Unit-V

Order statistics:

Let $X_1, X_2, \ldots, X_n$ be n independent identically distributed variables each with cumulative distribution function $F(x)$ and pdf $f(x)$.

If these variables are arranged in ascending order of magnitude and written $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ we called $X_{(r)}$ as the r^{th} order statistics $(r = 1, 2, \ldots, n)$

The $X_{(r)}$'s are independent because of inequality relation among them. Here $X_{(1)}$ is the smallest of $X_1, X_2 \ldots X_n$, $X_{(n)}$ is the largest of $X_1, X_2, \ldots X_n$ and $X_{(r)}$ is the r^{th} smallest of $X_1, X_2, \ldots, X_n$.

Distribution function of a single order statistics:

Let $F_r(x)$; $r = 1, 2, \ldots, n$ denote the distribution function of r^{th} order statistic $X_{(r)}$. Then the distribution function $F_n(x)$ of the largest order statistics $X_{(n)}$ is given by

$$F_n(x) = P[X_{(n)} \leq x] = P[X_i \leq x ; i = 1, 2, \ldots, n]$$

$$= P[X_1 \leq x \cap X_2 \leq x \cap X_3 \cap \ldots \cap X_n \leq x]$$

$$= P[X_1 \leq x] \cdot P[X_2 \leq x] \ldots P[X_n \leq x]$$

Since X_i's are independent

$$F_n(x) = [F(x)]^n$$

The distribution function of smallest order statistic is given by

$$F_1(x) = P[X_{(1)} \leq x]$$

$= 1 - P[x_i \le x;\ i = 1, 2, \dots, n]$

$= 1 - P[(X_1 > x) \cap (X_2 > x) \cap \dots \cap (X_n > x)]$

$= 1 - [p(x_1 > x) \cdot p(x_2 > x) \dots p(x_n > x)]$

$= 1 - \prod_{i=1}^{n} p(X_i > x)$

$= 1 - \prod_{i=1}^{n} [1 - p(X_i < x)]$

$F_1(x) = 1 - [1 - F(x)]^n$

In general the pdf of r^{th} order statistic,

$F_r(x) = p[X_r \le x]$ — ①

$= p[\text{Atleast } r \text{ of } x_i\text{'s are} \le x]$

$= \sum_{j=r}^{n} p[\text{Exactly } i \text{ of } n\ x_i\text{'s are} \le x]$

$= \sum_{j=r}^{n} \binom{n}{j} F^j_{(x)} [1 - F(x)]^{n-j}$ — ②

using binomial probability model, taking $r=1$ in equation ②. we get,

$F_1(x) = \sum_{j=1}^{n} \binom{n}{j} \cdot F_x^j [1 - F(x)]^{n-j}$

$= 1 - \left[\binom{n}{j} F_x^j (1 - F(x))^{n-j}\right]_{j=0}$

$= 1 - [1 - F(x)]^n$

If $r = n$, $F_n(x) = \binom{n}{1} F^n_{(x)} [1 - F(x)]^{n-n}$

$= F^n_{(x)}$

If it is continuous type equation ② can be written as,

$F_r(x) = I_{F(x)}(r, n-r+1)$

where, $I_p(a, b) = \frac{1}{\beta(a,b)} \int_a^b t^{a-1} (1-t)^{b-1} dt$ in the compute the β function.

$$f_r(x) = \frac{1}{\beta(r, n-r+1)} \int_{r}^{n-r+1} t^{r-1} (1-t)^{n-r} dt$$

This shows that the probability points of an order statistics can be obtained with the help of incomplete β function.

Pdf of single order statistics:

We assume that X_i's are i.i.d continuous random variable with pdf, If $f_r(x)$ denotes the pdf of $X_{(r)}$ then,

$$f_r(x) = \frac{d}{dx}\left[I_{F(x)}(r, n-r+1)\right]$$

$$= \frac{d}{dx}\left[\frac{1}{\beta(r, n-r+1)} \int_{0}^{f(x)} t^{r-1}(1-t)^{n-r+1} dt\right] \quad -①$$

