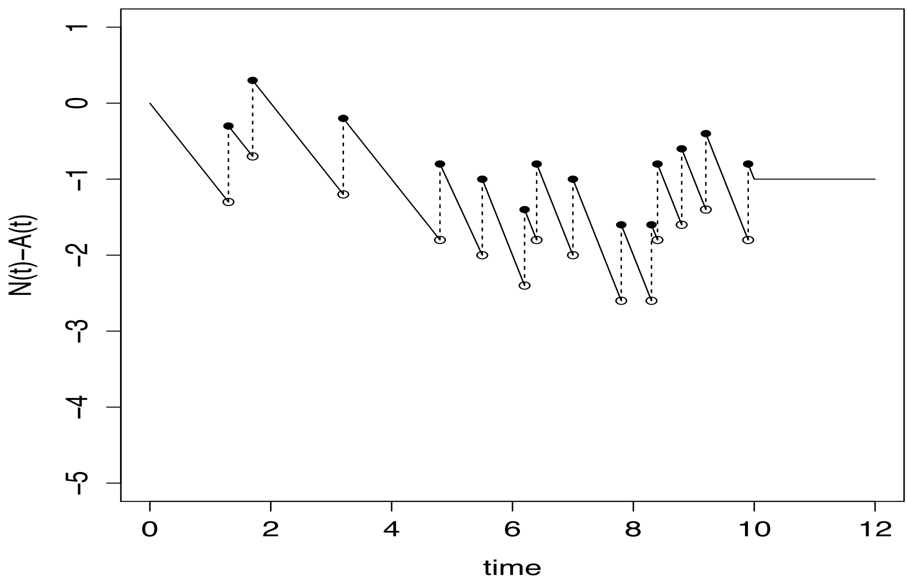
$$A(t) = \Lambda(t) = \begin{cases} t & \text{if } 0 \geq t < 5 \\ 2t - 5 & \text{if } 5 \geq t < 10 \\ 15 & t \geq 10. \end{cases}$$

iii. Sketch $N(t)$.
The counting process is $N(t) = \sum_{i=1}^n \mathbf{1}_{\{T_i \leq t\}}$



iv. Sketch the martingale associated with $N(t)$.

The martingale associated with $N(t)$ is $N(t) - A(t)$



(2) Let λ be any positive function and $\Lambda(t) = \int_0^t \lambda(s)ds$. Suppose X is a random variable with exponential distribution with parameter 1. Show that $\Lambda^{-1}(X)$ is a random variable with hazard rate $\lambda(t)$. We have $X \sim \exp(1)$. We want the distribution of $Y = \Lambda^{-1}(X)$.

cumintensproblem

If we let F_X and F_Y be the corresponding cdfs, we have $F_X(x) = \mathbb{P}\{X \leq x\} = 1 - e^{-x}$, so

$$\begin{aligned} F_Y(y) &= \mathbb{P}\{Y \leq y\} \\ &= \mathbb{P}\{\Lambda^{-1}(X) \leq y\} \\ &= \mathbb{P}\{X \leq \Lambda(y)\} \text{ because } \Lambda \text{ is strictly increasing} \\ &= 1 - e^{-\Lambda(y)}, \end{aligned}$$

which is the cdf of a random variable with hazard rate λ.

(3) Suppose X and Y are a pair of random variables with joint density $f(x, y)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a function such that $\mathbb{E}[|g(X)|] < \infty$, and let $\mathcal{Y}$ be the sigma algebra generated by Y. Show that $E[g(X) \mid \mathcal{Y}] = h(Y)$, where

$$h(y) := \frac{\int_{-\infty}^{\infty} g(x)f(x, y)dx}{\int_{-\infty}^{\infty} f(x, y)dx}.$$

Explain the relationship between this formula the conditional expectations you learned in prelims and Part A probability.

