Order Statistics

December 3, 2020

Consider a random sample $X_1, X_2, \ldots, X_n$ from a continuous distribution with PDF f with support $\mathcal{S} = (a, b)$ and CDF F. Define:

$$\begin{aligned}
X_{(1)} &= \text{ smallest of } X_1, X_2, \ldots, X_n \;=\; \min\{X_1, X_2, \ldots, X_n\} \\
X_{(2)} &= \text{ second smallest of } X_1, X_2, \ldots, X_n \\
&\vdots \\
X_{(n-1)} &= \text{ second largest of } X_1, X_2, \ldots, X_n \\
X_{(n)} &= \text{ largest of } X_1, X_2, \ldots, X_n \;=\; \max\{X_1, X_2, \ldots, X_n\}
\end{aligned}$$

$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ are known as the **order statistics** and represent $X_1, X_2, \ldots, X_n$ when the X_is have been arranged in ascending order.

1 Distributions of the maximum and minimum of a random sample

1.1 Maximum of a random sample

By the definition of a CDF, the CDF of $X_{(n)}$ is defined as

$$F_{\max}(x) \;=\; \mathbf{P}\left(X_{(n)} \leq x\right) \;=\; \mathbf{P}\left(\max\{X_1, X_2, \ldots X_n\} \leq x\right)$$

When the maximum of a sample is less than or equal to some value x, it must be that all (unordered) elements of the sample must also be simultaneously less than or equal to the value x.

$$\begin{aligned}
\mathbf{P}\left(\max\{X_1, X_2, \ldots X_n\} \leq x\right) &= \mathbf{P}\left(X_1 \leq x,\, X_2 \leq x,\, \ldots,\, X_n \leq x\right) && \text{(Def'n of maximum)} \\
&= \mathbf{P}\left(X_1 \leq x\right) \cdot \mathbf{P}\left(X_2 \leq x\right) \cdot \ldots \cdot \mathbf{P}\left(X_n \leq x\right) && \text{(Independence)} \\
&= \left(F(x)\right)^n && \text{(Identically distributed)}
\end{aligned}$$

The PDF of $X_{(n)}$ can be obtained by differentiating the CDF with respect to x.

$$f_{\max}(x) \;=\; \frac{dF_{\max}}{dx} \;=\; \frac{d}{dx}\left(F(x)\right)^n \;=\; n\left(F(x)\right)^{n-1} f(x)$$

for all values $x \in \mathcal{S}$, and zero otherwise.

1.2 Minimum of a random sample

The CDF of the minimum of a sample can be obtained similarly. We start by considering the same approach as above.

$$F_{\min}(x) \;=\; \mathbf{P}\left(X_{(1)} \leq x\right) \;=\; \mathbf{P}\left(\min\{X_1, X_2, \ldots, X_n\} \leq x\right)$$

However, claiming that the minimum of a random sample is less than or equal to some value x does not convey any additional information. Instead, let us rephrase the above equality using the complementary CDF.

$$\mathbf{P}\left(\min\left\{X_1, X_2, \ldots, X_n\right\} \leq x\right) \;=\; 1 - \mathbf{P}\left(\min\left\{X_1, X_2, \ldots, X_n\right\} > x\right)$$

We can now use the fact that if the minimum of a sample is greater than some value x, it must be that all (unordered) elements of the sample must also be simultaneously greater than the value x.

$$\begin{aligned} 1 - \mathbf{P}\left(\min\left\{X_1, X_2, \ldots, X_n\right\} > x\right) &= 1 - \mathbf{P}\left(X_1 > x,\ X_2 > x,\ \ldots,\ X_n > x\right) && \text{(Def'n of minimum)} \\ &= 1 - \mathbf{P}\left(X_1 > x\right) \cdot \mathbf{P}\left(X_2 > x\right) \cdot \ldots \cdot \mathbf{P}\left(X_n > x\right) && \text{(Independence)} \\ &= 1 - (1 - F(x))^n && \text{(Identically distributed)} \end{aligned}$$

The PDF of $X_{(1)}$ can be obtained by differentiating the CDF with respect to x.

$$f_{\min}(x) \;=\; \frac{dF_{\min}}{dx} \;=\; \frac{d}{dx}(1 - (1 - F(x))^n) \;=\; -n\,(1 - F(x))^{n-1}\,(-f(x)) \;=\; n\,(1 - F(x))^{n-1}\,f(x)$$

for all values $x \in \mathcal{S}$, and zero otherwise.

