XXXIII. Order Statistics

Premise

- **Example question**: How tall will the tallest student in my next semester's probability class be?
- **Setting**: We have independent random variables $Y_1, \ldots, Y_n$, with identical distributions $F(y)$ and densities $f(y) = F'(y)$.
- **Definition**: $Y_{(1)} := \min\{Y_1, \ldots, Y_n\}$
- **Definition**: $Y_{(n)} := \max\{Y_1, \ldots, Y_n\}$
- **Question**: What are the distributions and densities of $Y_{(1)}$ and $Y_{(n)}$?

I saw that I have 24 names registered, but I didn't see who they are. (The bigger the class, the more chance you have to have a tall student.)
Why it's unpleasant: $Y_{(1)}$ and Y_1 are different!

$$\begin{aligned} Y_1 &= \text{first} \\ Y_{(1)} &= \text{smallest} \\ Y_n &= \text{last} \\ Y_{(n)} &= \text{largest} \end{aligned}$$

Formulas

$$
\begin{aligned}
F_{Y_{(n)}}(y) &:= P\left(Y_{(n)} < y\right) \\
&= \boxed{F(y)^n} \\
f_{Y_{(n)}}(y) &= F'_{Y_{(n)}}(y) \\
&= \boxed{nF(y)^{n-1}f(y)} \\
F_{Y_{(1)}}(y) &:= P\left(Y_{(1)} < y\right) \\
&= \boxed{1 - [1 - F(y)]^n} \\
f_{Y_{(1)}}(y) &= F'_{Y_{(1)}}(y) \\
&= \boxed{n\left[1 - F(y)\right]^{n-1} f(y)}
\end{aligned}
$$

$$
\begin{aligned}
Y_{(1)} &= \text{smallest} \\
Y_{(n)} &= \text{largest} \\
F &= \text{distribution} \\
f &= \text{density}
\end{aligned}
$$

Example I

24 students in a class each write a term paper. The papers' lengths are uniformly distributed from 0 to 7 pages. Find the distribution and density functions for the length of the <u>longest</u> paper.

Uniform distribution: $f(y) = \frac{1}{7}, 0 \le y \le 7$. $F(y) = \frac{y}{7}$.

$$\begin{aligned} F_{Y_{(n)}}(y) &= F(y)^n \\ f_{Y_{(n)}}(y) &= nF(y)^{n-1}f(y) \\ F_{Y_{(n)}}(y) &= \boxed{\left(\frac{y}{7}\right)^{24}} \\ f_{Y_{(n)}}(y) &= 24\left(\frac{y}{7}\right)^{23}\frac{1}{7} = \boxed{\frac{24y^{23}}{7^{24}}, 0 \le y \le 7} \end{aligned}$$

Example II

As in Example I, 24 students in a class each write a term paper. The papers' lengths are uniformly distributed from 0 to 7 pages. Find the distribution and density functions for the length of the shortest paper.

Uniform distribution: $f(y) = \frac{1}{7}, 0 \le y \le 7$. $F(y) = \frac{y}{7}$.

$$\begin{aligned} F_{Y_{(n)}}(y) &= F(y)^n \\ f_{Y_{(n)}}(y) &= nF(y)^{n-1}f(y) \\ F_{Y_{(1)}}(y) &= \boxed{1 - \left(\frac{7-y}{7}\right)^{24}} \\ f_{Y_{(1)}}(y) &= 24\left(\frac{7-y}{7}\right)^{23}\frac{1}{7} = \boxed{\frac{24}{7^{24}}(7-y)^{23}, 0 \le y \le 7} \end{aligned}$$

Example III

As in Example I, 24 students in a class each write a term paper. The papers' lengths are uniformly distributed from 0 to 7 pages. Find the mean and variance for the length of the longest paper.

$$\begin{aligned}
\textbf{Mean}: E\left[Y_{(n)}\right] &:= \int_0^7 y f_{Y_{(n)}}(y)\,dy \\
&= \frac{24}{7^{24}} \int_0^7 y^{24}\,dy \\
&= \frac{24}{25 \cdot 7^{24}}\, y^{25}\Big|_{y=0}^{y=7} \\
&= \boxed{\frac{24 \cdot 7}{25}} = \boxed{\frac{7n}{n+1}}
\end{aligned}$$

Variance:

$$\begin{aligned}
\sigma^2 &= \left(\int_0^7 y^2 f_{Y_{(n)}}(y)\,dy\right) - \mu^2 \\
&= \left(\frac{24}{7^{24}} \int_0^7 y^{25}\,dy\right) - \left(\frac{24 \cdot 7}{25}\right)^2 \\
&= \text{Maybe skip these steps.} \\
&= \left(\frac{24}{7^{24} \cdot 26}\, y^{26}\Big|_{y=0}^{y=7}\right) - \left(\frac{24 \cdot 7}{25}\right)^2 \\
&= \frac{24 \cdot 7^{26}}{7^{24} \cdot 26} - \frac{24^2 \cdot 7^2}{25^2} \\
&= \frac{24 \cdot 7^2}{26} - \frac{24^2 \cdot 7^2}{25^2} \\
&= \frac{24 \cdot 7^2 \left(25^2 - 24 \cdot 26\right)}{26 \cdot 25^2} \\
&= \frac{24 \cdot 7^2 \left(625 - 624\right)}{26 \cdot 25^2} \\
&= \frac{24 \cdot 7^2}{26 \cdot 25^2} \\
&= \boxed{\frac{49n}{(n+1)^2(n+2)}}
\end{aligned}$$

Example IV

A basketball team has 10 players (including reserves). For each player, the time until her next injury follows an exponential distribution with a mean of 5 years. Find the distribution and density and distribution functions for the time until the team's first injury.

$Y_i \sim$ Exponential(5):

$$\begin{aligned}
f(y) &= \frac{1}{5}e^{-\frac{y}{5}} \\
F(y) &= \int_0^y \frac{1}{5}e^{-\frac{t}{5}}\,dt = -e^{-\frac{t}{5}}\Big|_{t=0}^{t=y} = 1 - e^{-\frac{y}{5}} \\
F_{Y_{(1)}}(y) &= 1 - [1 - F(y)]^n = 1 - \left(e^{-\frac{y}{5}}\right)^{10} = \boxed{1 - e^{-2y}}
\end{aligned}$$

We could take a derivative here, but let's practice our formula:

$$f_{Y_{(1)}}(y) = n\,[1 - F(y)]^{n-1}\,f(y) = 10\left(e^{-\frac{y}{5}}\right)^9 \frac{1}{5}e^{-\frac{y}{5}} = \boxed{2e^{-2y}}$$

Example V

As in Example IV, a basketball team has 10 players. For each player, the time until her next injury follows an exponential distribution with a mean of 5 years. Find the expected time until the team's first injury.

$Y_i \sim \text{Exponential}(5)$:

$$\begin{aligned} f(y) &= \frac{1}{5}e^{-\frac{y}{5}} \\ f_{Y_{(1)}}(y) &= n\left[1 - F(y)\right]^{n-1} f(y) = 10\left(e^{-\frac{y}{5}}\right)^9 \frac{1}{5}e^{-\frac{y}{5}} = \boxed{2e^{-2y}} \end{aligned}$$

This is Exponential$\left(\frac{1}{2}\right)$, so the expected time is $\boxed{\text{six months}}$. In general, if $Y_1, \ldots, Y_n \sim$ Exponential(β), then $\boxed{Y_{(1)} \sim \text{Exponential}\left(\frac{\beta}{n}\right)}$.