Republic of iraq
Ministry of Higher Education and
Scientific Research
University of Babylon
College of Education for Pure Sciences
Department of Mathematice

ON ORDER STATISTICS

A proposed research to the council of the college
Of Education for pure Sciences / University of Babylon
As part of the requirements for Bachelor's degree
By
Roqaya yass khudair

Supervised by
Asst.prof.Dr.Jinan Hamza Farhood

2023 A.D 1444.AH

بسم الله الرحمن الرحيم

﴿يَرفَعِ ٱللَّهُ ٱلَّذِينَ ءَامَنُوا مِنكُم وَٱلَّذِينَ أُوتُوا ٱلعِلمَ
دَرَجَٰت وَٱللَّهُ بِمَا تَعمَلُونَ خَبِير﴾

صدق الله العلي العظيم

سورة المجادلة [آية 11]

الاهداء

اهدي هذا الجهد المتواضع الى

الى من جعل الله الجنة تحت اقدامها والتي تعجز الكلمات عن وصفها

(امي الغالية)

الى رفيق دربي وحبيب قلبي سندي وشريك روحي

(زوجي العزيز)

الى كل من ساندني في مسيرتي الدراسية لهم كل المودة

الشكر والتقدير

اتقدم بالشكر أولاً للباري(عزّ وجل) الذي وفقني في انجاز هذا البحث فلابد لنا ونحن نخطو خطواتنا الأخيرة في الحياة الجامعية من وقفة نعود الى اعوام قضيناها في رحاب الجامعة مع اساتذتنا الاكارم الذين قدموا لنا الكثير باذلين بذلك جهوداً كبيرة في بناء جيل الغد لتبعث الأمة من جديد ...

اتقدم بجزيل الشكر والامتنان الى استاذتي العزيزة الفاضلة أ.م.د. جنان حمزه فرهود المشرفة على البحث

وقبل ان امضي اقدم اسمى آيات الشكر والامتنان والتقدير والمحبة الى الذين حملوا أقدس رسالة في الحياة ... الى الذين مهدوا لنا طريق العلم والمعرفة أساتذتي الافاضل.

رقية ياس

C

Contents

Introduction

We introduce in this research subject on order statistics such that suppose we have n independent, identically random varibles, and we are asked to calculate the expected value of the minimum or maximum of them. We can do this using order statistics. If we sort these random varibles from least to greatest, the K^{th} order statistic is the K^{th} variable in our list. We often use the notation X_i to refer to the i^{th} order statistic .for example, $X_{(1)}$ (the first order statistic) is the minimum of the random variables, $X_{(2)}$ (the second order statistic) is the second smallest, and so on, $X_{(n)}$ (the nth order statistic) is the maximum.

Since we are discussing a set of random variables which all share the same distribution, it is useful to refer to the probability density function (p.d.f.) of that distribution as $f(x)$ and the cumulative distribution function (c.d.f.) as $F(x)$. In order to answer most questions about X_i , we need to find the pdf for X_i . We do this by first finding the (c.d.f.) for X_i and then taking a derivative.

Also ,we explain some concepts about the order statistics such that the distribution of the minimum and distribution of the maximum and the joint distribution of the minmum and maximum.

In this research we offer some remarks , theorems and examples about the order statistics.

Order Statistics

Suppose that $X_{(1)}, X_{(2)}, \dots . X_{(n)}$ are n jointly distributed random variables from a continuous function with continuous c.d.f. $F_x\,(x)$ and p.d.f. $f_x\,(x)$. The corresponding order statistics are the X_i arranged in nondecreasing order. The smallest of the X_i is denoted by X_1 the second smallest is denoted by $X_2 \dots \dots$, and, finally the largest is denoted by X_n.

The order statistics of n-identically independently. distributed (iid) random variables $X_{(1)}, X_{(2)}, \dots . X_{(n)}$ p.d.f. $f(x)$ are the values placed in ascending order,

that is, $a < Y_1 < Y_2 < \dots < Y_n < b$ where $Y_1 = min\{X_i\}$, Y_2 the next $min\{X_i\}$, and $Y_n = max\,\{ X_i\}$.

