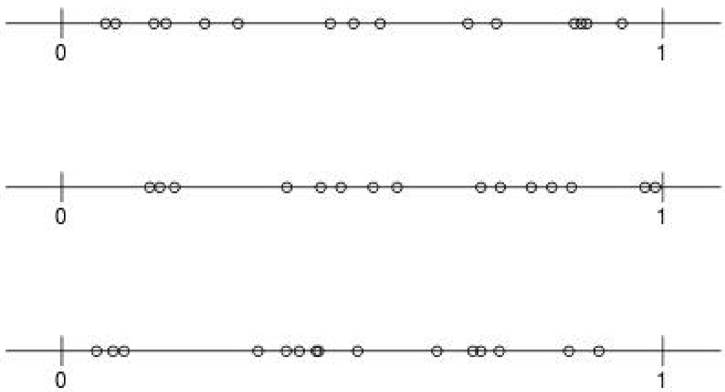
Order Statistics

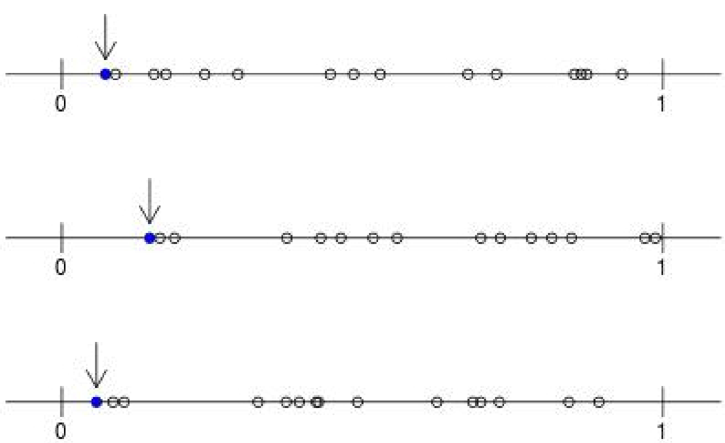
1 Introduction and Notation

Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 15 from the uniform distribution over the interval $(0, 1)$. Here are three different realizations realization of such samples.



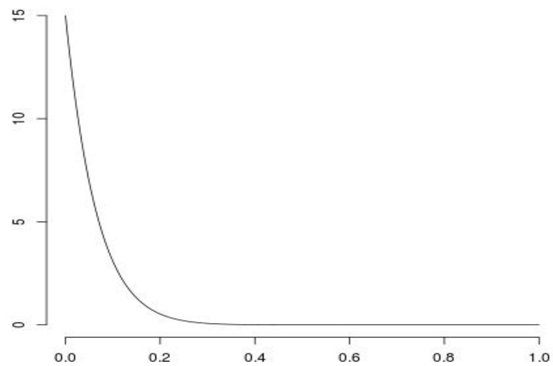
Because these samples come from a uniform distribution, we expect them to be spread out "randomly" and "evenly" across the interval $(0, 1)$. (You might think that you are seeing some sort of clustering but keep in mind that you are looking at a measly selection of only three samples. After collecting more samples I'm sure your view would change!)

Consider the single smallest value from each of these three samples, highlighted here.



Collect the minimums onto a single graph.

Not surprisingly, they are down towards zero! It would be pretty difficult to get a sample of 15 uniforms on $(0, 1)$ that has a **minimum** up by the right endpoint of 1. In fact, we will show that if we kept collecting minimums of samples of size 15, they would have a probability density function that looks like this.



Notation: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from some distribution. We denote the order statistics by

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

(Another commonly used notation is $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ for the min through the max, respectively.)

In what follows, we will derive the distributions and joint distributions for each of these statistics and groups of these statistics. We will consider **continuous random variables only**. Imagine taking a random sample of size 15 from the geometric distribution with some fixed parameter p. The chances are very high that you will have some repeated values and not see 15 distinct values. For example, suppose we observe 7 distinct values. While it would make sense to talk about the minimum or maximum value here, it would not make sense to talk about the 12th largest value in this case. To further confuse the matter, the next sample might have a different number of distinct values! Any analysis of the order statistics for this discrete distribution would have to be well-defined in what would likely be an ad hoc way. (For example, one might define them conditional on the number of distinct values observed.)