Trivially, $h(Y) \in \mathcal{Y}$, so we need to show that for any other random variable $Z \in \mathcal{Y}$, $\mathbb{E}[g(X)Z] = \mathbb{E}[h(Y)Z]$. Note that

$$h(y) := \frac{\int_{-\infty}^{\infty} f(x, y)g(x)dx}{f_Y(y)},$$

where $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$ is the marginal density of Y,

We may write a random variable $Z \in \mathcal{Y}$ as $z(Y)$, for some function $z : \mathbb{R} \to \mathbb{R}$. Thus $h(Y)Z = h(Y)z(Y)$, and

$$\begin{aligned} \mathbb{E}\big[h(Y)Z\big] &= \int_{-\infty}^{\infty} f_Y(y)h(y)z(y)dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)g(x)z(y)dxdy \\ &= \mathbb{E}\big[g(X)z(Y)\big] \\ &= \mathbb{E}\big[g(X)Z\big]. \end{aligned}$$

(4) Let $(N(t))$ be the counting process associated with the Poisson process with intensity λ, and let $(M(t)) = (N(t) - \lambda t)$ be the associated martingale. Show that λt is the compensator for M^2.

Trivially, λt is predictable. (It is both continuous and deterministic.) To prove that λt is a compensator for $M(t)^2$, we need to prove that $M(t)^2 - \lambda t$ is a martingale. Let $s < t$, then

$$
\begin{aligned}
M(t)^2 &= (N(t) - \lambda t)^2 \\
&= N(t)^2 - 2\lambda t N(t) + (\lambda t)^2 \\
&= \big(N(t) - N(s)\big)^2 + 2\big(N(t) - N(s)\big)\big(N(s) - \lambda t\big) + N(s)^2 + (\lambda t)^2
\end{aligned}
$$

Conditioned on $\mathcal{F}_s$, $N(t) - N(s)$ has Poisson distribution with parameter $\lambda(t - s)$, so

$$
\begin{aligned}
E\Big[\big[N(t) - N(s)\big]^2 \big| \mathcal{F}_s\Big] &= Var\Big[N(t) - N(s) \big| \mathcal{F}_s\Big] + E\Big[N(t) - N(s) \big| \mathcal{F}_s\Big]^2 \\
&= \lambda(t - s) + (\lambda(t - s))^2
\end{aligned}
$$

Thus

$$
\begin{aligned}
E\big[M(t)^2 - \lambda t | \mathcal{F}_s\big] &= \lambda(t - s) - \lambda t + [\lambda(t - s)]^2 + 2(\lambda(t - s))\big(N(s) - \lambda t\big) + N(s)^2 - 2\lambda t N(s) + (\lambda t)^2 \\
&= N(s)^2 - 2\lambda s N(s) + (\lambda s)^2 - \lambda s \\
&= M(s)^2 - \lambda s.
\end{aligned}
$$

(5) Later in the course we will discuss *current status data*, which is a form of extreme censoring. Individuals have an unobserved (assumed i.i.d.) event time U_i. What is observed is a census time C_i (independent of U_i), and $\delta_i = \mathbf{1}_{\{U_i \le C_i\}}$.

(a) Assuming U_i has density f_λ and cdf F_λ, write an expression for the log likelihood of U_i.

(b) Suppose the distribution is exponential, so $f_\lambda(u) = \lambda e^{-\lambda u}$. What is the relative efficiency of estimation (that is, ratio of expected Fisher information) based on the current status data, compared with complete observation of U_i?

(c) Suppose you can choose the distribution of C_i (but still independent of U_i). How would you maximise the expected information?

