Chapter 5. Multiple Random Variables

5.10: Order Statistics

Slides (Google Drive) | Alex Tsun | Video (YouTube)

We've talked a lot about the distribution of the sum of random variables, but what about the maximum, minimum, or median? For example, if there are 4 possible buses you could take, and the time until each arrives is independent with an exponential distribution, what is the expected time until the *first* one arrives? Mathematically, this would be $\mathbb{E}[\min\{X_1, X_2, X_3, X_4\}]$ if the arrival times were X_1, X_2, X_3, X_4.

In this section, we'll figure out how to find out the density function (and hence expectation/variance) of the minimum, maximum, median, and more!

5.10.1 Order Statistics

We'll first formally define order statistics.

Definition 5.10.1: Order Statistics

Suppose Y_1, ..., Y_n are iid *continuous* random variables with common PDF f_Y and common CDF F_Y. We define $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ to be the *sorted* version of this sample. That is,

$$Y_{\min} \equiv Y_{(1)} < Y_{(2)} < ... < Y_{(n)} \equiv Y_{\max}$$

$Y_{(1)}$ is the smallest value (the minimum), and $Y_{(n)}$ is the largest value (the maximum), and since they are so commonly used, they have special names $Y_{\min}$ and $Y_{\max}$ respectively.

Notice that we can't have equality because with continuous random variables, the probability that any two are equal is 0. So, we don't have to worry about any of these random variables being "less than or equal to" another.

Notice that each $Y_{(i)}$ is a random variable as well! We call $Y_{(i)}$ the **i-th order statistic**, i.e. the i-th smallest in a sample of size n. For example, if we had $n = 9$ samples, $Y_{(5)}$ would be the median value. We are interested in finding the distribution of each order statistic, and properties such as expectation and variance as well.

Why are order statistics important? Usually, we take the min, max or median of a set of random variables and do computations with them - so, it would be useful if we had a general formula for the PDF and CDF of the min or max.

We start with an example to find the distribution of $Y_{(n)} = Y_{\max}$, the largest order statistic. We'll then extend this to any of the order statistics (not just the max). Again, this means, if we were to repeatedly take the maximum of n iid RVs, what would the samples look like?

Example(s)

Let $Y_1, Y_2, \ldots, Y_n$ be iid continuous random variables with the same CDF F_Y and PDF f_Y. What is the distribution of $Y_{(n)} = Y_{\max} = \max\{Y_1, Y_2, \ldots, Y_n\}$ the largest order statistic?

Solution

We'll employ our typical strategy and work with probabilities instead of densities, so we'll start with the CDF:

$$\begin{aligned}
F_{Y_{\max}}(y) &= \mathbb{P}(Y_{\max} \le y) && \text{[def of CDF]} \\
&= \mathbb{P}\left(\bigcap_{i=1}^{n} \{Y_i \le y\}\right) && \text{[max is } \le y \text{ if and only if all are]} \\
&= \prod_{i=1}^{n} \mathbb{P}(Y_i \le y) && \text{[independence]} \\
&= \prod_{i=1}^{n} F_Y(y) && \text{[def of CDF]} \\
&= F_Y^n(y) && \text{[identically distributed, all have same CDF]}
\end{aligned}$$

We can differentiate the CDF to find the PDF:

$$\begin{aligned}
f_{Y_{\max}}(y) &= F'_{Y_{\max}}(y) \\
&= \frac{d}{dy}\left(F_Y^n(y)\right) \\
&= nF_Y^{n-1}(y) f_Y(y) && \left[\text{chain rule of calculus and } \frac{d}{dx}x^n = nx^{n-1}\right]
\end{aligned}$$

Let's take a step back and see what we just did here. We just computed the density function of the maximum of n iid random variables, denoted $Y_{\max} = Y_{(n)}$. We now need to find the density of any arbitrary ranked $Y_{(i)}$. □