2 The Distribution of the Minimum

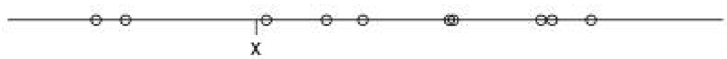
Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a continuous distribution with pdf f and cdf F. We will now derive the pdf for $X_{(1)}$, the minimum value of the sample. For order statistics, it is usually easier to begin by considering the cdf. The game plan will be to relate the cdf of the minimum to the behavior of the individual sampled values $X_1, X_2, \ldots, X_n$ for which we know the pdf and cdf.

The cdf for the minimum is

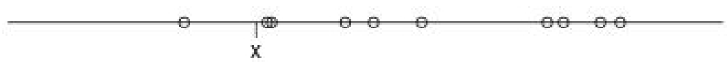
$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x).$$

Imagine a random sample falling in such a way that the minimum is below a fixed value x. It might look like this

or this



or this



or even this.

In other words,

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = P(\text{ at least one of } X_1, X_2, \ldots, X_n \text{ is } \leq x).$$

There are many ways for the individual X_i to fall so that the minimum is less than or equal to x. Considering all of the possibilities is a lot of work! On the other hand, the minimum is **greater** than x if and only if all the X_i are greater than x. So, it is easy to relate the probability $P(X_{(1)} > x)$ back to the individual X_i. Thus, we consider

$$\begin{aligned}
F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) \\
&= 1 - P(\, X_1 > x, X_2 > x, \ldots, X_n > x\,) \\
&= P(X_1 > x)P(X_2 > x) \cdots P(X_n > x) \quad \text{by independence} \\
&= 1 - [P(X_1 > x)]^n \qquad \text{because the } X_i \text{ are identically distributed} \\
&= 1 - [1 - F(x)]^n
\end{aligned}$$

So, we have that the pdf for the minimum is

$$\begin{aligned} \boxed{f_{X_{(1)}}(x)} &= \tfrac{d}{dx}F_{X_{(1)}}(x) = \tfrac{d}{dx}\left\{1 - [1 - F(x)]^n\right\} \\ &= \boxed{n[1 - F(x)]^{n-1}f(x)} \end{aligned}$$

Going back to the uniform example of Section 1, we had $f(x) = I_{(0,1)}(x)$ and

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ x & , \quad 0 \le x < 1 \\ 1 & , \quad x \ge 1. \end{cases}$$

The pdf for the minimum in this case is

$$f_{X_{(1)}}(x) = n[1 - x]^{n-1} I_{(0,\infty)}(x).$$

This is the pdf for the Beta distribution with parameters 1 and n. Thus, we can write

$$X_{(1)} \sim Beta(1, n).$$

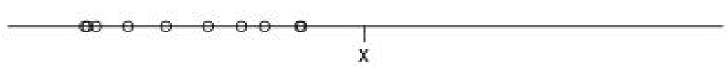
3 The Distribution of the Maximum

Again consider our random sample $X_1, X_2, \dots, X_n$ from a continuous distribution with pdf f and cdf F. We will now derive the pdf for $X_{(n)}$, the maximum value of the sample. As with the minimum, we will consider the cdf and try to relate it to the behavior of the individual sampled values $X_1, X_2, \dots, X_n$.

The cdf for the minimum is

$$F_{X_{(1)}}(x) = P(X_{(1)} \le x).$$

Imagine a random sample falling in such a way that the maximum is below a fixed value x. This will happen if and only if all of the X_i are below x.



Thus, we have

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \le x) \\ &= P(\, X_1 \le x, X_2 \le x, \dots, X_n \le x \,) \\ &= P(X_1 \le x)P(X_2 \le x) \cdots P(X_n \le x) \quad \text{by independence} \\ &= [P(X_1 \le x)]^n \qquad \text{because the } X_i \text{ are identically distributed} \\ &= [F(x)]^n. \end{aligned}$$

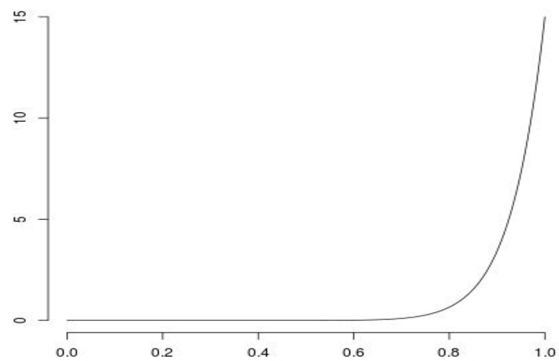
Take the derivative, we get the pdf for the maximum to be

$$\begin{aligned}
\boxed{f_{X_{(n)}}(x)} &= \tfrac{d}{dx}F_{X_{(1)}}(x) = \tfrac{d}{dx}[F(x)]^n \\
&= \boxed{n[F(x)]^{n-1}f(x)}
\end{aligned}$$

In the case of the random sample of size 15 from the uniform distribution on $(0, 1)$, the pdf is

$$f_{X(n)}(x) = nx^{n-1}\, I_{(0,1)}(x)$$

which is the pdf of the $Beta(n, 1)$ distribution.