Let us write,

$$g(t) = \int t^{r-1} (1-t)^{n-r} dt$$

$$= t^{r-1} (1-t)^{n-r}$$

$$= \int_{0}^{F(x)} t^{r-1} (1-t)^{n-r} dt$$

$$= \left| g(t) \right|_{0}^{F(x)}$$

$$= g\left[F(x) - g(0)\right]$$

$$\Rightarrow \frac{d}{dx} \int_{0}^{f(x)} t^{r-1} (1-t)^{n-r} dt$$

$$= g'\left[F(x)\right] \cdot f(x)$$

$$= \left[F(x)\right]^{r-1} \left[1 - F(x)\right]^{n-r} f(x)$$

Substitute in ①, we get

$$f_r(x) = \frac{1}{\beta(r, n-r+1)} F(x)^{r-1} \left[1 - F(x)\right]^{n-r} f(x)$$

Joint pdf of two order statistics:

Let us denote the joint pdf of $X_{(r)}$ and $X_{(s)}$

when $1 \le r \le s \le n$ by, then

$$f_{rs}(x,y) = \lim_{\substack{\partial x \to 0 \\ \partial y \to 0}} \frac{P[(x \le X_{(r)} \le x + \partial x) \cap (y \le X_{(s)} \le y + \partial y)]}{\partial x \, \partial y} \quad \text{↳ ①}$$

The event $E = (x \le X_{(r)} \le x + \partial x) \cap (y \le X_{(s)} \le y + \partial y)$

$X_i \le x$ for $r-1$ of X_i's

$x \le X_i \le x + \partial x$ for one X_i

$x + \partial x \le X_i \le y$ for $s-r-1$ of X_i's

$y < X_i \le y + \partial y$ for X_i and

$X_i > y + \partial y$ for $(n-s)$ X_i's

using multinomial probability law, we get

$$P(E) = P[(x \le X_{(r)} \le x + \partial x) \cap (y \le X_{(s)} \le y + \partial y)]$$

$$P(E) = \frac{n!}{(r-1)!\,1!\,(s-r-1)!\,1!\,(n-s)!} P_1^{r-1} P_2 P_3^{s-r-1} P_4 P_5^{n-s} \quad - ②$$

where,

$P_1 = P[X_1 \le x] = F(x)$

$P_2 = P[x < X_i \le x + \partial x] = F[(x + \partial x)] - [F(x)]$

$P_3 = P[x + \partial x < x_i \le y] = F(y) - F(x + \partial x)$

$P_4 = P[(y < x_i \le y + \partial y)] = F(y + \partial y) - F(y)$

$P_5 = P[X_i > y + \partial y] = 1 - P[X_i \le y + \partial y]$

$= 1 - F(y + \partial y)$

Substitute in ② and using ① we get

$$f_{rs}(x,y) = \lim_{\substack{\partial x \to 0 \\ \partial y \to 0}} \frac{P(E)}{\partial x \cdot \partial y}$$

$$f_{rs}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F_{(x)}^{r-1} \lim_{\partial x \to 0} \frac{F(x+\partial x) - F(x)}{\partial x}$$

$$\lim_{\partial x \to 0} [F(y) - F(x+\partial x)]^{s-r-1} \cdot \lim_{\partial y \to 0} \frac{F(y+\partial y) - F(y)}{\partial y}$$

$$\cdot \lim_{\partial y \to 0} \left[\frac{F(y+\partial x) - F(y)}{\partial y}\right]^{n-s}$$

$\therefore$ Which is the joint pdf of two order statistics

$$= \frac{n!}{(r-1)!(s-r-1)!} F_{(x)}^{r-1} f(x) [F(y) - F(x)]^{s-r-1} f(y) [1-F(y)]^{n-s}$$

Density function of $X_{(n)}$:

The density function of $X_{(n)}$

$$F[X_{(n)}] = \frac{d}{dx} [F(x)]^n$$

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

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

Problem:

1. For the exponential distribution $f(x) = e^{-x}$, $x \geq 0$; show that the cumulative distribution function of X_n is a random variable of size n is, $F_n(x) = (1-e^{-x})^n$, Hence prove that as $n \to \infty$, the c.d.f of $X_{(n)} - \log n$ tends to the limiting form $\exp[-\{\exp(-x)\}; -\infty < x < \infty]$

Hence $f(x) = e^{-x}$, $x \geq 0 \Rightarrow F(x) = p[X \leq x) = 1 - e^{-x}$ - ①

