Unit-IV

Order Statistics and Limit Theorems

4.1

Order Statistics

Independent and identically distributed random variables:
We say that $X_1, X_2, \dots, X_n$ are *independent* and *identically distributed* random variables (i.i.d.r.vs) if

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \text{(independent)} \qquad \dots (1)$$

and $F_{X_i}(x) = F(x)\ \forall\ i = 1,2, \dots, n$ (identically distributed) $\qquad \dots \quad (2)$

where $F_{X_i}(x)$ is the c.d.f. of X_i for $i = 1,2, \dots, n$ and $F_{X_1,\dots,X_n}(x_1, \dots, x_n)$ is the j.c.d.f. of $X_1, \dots, X_n$.

For continuous random variables, the c.d.fs are replaced with p.d.fs in equations (1) and (2) while for discrete random variables the c.d.fs are replaced with p.m.fs.

Definition: We say that $X_1, X_2, \dots, X_n$ is a random sample from a population with c.d.f. $F(x)$ (or p.d.f. $f(x)$ or p.m.f. $p(x)$) if $X_1, \dots, X_n$ are i.i.d.r.vs with common c.d.f. $F(x)$ (or p.d.f. $f(x)$ or p.m.f. $p(x)$).

Definition: Let $X_1, X_2, \dots, X_n$ be a random sample from a population with c.d.f. $F(x)$. Define

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

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

.

$X_{(r)} = r^{\text{th}}$ smallest of $X_1, X_2, \dots, X_n$,

.

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

The ordered values $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are known as the **order statistics** (o.s) of the n r.vs $X_1, X_2, \dots, X_n$.

Note:

1. o.s are r.vs themselves (as functions of $X_1, \dots, X_n$)
2. o.s. satisfy $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$
3. $X_1, X_2, \dots, X_n$ are i.i.d.r.vs but $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are neither independent nor identically distributed because of order restriction.

Distributions of $\mathbf{o.s.}$ in continuous case:

Let $X_1, X_2, \dots, X_n$ be a random sample from a continuous population with c.d.f. $F(x)$ and p.d.f. $f(x)$.

Marginal distributions:

1) The c.d.f. and p.d.f. of X_n, the n^{th} o.s. are given by

   $\boldsymbol{F_{X_{(n)}}(x) = [F(x)]^n}$ and $\boldsymbol{f_{X_{(n)}}(x) = n[F(x)]^{n-1}f(x)}$respectively.

2) The c.d.f. and p.d.f. of $X_{(1)}$, the first o.s. are given by

   $F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$ and $f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1}f(x)$ respectively.

3) The c.d.f. and p.d.f. of $X_{(j)}, 1 \le j \le n$, the j^{th} o.s. are given by

$$F_{X_{(j)}}(x) = \sum_{i=j}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i} \text{ and}$$

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

   respectively.

Joint distributions

4) For $1 \le i < j \le n$, the j.p.d.f. of $X_{(\,i\,)}$ and $X_{(\,j\,)}$ is given by
$f_{X_{(\,i\,)},X_{(\,j\,)}}(u, v)$
$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(u)]^{i-1}[F(v) - F(u)]^{j-i-1}[1 - F(v)]^{n-j} f(u)f(v)$
for $-\infty < u < v < \infty$

5) The j.p.d.f. of k-order statistics $X_{(j_1)}, X_{(j_2)}, \dots, X_{(j_k)}$ where $1 \le r_1 < r_2 < \cdots < r_k \le n$ and $1 \le k \le n$ is for $x_1 \le x_2 \le \cdots \le x_k$ given by
$f_{X_{(j_1)},\dots,X_{(j_k)}}(x_1, \dots, x_k)$

$$= \frac{n!}{(j_1 - 1)!\,(j_2 - j_1 - 1)!\,\dots(j_k - j_{k-1} - 1)!\,(n - j_k)!} \times$$
$$F^{j_1-1}(x_1)[F(x_2) - F(x_1)]^{j_2-j_1-1} \dots [F(x_k) - F(x_{k-1})]^{j_k-j_{k-1}-1} \times$$
$$\left[[1 - F(x_k)]^{n-j_k}\right] f(x_1)f(x_2) \dots f(x_k)$$