Not surprisingly, all most of the probability or "mass" for the maximum is piled up near the right endpoint of 1.

4 The Joint Distribution of the Minimum and Maximum

Let's go for the joint cdf of the minimum and the maximum

$$F_{X_{(1)},X_{(n)}}(x, y) = P(X_{(1)} \le x, X_{(n)} \le y).$$

It is not clear how to write this in terms of the individual X_i. Consider instead the relationship

$$P(X_{(n)} \le y) = P(X_{(1)} \le x, X_{(n)} \le y) + P(X_{(1)} > x, X_{(n)} \le y). \tag{1}$$

We know how to write out the term on the left-hand side. The first term on the right-hand side is what we want to compute. As for the final term,

$$P(X_{(1)} > x, X_{(n)} \le y),$$

note that this is zero if $x \ge y$. (In this case, $P(X_{(1)} \le x, X_{(n)} \le y) = P(X_{(n)} \le y)$ and (1) gives us only $P(X_{(n)} \le y) = P(X_{(n)} \le y)$ which is both true and uninteresting! So, we consider the case that $x < y$. Note then that

$$\begin{aligned}
P(X_{(1)} > x, X_{(n)} \le y) &= P(x < X_1 \le y, x < X_2 \le y, \ldots, x < X_n \le y) \\
&\overset{iid}{=} [P(x < X_1 \le y)]^n \\
&= [F(y) - F(x)]^n.
\end{aligned}$$

Thus, from (1), we have that

$$\begin{aligned} F_{X_{(1)},X_{(n)}}(x,y) &= P(X_{(1)} \leq x, X_{(n)} \leq y) \\ &= P(X_{(n)} \leq y) - P(X_{(1)} > x, X_{(n)} \leq y) \\ &= [F(y)]^n - [F(y) - F(x)]^n. \end{aligned}$$

Now the joint pdf is

$$\begin{aligned} \boxed{f_{X_{(1)},X_{(n)}}(x,y)} &= \tfrac{d}{dx}\tfrac{d}{dy}\left\{[F(y)]^n - [F(y) - F(x)]^n\right\} \\ &= \tfrac{d}{dx}\left\{n[F(y)]^{n-1}f(y) - n[F(y) - F(x)]^{n-1}f(y)\right\} \\ &= \boxed{n(n-1)[F(y) - F(x)]^{n-2}f(x)f(y).} \end{aligned}$$

This hold for $\boxed{x < y}$ and for x and y both in the support of the original distribution.

For the sample of size 15 from the uniform distribution on $(0,1)$, the joint pdf for the min and max is

$$f_{X_{(1)},X_{(n)}}(x,y) = 15 \cdot 14 \cdot [y-x]^{13}\, I_{(0,y)}(x)\, I_{(0,1)}(y).$$

A Heuristic:

Since $X_1, X_2, \ldots, X_n$ are assumed to come from a continuous distribution, the min and max are also **continuous** and **the joint pdf does not represent probability**– it is a surface under which volume represents probability. However, if we bend the rules and think of the joint pdf as probability, we can develop a heuristic method for remembering it.

Suppose (though it is not true) that

$$f_{X_{(1)},X_{(n)}}(x,y) = P(X_{(1)} = x, X_{(n)} = y).$$

This would mean that we need one value in the sample $X_1, X_2, \ldots, X_n$ to fall at x, one value to fall at y, and the remaining $n-2$ values to fall in between.

The "probability" one of the X_i is x is "like" $f(x)$. (Remember, we are bending the rules here in order to develop a heuristic. This probability is, of course, actually 0 for a continuous random variable.)

The "probability" one of the X_i is y is "like" $f(y)$.

The probability that one of the X_i is in between x and y is (actually) $F(y) - F(x)$.

The sample can fall many ways to give us a minimum at x and a maximum at y. For example, imagine that $n = 5$. We might get $X_3 = x$, $X_1 = y$ and the remaining X_2, X_4, X_5 in between x and y.

This would happen with "probability"

$$f(x)[F(y) - F(x)]^3 f(y).$$

Another possibility is that we get $X_5 = x$ and $X_2 = y$ and the remaining X_1, X_3, X_4 in between x and y.