The cdf of X_n is given by,

$$F_n(x) = p[X_{(n)} \leq x] = [F(x)]^n = [1-e^{-x}]^n \text{ - ②}$$

The cdf $G_n(\cdot)$ of $X_n - \log n$ is given by,

$$G_n(x) = P[X_{(n)} - \log n \leq x]$$

$$= P[X_n \le x + \log n]$$

$$= \{1 - e^{-x - \log n}\}^n \text{ from } ②$$

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

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

2. Show that for a random sample of size 2 from $N(0, \sigma^2)$ population.

For $n=2$, the pdf $f_1(x)$ of $X_{(1)}$ is given by

$$f_1(x) = \frac{1}{\beta(1,2)} \{1 - F(x)\} f(x) = 2\{1 - F(x)\} f(x) \; ; \quad -\infty < x < \infty$$

$$\text{where } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/\sigma^2} \quad [X \sim N(0, \sigma^2)]$$

$$\therefore E[X_{(1)}] = \int_{-\infty}^{\infty} x \cdot f_1(x) \, dx$$

$$= 2\int_{-\infty}^{\infty} \{1 - F(x)\} x f(x) \, dx \quad — ①$$

We have $\log f(x) = -\log(\sqrt{2\pi}, \sigma) - \frac{x^2}{2\sigma^2}$

Differentiating with respect to x, we get

$$\frac{f'(x)}{f(x)} = \frac{-x}{\sigma^2} \Rightarrow \int x f(x) \, dx$$

$$= -\sigma^2 \int f'(x) \, dx = -\sigma^2 f(x) \quad — ②$$

Integrating ① by parts and using ② we get

$$E[X_{(1)}] = 2\left[\{1 - F(x)\}\{-\sigma^2 f(x)\}\right]_{-\infty}^{\infty}$$

$$= 2\int_{-\infty}^{\infty} \{-\sigma^2 f(x)\}\{-f(x)\} dx$$

$$= -2\sigma^2 \int_{-\infty}^{\infty} [F(x)]^2 \, dx$$

$$E[X_{(1)}] = \frac{-1}{\pi} \int_{-\infty}^{\infty} [F(x)]^2 \, dx$$

$$= \frac{-1}{\pi} \int_{-\infty}^{\infty} e^{-x^2/\sigma^2} \, dx \qquad \because \int_{-\infty}^{\infty} e^{-a^2x^2} \, dx = \frac{\sqrt{\pi}}{a}$$

$$= \frac{-1}{\pi} \sqrt{\pi} \cdot \sigma = \frac{-\sigma}{\sqrt{\pi}}$$

3. Show that in odd samples of size n from $v[0,1]$ population the mean and variance of the distribution of median are $\frac{1}{2}$ and $\frac{1}{[4+(n+2)]}$ respectively.

We have, $f(x) = 1; \ 0 \le x \le 1$

$$F(x) = P[X \le x) = \int_0^x f(u) \, du = \int_0^1 1 \cdot du = u$$

Let $n = 2m+1$ (odd), where m is a positive integer ≥ 1 than median observation is $X_{(m+1)}$ taking $r = (m+1)$ in the pdf of median $X_{(m+1)}$ is given by

$$f_{(m+1)}(x) = \frac{1}{\beta(m+1, m+1)} x^m (1-x)^m \, dx$$

$$E[X_{(m+1)}] = \frac{1}{\beta(m+1, m+1)} \int_0^1 x \cdot x^m (-1-x)^m \, dx$$

$$= \frac{\beta(m+2, m+1)}{\beta(m+1, m+1)}$$

$$= \frac{\sqrt{m+2}\,\sqrt{m+1}}{\sqrt{m+1}\,\sqrt{m+1}\,\sqrt{2m+3}} \times \frac{\sqrt{2m+2}}{\sqrt{m+1}\,\sqrt{m+1}}$$

