# Lecture #19: Joint Density of the Order Statistic

**Reference.** §4.3 pages 109–113

We begin with a Stat 251 result.

**Theorem.** *If $X_1, \ldots, X_n$ are i.i.d. continuous random variables with common density $f$, then*

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i).$$

The joint distribution of the order statistic is closely related.

**Theorem.** *If $X_1, \ldots, X_n$ are i.i.d. continuous random variables with common density $f$, then the density function of the order statistic $(X_{(1)}, \ldots, X_{(n)})'$ is given by*

$$f_{X_{(1)},\ldots,X_{(n)}}(y_1, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad y_1 < y_2 < \cdots < y_n.$$

The proof of this theorem requires the multidimensional change-of-variables formula for many-to-one functions. We did not cover this in Chapter 1, and so we will not cover the proof.

**Note.** If we want any marginal, then we just integrate. If $j \neq k$, then

$$f_{X_{(j)},X_{(k)}}(y_j, y_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_{(1)},\ldots,X_{(n)}}(y_1, \ldots, y_n)\,\mathrm{d}y_1 \cdots \mathrm{d}y_{j-1}\,\mathrm{d}y_{j+1} \cdots \mathrm{d}y_{k-1}\,\mathrm{d}y_{k+1} \cdots \mathrm{d}y_n.$$

**Remark.** We derived the density of the $k$th order variable $X_{(k)}$ and the joint density of the extremes $(X_{(1)}, X_{(n)})'$ earlier. We could also find them as marginals; see page 111.

**Example.** Let $X_1, X_2, X_3$ be i.i.d. $U(0, 1)$ random variables.

**(a)** Compute the density of $(X_{(1)}, X_{(2)}, X_{(3)})'$.

**(b)** Compute the density of $(X_{(1)}, X_{(3)})'$.

**(c)** Compute the density of $(X_{(2)}, X_{(3)})'$.

**Solution.** Since $X_1, X_2, X_3$ have common density $f(x) = 1$, $0 < x < 1$, we conclude that

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = 1, \quad 0 < x_1, x_2, x_3 < 1.$$

**(a)** From the previous theorem we conclude

$$f_{X_{(1)},X_{(2)},X_{(3)}}(y_1, y_2, y_3) = 3! = 6, \quad 0 < y_1 < y_2 < y_3 < 1.$$

**(b)** We find

$$f_{X_{(1)},X_{(3)}}(y_1, y_3) = \int_{-\infty}^{\infty} f_{X_{(1)},X_{(2)},X_{(3)}}(y_1, y_2, y_3)\,\mathrm{d}y_2 = \int_{y_1}^{y_3} 6\,\mathrm{d}y_2 = 6(y_3 - y_1)$$

provided $0 < y_1 < y_3 < 1$.

**(c)** We find

$$f_{X_{(2)},X_{(3)}}(y_2, y_3) = \int_{-\infty}^{\infty} f_{X_{(1)},X_{(2)},X_{(3)}}(y_1, y_2, y_3)\,\mathrm{d}y_1 = \int_{0}^{y_2} 6\,\mathrm{d}y_1 = 6y_2$$

provided $0 < y_2 < y_3 < 1$.

Note that we could also find $f_{X_{(3)}}(y_3)$ by integrating either marginal. That is,

$$f_{X_{(3)}}(y_3) = \int_{-\infty}^{\infty} f_{X_{(1)},X_{(3)}}(y_1, y_3)\,\mathrm{d}y_1 = \int_{0}^{y_3} 6(y_3 - y_1)\,\mathrm{d}y_1 = 6y_3^2 - 3y_3^2 = 3y_3^2$$

provided $0 < y_3 < 1$, and

$$f_{X_{(3)}}(y_3) = \int_{-\infty}^{\infty} f_{X_{(2)},X_{(3)}}(y_2, y_3)\,\mathrm{d}y_2 = \int_{0}^{y_3} 6y_2\,\mathrm{d}y_2 = 3y_3^2$$

provided $0 < y_3 < 1$.

**Example.** Let $X_1, X_2, X_3$ be i.i.d. $U(0, 1)$ random variables.

**(a)** Compute $P\{X_2 + X_3 \le 1\}$.

**(b)** Compute $P\{X_{(2)} + X_{(3)} \le 1\}$.

**Solution. (a)** By the law of total probability,

$$P\{X_2 + X_3 \le 1\} = \int_0^1 P\{X_2 + X_3 \le 1 | X_3 = x\} f_{X_3}(x)\,\mathrm{d}x.$$

We know that $f_{X_3}(x) = 1$ for $0 < x < 1$. We also find

$$\begin{aligned} P\{X_2 + X_3 \le 1 | X_3 = x\} = P\{X_2 \le 1 - x | X_3 = x\} = P\{X_2 \le 1 - x\} &= \int_0^{1-x} f_{X_2}(y)\,\mathrm{d}y \\ &= \int_0^{1-x} 1\,\mathrm{d}y \\ &= 1 - x \end{aligned}$$

since $X_2$ and $X_3$ are independent. Thus, we have

$$P\{X_2 + X_3 \le 1\} = \int_0^1 P\{X_2 \le 1 - x\}\,\mathrm{d}x = \int_0^1 (1-x)\,\mathrm{d}x = \left(x - \frac{x^2}{2}\right)\Big|_0^1 = \frac{1}{2}.$$

