Chapter 3. part 4.

§3.7. Extrema and Order Statistics.

This section is concerned with ordering a collection of independent continuous r.v.s. $X_1, X_2, \cdots, X_n \overset{iid}{\sim} F$, cdf and f. pdf.

$X_{(1)}$ = smallest of $X_1, X_2, \cdots, X_n$;

$X_{(2)}$ = second smallest of $X_1, X_2, \cdots, X_n$.

....

$X_{(j)}$ = j-th smallest of $X_1, X_2, \cdots X_n$

....

r—demonstration

$X_{(n)}$ = largest of $X_1, X_2, \cdots X_n$. $X \leftarrow$ rexp(5, 0.1),

min(x), max(x), median(x)

$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, order statistics. Y=sort(x), Y[1], Y[2], ...

mean(x), , median(x) = quantile(X, 0.5), summary(x). boxplot(x) Q3 $\leftarrow$ quantile(X, 0.75)

Expl. $X_1, X_2, \cdots, X_5$ are lifetimes of 5 light bulbs. Assume $X_1, X_2, \cdots X_5 \overset{iid}{\sim} \exp(\lambda)$, $X_{(1)}$ is the shortest lifetime of 5. $X_{(5)}$ is the longest one. $X_{(3)}$ is the middle one.

$X_{(1)}$, $X_{(3)}$ and $X_{(5)}$ are all random variables, which depend on $X_1, \cdots X_5$.

e.g. $X_1 = 10$, $X_2 = 30$, $X_3 = 5$, $X_4 = 25$, $X_5 = 12$.

then $X_{(1)} = \min\{X_1, X_2, \cdots X_5\} = 5$

$X_{(5)} = \max\{X_1, X_2, \cdots, X_5\} = 30$.

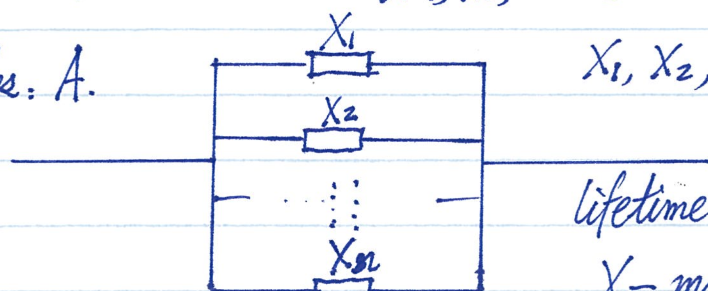
$X_{(3)}$ = third smallest $\{X_1, X_2, \cdots X_5\} = 12$.

also $X_{(2)} = 10$

$X_{(4)} = 25$.

R—demonstration

more examples: A.



$X_1, X_2, \cdots X_n$ lifetimes $\sim G(\alpha, \lambda)$.

lifetime of system is

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

What is the distribution of $X_{(n)}$? $X_{(n)} \sim f_{X_{(n)}}(t)$?

$P(X_{(n)} > 300) = ?$

B. 

$X_1, X_2, \cdots, X_n$ n components' lifetimes, $X_i \overset{iid}{\sim}$ lognormal(μ, σ^2).

What is the lifetime of the above system?

$X = \min\{X_1, X_2, \cdots, X_n\} = X_{(1)}$. find $P(X_{(1)} > 300)$?

$X_{(1)} \sim f_{X_{(1)}}(t)$?

C. $X_1, X_2, \cdots, X_n$, salaries of n people. a sample for certain group.
what is the median income?

if $n = 2k$, median $:= \frac{1}{2}(X_{(k)} + X_{(k+1)}) =: Y$, find $Y \sim f_Y(y)$?

$n = 2k+1$, median $= X_{(k+1)}$, what's distribution?

D. $X_1, X_2, \cdots X_n \overset{iid}{\sim}$ SAT scores.

range $= \max\{X_1, \cdots, X_n\} - \min\{X_1, X_2, \cdots, X_n\}$

$= X_{(n)} - X_{(1)}$. let $R := X_{(n)} - X_{(1)}$. find $R \sim f_R(r) = ?$

We need 1°. $(X_{(1)}, X_{(2)}, \cdots, X_{(n)}) \sim f_{X_{(1)}, \cdots, X_{(n)}}(x_1, x_2, \cdots, x_n) = ?$

2°. $X_{(1)} \sim f_{X_{(1)}}(x)$? In this book, $V = X_{(1)}$, $V \sim f_V(v)$.