Remark (1)

Although X_i are iid random variables, the random variables Y_i are neither independent nor identically distributed Thus,
the minimum of X_i is $Y_1 = min\,(X_1, X_2, \dots . X_n)$ and
the maximum of X_i $Y_n = max\,(X_1, X_2, \dots . X_n)$
The order statistics of the sample $X_1, X_2, \dots . X_n$ can also be denoted by $X_{(1)}, X_{(2)}, \dots . X_{(n)}$ where $X_{(1)} < X_{(2)} < \dots . < X_{(n)}$.

Here $X_{(k)}$ is the k^{th} order statistic and is equal to Y_k in definition one of the most commonly used order statistics is the median the value in the middle position in the sorted order of the Values

i.e, we denote the order statistics by:

$X_{(1)} = min(X_1, X_2, \dots . X_n)$

$X_{(2)}$ = the 2nd Smallest of($X_1, X_2, \dots . X_n$)

$\vdots \quad = \quad \vdots$

$X_{(n)} = max(X_1, X_2, \dots . X_n)$

Remark (2)

1- In the definition of order statistics it is not required that the $\{Y_i\}$ be independent identically distributed (iid).

2- the formal difintion above only works for continuous variables

3- We will assume that the n independent observations come from a continuous distribution, thereby making the probability zero that any tow observations are equal.

By difintion, every observation in a random variable has the same pdf.
For example. if a set of four measurements is taken from a normal distribution with $\mu = 80$ and $\sigma = 15$, then $f_{y1}(y)$, $f_{y2}(y), f_{y3}(y)$ and $f_{y4}(y)$ are all the Same - each is a normal pdf with $\mu = 80$ and $\sigma = 15$

The pdf describing an ordered observation, though ,is not the Same as the pdf describing a random observation. If a single observation is drawn from a normal distribution with $\mu = 80$ and $\sigma = 15$, it would not be surprising if that observation were to take on a value near 80.

On the other hand, if $n = 100$ observations is drawn. from that same distribution, we would not expect the smallest observation that is Y min- to be anywhere. near 80. common sense tells us that that smallest observation is likely to be much smaller than 80, just as the largest observation ,Y max, is likely to be much larger than 80.

The Distribution of the Minimum:

Suppose that $X_1, X_2, \dots . X_n$ 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 $X_1, X_2, \dots . X_n$ for which we know the pdf and cdf.

The cdf for the minimum $X_{(1)}$ is:

$$F_{x_{(1)}}(x) = p(x_{(1)} \le x)$$

$$F_{x_{(1)}}(x) = p\left(x_{(1)} \le x\right) = p(at\ least\ one\ of\ x_1, x_2 \dots \dots . x_n is \le x)$$

$$F_{x_{(1)}}(x) = p\left(x_{(1)} \le x\right) = 1 - p(x_{(1)} > x)$$

$$= 1 - p(x_1 > x, x_2 > x, \dots \dots \dots . x_n > x)$$

$$= 1 - p(x_1 > x)p(x_2 > x) \dots \dots \dots . (x_n > x)$$

$$= 1 - [p(x > x)]^n = 1 - [1 - p(x \le x)]^n$$

$$= 1 - [1 - f(x)]^n \text{ , } by\ independence$$

Take the derivative, we get the pdf for the minimum $X_{(1)}$ to be

$$fx_{(1)}(x) = \frac{d}{dx} fx_{(1)}(x) = \frac{d}{dx}\{1 - [1 - f(x)]^n\}$$

$$= n[1 - f(x)]^{n-1} f(x)$$

$$fx_{(1)}(x) = n(1 - f(x))^{n-1} f(x) \qquad \dots \dots .. (1)$$

Example (1):

Let $X_1, X_2, \dots . X_n$ be *independente* random variables uniformly distributed on the interval [0.1] . complete the pdf of