6) The j.p.d.f. of $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ is given by

$$f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n!\,f(x_1) \dots f(x_n) \;, & -\infty < x_1 < \cdots < x_n < \infty \\ \qquad 0 \qquad\quad , & otherwise \end{cases}$$

Distribution of Range: Let us obtain the p.d.f. of the r.v. $R_{ij} = X_{(j)} - X_{(i)}$ for $i < j$. The j.p.d.f. of $X_{(i)}$ and $X_{(j)}$ is given by

$f_{X_{(i)},X_{(j)}}(u, v)$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(u)]^{i-1}[F(v) - F(u)]^{j-i-1}[1 - F(v)]^{n-j} f(u)f(v) \dots .(1)$$

Let $R_{ij} = X_{(j)} - X_{(i)}$ and $X = X_{(i)} \;\; \Longrightarrow r_{ij} = v - u$ and $x = u$

$\Longrightarrow u = x$ and $v = r_{ij} + x$

The Jacobian of transformation is given by

$$J = \frac{\partial(u,v)}{\partial(x,r_{ij})} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial r_{ij}} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial r_{ij}} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \text{ and } |J| = 1 \qquad \ldots(2)$$

From (1) and (2), the j.p.d.f. of $X_{(i)}$ and R_{ij} is given by

$$f_{X_{(i)},R_{ij}}(x, r_{ij})$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(x)]^{i-1} [F(x + r_{ij}) - F(x)]^{j-i-1} \times$$

$$[1 - F(x + r_{ij})]^{n-j} f(x) f(x + r_{ij}) \quad \ldots (3)$$

From (2), the m.p.d.f. of R_{ij} is given by

$$f_{R_{ij}}(r_{ij}) = \int_{-\infty}^{\infty} f_{X_{(i)},R_{ij}}(x, r_{ij})\,dx \qquad \ldots (4)$$

Let $j = n$ and $i = 1$. Then the range is given by $W = X_{(n)} - X_{(1)}$. From (3) and (4), the p.d.f. of W is given by

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

The c.d.f. of w is given by

$$G(w) = P(W \le w) = \int_0^w g(u)\,du$$

$$= \int_0^w \left(n(n-1) \int_{-\infty}^{\infty} [F(x + u) - F(x)]^{n-2} f(x + u) f(x)\,dx \right) du$$

$$= n \int_{-\infty}^{\infty} f(x) \left[\int_0^w (n-1) f(x + u) [F(x + u) - F(x)]^{n-2} \right] dx$$

$$\Rightarrow G(w) = n \int_{-\infty}^{\infty} f(x) [F(x + w) - F(x)]^{n-1}\,dx$$

Example 1: Let X_1, X_2, X_3, X_4 be a random sample of size 4 from uniform $[0, \theta]$ distribution. Find the p.d.f. of $X_{(1)}, X_{(3)}$ and $X_{(4)}$.

Solution: Since each $X \sim U(0, \theta)$, its p.d.f. is given by $f(x) = \frac{1}{\theta}$, $0 < x < \theta$ and its c.d.f is given by

$$F(x) = P(X \leq x) = \int_0^x f(t)dt = \int_0^x \frac{1}{\theta} dt = \left[\frac{t}{\theta}\right]_0^x = \frac{x}{\theta}$$

The p.d.f. of $X_{(1)}$ is given by

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x) = 4\left(1 - \frac{x}{\theta}\right)^{4-1} \frac{1}{\theta}$$

$$\Rightarrow f_{X_{(1)}}(x) = \frac{4}{\theta}\left(1 - \frac{x}{\theta}\right)^3, 0 < x < \theta$$