3°. $X_{(n)} \sim f_{X_{(n)}}(x)$? $U = X_{(n)}$, $U \sim f_U(u)$.

4°. any k-th $X_{(k)}$. $X_{(k)} \sim f_{X_{(k)}}(x)$.

5°. any $(X_{(j)}, X_{(k)}) \sim f_{(X_{(j)}, X_{(k)})}(x, y) = ?$

Result 1. $X_1, X_2, \cdots X_n \sim F, f$. F is cdf, f is p.d.f.

$U = X_{(n)} = \max\{X_1, X_2, \cdots X_n\}$. find $f_U(u)$, $F_U(u)$.

Ans.

$$F_U(u) = P(U \le u) = P(X_{(n)} \le u) = P(X_1 \le u, X_2 \le u, \cdots X_n \le u)$$

$$\overset{iid}{=} P(X_1 \le u) \cdot P(X_2 \le u) \cdots P(X_n \le u) = [F(u)]^n.$$

$$f_U(u) = F_U'(u) = n [F(u)]^{n-1} \cdot f(u).$$

e.g. $X_1, \cdots X_n \sim \exp(\lambda)$ with $f(x) = \lambda e^{-\lambda x}$, $F(x) = 1 - e^{-\lambda x}$

then $U = X_{(n)} \sim F_U(u) = [1 - e^{-\lambda u}]^n$.

$$U = X_{(n)} \sim f_U(u) = n \cdot [1 - e^{-\lambda u}]^{n-1} \cdot \lambda e^{-\lambda u} = n\lambda e^{-\lambda u} \cdot [1 - e^{-\lambda u}]^{n-1}. \quad u > 0$$

R- simulation!

Result 2. $V = X_{(1)} = \min\{X_1, X_2, \cdots X_n\}$. find $F_V(v)$, $f_V(v)$.

Ans.

$$F_V(v) = P(V \le v) = P(X_{(1)} \le v) = 1 - P(X_{(1)} > v)$$

$$= 1 - P(X_1 > v, X_2 > v, \cdots X_n > v)$$

$$= 1 - [P(X_1 > v)]^n.$$

note $P(X_1 > v) = 1 - P(X_1 \le v)$
$= 1 - F(v)$

$$= 1 - [1 - F(v)]^n.$$

$$f_V(v) = F_V'(v) = -\{[1 - F(v)]^n\}_v' = -n [1 - F(v)]^{n-1} \cdot (-f(v))$$

$$= n \cdot f(v) \cdot [1 - F(v)]^{n-1}.$$

again, if $X_i \sim \exp(\lambda)$
with $f(x) = \lambda e^{-\lambda x}$, $x > 0$, $F(x) = 1 - e^{-\lambda x}$, $x > 0$

One has $F_V(v) = 1 - [1 - (1 - e^{-\lambda v})]^n = 1 - [e^{-\lambda v}]^n = 1 - e^{-\lambda n v}$.

$$f_V(v) = F_V'(v) = -e^{-\lambda n \cdot v}(-\lambda n) = \lambda n e^{-\lambda n v}, \quad v > 0.$$

$V \sim \exp(n\lambda)$.

R demonstration $\begin{cases} \lambda = 0.1 \\ n = 20 \end{cases}$ $n\lambda = 2$

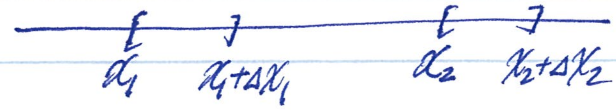
Result 3. $X_1, X_2, \cdots X_n \overset{iid}{\sim} f(x)$, then

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

$$x_1 \le x_2 \le \cdots \le x_n.$$

Proof. only for $n=2$ case here. show $(X_{(1)}, X_{(2)}) \sim f_{X_{(1)}X_{(2)}}(x_1, x_2) = 2f(x_1)f(x_2)$, $x_1 \le x_2$.