Theorem 5.10.1: Order Statistics

Suppose Y_1, ..., Y_n are iid *continuous* random variables with common PDF f_Y and common CDF F_Y. We define $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ to be the *sorted* version of this sample. That is,

$$Y_{\min} \equiv Y_{(1)} < Y_{(2)} < \ldots < Y_{(n)} \equiv Y_{\max}$$

. The density function of $Y_{(i)}$ is

$$f_{Y_{(i)}}(y) = \binom{n}{i-1, 1, n-i} \cdot [F_Y(y)]^{i-1} \cdot [1 - F_Y(y)]^{n-i} \cdot f_Y(y), y \in \Omega_Y$$

Now, using the same intuition as before, we'll use an informal argument to find the density of a general $Y_{(i)}$, $f_{Y_{(i)}}(y)$. For example, this might help find the distribution of the minimum $f_{Y_{(1)}}$ or the median.

Proof of Density of Order Statistics. The formula above may remind you of a multinomial distribution, and you would be correct! Let's consider what it means for $Y_{(i)} = y$ (the i-th smallest value in the sample of n to equal a particular value y).

- One of the values needs to be exactly y
- $i-1$ of the values need to be smaller than y (this happens for each with probability $F_Y(y)$)
- the other $n-i$ values need to be greater than y (this happens for each with probability $1 - F_Y(y)$

Now, we have 3 distinct types of objects: 1 that is exactly y, $i-1$ which are less than y and $n-i$ which are greater. Using multinomial coefficients and the above, we see that

$$f_{Y_{(i)}}(y) = \binom{n}{i-1, 1, n-i} \cdot [F_Y(y)]^{i-1} \cdot [1 - F_Y(y)]^{n-i} \cdot f_Y(y)$$

Note that this isn't a probability; it is a density, so there is something flawed with how we approached this problem. For a more rigorous approach, we just have to make a slight modification, but use the same idea.

Re-Proof (Rigorous) This time, we'll find $\mathbb{P}\left(y - \frac{\varepsilon}{2} \leq Y_{(i)} \leq y + \frac{\varepsilon}{2}\right)$ and use the fact that this is approximately equal to $\varepsilon f_{Y_{(i)}}(y)$ for small $\varepsilon > 0$ (Riemann integral (rectangle) approximation from 4.1).

We have very similar cases:

- One of the values needs to be between $y - \frac{\varepsilon}{2}$ and $y + \frac{\varepsilon}{2}$ (this happens with probability approximately $\varepsilon f_Y(y)$, again by Riemann approximation).
- $i-1$ of the values need to be smaller than $y - \frac{\varepsilon}{2}$ (this happens for each with probability $F_Y(y - \frac{\varepsilon}{2})$)
- the other $n-i$ values need to be greater than $y + \frac{\varepsilon}{2}$ (this happens for each with probability $1 - F_Y(y + \frac{\varepsilon}{2})$)

Now these are actually probabilities (not densities), so we get

$$\mathbb{P}\left(y - \frac{\varepsilon}{2} \leq Y_{(i)} \leq y + \frac{\varepsilon}{2}\right) \approx \varepsilon f_{Y_{(i)}}(y) = \binom{n}{i-1, 1, n-i} \cdot [F_Y(y)]^{i-1} \cdot [1 - F_Y(y)]^{n-i} \cdot (\varepsilon f_Y(y))$$

Dividing both sides by $\varepsilon > 0$ gives the same result as earlier!