The p.d.f of $X_{(3)}$ is given by

$$f_{X_{(3)}}(x) = \frac{4!}{2!\,1!}\left[\frac{x}{\theta}\right]^2 \left(1 - \frac{x}{\theta}\right)^1 \frac{1}{\theta} = \frac{12x^2(\theta - x)}{\theta^4}$$

$$\Rightarrow f_{X_{(3)}}(x) = \frac{12x^2(\theta - x)}{\theta^4}, 0 < x < \theta$$

The p.d.f of $X_{(4)}$ is given by

$$f_{X_{(4)}}(x) = n[F(x)]^{n-1} f(x) = 4\left(\frac{x}{\theta}\right)^3 \frac{1}{\theta} = \frac{4x^3}{\theta^4}$$

$$\Rightarrow f_{X_{(4)}}(x) = \frac{4x^3}{\theta^4}, 0 < x < \theta$$

Example 2: Let $X_1, X_2, \dots, X_n$ be i.i.d.r.v s with common p.d.f .

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

Find (i) p.d.f. of $X_{(j)}$, $1 \le j \le n$

(ii) j.p.d.f. of $X_{(j)}$ and $X_{(k)}$ for $1 \le j < k \le n$

(iii) p.d.f. of $R = X_{(n)} - X_{(1)}$

Solution: Given

p.d.f: $f(x) = 1 \, , 0 < x < 1$

c.g.f: $F(x) = \int_0^x f(t)dt = x \quad \Longrightarrow F(x) = x, 0 < x < 1$

(i) The pdf of $X_{(j)}$ is given by

$$f_{X_{(j)}}(x_j) = \frac{n!}{(j-1)!(x-j)!} x_j^{\,j-1}(1-x_j)^{n-j} \text{ for } 0 < x_j < 1 \, , 1 \le j \le n$$

(ii) The j.p.d.f. of $X_{(j)}$ and $X_{(k)}$ is given by

$$f_{X_{(j)},X_{(k)}}(x_j, x_k) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!} x_j^{\,j-1}(x_k - x_j)^{k-j-1}(1-x_k)^{n-k} \, ,$$

$$0 < x_j < x_k < 1 \text{ where } 1 \le j < k \le n$$

The j.p.d.f. of $X_{(1)}$ and $X_{(n)}$ is given by

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = n(n-1)(x_n - x_1)^{n-2} \, , 0 < x_1 < x_n < 1$$

(iii) The p.d.f. of $R = X_{(n)} - X_{(1)}$ is given by

$$g(w) = n(n-1)w^{n-2}(1-w), 0 < w < 1$$

Example 3: Let $X_{(1)}, X_{(2)}, X_{(3)}$ be the o.s. of i.i.d.r.vs X_1, X_2, X_3 with common p.d.f.

$$f(x) = \begin{cases} \beta\, e^{-x\beta}, x > 0, \beta > 0 \\ 0\,, otherwise \end{cases}$$

Let $Y_1 = X_{(3)} - X_{(2)}$ and $Y_2 = X_{(2)}$. Show that Y_1 and Y_2 are independent.

Solution: The c.d.f. is given by $F(x) = \int_0^x f(t)dt = 1 - e^{-x\beta}, x > 0$.

