MATH-449 - Biostatistics
EPFL, Spring 2021
Problem Set 1

1. A survival time T is exponentially distributed with rate parameter $\beta > 0$ if its survival function, $S(t) = P(T > t)$, takes the form $S(t) = e^{-\beta t}$ for $t \geq 0$.

   a) Find the density function $f(t) = -\frac{d}{dt}S(t)$.

   b) Find the hazard function and the cumulative hazard function.

   c) A waiting time T is *memoryless* if $P(T > t + s|T > t) = P(T > s)$ for all $t, s \geq 0$, i.e. if the waiting time distribution does not depend on how much time has already elapsed. Show that an exponentially distributed waiting time is memoryless.

   Solution

   a) $f(t) = -\frac{d}{dt}S(t) = \beta e^{-\beta t}$.

   b) Let $\alpha(t)$ denote the hazard function. From the relation $f(t) = \alpha(t)e^{-\int_0^t \alpha(s)ds}$ we see that $\alpha(t) = \beta$. The cumulative hazard function is $\int_0^t \alpha(s)ds = \beta t$.

   c) By Bayes' theorem we have

$$\begin{aligned} P(T > t + s|T > t) &= \frac{P(T > t + s, T > t)}{P(T > t)} \\ &= \frac{P(T > t + s)}{P(T > t)}, \end{aligned}$$

   where we used that the event $\{T > t + s, T > t\}$ is identical to the event $\{T > t + s\}$ to obtain the last line. Writing out the survival function of an exponentially distributed variable we get that the last line equals $e^{-\beta(t+s)}/e^{-\beta t} = e^{-\beta s} = P(T > s)$.

2. a) Find $E[T]$ when T is a Weibull distributed variable, i.e. when the hazard function of T is $\alpha(t) = \lambda k t^{k-1}$ for $\lambda, k > 0$‡

   b) (Exercise 1.3 in ABG 2008) Suppose T is a survival time with finite expectation. Show that†

$$E[T] = \int_0^\infty P(T > s)ds.$$

   Solution

   a) The density function is $\lambda k t^{k-1} e^{-\lambda t^k}$, which gives

$$E[T] = \int_0^\infty t \cdot \lambda k t^{k-1} e^{-\lambda t^k} dt.$$

   Performing the substitution $u = \lambda t^k$, $dt = \frac{du}{ku}\left(\frac{u}{\lambda}\right)^{1/k}$, gives

$$\begin{aligned} E[T] &= \int_0^\infty k \cdot u e^{-u} \cdot \frac{du}{ku}(\frac{u}{\lambda})^{1/k} \\ &= \lambda^{-1/k} \int_0^\infty u^{1/k} e^{-u} du \\ &= \lambda^{-1/k} \Gamma\big(1 - \frac{1}{k}\big). \end{aligned}$$

‡Hint: Express the solution using the gamma function, which is given by $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$.
†Hint: Write $T = \int_0^\infty I(T > u)du$.

b) Using the identiy $T = \int_0^\infty I(T > s)ds$, the result follows from

$$\begin{aligned} E[T] &= E\Big[\int_0^\infty I(T > s)ds\Big] \\ &= \int_0^\infty E[I(T > s)]ds \\ &= \int_0^\infty P(T > s)ds, \end{aligned}$$

where we used that $E[T] < \infty$ to exchange integration and expectation.

3. (Exercise 2.1 in ABG 2008) Let M_n be a discrete time martingale with respect to the filtration $\mathcal{F}_n$, for $n \in \{0, 1, 2, \cdots\}$. By definition of M being a martingale we have that $E[M_n|\mathcal{F}_{n-1}] = M_{n-1}$ for all $n \geq 1$. Show that this is equivalent to $E[M_n|\mathcal{F}_m] = M_m$ whenever $n \geq m \geq 0$.

   Solution For $n = m$ the result follows from the fact that M_n is adapted (i.e. measurable with respect to $\mathcal{F}_n$), and thus that $E[M_n|\mathcal{F}_n] = M_n$. For $n > m$ we have by the law of total expectation and the definition of a martingale that

$$\begin{aligned} E[M_n|\mathcal{F}_m] &= E\big[E[M_n|\mathcal{F}_{n-1}]|\mathcal{F}_m\big] \\ &= E[M_{n-1}|\mathcal{F}_m]. \end{aligned}$$

   The last step can repeated until we are left with $E[M_m|\mathcal{F}_m]$.