2 Distributions of the jth order statistic

2.1 CDF of the j^{th} order statistic

What about the distribution of $X_{(j)}$ where $1 < j < n$?

In order for $X_{(j)}$ to be less than or equal to some value x, it must be that exactly j of the (unordered) elements of the sample are also simultaneously less than or equal to the value x. Let us define a new random variable, N, such that

$$N \;=\; \text{the number of } X_i \leq x$$

Each X_i will then either be less than or equal to x with probability $F(x)$, or greater than x with probability $1 - F(x)$. Then $N \sim \text{Binomial}(n,\, p = F(x))$. From the PMF of the binomial distribution, we can obtain the CDF of $X_{(j)}$ as

$$F_{(j)}(x) \;=\; \mathbf{P}\left(X_{(j)} \leq x\right) \;=\; \sum_{k=j}^{n} \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}$$

A somewhat hand-wavey approach to finding the PDF of $X_{(j)}$ can be found starting on page 369 of Blitzstein and Hwang's *Introduction to Probability*. An even better hand-wavey approach can be found in section 6.6 of Ross' *A First Course in Probability*. We will take an alternative approach that requires a brief introduction to the **beta distribution** (which will be discussed in lecture soon).

2.2 A brief introduction to the beta distribution

The beta function, $B(a, b)$, is defined as

$$B(a, b) \;=\; \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} \;=\; \int_0^1 x^{a-1}(1 - x)^{b-1}\,dx$$

where $a, b > 0$ and $\Gamma(a) = (a - 1)!$ for $a \in \mathbb{Z}^+$. To make a proper probability density function out of this, the integral must integrate to 1. To achieve this, we can divide both sides by $B(a, b)$ such that

$$1 \;=\; \int_0^1 \frac{1}{B(a, b)}\, x^{a-1}\,(1 - x)^{b-1}\,dx \;=\; \int_0^1 \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}\,(1 - x)^{b-1}\,dx$$

Then

$$f(x) \;=\; \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad x \in (0,1), \quad a,b > 0$$

and zero otherwise, is the density of the beta distribution.

There is an interesting connection between a binomial sum and a beta integral.

Lemma 1 *For $0 < p < 1$, and $j, n \in \mathbb{Z}^+$ where $j \leq n$,*

$$\sum_{k=j}^{n} \binom{n}{k} p^k \, (1-p)^{n-k} \quad = \quad \int_0^p \frac{1}{B(j,\, n-j+1)} \, x^{j-1} \, (1-x)^{(n-j+1)-1} \, dx$$

or equivalently,

$$\sum_{k=j}^{n} \binom{n}{k} p^k \, (1-p)^{n-k} \quad = \quad \int_0^p \frac{n!}{(j-1)! \, (n-j)!} \, x^{j-1} \, (1-x)^{n-j} \, dx$$

This lemma will not be proven, but we will assume that it is true!