**(b)** Notice that we must have $X_{(2)} \le X_{(3)}$. This means that if $X_{(2)} > 1/2$, then $X_{(3)} \ge 1/2$, and so $X_{(2)} + X_{(3)}$ is necessarily greater than 1. Therefore, our intuition is that

$$P\{X_{(2)} + X_{(3)} \le 1\} = \int_0^{1/2} P\{X_{(2)} = x,\ x \le X_{(3)} \le 1 - x\}\,\mathrm{d}x.$$

Formally, we conclude that

$$\begin{aligned} P\{X_{(2)} + X_{(3)} \le 1\} = \int_0^{1/2} \int_x^{1-x} f_{X_{(2)},X_{(3)}}(x,y)\,\mathrm{d}y\,\mathrm{d}x &= \int_0^{1/2} \int_x^{1-x} 6x\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^{1/2} 6x(1-2x)\,\mathrm{d}x \\ &= \left(3x^2 - 4x^3\right)\Big|_0^{1/2} \\ &= \frac{1}{4}. \end{aligned}$$

**Example.** If $X_1, X_2, X_3, X_4$ are i.i.d. $U(0,1)$ random variables, determine $\mathbb{E}(X_{(4)}|X_{(1)})$.

**Solution.** Recall that to determine $\mathbb{E}(X_{(4)}|X_{(1)})$, we must compute $\mathbb{E}(X_{(4)}|X_{(1)} = y_1)$. We know from our results of Lecture #16 that

$$f_{X_{(1)},X_{(4)}}(y_1, y_4) = 4(4-1)f(y_1)f(y_4)[F(y_4) - F(y_1)]^{4-2} = 12(y_4 - y_1)^2$$

if $0 < y_1 < y_4 < 1$. Therefore,

$$f_{X_{(4)}|X_{(1)}=y_1}(y_4) = \frac{f_{X_{(1)},X_{(4)}}(y_1, y_4)}{f_{X_{(1)}}(y_1)} = \frac{12(y_4 - y_1)^2}{4(1-y_1)^3} = 3(y_4 - y_1)^2(1-y_1)^{-3}$$

if $0 < y_1 < 1$, and so

$$\begin{aligned} \mathbb{E}(X_{(4)}|X_{(1)} = y_1) = \int_{-\infty}^{\infty} y_4 f_{X_{(4)}|X_{(1)}=y_1}(y_4)\,\mathrm{d}y_4 &= \int_{y_1}^1 3y_4(y_4 - y_1)^2(1-y_1)^{-3}\,\mathrm{d}y_4 \\ &= 3(1-y_1)^{-3} \int_{y_1}^1 y_4^3 - 2y_1 y_4^2 + y_1^2 y_4\,\mathrm{d}y_4 \\ &= 3(1-y_1)^{-3} \left(\frac{1}{4}y_4^4 - \frac{2}{3}y_1 y_4^3 + \frac{1}{2}y_1^2 y_4^2\right)\Big|_{y_4=y_1}^{y_4=1} \\ &= \frac{3 - 8y_1 + 6y_1^2 - y_1^4}{4(1-y_1)^3}. \end{aligned}$$

We conclude that

$$\mathbb{E}(X_{(4)}|X_{(1)}) = \frac{3 - 8X_{(1)} + 6X_{(1)}^2 - X_{(1)}^4}{4(1 - X_{(1)})^3}.$$

Suppose that we also wish to compute

$$\mathbb{E}\left(\frac{3 - 8X_{(1)} + 6X_{(1)}^2 - X_{(1)}^4}{4(1 - X_{(1)})^3}\right).$$

Using techniques of Stat 251, we know that $f_{X_{(1)}}(y_1) = 4(1 - y_1)^3$, $0 < y_1 < 1$, and so

$$\begin{aligned}
\mathbb{E}\left(\frac{3 - 8X_{(1)} + 6X_{(1)}^2 - X_{(1)}^4}{4(1 - X_{(1)})^3}\right) &= \int_{-\infty}^{\infty} \frac{3 - 8y_1 + 6y_1^2 - y_1^4}{4(1 - y_1)^3} \cdot f_{X_{(1)}}(y_1)\,\mathrm{d}y_1 \\
&= \int_0^1 \frac{3 - 8y_1 + 6y_1^2 - y_1^4}{4(1 - y_1)^3} \cdot 4(1 - y_1)^3\,\mathrm{d}y_1 \\
&= \int_0^1 3 - 8y_1 + 6y_1^2 - y_1^4\,\mathrm{d}y_1 \\
&= \left(3y_1 - 4y_1^2 + 2y_1^3 - \frac{1}{5}y_1^5\right)\Bigg|_0^1 \\
&= 3 - 4 + 2 - \frac{1}{5} = \frac{4}{5}.
\end{aligned}$$

However, the easy way is to observe that $\mathbb{E}(\mathbb{E}(X_{(4)}|X_{(1)})) = \mathbb{E}(X_{(4)})$. Since $X_{(4)} \in \beta(4, 1)$, we know that $\mathbb{E}(X_{(4)}) = 4/5$. Thus, we conclude as before that

$$\mathbb{E}\left(\frac{3 - 8X_{(1)} + 6X_{(1)}^2 - X_{(1)}^4}{4(1 - X_{(1)})^3}\right) = \frac{4}{5}.$$

**Read.** Example 2.3 on page 108 does a similar calculation for three i.i.d. Exp(1) random variables.