5.4 Order Statistics

Sample values such as the smallest, largest, or middle observation from a random sample can provide additional summary information.

Definition 5.4.1:

The order statistics of a random sample $X_1, X_2, \dots, X_n$ are the sample values placed in acending order. They are denoted by $X_{(1)}, X_{(2)}, \dots, X_{(n)}$

The order statistic are random variables that satisfy $X_{(1)} \le X_{(2)} \le X_{(3)} \le \dots \le X_{(n)}$

$$X_{(1)} = \min_{1 \le i \le n} X_i$$

$$X_{(2)} = \text{Second Smallest } X_i$$

$$\vdots$$

$$X_{(n)} = \max_{1 \le i \le n} X_i$$

Sample range: $R = X_{(n)} - X_{(1)}$

distance b/w the smallest and Largest observation's.

Sample median: M, is a number s.t approximately one-half of the observation's are less than M and one-half are greater.

$$M = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & n = \text{odd} \\ \dfrac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} & n = \text{even} \end{cases}$$

For any number p b/w 0,1, the $(100p)^{th}$ sample percentile is the observation s.t approximately np of the observation's are less than this observation and $n(1-p)$ of the observation's are greater.

Ex: $p = 0.5 \Rightarrow 50^{th}$ percentile

$\Rightarrow 50^{th}$ sample percentile = sample median

Definition 5.4.2:

The notation $\{b\}$, when appearing in a subscript, is defined to be the number b rounded to the nearest integer in the usual way. More precisely, if i is an integer and

$$i - 0.5 \leq b < i + 0.5, \text{ then } \{b\} = i$$

$(100p)^{th}$ Sample Percentile is $X_{(\{np\})}$

if $\frac{1}{2n} < p < 0.5$ and $X_{(n+1-\{n(1-p)\})}$

if $0.5 < p < 1 - \frac{1}{2n}$.

Ex: $n = 12$ (12 samples)

65^{th} Sample Percentile ?

$\Rightarrow 12 \times (1 - 0.65) = 4.2$ and $12 + 1 - 4 = 9$

$\Rightarrow 65^{th}$ sample Percentile is $X_{(9)}$

Theorem 5.4.3 :- Let $X_1, X_2, \dots X_n$ be a random sample from a discrete distribution with pmf $f_{X_i}(x_i) = p_i$ where $x_1 < x_2 < \dots$ are the possible value of X in ascending order