2.3 PDF of the jth order statistic

Applying the above lemma to the CDF of $X_{(j)}$ and replacing the x in the integral with a dummy variable, t, we have

$$F_{(j)}(x) \;=\; \sum_{k=j}^{n} \binom{n}{k} (F(x))^k (1-F(x))^{n-k} \;=\; \int_0^{F(x)} \frac{n!}{(j-1)! \, (n-j)!} \, t^{j-1} \, (1-t)^{n-j} \, dt$$

To find the PDF of $X_{(j)}$, we differentiate the CDF with respect to x and apply the **Fundamental Theorem of Calculus** to obtain

$$\begin{aligned}
f_{(j)}(x) &= \frac{d}{dx} F_{(j)}(x) \\
&= \frac{d}{dx} \int_0^{F(x)} \frac{n!}{(j-1)! \, (n-j)!} \, t^{j-1} \, (1-t)^{n-j} \, dt \\
&= \frac{n!}{(j-1)! \, (n-j)!} \, (F(x))^{j-1} \, (1-F(x))^{n-j} \, \frac{dF(x)}{dx} \\
&= \frac{n!}{(j-1)! \, (n-j)!} \, (F(x))^{j-1} \, (1-F(x))^{n-j} \, f(x)
\end{aligned}$$

for $x \in \mathcal{S}$, and zero otherwise. An alternative way of interpreting this PDF is that we require exactly $j-1$ X_is to be less than or equal to x, exactly $n-j$ X_is to be greater than x, and exactly one X_i equal to x. The constant out front then arises due to the possible number of partitions that can be made.

3 Joint distributions of order statistics

3.1 Joint distribution of two order statistics

Suppose we now have two order statistics, $X_{(i)}$ and $X_{(j)}$, from a sample of size n and $i < j$. Let x_i and x_j be the realized values of the i^{th} and j^{th} order statistics, respectively. Then their joint PDF is given by

$$f_{X_{(i)},\,X_{(j)}}(x_i,\,x_j) \;=\; \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\,(F(x_i))^{i-1}\,(F(x_j)-F(x_i))^{j-i-1}\,(1-F(x_j))^{n-j}\,f(x_i)\,f(x_j)$$

for $a < x_i < x_j < b$, and zero otherwise.

Using an interpretation similar to the that of the PDF of a single order statistic, we now require $i-1$ elements less than or equal to x_i, $n-j$ elements greater than x_j, $j-i-1$ elements between x_i and x_j, and exactly two elements equal to x_i and x_j.

3.2 Joint distribution of all order statistics

For simplicity, first suppose that we have a random sample of size $n = 2$. Then the joint distribution of the order statistics is a transformation of the joint distribution of the unordered elements of the sample using:

$$\min\{X_1,\,X_2\} \;=\; X_{(1)} \quad \text{and} \quad \max\{X_1,\,X_2\} \;=\; X_{(2)}$$

Case 1: $X_1 \;=\; X_{(1)},\, X_2 \;=\; X_{(2)}$

$$\det(\mathbf{J}) \;=\; \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \;=\; 1$$

Case 2: $X_1 \;=\; X_{(2)},\, X_2 \;=\; X_{(1)}$

$$\det(\mathbf{J}) \;=\; \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \;=\; -1$$

Applying the transformation,

$$\begin{aligned} f_{X_{(1)},\,X_{(2)}}(x_1,\,x_2) &= f(x_1)\cdot f(x_2)\cdot|1| \,+\, f(x_2)\cdot f(x_1)\cdot|-1| \\ &= 2\cdot f(x_1)\cdot f(x_2) \\ &= 2!\cdot f(x_1)\cdot f(x_2) \end{aligned}$$

for $a < x_1 < x_2 < b$, and zero otherwise. For arbitrary n, there will be $n!$ cases to consider and the determinants of the $n!$ Jacobians will all be ± 1.

Thus, the joint distribution of n order statistics is given by:

$$\begin{aligned} f(x_1,\,x_2,\,\ldots,\,x_n) &= \sum_{i=1}^{n!} f(x_1)\cdot f(x_2)\cdots f(x_n)\cdot|\det(\mathbf{J}_i)| \\ &= n!\cdot f(x_1)\cdot f(x_2)\cdots f(x_n) \end{aligned}$$

for $a < x_1 < x_2 < \ldots < x_n < b$, and zero otherwise. As this is a joint distribution for all n order statistics, if we require the joint distribution of $k < n$ order statistics, we will need to integrate out the unwanted variables.