(a)

$$
\ell(\lambda) = \sum_{i=1}^{n} \delta_i \log F_\lambda(C_i) + (1 - \delta_i) \log(1 - F_\lambda(C_i)).
$$

(b)

$$
\begin{aligned}
\ell(\lambda) &= \sum_{i=1}^{n} \delta_i \log\left(1 - e^{-\lambda C_i}\right) - \lambda \sum_{i=1}^{n} (1 - \delta_i) C_i. \\
\ell'(\lambda) &= \sum_{i=1}^{n} \delta_i \frac{C_i}{e^{\lambda C_i} - 1} - \sum_{i=1}^{n} (1 - \delta_i) C_i). \\
\ell''(\lambda) &= -\sum_{i=1}^{n} \delta_i \frac{C_i^2 e^{\lambda C_i}}{(e^{\lambda C_i} - 1)^2}.
\end{aligned}
$$

The expected information is thus

$$\begin{aligned}\mathcal{I}(\lambda) &= n\mathbb{E}\left[\frac{C^2 e^{\lambda C}}{(e^{\lambda C}-1)^2}\mathbf{1}_{U<C}\right]\\ &= n\mathbb{E}\left[\frac{C^2 e^{\lambda C}}{(e^{\lambda C}-1)^2}\mathbb{P}\{U<C\,|\,C\}\right]\\ &= n\mathbb{E}\left[\frac{C^2 e^{\lambda C}}{(e^{\lambda C}-1)^2}\left(1-e^{-\lambda C}\right)\right]\\ &= n\mathbb{E}\left[\frac{C^2}{e^{\lambda C}-1}\right].\end{aligned}$$

With complete observation the expected information is n/λ^2. Thus, the relative efficiency is

$$\frac{\lambda^2 C^2}{e^{\lambda C}-1}.$$

(c) The function $x^2/(e^x-1)$ has a unique maximum value about 0.628 at $x_0 \approx 1.544$. Thus the relative efficiency is no more than 0.628, and may be made as close as we like to that value by making C close to being deterministically $1.544/\lambda$. Of course, we can't do that without knowing λ, in which case we wouldn't need to do the experiment. One could imagine using an adaptive procedure, where $\hat{\lambda}^{(k)}$ is the MLE based on the first k observations, and then we choose $C_{k+1} = 1.544/\hat{\lambda}^{(k)}$.

(6) Suppose we have an inhomogeneous Poisson process $N(t)$, whose intensity starts out as either 1 or 2, each with probability $1/2$, and immediately after each event the intensity is determined again by an independent coin flip. Let $\lambda(t)$ be the intensity at time t.

(a) Suppose $\lambda(t)$ is observed (i.e., $\lambda(t) \in \mathcal{F}_t$). Find the compensator (i.e., the cumulative intensity process) of $N(t)$. Find the predictable variation process and the optional variation process for the martingale $M(t)$ obtained by subtracting the compensator from $N(t)$. Compute $\mathrm{Var}(M(t))$.

Let $(\mathcal{G}_t)$ be the natural filtration of $(N(t))$, and let $\mathcal{F}_t = \mathcal{G}_t \vee \langle\{\lambda(s) : s \le t\}\rangle$; that is, $\mathcal{F}_t$ is obtained from $\mathcal{G}_t$ by adding the variables $\lambda(s)$ for $s \le t$. $\lambda^{\mathcal{F}}(t) = \lambda(t) \in \mathcal{F}_t$, so

$$\Lambda^{\mathcal{F}}(t) = \int_0^t \lambda(s)ds = t + \int_0^t \mathbf{1}_{\{\lambda(s)=2\}}ds.$$

Defining the martingale $M(t) = N(t) - \Lambda(t)$,

$$\begin{aligned}d\langle M\rangle(t) &= \mathrm{Var}(dM(t)|\mathcal{F}_t)\\ &= \lambda(t)dt\\ \langle M\rangle(t) &= \Lambda(t)\end{aligned}$$

$$\begin{aligned}[M](t) &= \sum_{T_i \le t} dM(t)^2 \\ &= N(t) \qquad \text{since the jumps have size 1.}\end{aligned}$$

$$\mathrm{Var}\big([M(t)]\big) = \mathbb{E}\Big[\langle M\rangle(t)\Big] = \mathbb{E}\Big[\int_0^t \lambda(s)ds\Big] = \int_0^t \mathbb{E}\big[\lambda(s)\big]ds.$$