$X_{(1)} = min\,\{ X_1, X_2, \dots . X_n\}$

Solution: since

The CDF of x is:

$$f(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} 1dt = t|_0^x = x$$

So the pdf of $X_{(1)}$ is

$$fx_{(1)}(x) = n(1 - Fx(x))^{n-1} fx(x)$$

$$= n(1 - x)^{n-1}$$

This is the pdf for the beta distribution with parameters land n. Thus we can write $X_{(1)} \sim Beta(1, n)$

Example (2):

Support $X_1, X_2, \dots . X_n$ are iid exponential with mean $\beta > 0$. Recall the exponential $({}^{1}/_{\beta})$, the pdf $f_x(x)$ is given by

$$f_x(x) = \begin{cases} \dfrac{1}{\beta} e^{-x/\beta \;, x>0} \\ 0, other\ wise \end{cases}$$

Find the pdf of $X_{(1)}$, the minimum order statistics

Solution:

The pdf of $X_{(1)}$, the minimum order statistics is

$$f_{X_{(1)}}(x) = n(1 - F_x(x))^{n-1} fx(x)$$

$$Fx_{(x)} = \begin{cases} 0 & , x \leq 0 \\ 1 - e^{-x/\beta}, & , x > 0 \end{cases}$$

$$f_{X_{(1)}}(x) = n(1 - F_x(x))^{n-1} f_x(x)$$

$$= n\left(\frac{1}{\beta} e^{-x/\beta}\right)\left[1 - \left(1 - e^{-x/\beta}\right)\right]^{n-1}$$

$$= \frac{n}{\beta} e^{-x/\beta}\left(e^{-x/\beta}\right) = \frac{n}{\beta}(e^{-x/\beta})^n = \frac{n}{\beta} e^{-nx/\beta}$$

$$fx(1)(x) = \begin{cases} \frac{n}{\beta} e^{-nx/\beta} & , x > 0 \\ 0, other\ wise \end{cases}$$

There fore $X_1, X_2, \dots. X_n \sim$ iid exponential $({}^{1}/_{\beta}) \rightarrow X_{(1)} \sim$exponential $({}^{n}/_{\beta})$

Example (3):

Let $X_1, X_2, \dots. X_n$ be $independente$ exponential random variable with mean1 .compute the pdf of $X_{(1)} = \ min\{ X_1, X_2, \dots. X_n \}$

Solution: since

$$f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_x(x)$$

The pdf of x is $f(x) = e^{-x}$

The CDF of x is:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} e^{-t}dt = -e^{-t}|_{0}^{x} = 1 - e^{-x}, x > 0$$

$$f_{X_{(1)}}(x) = n(1 - F_{X_1}(x))^{n-1} fx(x)$$

$= n(1-(1-e^{-x}))^{n-1}e^{-x} = ne^{-nx}$ For $t > 0$ and 0 other wise

The Distribution of the Maximum:

Again consider our random sample $X_1, X_2, \ldots . X_n$ 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 $X_1, X_2, \ldots . X_n$.

The cdf for the maximum $X_{(n)}$ is:

$$F_{x_{(n)}}(x) = p(x_n \leq x) = p(\max\{X_1, X_2, \ldots . X_n\} \leq \mathrm{x})$$

$$= P(X_1 \leq x, x_2 \leq x, \ldots \ldots .., x_n \leq x)$$

$$= (x_1 \leq x)p(x_2 \leq x) \ldots \ldots .. p(x_n \leq x)$$

$$= [p(x \leq x)]^n = [f(x)]^n, by\ independence$$

$Take\ the\ derivative\ , we\ get\ the\ pdf\ for\ the\ maximum\ X_{(n)}\ to\ be$

$$f_{x_{(n)}}(x) = \frac{d}{dx} Fx_{(n)}(x) = \frac{d}{dx}[Fx(x)]^n = n[F(x)]^{n-1} f(x)$$

$f_{x_{(n)}}(x) = n[Fx(x)]^{n-1} fx(x)$ ……. (2)

Example (4):

let $X_1, X_2, \dots . X_n$ be independent random varibles uniformly distributed on the interval [0,1]. Compute. the pdf of $X_{(n)} \ = \ max\,\{\, X_1, X_2, \dots . X_n\}$

