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.

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[†]

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

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 $n \geq m \geq 0$.

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 $E(M_n^2 - \langle M \rangle_n | \mathcal{F}_{n-1}) = M_{n-1}^2 - \langle M \rangle_{n-1}$

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.

   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 $N^c(t) = \sum_{i=1}^n N_i^c(t)$.

[‡]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 $T = \int_0^\infty I(T > u)du$.

[§]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.

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?