4. (Exercise 2.4 in ABG 2008) Let M be as in the previous problem, and suppose $M_0 = 0$. Prove that $M^2 - \langle M \rangle$ is a martingale with respect to the filtration $\mathcal{F}$, that is, that $M_0^2 - \langle M \rangle_0 = 0$ and that $E[M_n^2 - \langle M \rangle_n|\mathcal{F}_{n-1}] = M_{n-1}^2 - \langle M \rangle_{n-1}$

   Solution We have $\langle M \rangle_n = \sum_{i=1}^n E[(M_i - M_{i-1})^2|\mathcal{F}_{i-1}]$ and $\langle M \rangle_0 = 0$ by definition, which means that $M_0^2 - \langle M \rangle_0 = 0$. For $n > 0$,

$$\begin{aligned} E\big[M_n^2 - \langle M \rangle_n|\mathcal{F}_{n-1}\big] &= E\Big[(M_n - M_{n-1} + M_{n-1})^2 - E[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}] - \langle M \rangle_{n-1}\Big|\mathcal{F}_{n-1}\Big] \\ &= E\Big[(M_n - M_{n-1})^2 + M_{n-1}^2 - E[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}]\Big|\mathcal{F}_{n-1}\Big] - \langle M \rangle_{n-1} \\ &= E\Big[(M_n - M_{n-1})^2 - E[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}]\Big|\mathcal{F}_{n-1}\Big] + M_{n-1}^2 - \langle M \rangle_{n-1} \\ &= M_{n-1}^2 - \langle M \rangle_{n-1} \end{aligned}$$

   where we used that $E[(M_n - M_{n-1})M_{n-1}] = 0$, and that $\langle M \rangle_{n-1}$ and M_{n-1}^2 are $\mathcal{F}_{n-1}$-measurable.

5. Suppose we have n independent survival times $\{T_i\}_{i=1}^n$, where T_i corresponds to the time of death of individual i. Suppose we somehow could observe each individual from $t = 0$ up to his/her time of death.

   In the lectures you learned that a counting process $\{N(t)\}_{t \geq 0}$ is an increasing right-continuous integer-valued stochastic process such that $N(0) = 0$. Write down the counting process N_i^c (that "counts" the death of individual i) in terms of T_i.

   Solution N_i^c has at most one jump, at time T_i. As $N_i^c(t)$ is right-continuous it must take the form $N_i^c(t) = I(T_i \leq t)$

   You will also learn about the *intensity process* λ of a counting process N with respect to a filtration $\mathcal{F}$. It is informally defined through the relationship $\lambda(t)dt = E[dN(t)|\mathcal{F}_t]$.

   In general, if the intensity $\lambda(t)$ of a counting process $N(t)$ with respect to $\mathcal{F}_t$ can be written on the form

$$\lambda(t) = \alpha(t) \cdot Y(t),$$

where α is an unknown deterministic function and Y is an $\mathcal{F}_t$-predictable[§] function that does not depend on α, $N(t)$ is said to satisfy the *multiplicative intensity model*[*].

6. (Exercise 1.10 in ABG 2008) Consider the scenario in Exercise 5, and let $\mathcal{F}_t^c$ be the filtration generated by $\{N_i^c(s), s \leq t, i = 1, \cdots, n\}$. In the lectures we will see that the intensity of N_i^c with respect to $\mathcal{F}^c$ in this case is $\lambda_i^c(t) = E[dN_i^c(t)|\mathcal{F}_t^c] = \alpha_i(t)Y_i(t)$, where $\alpha_i(t)$ is the hazard function of individual i and $Y_i(t) = I(T_i \geq t)$. Consider the aggregated counting process $N^c(t) = \sum_{i=1}^n N_i^c(t)$.

   i) Let $\{\eta_i(t)\}_{i=1}^n$ be known, positive, continuous functions. Find the intensity process of N^c with respect to $\mathcal{F}_t^c$ when α_i take the following forms:

      a) $\alpha_i(t) = \alpha(t)$
      b) $\alpha_i(t) = \eta_i(t)\alpha(t)$
      c) $\alpha_i(t) = \alpha(t) + \eta_i(t)$

   ii) For which of the three cases in i) does N^c satisfy the multiplicative intensity model?

Solution We have

$$\begin{aligned}\lambda^c(t)dt &= E[dN^c(t)|\mathcal{F}_t^c] \\ &= \sum_{i=1}^n E[dN_i^c(t)|\mathcal{F}_t^c] \\ &= \sum_{i=1}^n \alpha_i(t)Y_i(t)dt.\end{aligned}$$

Define $Y(t) = \sum_{i=1}^n Y_i(t)$ and $Y_\eta(t) = \sum_{i=1}^n \eta_i(t)Y_i(t)$. Note first that Y_i is left-continuous. Also, all the information needed to determine whether T_i have happened at t is contained in $\mathcal{F}_t^c$. We thus have

i) a) $\lambda^c(t) = \alpha(t)Y(t)$
   b) $\lambda^c(t) = \alpha(t)Y_\eta(t)$
   c) $\lambda^c(t) = \alpha(t)Y(t) + Y_\eta(t)$.

ii) a) Yes
   b) Yes
   c) No. For any representation $\lambda^c(t) = \tilde{\alpha}(t)\tilde{Y}(t)$, either $\tilde{\alpha}$ will not be deterministic or $\tilde{Y}$ will be a function of α.

[§]Recall that this holds when Y is left-continuous and adapted to $\mathcal{F}$, i.e. that all the information needed to know the value of Y at time t is contained in $\mathcal{F}_t$.

[*]We will later derive estimators for the unknown function α under the multiplicative intensity model.