Solution: Since

$$f_{x_{(n)}}(x) = n(F_x(x))^{n-1} f_x(x)$$

The CDF of x is :

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} 1dt = t|_{0}^{x} = x$$

So The pdf of $X_{(n)}$ is :

$$f_{x_{(n)}}(x) = n(F_x(x))^{n-1} f_x(x)$$

$$= nx^{n-1}$$

Which is the pdf of the $Beta\ (n, 1)$ distribution

Example (5):

Suppose $X_1, X_2, \dots . X_n$ are iid exponential with mean $\beta > 0$, recall the exponential $({}^{1}/_{\beta})$, the pdf $f_x(x)$ is given by

$$f_x(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} \ , x > 0 \\ 0, other\ wise \end{cases}$$

Find the pdf of $X_{(n)}$, the maximum order statistics.

Solution:

the pdf of $X_{(n)}$ the maximum order statistics is

$$f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_x(x)$$

$$= n(1 - e^{-x/\beta})^{n-1} \left({}^{1}/_{\beta}\, e^{-x/\beta} \right) = \frac{n}{\beta} e^{-x/\beta} (1 - e^{-x/\beta})^{n-1}$$

$$\therefore f_{X_{(n)}}(x) = \begin{cases} \frac{n}{\beta} e^{-x/\beta} (1 - e^{-x/\beta})^{n-1} & , x > 0 \\ 0 & other\ wise \end{cases}$$

The Joint Distribution of the Minimum and Maximum

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

$$F_{X_{(1)}, X_{(n)}}(x, y) = p(X_{(1)} \leq x, X_{(n)} \leq y)$$

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

$$p\big(X_{(n)} \leq y\big) = p\big(X_{(1)} \leq x, X_{(n)} \leq y\big) + p(X_{(1)} \leq x, X_{(n)} \leq y) \quad \ldots \quad 1$$

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,

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

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

$$P\left(X_{(1)} > x, X_{(n)} \leq y\right) = P(x < X_1 \leq y, x < X_2 \leq y, \dots\dots, x < X_n \leq y)$$

$$\underset{\text{iid}}{=} [P(x < X_1 \leq y)]^n$$

$$= [F(y) - F(x)]^n.$$

Thus from (1), we have that

$$F_{X(1),X(n)}(x, y) = P(X_{(1)} \leq x, X_{(n)} \leq y)$$

$$= P\left(X_{(n)} \leq y\right) - P(X_{(1)} > x, X_{(n)} \leq y)$$

$$= [F(y)]^n - [F(y) - F(x)]^n.$$

Now the joint pdf is

$$\boxed{f_{X(1),X(n)}(x, y)} = \frac{d}{dx}\frac{d}{dx}\{[F(y)]^n - [F(y) - F(x)]^n\}$$

$$= \frac{d}{dx}\{n[F(y)]^{n-1}f(y) - n[F(y) - F(x)]^{n-1}f(y)\}$$

$$= \boxed{n(n-1)[F(y) - F(x)]^{n-2}f(x)f(y).} \quad \dots\dots.(3)$$

This hold for 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

$$f_{X(1),X(n)}(x, y) = 15.14.[y - x]^{13} I_{(0,y)}(x) I_{(0,1)}(y).$$

A Heuristic:

Since $X_1, X_2, \dots . X_n$ 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!

Example (6):

Consider 15 random variables $X_1, X_2, \ldots. X_{15}$ all having the $u[0,1]$ distribution . Find the joint pdf for the max and min statistics.

Solution: The pdf and cdf of the r.v. $X_i$are

$$f_{X_i}(x) = \begin{cases} 1 & ,0 \le x \le 1 \\ 0 & ,otherwise \end{cases} \quad \text{and} \quad F_{X_i}(x) = \begin{cases} 0 & ,x \le 0 \\ x & ,0 \le x \le 1 \\ 1 & ,x \ge 1 \end{cases}$$

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

$$= 15(14)\,(y-x)^3\,(1)\,(1)$$

$$= 210(y-x)^3 \qquad ,0 < x < y < 1$$

The Joint Distribution for All of the Order Statistics

We wish now to find the pdf

$$f_{X_{(1)}, X_{(2)}, \ldots \ldots, X_n}(x_1, x_2, \ldots \ldots., x_3)$$

This time, we will start with the heuristic aid.

