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* $\boxed{\textit{independent.}}$

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)?

→: $T = \min(T_g, T_b)$ $\quad S_T = [0, +\infty)$

$$F_T(t) = \mathbb{P}[T \le t] = \mathbb{P}[\min(T_g, T_b) \le t] = 1 - \mathbb{P}[\min(T_g, T_b) > t]$$
$$= 1 - \mathbb{P}[T_g > t, T_b > t] = 1 - \mathbb{P}[T_g > t] \cdot \mathbb{P}[T_b > t]$$
$$= 1 - e^{-t/\tau_g} \cdot e^{-t/\tau_b} = 1 - e^{-t(1/\tau_g + 1/\tau_b)}$$

$T \sim E(\tau)$ w/ $\tau = 1/(1/\tau_g + 1/\tau_b) = 1/\tfrac{1}{2} = 2$ $\square$

$\boxed{10/18/2024.}$

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

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

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

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

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

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

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

$\longrightarrow$: For $y \in \mathbb{R}$: $F_{Y_{(n)}}(y) = \mathbb{P}[Y_{(n)} \leq y] = \mathbb{P}[\max(Y_1, \dots, Y_n) \leq y]$

$$= \mathbb{P}[Y_1 \leq y, Y_2 \leq y, \dots, Y_n \leq y] = \text{(independence)}$$

$$= \mathbb{P}[Y_1 \leq y] \cdot \mathbb{P}[Y_2 \leq y] \cdots \mathbb{P}[Y_n \leq y] = \left(\begin{smallmatrix}\text{ident.}\\ \text{dist'd}\end{smallmatrix}\right)$$

$$= (\mathbb{P}[Y_1 \leq y])^n = \boxed{(F_Y(y))^n} \quad \square$$

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 $g_{(n)}$?*

$\longrightarrow$: For all y where F_Y is differentiable:

$$g_{(n)}(y) = \frac{d}{dy} F_{Y_{(n)}}(y) = \frac{d}{dy}\left((F_Y(y))^n\right) =$$

$$= n \cdot f_Y(y) \cdot (F_Y(y))^{n-1}$$

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

$\longrightarrow$: For $y \in \mathbb{R}$: $F_{Y_{(1)}}(y) = \mathbb{P}[Y_{(1)} \leq y] = \mathbb{P}[\min(Y_1, Y_2, \dots, Y_n) \leq y]$

$$= 1 - \mathbb{P}[\min(Y_1, Y_2, \dots, Y_n) > y] = 1 - \mathbb{P}[Y_1 > y, \dots, Y_n > y]$$

independence $= 1 - \mathbb{P}[Y_1 > y] \cdots \mathbb{P}[Y_n > y]$

identic. dist'n $= 1 - (\mathbb{P}[Y_1 > y])^n$

$$= \boxed{1 - (1 - F_Y(y))^n}$$

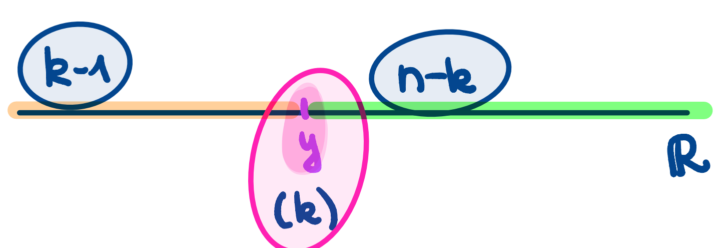
Problem 11.5. *Assume that the random sample comes from a density f_Y. Is the r.v. $Y_{(1)}$ continuous? If so, what is its density $g_{(1)}$?*

$\longrightarrow$:

$$g_{(1)}(y) = \frac{d}{dy} F_{Y_{(1)}}(y) =$$

$$= \frac{d}{dy}\left(1 - (1 - F_Y(y))^n\right) = + n(1 - F_Y(y))^{n-1} \cdot (+1) \cdot f_Y(y)$$

$$= n f_Y(y) (1 - F_Y(y))^{n-1}$$



Theorem 11.2. *Lt $Y_1, \dots, Y_n$ 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,*

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

Checking!

$$g_{(1)}(y) = \frac{n!}{(n-1)!}\, n \; \cancel{(F_Y(y))^0}\, f_Y(y)\, (1 - F_Y(y))^{n-1} \qquad ✓$$

$$g_{(n)}(y) = \frac{n!}{(n-1)!}\, n \; (F_Y(y))^{n-1} f_Y(y)\, \cancel{(1 - F_Y(y))^{n-n}} \qquad ✓$$

M378K Introduction to Mathematical Statistics
Problem Set #12
Statistics.

Definition 12.1. *A* random sample *of size n from distribution D is a random vector*

$$(Y_1, Y_2, \dots, Y_n)$$

such that

1. *$Y_1, Y_2, \dots, Y_n$ are independent, and*
2. *each Y_i has the distribution D.*

Example 12.2. Quality control. *Times until a breaker trips under a particular load are modeled as exponential. The intended procedure is to choose n breakers at random from the assembly line, subject them to the load, and measure the time it takes for them to trip. The lifetime of a specific breaker indexed by i is a random variable Y_i with an exponential distribution with an unknown parameter $\theta = \tau$. Independence of Y_i, $i = 1, \dots, n$ is assured by the random choice of breakers to test.*

Definition 12.3. *A* statistic *is a function of the (observable) random sample and known constants.*

Problem 12.1. *Give at least three examples of statistics of a certain random sample $Y_1, Y_2, \dots, Y_n$.*

$\rightarrow$:

- $\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$
- $S'^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$
- $\ln\left((Y_1 \cdots Y_n)^{1/n}\right)$ geometric average

Not a statistic.
$\tau + SD(Y_1, \dots, Y_n)$

Remark 12.4. *Statistics are random variables in their own right. We call their probability distributions* sampling distributions.

Example 12.5. Quality control, cont'd. *Let the random variable Y be the minimum of random variables $Y_1, \dots, Y_n$, i.e., the shortest time until the breaker is tripped in the sample. We can write*

$$Y = \min(Y_1, \dots, Y_n).$$

What is another name for this random variable? $Y_{(1)}$ first order statistic

$$F_Y(y) = 1 - e^{-y/\tau}$$

Then, the sampling distribution *of Y can be figured out by looking at its cumulative distribution function. We have ...*

$$g_{(1)}(y) = n \cdot f_Y(y)\,(1 - F_Y(y))^{n-1} = n \cdot \frac{1}{\tau} e^{-y/\tau} \left(e^{-y/\tau}\right)^{n-1}$$

$$= \left(\frac{n}{\tau}\right) e^{-\frac{ny}{\tau}} \qquad Y_{(1)} \sim E(\tau/n)$$

Problem 12.2. *Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. What is the sampling distribution of*

$$\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k \quad ?$$