M378K Introduction to Mathematical Statistics
Problem Set #11
Order Statistics.

Problem 11.1. *An insurance company is handling claims from two categories of drivers: the good drivers and the bad drivers. The waiting time for the first claim from a* **good** *driver is modeled by an exponential random variable T_g with mean 6 (in years). The waiting time for the first claim from a* **bad** *driver is modeled by an exponential random variable T_b with mean 3 (in years). We assume that the random variables T_g and T_b are independent.*

What is the distribution of the waiting time T until the first claim occurs (regardless of the type of driver this claim was filed by)?

Solution: Formally, we have that $T = \min\{T_b, T_g\}$, and the image of T is $[0, \infty)$. Let us calculate the cdf of the random variable T. For every $t \geq 0$, we have

$$F_T(t) = \mathbb{P}[T \leq t] = 1 - \mathbb{P}[T > t] = 1 - \mathbb{P}[\min\{T_b, T_g\} > t] = 1 - \mathbb{P}[T_b > t, T_g > t].$$

Due to the independence of T_b and T_g and the fact that $T_b \sim E(\tau_b)$ and $T_g \sim E(\tau_g)$, with $\tau_g = 6$ and $\tau_b = 3$, we can write

$$\mathbb{P}[T_b > t, T_g > t] = \mathbb{P}[T_b > t]\mathbb{P}[T_g > t] = e^{-\frac{t}{\tau_b}} e^{-\frac{t}{\tau_b}} = e^{-\left(\frac{1}{\tau_g}+\frac{1}{\tau_b}\right)t}.$$

So, $F_T(t) = 1 - e^{-\left(\frac{1}{\tau_g}+\frac{1}{\tau_b}\right)t}$, and $T \sim E\left(1/\left(\frac{1}{\tau_g} + \frac{1}{\tau_b}\right)\right)$, i.e., $T \sim E(\tau = 2)$.

Definition 11.1. *Let $Y_1, \ldots, Y_n$ be a* **random sample**. *The random sample ordered in an increasing order is called an* order statistic *and denoted by*

$$Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}.$$

Question Write $Y_{(1)}$ as a function of $Y_1, Y_2, \ldots, Y_n$.
Solution:

$$Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$$

Question Write $Y_{(n)}$ as a function of $Y_1, Y_2, \ldots, Y_n$.
Solution:

$$Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$$

Problem 11.2. *What is the distribution function of the random variable $Y_{(n)}$?*

Solution: Let $y \in \mathbb{R}$. Then,

$$\begin{aligned}
F_{Y_{(n)}}(y) &= \mathbb{P}[Y_{(n)} \leq y] \\
&= \mathbb{P}[\max(Y_1, Y_2, \ldots, Y_n) \leq y] \\
&= \mathbb{P}[Y_1 \leq y, Y_2 \leq y, \ldots, Y_n \leq y] \\
&= \mathbb{P}[Y_1 \leq y]\mathbb{P}[Y_2 \leq y] \ldots \mathbb{P}[Y_n \leq y] \\
&= (\mathbb{P}[Y_1 \leq y])^n = (F_Y(y))^n.
\end{aligned}$$

Problem 11.3. *Assume that the random sample comes from a density f_Y. Is the r.v. $Y_{(n)}$ continuous? If so, what is its density $g_{(n)}$?*

Solution: For every y where F_Y is differentiable, we have

$$g_{(n)}(y) = f_{Y_{(n)}}(y) = \frac{d}{dy} F_{Y_{(n)}}(y) = \frac{d}{dy}\left((F_Y(y))^n\right) = n f_Y(y)(F_Y(y))^{n-1}.$$

Problem 11.4. *What is the distribution function of the random variable $Y_{(1)}$?*

Solution: For $y \in \mathbb{R}$, we have that

$$\begin{aligned}
F_{Y_{(1)}}(y) &= \mathbb{P}[Y_{(1)} \leq y] \\
&= \mathbb{P}[\min(Y_1, \ldots, Y_n) \leq y] \\
&= 1 - \mathbb{P}[\min(Y_1, \ldots, Y_n) > y] \\
&= 1 - \mathbb{P}[Y_1 > y, Y_2 > y, \ldots, Y_n > y] \\
&= 1 - \mathbb{P}[Y_1 > y]\mathbb{P}[Y_2 > y] \ldots \mathbb{P}[Y_n > y] \\
&= 1 - (\mathbb{P}[Y_1 > y])^n \\
&= 1 - (1 - \mathbb{P}[Y_1 \leq y])^n \\
&= 1 - (1 - F_Y(y))^n
\end{aligned}$$

Problem 11.5. *Assume that the random sample comes from a density f_Y. Is the r.v. $Y_{(1)}$ continuous? If so, what is its density $g_{(1)}$?*

Solution: For every y where F_Y is differentiable, we have

$$g_{(1)}(y) = f_{Y_{(1)}}(y) = \frac{d}{dy} F_{Y_{(1)}}(y) = \frac{d}{dy}\left(1 - (1 - F_Y(y))^n\right) = n f_Y(y)(1 - F_Y(y))^{n-1}.$$

Theorem 11.2. *Lt $Y_1, \ldots, Y_n$ be independent, identically distributed random variables with the common cumulative distribution function F_Y and the common probability density function f_Y. Let $Y_{(k)}$ denote the k^{th} order statistic and let $g_{(k)}$ denote its probability density function. Then,*

$$g_{(k)}(y) = \frac{n!}{(k-1)!(n-k)!}(F_Y(y))^{k-1} f_Y(y)(1 - F_Y(y))^{n-k} \quad \textit{for all } y \in \mathbb{R}.$$