Define

$$P_0 = 0$$

$$P_1 = p_1$$

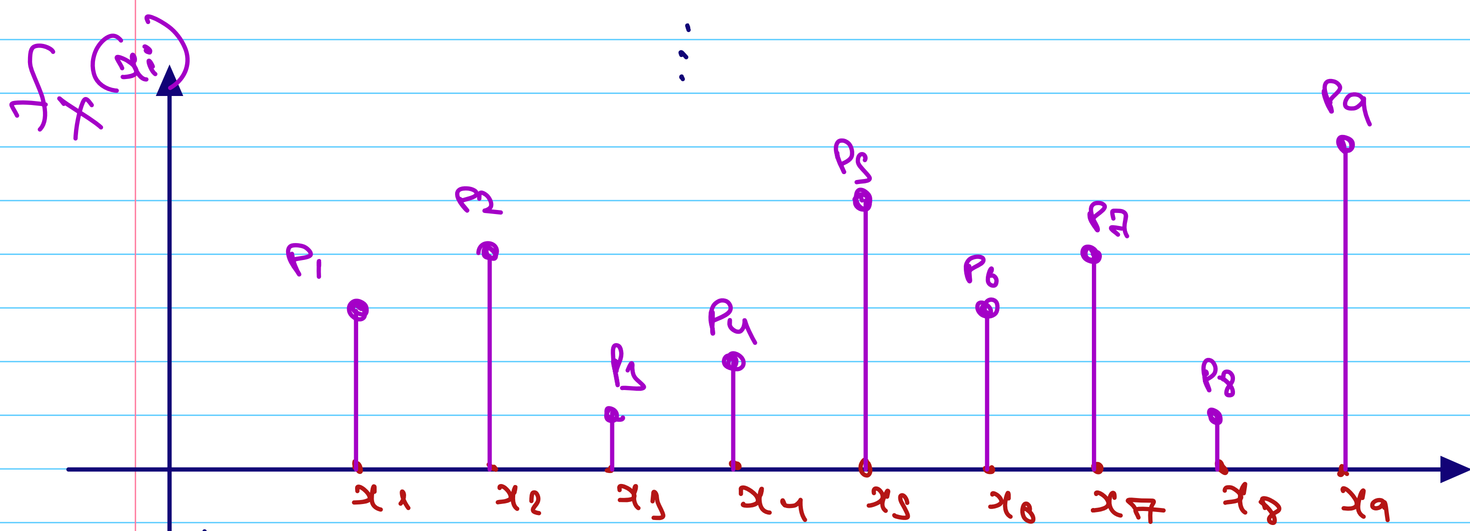
$$P_2 = p_1 + p_2$$

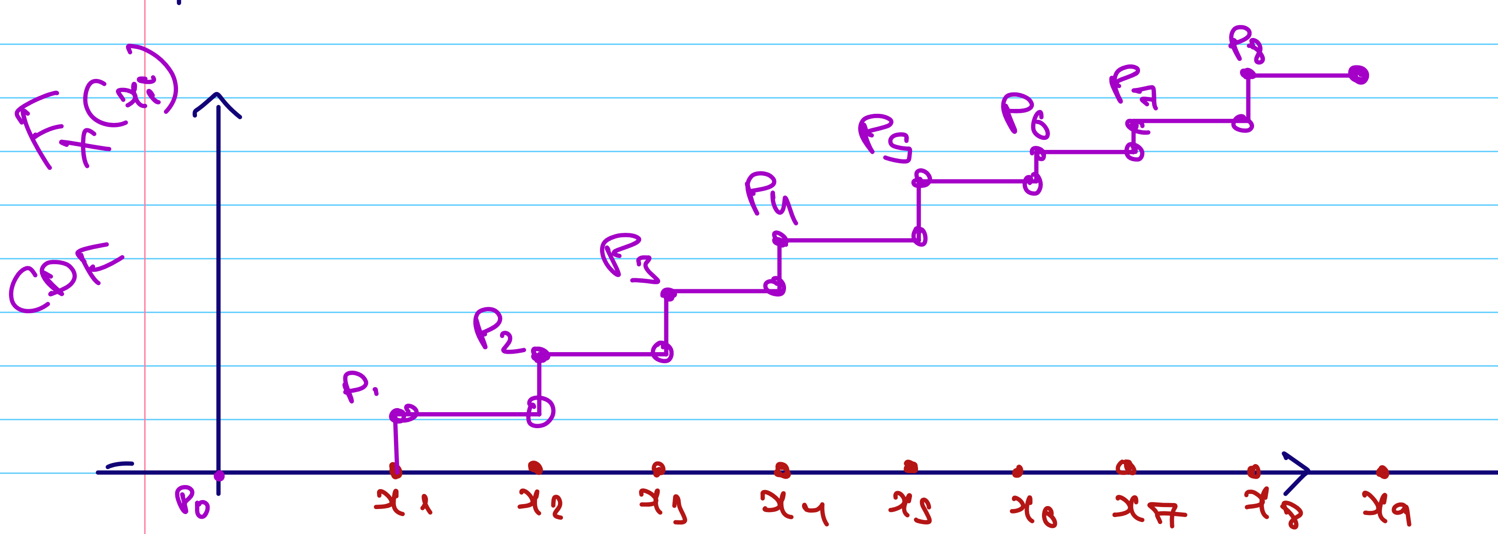
$$P_3 = p_1 + p_2 + p_3$$

$$\vdots$$

$$P_i = p_1 + p_2 + \dots + p_i$$

$$\vdots$$





Let $X_{(1)}, X_{(2)}, \dots X_{(n)}$ denote the order statistics from the sample, then.

$$\mathbb{P}(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} p_i^k (1-p_i)^{n-k}$$

and

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

Proof:

① fix $i \in \mathbb{N}$

② Y is a r.v that counts the number of $X_1, X_2, \dots, X_n$ that are less than or equal to x_i.

$\Rightarrow$ for each of $X_1, X_2, X_3 \dots X_n$, call the event $\{X_j \leq x_i\}$ a "success" and $\{X_j > x_i\}$ a "failure".

$\Rightarrow$ Y = number of success in n trails.

The Probability of success is the same value $= P_i = P(X_j \leq x_i)$, for each trail, since $X_1, X_2, \ldots X_n$ iid X.