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

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

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 is 3! Possibilities to list. The "probability" for each is $f(x_1)f(x_2)f(x_3)$ thus,

$$\begin{aligned} f_{X_{(1)}, X_{(2)}, X_{(3)}}(x_1, x_2, x_3)"="P(X_{(1)} = x_1, X_{(2)} = x_2, X_{(3)} = x_3) \\ = 3!\, f(x_1)f(x_2)f(x_3). \end{aligned}$$

For general n, we have

$$\boxed{f_{X_{(1)}, X_{(2)}, \ldots, X_n}(x_1, x_2, \ldots., x_n)}" = "P(X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(n)} = x_n)$$

$$= \boxed{n!\, f(x_1)f(x_2) \ldots f(x_n)}$$

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,

$\mathrm{P}(X_{(1)} \leq x_1, X_{(2)} \leq x_2, X_n \leq x_n)$.

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

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

For values $y_1 < x_1 \leq y_2 < x_2 \leq y_3 < x_3 \leq \cdots \leq y_n < x_n$.

This can happen if

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

Or if

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

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

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

Note that

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

So,

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

The left –hand side is

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

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

$$f x_{(1),} X_{(2)}, \dots, X_{(n)}(x_1, x_2, \dots, x_n)$$

Differentiation both side of (2) with respect to $x_1, x_2, \dots, x_n$ gives us

$f_{x_{(1)},x_{(2)},\dots,x_{(n)}}(x_1, x_2, \dots, x_n) = n!\, f(x_1) f(x_2) \dots f(x_n)$ *.......(4)*

Which hold for $x_1 < x_2 < \cdots, x_n$ and all X_i in the support of the original distribution. The pdf is zero otherwise.

Example (7):

Let X_1, X_2, X_3, X_4 be a random sample from uniform distribution on the interval [0,5] if $y_1 < y_2 < y_3 < y_4$

be the order statistic of the random sample $(x_1, x_2, \dots, x_4)$

Find the joint pdf $(y_1, y_2, \dots, y_4)$

Solution: $g\ (y_1\ , y_2\ , \dots, y_4\ ) = n!\, f(y_1), f(y_2), \dots, f(y_4)$

$$since\ f(x) = \frac{1}{b-a} = \frac{1}{5-0} = \frac{1}{5}$$

$$g(y_1, y_2, \dots, y_4) = 4!\, f(y_1), f(y_2), \dots, f(y_4)$$

$$= 24.\frac{1}{5}.\frac{1}{5}.\frac{1}{5}.\frac{1}{5}$$

$$= \frac{24}{625} \qquad ,0 < y_1 < y_2 < y_3 < y_4 < 5$$

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

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

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

$$f_{X_{(2)},\ldots X_{(n)}}(x_2,\ldots\ldots,x_n) = \int_{-\infty}^{x_2} f_{X_{(1)},\ldots X_{(n)}},(x_1,\ldots\ldots,x_n)dx_1$$

$$= \int_{-\infty}^{x_2} n!\, f x_{(1)} f x_{(2)} \ldots\ldots f x_{(n)}) dx_1$$

$$= n!\, f x_{(2)} \ldots\ldots f x_{(n)}) \int_{-\infty}^{x_2} f(x_1)\, dx_1$$

$$= n!\, f x_{(2)} \ldots\ldots f x_{(n)}) F(x_2) dx_1$$

for $x_2 < x_3 < \cdots < x_n$

Now lets integrate out x_2 which from $-\infty\ to\ x_3$

$$f_{X_{(3)},\ldots X_{(n)}}(x_3,\ldots\ldots,x_n) = \int_{-\infty}^{x_3} f_{X_{(2)},\ldots X_{(n)}},(x_2,\ldots\ldots,x_n)dx_2$$

