Order Statistics

1 Motivation & Preliminaries

- A special way that we can transform random variables is by taking a bunch of random variables and finding their maximum and minimum value
- This "bunch" of random variables would represent our real world data collection
- Because our data is a collection of random variables, the max and the minimum values of the measurements will also be random variables
- Therefore, the max and min will have their own distributions, which we can find if we know the distribution of our data
- For the sections below, Let $X_1, X_2, ..., X_n$ be i.i.d. (i.i.d. stands for (mutually) independent and identically distributed) with pdf $f(x)$ and cdf $F(x)$
- Saying that $X_1, ..., X_n$ are i.i.d. means that each random variable has the same probability distribution as the other random variables (i.e. they all have the same distribution) AND all of the random variables are independent from each other.
- For Notational clarity, we introduce the following notation

Definition 1 (Order Statistics) *Let $X_1, \ldots, X_n$ be i.i.d., then the jth largest value among $X_1, \ldots, X_n$ is denoted $X_{(j)}$ and is called the jth order statistic.*

2 Maximum

Using the above definition, the maximum value of from $X_1, X_2, ..., X_n$ ($MAX(X_1, \ldots, X_n)$) is the nth or max order statistic and is denoted $X_{(n)}$. What we want to know is what the distribution of $X_{(n)}$ is. We will proceed using the df transformation technique

$$
\begin{aligned}
\rightarrow F_{X_{(n)}}(x) &= P(X_{(n)} < x) \\
&= P(X_1 < x, X_2 < x, ..., X_n < x) \text{ (Because the max being less than x is equivalent to all of} \\
&= P(X_1 < x)P(X_2 < x)...P(X_n < x) \text{ (Because } X_1, ..., X_n \text{ are i.i.d.)} \\
&= F(x)F(x)...F(x) \\
&= F^n(x) \\
\Rightarrow f_{X_{(n)}}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_{X_{(n)}}(x) \\
&= nF^{n-1}(x)f(x)
\end{aligned}
$$

3 Minimum

Using the same definition,the minimum value of from $X_1, X_2, ..., X_n$ ($MIN(X_1, \ldots, X_n)$) is the 1st or min order statistic and is denoted $X_{(1)}$. Again, we want to know is what the distribution of $X_{(1)}$ is. We will proceed using the df transformation technique

$$
\begin{aligned}
\rightarrow F_{X_{(1)}}(x) &= P(X_{(1)} < x) \\
&= 1 - P(X_{(1)} > x) \\
&= 1 - P(X_1 > x, X_2 > x, ..., X_n > x) \text{ (Because the min being greater than x is equivalent to} \\
&= 1 - P(X_1 > x)P(X_2 > x)...P(X_n > x) \text{ (Because } X_1, ..., X_n \text{ are i.i.d.)} \\
&= 1 - (1 - P(X_1 < x))(1 - P(X_2 < x))...(1 - P(X_n < x)) \\
&= 1 - (1 - F(x))(1 - F(x))...(1 - F(x)) \\
&= 1 - (1 - F(x))^n \\
\Rightarrow f_{X_{(n)}}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_{X_{(n)}}(x) \\
&= n(1 - F(x))^{n-1} f(x)
\end{aligned}
$$

4 Examples

1. Let X_1, X_2 be distributed i.i.d. U$(0, 1)$. Find the pdf of $X_{(2)}$.

$$
\begin{aligned}
f_{X_{(n)}}(x) &= nF^{n-1}(x)f(x) \\
&= \begin{cases} 2(x)^1(1) & 0 < x < 1 \\ 0 & else \end{cases} \\
&= \begin{cases} 2x & 0 < x < 1 \\ 0 & else \end{cases}
\end{aligned}
$$

2. Let X_1, X_2 be distributed i.i.d. Exp(1). Find the pdf of $X_{(1)}$.

$$
\begin{aligned}
f_{X_{(1)}}(x) &= n(1 - F(x))^{n-1} f(x) \\
&= \begin{cases} 2(1 - (1 - e^{-x}))^1 e^{-x} & 0 < x < \infty \\ 0 & else \end{cases} \\
&= \begin{cases} 2e^{-2x} & 0 < x < \infty \\ 0 & else \end{cases}
\end{aligned}
$$

5 Exercises

1. Let X_1, X_2, X_3 be distributed i.i.d. U$(0,1)$. Find the pdf of $X_{(3)}$.
2. Let X_1, X_2, X_3 be distributed i.i.d. Exp(1). Find the pdf of $X_{(1)}$.

6 Solutions

1.

$$\begin{aligned} f_{X_{(n)}}(x) &= nF^{n-1}(x)f(x) \\ &= \begin{cases} 3(x)^2(1) & 0 < x < 1 \\ 0 & else \end{cases} \\ &= \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & else \end{cases} \end{aligned}$$

2.

$$\begin{aligned} f_{X_{(1)}}(x) &= n(1 - F(x))^{n-1}f(x) \\ &= \begin{cases} 3(1 - (1 - e^{-x}))^2 e^{-x} & 0 < x < \infty \\ 0 & else \end{cases} \\ &= \begin{cases} 3e^{-3x} & 0 < x < \infty \\ 0 & else \end{cases} \end{aligned}$$