for any $x_1 \le x_2$,



$$P(x_1 \le X_{(1)} \le x_1+\Delta x_1,\ x_2 \le X_{(2)} \le x_2+\Delta x_2) = \binom{2}{1} P(x_1 \le X_{i_1} \le x_1+\Delta x_1) \cdot \binom{1}{1} P(x_2 \le X_{i_2} \le x_2+\Delta x_2)$$

$\uparrow$ indicate any one of two r.v's

$$= 2\int_{x_1}^{x_1+\Delta x_1} f(x)dx \int_{x_2}^{x_2+\Delta x_2} f(y)dy$$

$$= \int_{x_1}^{x_1+\Delta x_1}\int_{x_2}^{x_2+\Delta x_2} 2f(x)f(y)dxdy.$$

however, left side of above equality also

$$= \int_{x_1}^{x_1+\Delta x_1}\int_{x_2}^{x_2+\Delta x_2} f_{(X_{(1)},X_{(2)})}(x,y)dxdy$$

let $\Delta x_1 \to 0$, $\Delta x_2 \to 0$

thus

left side $= f_{X_{(1)},X_{(2)}}(x_1, x_2) \cdot \Delta x_1 \Delta x_2$.

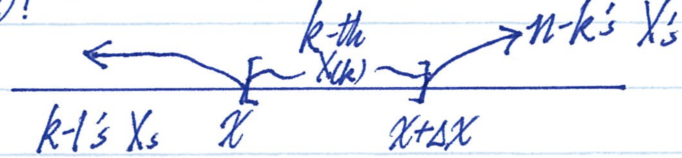
right side $= 2f(x_1) \cdot f(x_2) \Delta x_1 \Delta x_2$.

thus, $f_{X_{(1)}X_{(2)}}(x_1 x_2) = 2f(x_1)f(x_2)$.

$x_1 \le x_2$.

Result 4. $X_{(k)} = k$-th smallest of $\{X_1, X_2, \cdots X_n\}$.

then $$X_{(k)} \sim f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x) \cdot F(x)^{k-1} \cdot [1-F(x)]^{n-k}.$$



Proof: $P(x \le X_{(k)} \le x+\Delta x)$

$$= \binom{n}{k-1}\left[P(X \le x)\right]^{k-1} \cdot \binom{n-k+1}{1} P(x \le X \le x+\Delta x) \cdot \binom{n-k}{n-k}\left[P[X > x+\Delta x)\right]^{n-k}$$

$$= \frac{n!}{(k-1)!(n-k+1)!}\left[F(x)\right]^{k-1} \cdot (n-k+1) \cdot \int_x^{x+\Delta x} f(y)dy \cdot \frac{(n-k)!}{(n-k)!\,0!}\left[1-F(x+\Delta x)\right]^{n-k}$$

$$= \frac{n!}{(k-1)!\,(n-k)!}\left[F(x)\right]^{k-1}\int_x^{x+\Delta x} f(y)\,dy\cdot\left[1-F(x+\Delta x)\right]^{n-k}.$$

note $P(x \le X_{(k)} \le x+\Delta x) = \int_x^{x+\Delta x} f_{X_{(k)}}(y)\,dy$

thus $$\int_x^{x+\Delta x} f_{X_{(k)}}(y)\,dy = \frac{n!}{(k-1)!\,(n-k)!}\left[F(x)\right]^{k-1}\cdot\int_x^{x+\Delta x} f(y)\,dy\cdot\left[1-F(x+\Delta x)\right]^{n-k}.$$

let $\Delta x \to 0$, $$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!}\left[F(x)\right]^{k-1}\cdot f(x)\cdot\left[1-F(x)\right]^{n-k}.$$

Expl. $X_1, X_2, X_3 \overset{iid}{\sim} U[0,1]$, find $P\left(\frac{1}{4} < X_{(2)} < \frac{3}{4}\right)$

Ans: $n=3$, $k=2$. $f(x)=1$, $F(x)=x$,

$$f_{X_{(2)}}(x) = \frac{3!}{1!\,(3-2)!}\,x^{2-1}\cdot 1\cdot(1-x)^{3-2} = 6\cdot x\cdot(1-x), \quad 0<x<1.$$

i.e. $$f_{X_{(2)}}(x) = \begin{cases} 6x(1-x), & 0<x<1 \\ 0, & \text{else.}\end{cases}$$