$$= n!\, f(x_3) \ldots\ldots f(x_n) \int_{-\infty}^{x3} F(x_2)\, f(x_2) dx_2$$

$$= n!\, f(x_3) \ldots\ldots f(x_n) \frac{1}{2} [F(x_2)]^2 \Big|_{x_{(2)} = -\infty}^{x_{(2)} = x_{(3)}}$$

$$= n!\, f(x_3) \ldots\ldots f(x_n) \frac{1}{2} ([F(x_3)]^2 - [F(-\infty)]^2)$$

$$= \frac{n!}{2} f(x_3) \ldots\ldots f(x_n) [F(x_3)]^2$$

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

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

Continue untilwe reach $X_{(i)}$:

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

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

Now, we start integrating off x is from the other side

$$f_{X_{(i)},\dots X_{(n-1)}}(x_i, \dots\dots, x_{n-1}) = \int_{x_{n-1}}^{\infty} f_{X_{(i)},\dots X_{(n-1)}}(x_i, \dots\dots, x_{n-1})\, dx_n$$

$$= \frac{n!}{(i-1)!} f(x_i) \dots\dots f(x_{n-1}) \left[F(x_i)\right]^{i-1} \int_{x_{n-1}}^{\infty} f\left(x_n\right) dx_n$$

$$= \frac{n!}{(i-1)!} f(x_i) \dots\dots f(x_{n-1}) \left[F(x_i)\right]^{i-1} \left[1 - F(x_{n-1})\right]$$

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

$$f_{X_{(i)},\dots X_{(n-2)}}(x_i, \dots\dots, x_{n-2}) = \int_{x_{n-2}}^{\infty} f_{X_{(i)},\dots X_{(n-1)}}(x_i, \dots\dots, x_{n-1})\, dx_{n-1}$$

$$= \frac{n!}{(i-1)!} f(x_i) \dots\dots f(x_{n-2})\, [F(x_i)]^{i-1} \int_{x_{n-2}}^{\infty} f(x_{n-1})\, [1 - F\,(x_{n-1})]\, dx_{n-1}$$

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

$$f x_i \dots . x_{(n-2)}(x_i, \dots, x_{n-2}) = \frac{n!}{(i-1)!} f(x_i) \dots\dots f(x_{n-2})\, [F(x_i)]^{i-1} \left\{-\frac{1}{2}[1 - F\,(x_{n-1})]^2\right\}_{x_{n-1} = x_{n-2}}^{x_{n-1} = \infty}$$

$$= \frac{n!}{2(i-1)!} f(x_i) \dots\dots f(x_{n-2})\, [F(x_i)]^{i-1} [1 - F\,(x_{n-2})]^2$$

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

$$\int_{x_{n-3}}^{\infty} f(x_{n-2})\, [1 - F\,(x_{n-2})]^2\, dx_{n-2} = -\frac{1}{3}[1 - F\,(x_{n-2})]^3 \Big|_{x_{n-2} = x_{n-3}}^{x_{n-2} = \infty}$$

$$= \frac{1}{3}[1 - F\,(x_{n-3})]^3$$

Thus

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

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

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

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

For -∞ $< X_{(i)} < \infty(\leftarrow \textit{this may be further restricted by indicators in } f(x_i)$

Theorem (1): Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of continuous random variables drawn from a distribution having pdf $f_Y(y)$ and cdf $F_Y(y)$. The pdf of the ith order statistic is given by

$$f_{Y_i}(y) = \frac{n!}{(n-1)!\,(i-1)!}[F_Y(y)]^{i-1}\,[1 - F_Y(y)]^{n-i} f_Y(y)$$

For $1 \leq i \leq n$.

Proof We will give a heuristic argument that draws on the similarity between the statement of Theorem (1) and the binomial distribution. For a formal induction proof verifying the expression given for $f_Y(y)$.