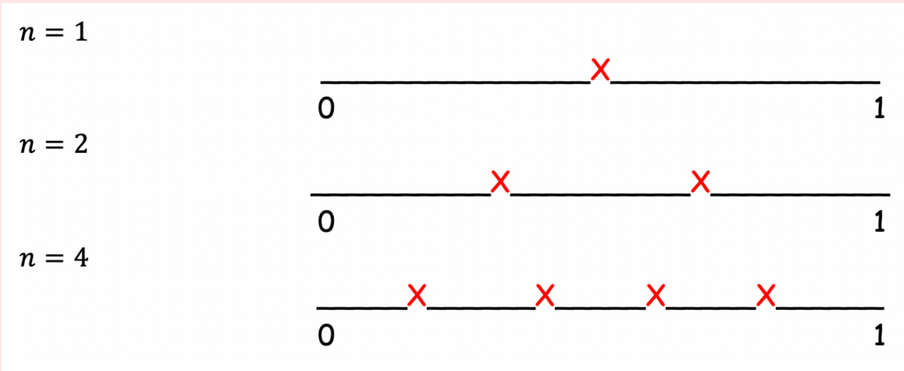
□

Let's verify this formula with our maximum that we derived earlier by plugging in n for i:

$$f_{Y_{\max}}(y) = f_{Y_{(n)}}(y) = \binom{n}{n-1, 1, 0} \cdot [F_Y(y)]^{n-1} \cdot [1 - F_Y(y)]^0 \cdot f_Y(y) = nF_Y^{n-1}(y) f_Y(y)$$

Example(s)

If $Y_1, ..., Y_n$ are iid Unif$(0,1)$, where do we "expect" the points to end up? That is, find $\mathbb{E}\left[Y_{(i)}\right]$ for any i. You may find this picture with different values of n useful for intuition.



Solution

Intuitively, from the picture, if $n = 1$, we expect the single point to end up at $\frac{1}{2}$. If $n = 2$, we expect the two points to end up at $\frac{1}{3}$ and $\frac{2}{3}$. If $n = 4$, we expect the four points to end up at $\frac{1}{5}$, $\frac{2}{5}$, $\frac{3}{5}$ and $\frac{4}{5}$.

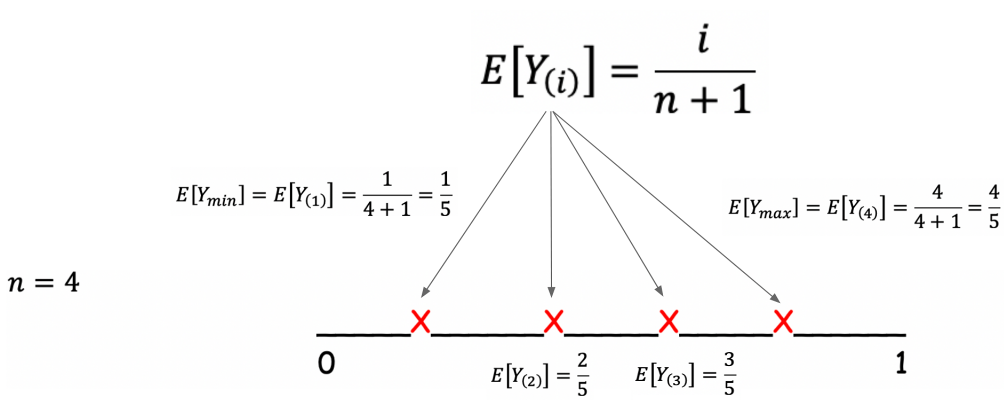
Let's prove this formally. Recall, if $Y \sim \text{Unif}(0,1)$ (continuous), then $f_Y(y) = 1$ for $y \in [0,1]$ and $F_Y(y) = y$ for $y \in [0,1]$. By the order statistics formula,

$$\begin{aligned} f_{Y_{(i)}}(y) &= \binom{n}{i-1, 1, n-i} \cdot [F_Y(y)]^{i-1} \cdot [1 - F_Y(y)]^{n-i} \cdot f_Y(y) \\ &= \binom{n}{i-1, 1, n-i} \cdot [y]^{i-1} \cdot [1-y]^{n-i} \cdot 1 \end{aligned}$$

Using the PDF, we find the expectation:

$$\mathbb{E}\left[Y_{(i)}\right] = \int_0^1 y \binom{n}{i-1, 1, n-i} \cdot [y]^{i-1} \cdot [1-y]^{n-i} dy = \frac{i}{n+1}$$

Here is a picture which may help you figure out what the formulae you just computed mean!