$P\left(X_{(2)} < \frac{1}{4}\right) = \int_0^{\frac{1}{4}} f_{X_{(2)}}(x)\,dx = \frac{1}{8} = 0.125$

$$P\left(\frac{1}{4} < X_{(2)} < \frac{3}{4}\right) = \int_{1/4}^{3/4} 6\cdot x\cdot(1-x)\,dx = \cdots = \frac{11}{16} = 0.6875.$$

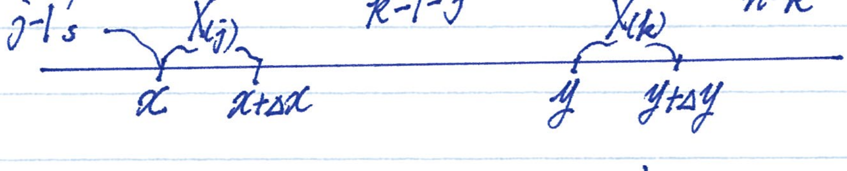
in R: $1 - \text{sum}(\text{med} < \frac{1}{4})/k - \text{sum}(\text{med} > \frac{3}{4})/k$

Result 5. $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x), F(x)$.

then $(X_{(j)}, X_{(k)}) \sim f_{(X_{(j)}, X_{(k)})}(x,y) = \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} * \left[F(x)\right]^{j-1} *$

$(j<k)$ $$\left[F(y)-F(x)\right]^{k-j-1} * \left[1-F(y)\right]^{n-k}\cdot f(x)\cdot f(y). \quad x \le y.$$

Proof:



$P(x \le X_{(j)} \le x+\Delta x,\ y \le X_{(k)} \le y+\Delta y) =$

$$\binom{n}{j-1}\left[P(X\le x)\right]^{j-1}\cdot\binom{n-j+1}{1}P(x\le X\le x+\Delta x)\cdot\binom{n-j}{k-j-1}\left[P(x+\Delta x\le X\le y)\right]^{k-j-1}$$

$$*\binom{n-k+1}{1}P(y\le X\le y+\Delta y)\cdot\left[1-F(y+\Delta y)\right]^{n-k}$$

$$=\underbrace{\binom{n}{j-1}\cdot\binom{n-j+1}{1}\cdot\binom{n-j}{k-j-1}\binom{n-k+1}{1}}\cdot\left[F(x)\right]^{j-1}\cdot\int_x^{x+\Delta x}f(u)du\left[F(y)-F(x+\Delta x)\right]^{k-j-1}$$

$$\int_y^{y+\Delta y}f(u)du\cdot\left[1-F(y+\Delta y)\right]^{n-k}$$

$$\frac{n!}{(j-1)!\cancel{(n-j+1)!}}\frac{\cancel{(n-j+1)!}}{1\cdot\cancel{(n-j)!}}\frac{\cancel{(n-j)!}}{(k-j-1)!\cancel{(n-k+1)!}}\cdot\frac{\cancel{(n-k+1)!}}{1\cdot(n-k)!}$$

$$=\frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!}$$

let $\Delta x\to 0$, $\Delta y\to 0$

thus, right hand side $=\dfrac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!}\left[F(x)\right]^{j-1}\cdot f(x)\Delta x\left[F(y)-F(x)\right]^{k-j-1}$

$$*f(y)\cdot\Delta y\cdot\left[1-F(y)\right]^{n-k}.$$

left hand side $=\displaystyle\int_x^{x+\Delta x}\int_y^{y+\Delta y}f_{X_{(j)},X_{(k)}}(u,v)\,dv\,du$

$$=f_{X_{(j)},X_{(k)}}(x,y)\cdot\Delta x\cdot\Delta y$$

Thus, $f_{X_{(j)},X_{(k)}}(x,y)=\dfrac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!}\left[F(x)\right]^{j-1}\cdot\left[F(y)-F(x)\right]^{k-j-1}\cdot\left[1-F(y)\right]^{n-k}\cdot f(x)\cdot f(y)$

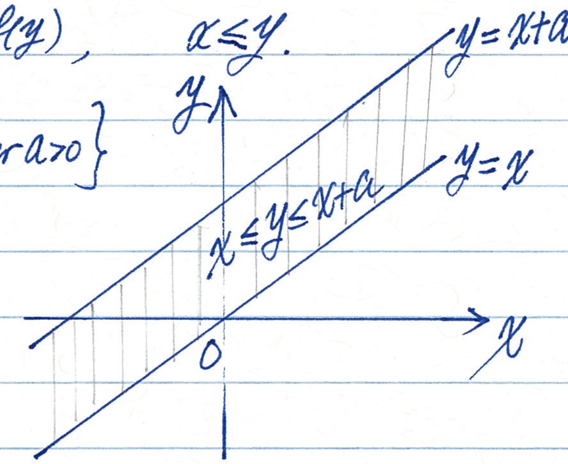
Example. Distribution of the range of a random sample.