This would also happen with "probability"

$$f(x)[F(y) - F(x)]^3 f(y).$$

We have to add this "probability" up as many times as there are scenarios. So, let's count them. There are 5! different ways to lay down the X_i. For each one, there are 3! different ways to lay down the remaining values in between that will result in the same min and max. So, we need to divide these redundancies out for a total of $5!/3! = (5)(4)$ ways to get that min at x and max at y.

In general, for a sample of size n, there are $n!$ different ways to lay down the X_i. For each one, there are $(n-2)!$ different ways that result in the same min and max. So, there are a total of $n!/(n-2)! = n(n-1)$ ways to get that

Thus, the "probability" of getting a minimum of x and a maximum of y is

$$n(n-1)f(x)[F(y) - F(x)]^{n-2} f(y),$$

which looks an awful lot like the formula we derived above!

5 The Joint Distribution for <u>All</u> of the Order Statistics

We wish now to find the pdf

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n).$$

This time, we will <u>start</u> with the heuristic aid.

Suppose that $n = 3$ and we want to find

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3) \quad \text{``="} \quad P(X_{(1)} = x_1, X_{(2)} = x_2, X_{(3)} = x_3).$$

The first thing to notice is that this probability will be 0 if we don't have $x_1 < x_2 < x_3$. (Note that we need strict inequalities here. For a continuous distribution, we will never see repeated values so the minimum and second smallest, for example, could not take on the same value.)

Fix values $x_1 < x_2 < x_3$. How could a sample of size 3 fall so that the minimum is x_1, the next smallest is x_2, and the largest is x_3? We could observe

$$X_1 = x_1, X_2 = x_2, X_3 = x_3,$$

or

$$X_1 = x_2, X_2 = x_1, X_3 = x_3,$$

or

$$X_2 = x_2, X_2 = x_3, X_3 = x_1,$$

or...

There are 3! possibilities to list. The "probability" for each is $f(x_1)f(x_2)f(x_3)$. Thus,

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3) \quad \text{``="} \quad P(X_{(1)} = x_1, X_{(2)} = x_2, X_{(3)} = x_3) = 3! f(x_1)f(x_2)f(x_3).$$

For general n, we have

$$\boxed{f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n)} \quad \text{“=”} \quad P(X_{(1)} = x_1, X_{(2)} = x_2, \ldots X_{(n)} = x_n)$$

$$= \quad \boxed{n! f(x_1) f(x_2) \cdots f(x_n)}$$

which holds for $\boxed{x_1 < x_2 < \cdots < x_n}$ with all x_i in the support for the original distribution. The joint pdf is zero otherwise.

The Formalities:

The joint cdf,

$$P(X_{(1)} \le x_1, X_{(2)} \le x_2, \ldots, X_{(n)} \le x_n),$$

is a little hard to work with. Instead, we consider something similar:

$$P(y_1 < X_{(1)} \le x_1, y_2 < X_{(2)} \le x_2, \ldots, y_n < X_{(n)} < x_n)$$

for values $y_1 < x_1 \le y_2 < x_2 \le y_3 < x_3 \le \cdots \le y_n < x_n$.

This can happen if

$$y_1 < X_1 \le x_1, \; y_2 < X_2 \le x_2, \; \ldots, \; y_n < X_n < x_n,$$

or if

$$y_1 < X_5 \le x_1, \; y_2 < X_3 \le x_2, \; \ldots, \; y_n < X_{n-2} < x_n,$$

or...

Because of the constraints on the x_i and y_i, these are disjoint events. So, we can add these $n!$ probabilities, which will all be the same, together to get

$$P(y_1 < X_{(1)} \le x_1, \ldots, y_n < X_{(n)} < x_n) = n! \, P(y_1 < X_1 \le x_1, \ldots, y_n < X_n < x_n).$$

Note that

$$P(y_1 < X_1 \le x_1, \ldots, y_n < X_n < x_n) \stackrel{indep}{=} \prod_{i=1}^{n} P(y_i < X_i \le x_i) = \prod_{i=1}^{n} [F(x_i) - F(y_i)].$$

So,

$$P(y_1 < X_{(1)} \le x_1, \ldots, y_n < X_{(n)} < x_n) = n! \prod_{i=1}^{n} [F(x_i) - F(y_i)] \tag{2}$$

The left-hand side is

$$\int_{y_n}^{x_n} \int_{y_{n-1}}^{x_{n-1}} \cdots \int_{y_1}^{x_1} f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(u_1, u_2, \ldots, u_n) \, du_1 du_2 \ldots, du_n.$$

Taking derivatives $\frac{d}{dx_1} \frac{d}{dx_2} \cdots \frac{d}{dx_n}$ gives

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n)$$

Differentiating <u>both sides</u> of (2) with respect to $x_1, x_2, \ldots, x_n$ gives us

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n) = n! f(x_1) f(x_2) \cdots f(x_n)$$

which holds for $x_1 < x_2 < \cdots, x_n$ and all x_i in the support of the original distribution. The pdf is zero otherwise.