Recall the derivation of the binomial probability function,

$p_x(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, where X is the number of successes in n independent trials, and p

is the probability that any given trial ends in success. Central to that derivation was the recognition that the event "$X = k$" is actually a union of all the different (mutually exclusive) sequences having exactly k successes and $n - k$ failures. Because the trials are independent, the probability of any such sequence is $p^k(1 - p)^{n-k}$ and the number of such sequences is $n!/[k!\,(n-k)!]$ *(or* $\binom{n}{k}$).

probability that $X = k$ is the product $\binom{n}{k} p^k (1 - p)^{n-k}$

Here we are looking for the pdf of the ith order statistic at some point y— that is $f_Y(y)$. As was the case with the binomial, that pdf will reduce to a combinatorial term times the probability associated with an intersection of independent events The only fundamental difference is that Y_i is a continuous random variable, whereas the binomial X is discrete, which means that what we find here will be a probability denstry function.

$$f_{Y_i}(y) = \frac{n!}{(n-1)!(i-1)!}[F_Y(y)]^{i-1}\,[1-F_Y(y)]^{n-i}\,f_Y(y) \quad \ldots\ldots(5)$$

Example (8):

Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statitics on of a random samples of size (4) from a distribution having the following

p.d.f

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & o.w \end{cases}$$

Find the p.d.f of y3

Solution : $n = 4$ and $i = 3$

$$g(Y_i) = \frac{n!}{(i-1)!\,(n-i)!}[F(y_i)]^{i-1}\,[1-F(y_i)]^{n-i}F(y_i)$$

$$g(y_3) = \frac{4!}{2!\;\;1!}[F(y_3)]^2\,[1-F(y_3)]^1\,f(y_3)$$

$$F(y_3) = \int_0^{y_3} f(u)du = \int_0^{y_3} 24\,du = 4^2\ \mathrm{I}_0^{y_3} = y_3^2$$

$$g(y_3) = 12\,(y_3^2)^2.(1-y_3^2)^1.2y_3 = 24\,y_3^5(1-y_3^2)$$

$$\therefore\ g(y_3) = 24\,y_3^5\,.(1-y_3^2) \qquad 0 < y_3 < 1$$

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

The joint pdf for all of the order statistics and integrate out The result will be .

$$f x_{(i)},X_j\left(x_j, x_j\right) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(xi)]^{\,i-1} f\,(xi)\,[F(xi)]^{\,j-i-1} f(xi)[1 - F(xj)]^{n-j} \quad .....(6)$$

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

Example (9):

Let $X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)} < X_{(5)}$ be order statistics on of a random samples of size 5 from a distribution having probability density function

$$f(x) = \begin{cases} 2x & , 0 < x < 1 \\ 0 & , otherwise \end{cases}$$

Find the joint pdf of $X_{(2)}$ and $X_{(4)}$

Solution :

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

$$f_{X_{(2)},X_{(4)}}(x_2, x_4) = \frac{5!}{1!\,1!\,1!}[x_2^2]\,[x_4^2 - x_2^2][1 - x_4^2]\,(2x_2)(2x_4)$$

$$= 480\,[x_2^3]\,[x_4][x_4^2 - x_2^2]\,[1 - x_4^2], \qquad 0 < x_2 < x_4 < 1$$

References :

1. Border, K. C, (lecture15:Orders Statistics, Conditional Expectations), Introduction to probability and Statistics, Department of Mathematics, p. 1-22, 2021.
2. kschischang, F.R. (Order Statistics) , 2019.
3. Mohammed, A.T, (lectures on mathematical statistics), university of Baghdad , 2022.
4. Sheffield, S. , (Conditional probability, Order Statistics, expectations of Sums).
5. Tsun, A., （Chapter 5,Multiple Random Variables), 5.10: Order Statistics, probability and Statistics with Applications to Computing 5.10,p.1-5.
6. https: academics. Su.edu. krd (order Statistics Definition)
7. https:// www.colorado.edu (order statistics 1 Introduction and Notation)
8. .http:// www.math.ntu.edu.tw(Order Statistics. Chapter 2)
9. http:// www.maths.dur.ac.uk (order Statistics)
10. http://www2 .stat.duke.edu (lecture 15: order statistics)