Then the j.p.d.f. of $X_{(2)}$ and $X_{(3)}$ is given by

$$f_{X_{(2)},X_{(3)}}(x, y) = \frac{3!}{1!\,0!\,0!}\left(1 - e^{-\beta x}\right)\beta e^{-\beta x}\beta e^{-\beta y}, 0 < x < y < \infty$$

Here $y_1 = y - x$ and $y_2 = x$

$\Rightarrow x = y_2$ and $y = y_1 + y_2$

The Jacobian of transformation is given by

$$J = \frac{\partial(x,y)}{\partial(y_1,y_2)} = \begin{vmatrix} \frac{\partial x}{\partial y_1} & \frac{\partial x}{\partial y_2} \\ \frac{\partial y}{\partial y_1} & \frac{\partial y}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 \text{ and } |J| = 1$$

The j.p.d.f. of Y_1 and Y_2 is given by

$$f(y_1, y_2) = 3!\,\beta^2\left(1 - e^{-\beta y_2}\right)e^{-\beta y_2}e^{-\beta(y_1+y_2)}\,, 0 < y_1 < \infty\,, 0 < y_2 < \infty \quad .... (1)$$

The m.p.d.f. of Y_2 is given by

$$f_2(y_2) = 3!\,\beta e^{-2\beta y_2}\left(1 - e^{-\beta y_2}\right), 0 < y_2 < \infty \quad .... (2)$$

and the m.p.d.f. of Y_1 is given by

$$f_1(y_1) = \beta e^{-\beta y_1},\ 0 < y_1 < \infty \quad .... (3)$$

From (1), (2) and (3), Y_1 and Y_2 are independent.

Example 4: Let $X_1, X_2, \dots, X_n$ be a random samples from a population with continuous density. Show that $Y = min(X_1, X_2, \dots, X_n)$ is exponential with parameter $n\lambda$ iff each X_i is exponential with parameter λ.

Solution: Let X_i be the i.i.d exponential variates with parameter λ and p.d.f.

$$f(x) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

and $F(x) = P(X \le x) = \int_0^x f(u)du = \lambda \int_0^x e^{-\lambda u} du = 1 - e^{-\lambda x}$

The distribution function of $Y_1 = min(X_1, \dots, X_n)$ is given by

$$F_{Y_1}(y) = 1 - [1 - F(x)]^n = 1 - \left[1 - 1 + e^{-\lambda x}\right]^n = 1 - e^{-(n\lambda)x}$$

which is the distribution function of exponential distribution with parameter $n\lambda$. Thus, $Y_1 \sim exp(n\lambda)$.

Converse: Let $Y_1 = min(X_1, \dots, X_n) \sim exp(n\lambda)$

Then $P(Y_1 \le y) = 1 - e^{-n\lambda y}$

Now, $P(Y_1 \ge y) = 1 - P(Y_1 \le y) = 1 - \left(1 - e^{-n\lambda y}\right) = e^{-n\lambda y}$
$\Rightarrow P[min(X_1, \dots, X_n) \ge y] = e^{-n\lambda y}$

$\Rightarrow P[X_1 \ge y\,,\ X_2 \ge y\,, \dots.., X_n \ge y] = e^{-n\lambda y}$

$\Rightarrow \prod_{i=1}^{n} P(X_i \ge y) = e^{-n\lambda y}$ $\qquad (\because \quad Xs\ are\ i.d.d)$

$\Rightarrow P(X_i \ge y) = e^{-\lambda y} \quad \Rightarrow P(X_i \le y) = 1 - e^{-\lambda y}$

which is $exp(\lambda)$ distribution. Thus, X_i's are i.d.d $Exp(\lambda)$.

Example 5: For exponential distribution $f(x) = e^{-x}, x \geq 0$, show that the c.d.f. of $X_{(n)}$ in a random sample of size n is $F_n(x) = (1 - e^{-x})^n$. Hence prove that as $n \longrightarrow \infty$, the c.d.f. of $X_n - \ln n$ tends to the limiting form

$$\boldsymbol{exp(-exp(-x)), -\infty < x < \infty.}$$

Solution: Here $f(x) = e^{-x},\ x \geq 0 \ \Longrightarrow F(x) = P(X \leq x) = 1 - e^{-x}$.