□

Now let's do our bus example from earlier.

Example(s)

At 5pm each day, four buses make their way to the HUB bus stop. Each bus would be acceptable to take you home. The time in hours (after 5pm) that each arrives at the stop is independent with $Y_1, Y_2, Y_3, Y_4 \sim \text{Exp}(\lambda = 6)$ (on average, it takes 1/6 of an hour (10 minutes) for each bus to arrive).

1. On Mondays, you want to get home ASAP, so you arrive at the bus stop at 5pm sharp. What is the expected time until the *first* one arrives?

2. On Tuesdays, you have a lab meeting that runs until 5:15 and are worried you may not catch any bus. What is the probability you miss all the buses?

Solution The first question asks about the smallest order statistic $Y_{(1)} = Y_{\min}$ since we care about the first bus. The second question asks about the largest order statistic $Y_{(4)}$ since we care about the last bus. Let's compute the general formula for order statistics first so we can apply it to both parts of the problem.

Recall, if $Y \sim \text{Exp}(\lambda = 6)$ (continuous), then $f_Y(y) = 6e^{-6y}$ for $y \in [0, \infty)$ and $F_Y(y) = 1 - e^{-6y}$ for $y \in [0, \infty)$. By the order statistics formula,

$$\begin{aligned} f_{Y_{(i)}}(y) &= \binom{n}{i-1, 1, n-i} \cdot [F_Y(y)]^{i-1} \cdot [1 - F_Y(y)]^{n-i} \cdot f_Y(y) \\ &= \binom{n}{i-1, 1, n-i} \cdot [1 - e^{-6y}]^{i-1} \cdot [e^{-6y}]^{n-i} \cdot 6e^{-6y} \end{aligned}$$

1. For the first part, we want $\mathbb{E}\left[Y_{(1)}\right]$, so we plug in $i = 1$ (and $n = 4$) to the above formula to get:

$$f_{Y_{(1)}}(y) = \binom{4}{1-1, 1, 4-1} \cdot [1 - e^{-6y}]^{1-1} \cdot [e^{-6y}]^{4-1} \cdot 6e^{-6y} = 4[e^{-18y}]6e^{-6y} = 24e^{-24y}$$

   Now we can use the PDF to find the expectation normally. However, notice that the PDF is that of an $\text{Exp}(\lambda = 24)$ distribution, so it has expectation 1/24. That is, the expected time until the first bus arrives is 1/24 an hour, or 2.5 minutes.

   Let's talk about something amazing here. We found that $\min\{Y_1, Y_2, Y_3, Y_4\} \sim \text{Exp}(\lambda = 4 \cdot 6)$; that the minimum of exponentials is distributed as an exponential with the sum of the rates! Why might this be true? If we have $Y_1, Y_2, Y_3, Y_4 \sim \text{Exp}(6)$, that means on average, 6 buses of each type arrive each hour, for a total of 24. That just means we can model our waiting time in thie regime with an average of 24 buses per hour, to get that the time until the first bus has an $\text{Exp}(6 + 6 + 6 + 6)$ distribution!

2. For finding the maximum, we just plug in $i = n = 4$ (and $n = 4$), to get

$$f_{Y_{(4)}}(y) = \binom{4}{4-1, 1, 4-4} \cdot [1 - e^{-6y}]^{4-1} \cdot [e^{-6y}]^{4-4} \cdot 6e^{-6y} = [1 - e^{-6y}]^3 6e^{-6y}$$

   Unfortunately, this is as simplified as it gets, and we don't get the nice result that the maximum of exponentials is exponential. To find the desired quantity, we just need to compute the probability the last bus comes before 5:15 (which is 0.25 hours - be careful of units!):

$$\mathbb{P}\left(Y_{\max} \leq 0.25\right) = \int_0^{0.25} f_{Y_{\max}}(y) dy = \int_0^{0.25} [1 - e^{-6y}]^3 6e^{-6y} dy$$

□