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 $X_1, X_2, \ldots, X_n$ are the sample values placed in acending order. They are denoted by $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$

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

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

$$X_{(2)} = \text{second smallest } X_i$$

$$\vdots$$

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

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 ?

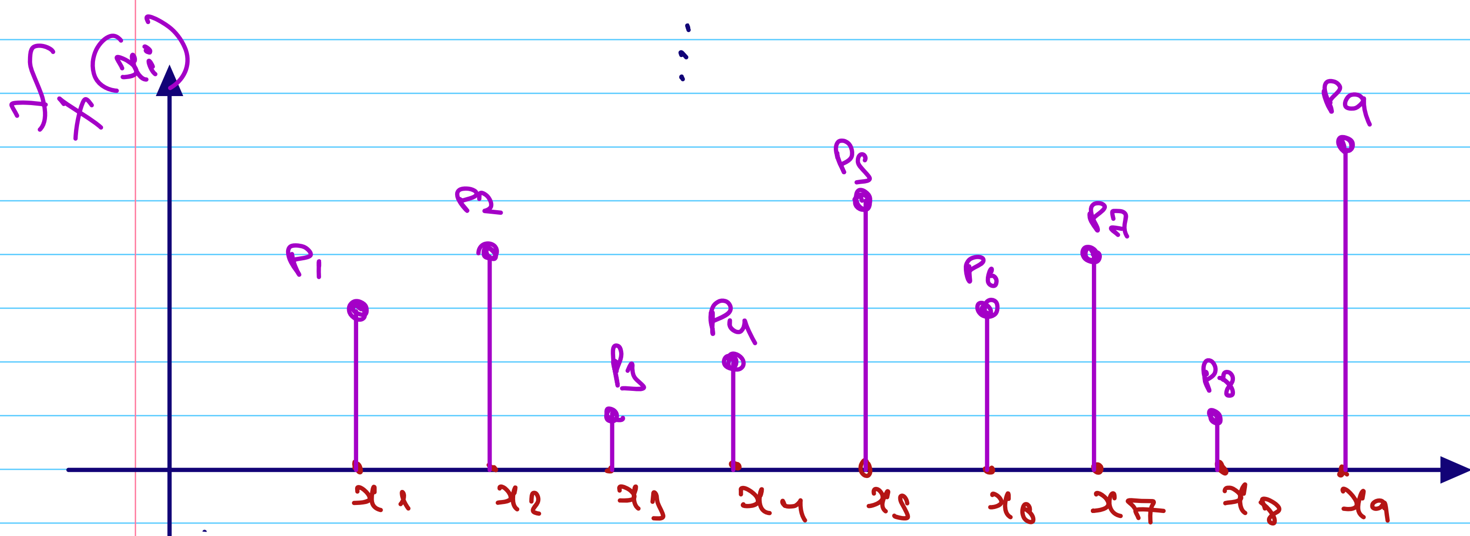
$\Rightarrow$ $12 \times (1 - 0.65) = 4.2$ and $12 + 1 - 4 = 9$

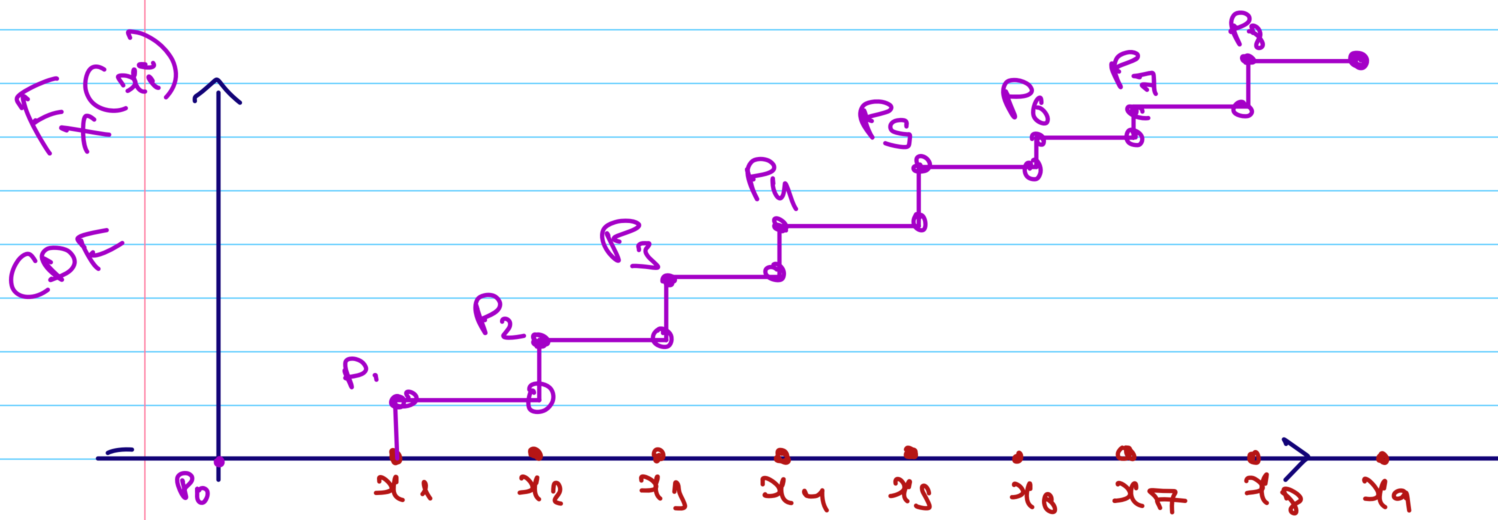
$\Rightarrow$ 65^{th} sample Percentile is $X_{(9)}$

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 $X_1, X_2, \ldots X_n \overset{iid}{\sim} X$.

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

(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

$x_1 \quad x_2 \quad x_3 \quad x_4 \mid x_5 \quad x_6 \quad x_7 \quad x_8$

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

the Probability that two X_i's same

$= 0$

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

Theorem 5.4.4:

Let $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots X_n$ 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 $X_1, \ldots X_n$ less than or equal to x.

② then $\{X_j \le x\}$.

$$\Rightarrow \mathbb{P}(X_j \le x) = \mathbb{P}(Y \ge j)$$

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

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

$$f_{X_{(j)}}(x) = \frac{d}{dx} F_{X_{(j)}}(x)$$

give the above formula.