$X_1, X_2, \cdots, X_n$ iid f. $R = X_{(n)} - X_{(1)}$, find $R \sim f_R(r) = ?$

Ans: $j=1$, $k=n$.

$$f_{(X_{(1)}, X_{(n)})}(x,y) = \frac{n!}{(1-1)!\,(n-1-1)!\,(n-n)!}[F(x)]^{1-1}\cdot[F(y)-F(x)]^{n-1-1}[1-F(y)]^{n-n}f(x)f(y)$$

$$= n\cdot(n-1)\,[F(y)-F(x)]^{n-2}\cdot f(x)f(y), \quad x \le y.$$



$R = X_{(n)} - X_{(1)}$,

$$F_R(a) = P(R \le a) = P(X_{(n)} - X_{(1)} \le a) \quad \{\text{only consider } a>0\}$$

$$= \iint_{y-x\le a} f_{(X_{(1)},X_{(n)})}(x,y)\,dx\,dy \qquad \{\text{here } x \le y\}$$

$$= \int_{-\infty}^{\infty}\int_{x}^{x+a} n(n-1)[F(y)-F(x)]^{n-2} f(x)f(y)\,dy\,dx$$

$$= \int_{-\infty}^{\infty} n\cdot f(x)\left\{\int_{x}^{x+a}(n-1)[F(y)-F(x)]^{n-2}f(y)\,dy\right\}dx$$

For the above inner integal $\int_{x}^{x+a}(n-1)[F(y)-F(x)]^{n-2}f(y)\,dy$,

Let $u = F(y)-F(x)$, $du = F'(y)dy = f(y)dy$.

thus, above inner integal $= \int_{F(x)-F(x)}^{F(x+a)-F(x)}(n-1)u^{n-2}\,du = u^{n-1}\Big|_{0}^{F(x+a)-F(x)}$

$= [F(x+a)-F(x)]^{n-1}$.

Thus, $F_R(a) = \int_{-\infty}^{\infty} n\cdot f(x)\cdot[F(x+a)-F(x)]^{n-1}\,dx$.

Let's consider case $X_1, X_2, \cdots X_n \overset{iid}{\sim} U[0,1]$. Then $f(x) = \begin{cases}1, & x\in[0,1]\\ 0, & \text{else}.\end{cases}$

$$F(x)=P(X\le x)=\begin{cases}0, & x<0\\ x, & 0\le x\le 1\\ 1, & x>1.\end{cases}$$

$$F_R(a)=P(R\le a)=P(X_{(n)}-X_{(1)}\le a)=\begin{cases}0, & a\le 0\\ 1, & a\ge 1\end{cases}$$

So, we only need to consider $a\in[0,1]$ below.

$0\le a\le 1$ $\quad F_R(a)=\int_{-\infty}^{\infty} n\cdot f(x)\left[F(x+a)-F(x)\right]^{n-1}dx$

$$=\int_0^1 n\cdot 1\cdot\left[F(x+a)-x\right]^{n-1}dx \qquad \left\{\text{from } f(x)=\begin{cases}1, & x\in[0,1]\\ 0, & \text{else}\end{cases}\right. \text{ any } F(x)=x,\ x\in[0,1]\ \Big\}$$

$$=\int_0^{1-a} n\left[F(x+a)-x\right]^{n-1}dx+\int_{1-a}^{1} n\left[F(x+a)-x\right]^{n-1}dx$$

$$=\int_0^{1-a} n\left[x+a-x\right]^{n-1}dx+\int_{1-a}^{1} n\left[1-x\right]^{n-1}dx$$