6 The Distribution of $X_{(i)}$

We can get the marginal pdf for the ith order statistic $X_{(i)}$, by taking the joint pdf for all order statistics from Section 5 and integrating out the unwanted x_j.

Let's start by integrating out x_1. Since the support of the joint pdf for the order statistics includes the constraint $x_1 < x_2 < \cdots < x_n$, limits of integration are $-\infty$ to x_2.

$$\begin{aligned}
f_{X_{(2)},\ldots,X_{(n)}}(x_2,\ldots,x_n) &= \int_{-\infty}^{x_2} f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1,x_2,\ldots,x_n)\,dx_1 \\
&= \int_{-\infty}^{x_2} n!f(x_1)f(x_2)\cdots f(x_n)\,dx_1 \\
&= n!f(x_2)\cdots f(x_n)\int_{-\infty}^{x_2} f(x_1)\,dx_1 \\
&= n!f(x_2)\cdots f(x_n)F(x_2)
\end{aligned}$$

for $x_2 < x_3 < \cdots < x_n$.

Now let's integrate out x_2 which goes from $-\infty$ to x_3.

$$\begin{aligned}
f_{X_{(3)},\ldots,X_{(n)}}(x_3,\ldots,x_n) &= \int_{-\infty}^{x_3} f_{X_{(2)},\ldots,X_{(n)}}(x_2,\ldots,x_n)\,dx_2 \\
&= n!f(x_3)\cdots f(x_n)\int_{-\infty}^{x_3} \underbrace{F(x_2)}_{u}\,\underbrace{f(x_2)\,dx_2}_{du} \\
&= n!f(x_3)\cdots f(x_n)\ \tfrac{1}{2}[F(x_2)]^2\Big|_{x_2=-\infty}^{x_2=x_3} \\
&= n!f(x_3)\cdots f(x_n)\tfrac{1}{2}([F(x_3)]^2 - [\underbrace{F(-\infty)}_{0}]^2) \\
&= \tfrac{n!}{2}f(x_3)\cdots f(x_n)[F(x_3)]^2
\end{aligned}$$

which holds for $x_3 < x_4 < \cdots < x_n$.

The next time through, we will integrate out x_3 from $-\infty$ to x_4. Using $u = F(x_3)$ and $du = f(x_3)\,dx_3$, we get

$$f_{X_{(4)},\ldots,X_{(n)}}(x_4,\ldots,x_n) = \frac{n!}{(3)(2)}f(x_4)\cdots f(x_n)[F(x_4)]^3.$$

Continue until we reach $X_{(i)}$:

$$f_{X_{(i)},\ldots,X_{(n)}}(x_i,\ldots,x_n) = \frac{n!}{(i-1)!}f(x_i)\cdots f(x_n)[F(x_i)]^{i-1}$$

which holds for $x_i < x_{i+1} < \cdots < x_n$.

Now, we start integrating off x's from the other side.

$$\begin{aligned}
f_{X_{(i)},\ldots,X_{(n-1)}}(x_i,\ldots,x_{n-1}) &= \int_{x_{n-1}}^{\infty} f_{X_{(i)},\ldots,X_{(n-1)}}(x_i,\ldots,x_n)\,dx_n \\
&= \tfrac{n!}{(i-1)!}f(x_i)\cdots f(x_{n-1})[F(x_i)]^{i-1}\int_{x_{n-1}}^{\infty} f(x_n)\,dx_n \\
&= \tfrac{n!}{(i-1)!}f(x_i)\cdots f(x_{n-1})[F(x_i)]^{i-1}[1-F(x_{n-1})]
\end{aligned}$$

for $x_i < x_{i+1} < \cdots, x_{n-1}$.