$\Rightarrow$ $Y \sim \text{binomial}(n, P_i)$

(3) The event $\{X_{(j)} \leq x_i\}$ is equivalent to the event $\{Y \geq j\}$, that is at least j samples value are less than or equal to x_i

$$x_1 \quad x_2 \quad x_3 \quad x_4 \;|\; x_5 \quad x_6 \quad x_7 \quad x_8$$

drawn 10 samples :-

$X_{(6)} \leq x_4 \Rightarrow$ at least 6 success

$\Rightarrow \{Y \geq 6\}$

$$\mathbb{P}(X_{(j)} \le x_i) = \mathbb{P}(Y \geq j)$$

$$= \sum_{k=j}^{n} \binom{n}{k} p_i^k (1-p_i)^{n-k}$$

$\mathbb{P}(X_{(j)} = x_i) \Rightarrow$ exatly j sucess

& $n-j$ failure's

$$\Rightarrow \mathbb{P}(X_{(j)} \le x_i) - \mathbb{P}(X_{(j)} \le x_{i-1})$$

if $X_1, X_2, \dots X_n \overset{iid}{\sim} X$ (continuous case)

the Probability that two X_i's same

$= 0$

$$\Rightarrow \mathbb{P}(X_{(1)} < X_{(2)} < X_{(3)} \dots < X_{(n)}) = 1$$

Theorem 5.4.4:

Let $X_{(1)}, X_{(2)}, \dots X_{(n)}$ denote the order statistics of a random sample, $X_1, \dots X_n$ from a continuous population with CDF $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}$$

Proof:

1. Y = number of $X_1, \dots X_n$ less than or equal to x.

2. then $\{X_j \leq x\}$.

$$\Rightarrow \mathbb{P}(X_j \leq x) = \mathbb{P}(Y \geq j)$$

$$\Rightarrow Y \sim \text{binomial}\,(n, F_X(x))$$

$$F_{X_{(j)}}(x) = \mathbb{P}(Y \geq j) = \sum_{k=j}^{n} \binom{n}{k} F_X(x)^k \, (1 - F_X(x))^{n-k}$$

$$f_{X_{(j)}}(x) = \frac{d}{dx} F_{X_{(j)}}(x)$$ give the above formula.

$$= \frac{d}{dx}\left[\sum_{k=j}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k} \right]$$

$$= \sum_{k=j}^{n} \left[\binom{n}{k} k \cdot \left[F_X(x)\right]^{k-1} \cdot f_X(x) \left[1 - F_X(x)\right]^{n-k} - \binom{n}{k} \left[F_X(x)\right]^{k} \cdot (n-k) \left(1 - F_X(x)\right)^{n-k-1} \cdot f_X(x) \right]$$

$$= \binom{n}{j} j \cdot \left[F_X(x)\right]^{j-1} \cdot f_X(x) \left[1 - F_X(x)\right]^{n-j}$$

$$+ \sum_{k=j+1}^{n} \binom{n}{k} k \cdot \left[F_X(x)\right]^{k-1} \cdot \left[1 - F_X(x)\right]^{n-k} \cdot f_X(x)$$

$$- \sum_{k=j}^{n-1} \binom{n}{k} (n-k) \left[F_X(x)\right]^{k} \left[1 - F_X(x)\right]^{n-k-1} f_X(x)$$

$$= \binom{n}{j} \cdot j \cdot \left[F_X(x)\right]^{j-1} \cdot f_X(x) \left[1 - F_X(x)\right]^{n-j}$$

$$+ \sum_{k=j}^{n-1} \binom{n}{k+1} (k+1) \left[F_X(x)\right]^{k} \cdot \left[1 - F_X(x)\right]^{n-k-1} \cdot f_X(x)$$

$$- \sum_{k=j}^{n-1} \binom{n}{k} (n-k) \left[F_X(x)\right]^{k} \cdot \left[1 - F_X(x)\right]^{n-k-1} \cdot f_X(x)$$

$$= \binom{n}{j} \cdot j \cdot \left[F_X(x)\right]^{j-1} \cdot f_X(x) \left[1 - F_X(x)\right]^{n-j}$$

$$+ \sum_{k=j}^{n} \underbrace{\left[\binom{n}{k+1}(k+1) - \binom{n}{k}(n-k)\right]}_{\frac{n!}{k!(n-k-1)!} - \frac{n!}{k!(n-k-1)!} = 0} \left[F_X(x)\right]^{k} \cdot \left[1 - F_X(x)\right]^{n-k-1} \cdot f_X(x)$$

$$\boxed{f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}}$$

Example 5.4.5 (Uniform order statistic pdf)

$X_1, X_2, \ldots X_n \sim \text{Unif}(0,1)$

$$\Rightarrow \quad f_X(x) = 1 \quad , \quad F_X(x) = x \quad \forall x \in (0,1)$$

$$\Rightarrow \quad f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} \, x^{j-1} \cdot (1-x)^{n-j}$$

$$\Rightarrow \quad f_{X_{(j)}}(x) = \frac{\Gamma(n+1)}{\Gamma(j)\,\Gamma(n-j+1)} \, x^{j-1} (1-x)^{(n-j+1)-1}$$

$$\Rightarrow \quad X_{(j)} \sim \text{Beta}(j, n-j+1)$$

$$\Rightarrow \quad \mathbb{E}\left[X_{(j)}\right] = \frac{j}{n+1} \qquad \text{Var}(X_{(j)}) = \frac{j(n-j+1)}{(n+1)^2 (n+2)}$$