4 Examples

4.1 Distribution of a minimum of independent exponential RVs

Suppose $X_1 \sim \text{Exp}(\lambda_1)$, $X_2 \sim \text{Exp}(\lambda_2)$, and $X_1 \perp X_2$. Let $X = \min\{X_1, X_2\}$. Show that $X \sim \text{Exp}(\lambda_1 + \lambda_2)$.

$$\begin{aligned}
\mathbf{P}(X > t) &= \mathbf{P}(\min\{X_1, X_2\} > t) \\
&= \mathbf{P}(X_1 > t, X_2 > t) \\
&= \mathbf{P}(X_1 > t) \cdot \mathbf{P}(X_2 > t) \\
&= e^{\lambda_1 t} \cdot e^{\lambda_2 t} \\
&= e^{(\lambda_1 + \lambda_2)t}
\end{aligned}$$

Therefore, $X \sim \text{Exp}(\lambda_1 + \lambda_2)$, as desired.

4.2 Distribution of the range: example 1

Given a random sample of size n from a continuous distribution with PDF f and CDF F, define the range as

$$R := X_{(n)} - X_{(1)}.$$

Find the CDF and PDF of R.

Assume the joint density of $X_{(1)}$ and $X_{(n)}$ will have support $-\infty < x_1 < x_n < \infty$. The region $x_n - x_1 \leq a$ is equivalent to $x_n \leq x_1 + a$.

$$\begin{aligned}
\mathbf{P}(R \leq a) &= \mathbf{P}\left(X_{(n)} - X_{(1)} \leq a\right) \\
&= \iint\limits_{x_n - x_1 \leq a} f_{X_{(1)}, X_{(n)}}(x_1, x_n)\, dx_1\, dx_n \\
&= \int_{-\infty}^{\infty} \int_{x_1}^{x_1+a} \frac{n!}{(n-2)!} (F(x_n) - F(x_1))^{n-2} f(x_1) f(x_n)\, dx_n\, dx_1 \qquad (*)
\end{aligned}$$

Let $y = F(x_n) - F(x_1)$ and $dy = f(x_n)\, dx_n$. Then

$$\begin{aligned}
\int_{x_1}^{x_1+a} (F(x_n) - F(x_1))^{n-2} f(x_n)\, dx_n &= \int_{0}^{F(x_1+a)-F(x_1)} y^{n-2}\, dy \\
&= \frac{1}{n-1} (F(x_1 + a) - F(x_1))^{n-1}
\end{aligned}$$

Plugging this result into $(*)$, we obtain

$$\mathbf{P}(R \leq a) = n \int_{-\infty}^{\infty} (F(x_1 + a) - F(x_1))^{n-1} f(x_1)\, dx_1$$

This equation can be evaluated explicitly only in a few cases. One such case is when the X_is come from a uniform distribution on (0,1). Then for $0 < a < 1$,

$$\mathbf{P}(R \le a) = n \int_0^1 (F(x_1 + a) - F(x_1))^{n-1} f(x_1)\, dx_1$$

$$= n \int_0^{1-a} a^{n-1}\, dx_1 \; + \; n \int_{1-a}^1 (1 - x_1)^{n-1}\, dx_1$$

$$= n(1-a)a^{n-1} + a^n$$

Differentiating with respect to a yields the density

$$f_R(a) = n(n-1)\, a^{n-2}\, (1-a)$$

$$= \frac{n!}{(n-2)!\, 1!}\, a^{n-2}\, (1-a)$$

$$= \frac{(n-1+2-1)!}{(n-1-1)!\,(2-1)!}\, a^{(n-1)-1}\, (1-a)^{2-1}$$

$$= \frac{\Gamma(n-1+2)}{\Gamma(n-1)\Gamma(2)}\, a^{(n-1)-1}\, (1-a)^{2-1}$$

$$= \frac{1}{B(n-1,\, 2)}\, a^{(n-1)-1}\, (1-a)^{2-1}$$

We recognize this as the density of the beta distribution with parameters $n-1$, 2.