$$= \frac{m+1}{2m+2} = \frac{1}{2}$$

$$E[X^2_{(m+1)}] = \int_0^1 x^2 f_{(m+1)}(x)\, dx$$

$$= \frac{1}{\beta(m+1, m+1)} \int_0^1 x^{m+2} (1-x)^m \, dx$$

$$= \frac{\beta(m+3, m+1)}{\beta(m+1, m+1)} = \frac{m+2}{2(2m+3)}$$

$$V[X_{(m+1)}] = E[X^2_{(m+1)}] - [E(X_{(m+1)})]^2$$

$$= \frac{m+2}{2(2m+3)} - 1/4$$

$$= \frac{1}{4(2m+3)}$$

$$= \frac{1}{4(n+2)}$$

4. Let $X_1, X_2, \ldots, X_n$ be i.i.d non-negative random variable of the continuous type with pdf $f(\cdot)$ and distribution function $F(\cdot)$. a) If $E|X| < \infty$, show that $E|X_r| < \infty$

b) Write $M_n = X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ show that

$$E(M_0) = E(M_{n-1}) + \int_0^\infty F^{n-1}(x)[1 - F(x)]\, dx; n = 2, 3, \ldots$$

Hence the Evaluate $E(\mu_0)$ if $X_1, X_2, \ldots, X_n$ have common distribution function $F(x) = x$; $0 < x < 1$

a) $E|X_r| = \int_0^\infty |x| f_r(x)\, dx$ ($\because$ x is non-negative continuous random variables)

$$= \int_0^\infty |x| \cdot \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x) [1-F(x)]^{n-r}\, dx$$

$$= n\binom{n-1}{r-1} \int_0^\infty |x| f(x)\, dx \leq n \binom{n-1}{r-1} E|X|$$

Hence $E|X_r| < \infty$ if $E|X| < \infty$

b) The pdf $f_n(x)$ of $M_n = X_{(n)}$ is given by

$$f_n(x) = n\,[F(x)]^{n-1} f(x) \qquad (\because x \geq 0 \text{ as})$$

$$E(M_n) = \lim_{a\to\infty} \int_0^{\infty} x\, f_n(x)\, dx$$

$$E(M_n) = \lim_{a\to\infty} n \int_0^{\infty} x\, f_n(x)\, dx$$

$$= \lim_{a\to\infty} n \int_0^{\infty} x\,[F_x]^{n-1} f(x)\, dx$$

Integrating by parts we get

$$E(M_n) = n.\lim_{a\to\infty} \left\{ \left| x.\frac{F^n(x)}{n} \right|_0^a - \int_0^a \frac{F^n(x)}{n}\, 1.\,dx \right\}$$

$$= \lim_{a\to\infty} \left\{ a.F^n(a) - \int_0^a F^n(x)\, dx \right\}$$

$$= \lim_{a\to\infty} \int_0^a \left[1 - F^n(x)\, dx - a + aF^n(a) \right]$$

$$= \lim_{a\to\infty} \left[\int_0^a \{1 - F^n(a) - 1\} dx + aF^n(a) \right]$$

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

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

Since $E(M_n)$ exists by part (a)

$$a.p(M_n > a) = a\left[1 - p(M_n \leq a)\right] = a\left[1 - F^n(a)\right] \to 0 \text{ as } a \to \infty$$

$$E(M_n) = \lim_{a\to\infty} \int_0^a \left[1 - F^n(x) \right] dx$$

$$= \int_0^{\infty} 1 - F^n(x)\, dx \quad \text{—①}$$

$$= \int_0^{\infty} \left[1 - F^{n-1}(x)\, F(x) \right] dx$$

$$E(M_n) = \int_0^{\infty} \left\{ 1 - F^{n-1}(x) \right\} \left\{ 1 - \left[1 - F(x) \right] \right\} dx$$

$$= \int_0^\infty \{1 - F^{n-1}(x)\} dx + \int_0^\infty F^{n-1}(x) (1 - F(x)) dx$$

$$= E(M_{n-1}) + \int_0^\infty F^n(x) \{1 - F(x)\} dx \quad -② \text{, From } ①$$

If $X \sim U(0,1)$ then $f(x) = 1; \ 0 < x < 1$ and

$F(x) = x, \ 0 < x < 1$

Substituting in ②, we get

$$E(M_n) - E(M_{n-1}) = \int_0^1 x^{n-1}(1-x) dx = \frac{1}{n} - \frac{1}{n-1} \quad - ③$$

Changing n to $n-1, n-2, \dots,$ in ③ we get respectively

$$E(M_{n-1}) - E(M_{n-2}) = \frac{1}{n-1} - \frac{1}{n}$$

$$E(M_2) - E(M_1) = \frac{1}{2} - 1$$

$$E(M_1) - E(M_0) = 1 - \frac{1}{2} = \frac{1}{2}$$

Adding ③ and the above equations and nothing that $E(M_0) = 0$ we get

5.