Theorem 5.4.6 :

Let $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ denote the order statistic of a random sample, $X_1, X_2, \ldots X_n$ from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$

$1 \le i \le j \le n$ is

$$f_{X_{(i)}, X_{(j)}}(u, v) = \frac{n!}{(i-1)!\,(j-1-i)!\,(n-j)!} f_X(u)\, f_X(v) \left[F_X(u)\right]^{i-1}$$

$$\times \left[F_X(v) - F_X(u)\right]^{j-1-i} \left[1 - F_X(v)\right]^{n-j}$$

$$-\infty < u < v < \infty$$

The joint pdf of all order statistics :

$$f_{X_{(1)} X_{(2)} \ldots X_{(n)}}(x_1, x_2, \ldots x_n) = \begin{cases} n!\, f_X(x_1)\, f_X(x_2) \ldots f_X(x_n) \\ 0 & \text{o.w} \end{cases}$$

Example 5.4.7: Distribution of the midrange and range.

$X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{unif}(0,a)$ and let $X_{(1)}, \ldots X_{(n)}$ denote the order statistics. The Range $R = X_{(n)} - X_{(1)}$

The midrange, $V = \dfrac{X_{(1)} + X_{(n)}}{2}$

The Joint pdf of $X_{(1)}, X_{(n)}$

$$f_X(x) = \frac{1}{a} \qquad 0 < x < a$$

$$F_X(x) = \frac{x}{a} \qquad 0 < x < a$$

$$f_{X_{(1)}, X_{(n)}}(x_1, x_n) = \frac{n!}{(n-2)!} \left(\frac{1}{a}\right) \cdot \left(\frac{1}{a}\right) \left(\frac{x_1}{a}\right)^{1-1} \times \left[\frac{x_n}{a} - \frac{x_1}{a}\right]^{n-2} \left(1 - \frac{x_n}{a}\right)^{n-n}$$

$$f_{X_{(1)}, X_{(n)}}(x_1, x_n) = \frac{n(n-1)}{a^2} \left(\frac{x_n}{a} - \frac{x_1}{a}\right)^{n-2}$$

$$0 < x_1 < x_n < a$$

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

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

$$X_{(1)} = V - \frac{R}{2}, \quad X_{(n)} = V + \frac{R}{2} \quad \Bigg\} \quad J = \begin{vmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{vmatrix} = 1$$

$$0 < x_1 < x_n < a \Rightarrow \quad 0 < r < a$$

$$\frac{r}{2} < v < a - \frac{r}{2}$$

$$f_{R,V}(r,v) = \frac{n(n-1)}{a^2}\left(\frac{v}{a} + \frac{R}{2a} - \frac{v}{a} + \frac{R}{2a}\right)^{n-2}$$

$$\Rightarrow \quad f_{R,V}(r,v) = \frac{n(n-1)}{a^2} \cdot \left(\frac{r}{a}\right)^{n-2}$$

$$\Rightarrow \quad f_{R,V}(r,v) = \frac{n(n-1)}{a^n} r^{n-2}$$

$$0 < r < a$$

$$\frac{r}{2} < v < a - \frac{r}{2}$$

The marginal pdf of R

$$f_R(r) = \int_{\frac{r}{2}}^{a-\frac{r}{2}} \frac{n(n-1)}{a^n} r^{n-2} \, dv$$

$$\Rightarrow \quad f_R(r) = \frac{n(n-1)}{a^n}(a-r)\cdot r^{n-2} \qquad 0<r<a$$

<u>Marginal pdf of v</u>

$$f_V(v) = \int_R \frac{n(n-1)}{a^n} r^{n-2}\, dr$$

$$\Rightarrow \quad f_V(v) = \int_0^{2v} \frac{n(n-1)}{a^n} r^{n-2}\, dr \qquad 0<v\le a/2$$

$$= \int_0^{2(a-v)} \frac{n(n-1)}{a^n} r^{n-2}\, dr \qquad \frac{a}{2}<v\le a$$

$$\Rightarrow \quad f_V(v) = \frac{n(2v)^{n-1}}{a^n} \qquad 0<v\le \frac{a}{2}$$

$$= \frac{n\,[2(a-v)]^{n-1}}{a^n} \qquad \frac{a}{2}<v\le a$$