$$
\begin{aligned}
f_{X_{(i)},\ldots,X_{(n-2)}}(x_i,\ldots,x_{n-2}) &= \textstyle\int_{x_{n-2}}^{\infty} f_{X_{(i)},\ldots,X_{(n-1)}}(x_i,\ldots,x_{n-1})\,dx_{n-1} \\
&= \tfrac{n!}{(i-1)!} f(x_i)\cdots f(x_{n-2})[F(x_i)]^{i-1} \textstyle\int_{x_{n-2}}^{\infty} f(x_{n-1})[1-F(x_{n-1})]\,dx_{n-1}
\end{aligned}
$$

Letting $u = 1 - F(x_{n-1})$ and $du = -f(x_{n-1})\,dx_{n-1}$, we get

$$
\begin{aligned}
f_{X_{(i)},\ldots,X_{(n-2)}}(x_i,\ldots,x_{n-2}) &= \tfrac{n!}{(i-1)!} f(x_i)\cdots f(x_{n-2})[F(x_i)]^{i-1} \left\{-\tfrac{1}{2}[1-F(x_{n-1})]^2\right\}_{x_{n-1}=x_{n-2}}^{x_{n-1}=\infty} \\
&= \tfrac{n!}{2(i-1)!} f(x_i)\cdots f(x_{n-2})[F(x_i)]^{i-1}[1-F(x_{n-2})]^2
\end{aligned}
$$

for $x_i < x_{i+1}, \cdots < x_{n-2}$.

The next time through we will integrate out x_{n-2} from x_{n-3} to ∞. Note that

$$
\begin{aligned}
\textstyle\int_{x_{n-3}}^{\infty} f(x_{n-2})[\underbrace{1-F(x_{n-2})}_{u}]^2\,dx_{n-2} &= -\tfrac{1}{3}[1-F(x_{n-2})]^3\Big|_{x_{n-2}=x_{n-3}}^{x_{n-2}=\infty} \\
&= \tfrac{1}{3}[1-F(x_{n-3})]^3.
\end{aligned}
$$

Thus,

$$
f_{X_{(i)},\ldots,X_{(n-3)}}(x_i,\ldots,x_{n-3}) = \frac{n!}{(3)(2)(i-1)!} f(x_i)\cdots f(x_{n-3})[F(x_i)]^{i-1}[1-F(x_{n-3})]^3
$$

for $x_i < x_{i+1} < \cdots < x_{n-3}$.

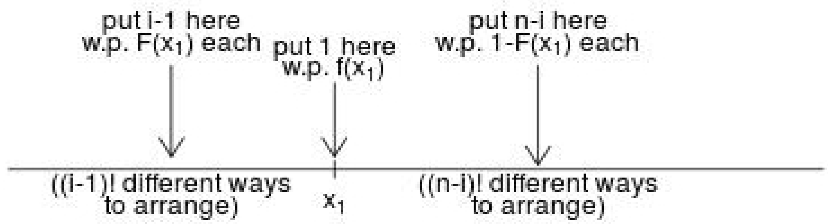
Continuing all the way down to the marginal pdf for $X_{(i)}$ alone, we get

$$
\boxed{f_{X_{(i)}} = \tfrac{n!}{(n-i)!(i-1)!}[F(x_i)]^{i-1} f(x_i)[1-F(x_{n-3})]^{n-i}}
$$

for $-\infty < x_i < \infty$. ($\leftarrow$ This may be further restricted by indicators in $f(x_i)$.)

The Heuristic:

We once again will think of the continuous random variables $X_1, X_2, \ldots, X_n$ as discrete and $f_{X_{(i)}}(x_i)$ as the "probability" that the ith order statistic is at x_i. First not that there are $n!$ different ways to arrange the $x's$. We need to put 1 at x_i, which will happen with "probability" $f(x_i)$. We need to put $i-1$ below x_i, which will happen with probability $[F(x_i)]^{i-1}$ and we need to put $n-i$ above x_i, which will happen with probability $[1-F(x_i)]^{n-i}$. There are $(i-1)!$ different ways to arrange the x's chosen to go below x_i. These arrangements are redundant and need to be divided out. Hence, we have $(i-1)!$ in the denominator. There are $(n-i)!$ different ways to arrange the x's chosen to go above x_i. These arrangements are also redundant and need to be divided out. Thus, we also have $(n-i)!$ in the denominator.



7 The Joint Distribution of $X_{(i)}$ and $X_{(j)}$ for $i < j$

As in Section 6, one could start with the joint pdf for all of the order statistics and integrate out the unwanted ones. The result will be

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

for $-\infty < x_i < x_j < \infty$.

Can you convince yourself of this heuristically?