$$\left\{\begin{array}{l}\text{if } 0\le x\le 1-a,\ a\le x+a\le 1,\ \text{then } F(x+a)=x+a.\\ \text{if } x\ge 1-a,\ x+a\ge 1,\ \text{then } F(x+a)=1\end{array}\right\}$$

$$=\int_0^{1-a} n a^{n-1}dx+n\int_{1-a}^{1}[1-x]^{n-1}dx$$

$$=n\cdot a^{n-1}\cdot(1-a)+n\cdot\left.\frac{-(1-x)^n}{n-1+1}\right|_{1-a}^{1}$$

$$=na^{n-1}(1-a)+n\frac{1}{n}\left[0+(1-1+a)^n\right]$$

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

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

thus, $f_R(a)=F_R'(a)=n(n-1)a^{n-2}-n(n-1)a^{n-1}=n(n-1)a^{n-2}(1-a),\ 0\le a\le 1.$

Put together. $R\sim f_R(a)=\begin{cases}n(n-1)a^{n-2}(1-a), & 0\le a\le 1\\ 0, & \text{else.}\end{cases}$

i.e. $R\sim \text{Beta}(n-1,2)$.

Another solution: $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$, $(X_1, X_2) \sim f_{(X_1,X_2)}(x_1, x_2)$.

$$\begin{cases} x_1 = h_1(y_1, y_2) \\ x_2 = h_2(y_1, y_2) \end{cases} \qquad J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix}$$

Then, $f_{(Y_1,Y_2)}(y_1, y_2) = f_{(X_1,X_2)}(x_1, x_2) \cdot |J(x_1, x_2)|^{-1}$.

here $f_{(X_{(1)}, X_{(n)})}(x_1, x_2) = n(n-1)(x_2 - x_1)^{n-2}$, $0 \le x_1 \le x_2 \le 1$.

Now, let $\begin{cases} Y_1 = X_{(n)} - X_{(1)} \\ Y_2 = X_{(1)} \end{cases}$ i.e. $Y_1 = X_2 - X_1$, $Y_2 = X_1$. $\begin{cases} y_1 = g_1(x_1, x_2) = x_2 - x_1 \\ y_2 = g_2(x_1, x_2) = x_1 \end{cases}$

$$J(x_1, x_2) = \begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} = -1. \quad |J(x_1, x_2)| = |-1| = 1.$$

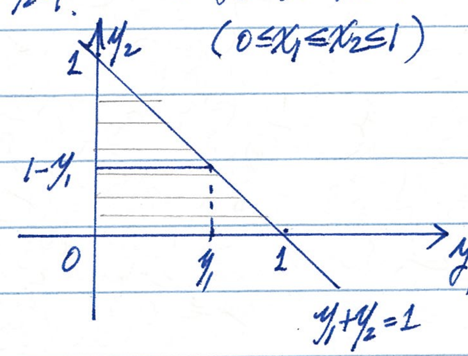
$$\begin{cases} x_1 = y_2 \\ x_2 = y_1 + y_2 \end{cases}$$

From above formula $f_{Y_1,Y_2}(y_1, y_2) = n(n-1)(x_2 - x_1)^{n-2} \cdot 1$

$= n(n-1) \cdot y_1^{n-2}$, $0 \le y_2 \le y_1 + y_2 \le 1$.

$0 \le y_2 \le y_1 + y_2 \le 1$
$(0 \le x_1 \le x_2 \le 1)$

i.e. $f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} n(n-1) y_1^{n-2}, & 0 \le y_1 \le 1, \ 0 \le y_1 + y_2 \le 1 \\ 0, & \text{else.} \end{cases}$



$0 \le y_1 \le 1$

$$f_{Y_1}(y_1) = \int f_{Y_1,Y_2}(y_1, y_2)\, dy_2 = \int_0^{1-y_1} n \cdot (n-1) y_1^{n-2}\, dy_2$$

$$= n(n-1) y_1^{n-2} \cdot (1 - y_1).$$

Put together. $X_{(n)} - X_{(1)} = Y_1 \sim f_{Y_1}(y) = \begin{cases} n(n-1) y^{n-2}(1-y), & 0 \le y \le 1 \\ 0, & \text{else} \end{cases}$

i.e. $R = X_{(n)} - X_{(1)} \sim \text{Beta}(n-1, 2)$.