You might think that $\mathbb{E}[\lambda(s)] = 3/2$, since each time a new λ is chosen, it has equal chances of being 1 or 2. But in fact, when λ takes on the value 2 it spends only half as long on average in that state as when it takes on the value 1, so we may expect that $\mathbb{E}[\lambda(t)] \xrightarrow{t\to\infty} 4/3$. Formally, we observe that $\lambda(t)$ is a Markov chain on the states $\{1,2\}$, with Q-matrix

$$Q = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}.$$

The transition probabilities at time t are given by

$$e^{tQ} = \frac{1}{3}\begin{pmatrix} 2+e^{-3t} & 1-e^{-3t} \\ 2-2e^{-3t} & 1+2e^{-3t} \end{pmatrix}.$$

Given that we start in the distribution $(1/2, 1/2)$, the distribution at time t is $(2/3 - e^{-3t}/6, 1/3 + e^{-3t}/6)$. Thus

$$\mathbb{E}\big[\lambda(t)\big] = \frac{4}{3} + \frac{e^{-3t}}{6},$$

and finally

$$\mathrm{Var}\big(M(t)\big) = \int_0^t \left(\frac{4}{3} + \frac{e^{-3s}}{6}\right) ds = \frac{4}{3}t + \frac{1}{18}\left(1 - e^{-3t}\right).$$

(b) Now suppose $\lambda(t)$ is unobserved (i.e., $\lambda(t) \notin \mathcal{F}_t$). Find the compensator. Find the predictable variation process and the optional variation process of $\widetilde{M}$, where $\widetilde{M}$ is obtained by subtracting the compensator from $N(t)$. Is $\mathrm{Var}(\widetilde{M}(t)) = \mathrm{Var}(M(t))$? Why or why not?

By the Innovation Theorem,

$$\begin{aligned}\lambda^{\mathcal{G}}(t) &= \mathbb{E}\big[\lambda^{\mathcal{F}}(t) \,\big|\, \mathcal{G}_t\big] \\ &= 1 + \mathbb{P}\big\{\lambda(t) = 2 \,\big|\, \mathcal{G}_t\big\}\end{aligned}$$

We know that $\lambda(t)$ is independent of anything that happened before time $T_*(t)$, so conditioning on $\mathcal{G}_t$ is equivalent to conditioning on the stopping time $T_*(t) :=$ last jump before time t, or, equivalently, conditioning on $\tau(t) := t - T_*(t) \in \mathcal{G}_t$. That is, if we write $T^*(t)$ for the first event time after t,

$$\begin{aligned}\mathbb{P}\{\lambda(t) = 2 \,\big|\, \mathcal{G}_t\} &= \mathbb{P}\{\lambda(t) = 2 \,\big|\, \tau\} \\ &= \mathbb{P}\{\lambda(T_*) = 2 \,\big|\, T^* - T_* > \tau\} \\ &= \frac{1/2 \cdot e^{-2\tau}}{1/2 \cdot e^{-\tau} + 1/2 \cdot e^{-2\tau}} \\ &= \frac{e^{-\tau}}{e^{-\tau} + 1}.\end{aligned}$$

Note that

$$\int_0^t \frac{e^{-s}}{e^{-s} + 1} d\tau = \log \frac{2}{1 + e^{-t}},$$

so if we define the inter-event times $\tau_i := T_i - T_{i-1}$, we get

$$\Lambda(t) = \log \frac{2}{1 + e^{-\tau(t)}} + \sum_{i:T_i \le t} \log \frac{2}{1 + e^{-\tau_i}}.$$

This is the compensator.
Thus $\widetilde{M}(t) = N(t) - \Lambda(t)$ is a martingale, which has predictable variation $\Lambda(t)$.
The optional variation does not depend on the choice of σ-algebra, and is still $[\widetilde{M}](t) = N(t)$.
Despite the fact that $\widetilde{M}$ is a very different process from M, the variance of $M(t)$ is the same as that of $\widetilde{M}(t)$, since both are equal to $\mathbb{E}[N(t)]$.