5. a) Find the pdf of $X_{(r)}$ in a random sample of size n from the exponential distribution, $f(x) = \alpha e^{-\alpha x}; \ \alpha > 0$

b) Show that $X_{(r)}$ and $W_{rs} = X_{(s)} - X_{(r)}$; $r < s$ are independently distributed.

c) What is the distribution $W_1 = X_{(r+1)} - X_{(r)}$

a) Hence, $F(x) = P(X \leq x) = \int_0^x \alpha e^{-\alpha u} du = 1 - e^{-\alpha u} \quad -①$

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

$$f_r(x) = \frac{1}{\beta(r, n-r+1)} [F(x)]^{r-1} [1 - F(x)]^{n-r} f(x)$$

$$f_r(x) = \frac{1}{B(r, n-r+1)} (1-e^{-\alpha x})^{r-1} \cdot e^{-\alpha x(n-1)} \cdot \alpha e^{-\alpha x}$$

$$= \frac{1}{\beta(r, n-r+1)} \alpha e^{-\alpha x(n-r+1)} (1-e^{-\alpha x})], x>0$$

b) The joint pdf of $X_{(r)}$ and $W_{rs} = X_{(s)} - X_{(1)}$ is given by

$$g(x, W_{rs}) = C_{rs} F^{r-1}(x) f(x) [(F(x) + W_{rs}) - F(x)]^{s-r-1}$$

$$f(x + W_{rs}) [1 - F(x + W_{rs})]^{n-r}$$

$$= \frac{n!}{(r-1)!(n-r)!} * \frac{(n-1)!}{(s-r-1)!(n-s)!} \cdot [1-e^{-\alpha x}] \alpha e^{-\alpha x}$$

$$[e^{-\alpha x} - e^{-\alpha(x + W_{rs})}]^{s-r-1} \alpha e^{-\alpha(x + W_{rs})} \cdot \frac{1}{\beta(r, n-r+1)}$$

$$\alpha e^{-(n-r+1)\alpha + W_{rs}} (1 - e^{-\alpha W_{rs}})^{rs-r+1} \quad - (2)$$

$\Rightarrow X_r$ and W_{rs} are independently distributed

c) Taking, $s = r+1$ in (2) the pdf of $W_1 = X_{(r-1)} - X_{(r)}$

$$g(w_1) = \frac{1}{\beta(1, n-r)} \alpha e^{-\alpha(n-1) w}$$

$$= (n-r)\alpha e^{-(n-r)\alpha W} ; W_1 \geq 0$$

Which shows that W_1 has an exponential distributi with parameter $(n-r)\alpha$

Joint pdf of k-order statistics:

The joint pdf of k-order statistics $X_{(1)}, X_{(2)}$ ... $X_{(rk)}$ where $1 \leq r_1 \leq r_2 \leq \ldots \leq r_k \leq n$ and

$1 \le k \le n$ for $x_1 \le x_2 \le \dots \le x_k$ is given by

$$f_{r_1, r_2, \dots, r_k}(x_1, x_2, \dots, x_k) = \frac{n!}{(r-1)!(r_2-r_1-1)! \dots (r_k - r_{k-1} - 1)(n-r_k)!}$$

$$[F(x_1)]^{r-1} \times f(x_1) \times [f(x_2) - f(x_1)]^{r_2-r_1-1} \times f(x_2) \times \dots \times f(x_r)$$

$$[1 - F(x_k)^{n-k}] \quad - (3)$$

Joint pdf of all n-order statistics:

In particular the joint pdf of all the n-order statistics is obtain on taking $k=n$ in (3)

$\Rightarrow r_i \simeq i$ for $(i = 1, 2, \dots, n)$

Hence joint pdf of n order statistics is given by,

$$f_{r_1, r_2, \dots, r_n}(x_1, x_2, \dots, x_k) = n! \, f(x_1) \, f(x_2) \dots f(x_n).$$