The c.d.f of $X_{(n)}$ is given by $F_{X_{(n)}} = [F(x)]^n = (1 - e^{-x})^n$

The c.d.f. of $X_{(n)} \ - \ln n$ is given by

$$G_n(x) = P\left[X_{(n)} - \ln n \leq x\right] = P\left[X_{(n)} \leq x + \ln n\right]$$

$$= \left[1 - e^{-(x + \ln n)}\right]^n = [1 - e^{-x}e^{-\ln n}]^n$$

$$\Longrightarrow \quad G_n(x) = \left(1 - \frac{e^{-x}}{n}\right)^n$$

$$\Longrightarrow \lim_{n\to\infty} G_n(x) = \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n = e^{-e^{-x}} \qquad \left(\because \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x\right)$$

Distribution of O.S. in discrete case: In discrete case there is no magic formula to compute the distribution of any $X_{(j)}$ or any of the joint distributions. A direct computation is the best course of action.

Let $X_1, X_2, \ldots X_n$ be a random sample, from a population with p.m.f. $p(x_i) = P(X = x_i)$ for $i = 1, 2, \ldots$

Let $r_i = \sum_{k=1}^{i} p(x_k)$. Then $P\left(X_{(j)} \leq x_i\right) = \sum_{k=j}^{n} \binom{n}{k} r_i^k (1 - r_i)^{n-k}$

$$P\left[X_{(j)} = x_i\right] = \sum_{k=j}^{n} \binom{n}{k} \left[r_i^k (1 - r_i)^{n-k} - r_{i-1}^k (1 - r_{i-1})^{n-k}\right]$$

Example 6: Let $X_1, X_2, \dots X_n$ are i.i.d.r.vs with common geometric p.m.f. given by

$$p_k = P(X = k) = pq^{k-1}, \;\; k = 1, 2, \dots, \;\; 0 < p < 1, \; q = 1 - p$$

(i) Find p.m.f. of $X_{(r)}, 1 \le r \le n$ and

(ii) Show that X_1 and $X_{(2)} - X_{(1)}$ are independent random variables and $X_{(2)} - X_{(1)}$ has a geometric distribution.

Solution:

(i) For any integer $x \ge 1$ and $r \ge 1$,

$$P[X_{(r)} = x] = P[X_{(r)} \le x] - P[X_{(r)} \le (x-1)]$$

Now $P(X_{(r)} \le x) = P[at\ least\ \ r\ of\ X\ s\ are\ \le x]$

$$= \sum_{i=1}^{r} \binom{n}{i} \left[P(X_1 \le x) \right]^i \left[P(X_1 > x) \right]^{n-i}$$

and $P(X_1 \ge x) = \sum_{k=x}^{\infty} pq^{k-1} = (1-p)^{x-1} = q^{x-1}$

It follows that, $P\left[X_{(r)} = x \right] = \sum_{i=r}^{n} \binom{n}{i} q^{(x-1)(n-i)} \left[q^{n-i} \left(1 - q^x\right)^i - \left(1 - q^{x-1}\right)^i \right]$

$x = 1, 2, \dots.$

(ii) Let $n = r = 2$. Then , $P[X_{(2)} = x] = pq^{x-1} \left(pq^{x-1} + 2 - 2q^{x-1}\right), x \ge 1$

Also, for integers $x, y \ge 1$, we have $P[X_{(1)} = x, \; X_{(2)} - X_{(1)} = y]$

$$= P[X_{(1)} = x, X_{(2)} = x + y]$$
$$= P[X_1 = x, X_2 = x + y] + P[X_1 = x + y, X_2 = x]$$
$$= 2pq^{x-1}pq^{x+y-1} = 2pq^{2x-2}pq^{y}$$
$$= P(X_{(1)} = x) P(X_{(2)} = y)$$

and $P(X_{(1)} = 1, \; X_{(2)} - X_{(1)} = 0) = P(X_{(1)} = X_{(2)} = 1) = p^2$

It follows that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent random variables and, moreover, that $X_{(2)} - X_{(1)}$ a geometric distribution.