In general, the density of R is found as

$$f_{R_n}(r) \;=\; n(n-1) \int_{-\infty}^{\infty} (F(u+r) - F(u))^{n-2} f(u+r) f(u)\, du$$

for $r > 0$.

4.3 Distribution of the range: example 2

Consider a random sample of size n from an Exp(1) distribution. Determine

(a) $f_{X_{(1)},\, X_{(n)}}(x_1,\, x_n)$

$$f_{X_{(1)},\, X_{(n)}}(x_1,\, x_n) = n(n-1)(1 - e^{-x_n} - (1 - e^{-x_1}))^{n-2}\, e^{-x_1} e^{-x_n}$$

$$= n(n-1)(e^{-x_1} - e^{-x_n})^{n-2}\, e^{-(x_1 + x_n)}$$

for $0 < x_1 < x_n < \infty$, and zero otherwise.

(b) $f_{R_n}(r)$

$$f_{R_n}(r) = n(n-1)\int_0^\infty \left(e^{-u} - e^{-(u+r)}\right)^{n-2} e^{-(2u+r)}\,du$$

$$= n(n-1)\int_0^\infty e^{-u(n-2)}\,(1-e^{-r})^{n-2}\,e^{-2u+r}\,du$$

$$= n(n-1)(1-e^{-r})^{n-2}\,e^{-r}\int_0^\infty e^{nu}\,du$$

$$= (n-1)(1-e^{-r})^{n-2}\,e^{-r}$$

for $r > 0$, and zero otherwise.

4.4 Conditional expectation of order statistics

Suppose we have a random sample of size $n = 3$ from Exp(1). Compute $\mathbf{E}\left(X_{(3)} \,|\, X_{(1)} = x\right)$.

The joint density of $X_{(1)}$, $X_{(3)}$ is

$$f_{X_{(1)},X_{(3)}}(x_1,\,x_3) \;=\; 3!\,(e^{-x_1} - e^{-x_3})\,e^{-(x_1+x_3)}$$

for $0 < x_1 < x_3 < \infty$, and zero otherwise.

The conditional distribution is found as:

$$f_{X_{(3)}\,|\,X_{(1)}=x_1}(x_3) = \frac{f_{X_{(1)},X_{(3)}}(x_1,\,x_3)}{f_{X_{(1)}}(x_1)}$$

$$= \frac{3!(e^{-x_1} - e^{-x_3})\,e^{-(x_1+x_3)}}{3e^{-3x_1}}$$

$$= 2(e^{-x_1} - e^{-x_3})\,e^{2x_1-x_3}$$

for $0 < x_1 < x_3 < \infty$.

The conditional expectation is

$$\mathbf{E}\left(X_{(3)}\,|\,X_{(1)} = x_1\right) = \int_{x_1}^\infty 2x_3\,(e^{-x_1} - e^{-x_3})\,e^{2x_1-x_3}\,dx_3$$

Make the substitution: $u = x_3 - x_1$, $du = dx_3$.

$$= \int_0^\infty 2(u+x_1)(e^{x_1} - e^{-(u+x_1)})\,e^{2x_1-u-x_1}\,du$$

$$= 2\int_0^\infty (u+x_1)(1-e^{-u})\,e^{-u}\,du$$

$$= 2\int_0^\infty u(e^{-u} - e^{-2u})\,du \;+\; 2x_1\int_0^\infty e^{-u} - e^{-2u}\,du$$

$$= 2\left(1 - \frac{1}{2}\cdot\frac{1}{2}\right) \;+\; 2x_1\left(1 - \frac{1}{2}\right)$$

$$= x_1 + \frac{3}{2}$$