

Order Statistics

GATE 2021 (ST), Q.17 (STATISTICS SECTION)

Suraj Telugu - CS20BTECH11050

IITH

May 29, 2021

Topics covered:

Prerequisites:

- Definition of order statistics
- Range, median and an example
- Theorem 1 and its proof (CDF of k^{th} order statistic)
- Theorem 2 and its proof (PDF of k^{th} order statistic)
- Uniform order statistics
- Introducing Beta function
- Beta distribution and its features

Gate Problem

- Problem
- Solution method 1
- Solution method 2

Introduction to Order statistics

Definition

- For a given statistical sample $\{X_1, X_2, \dots, X_n\}$, The order statistics is obtained by sorting the sample in ascending order
- The ordered sample values are denoted as $\{X_{(1)}, X_{(2)}, \dots, X_{(n)}\}$
- $X_{(1)}$ is the minimum and $X_{(n)}$ is the maximum of the given sample
- The k^{th} smallest value $X_{(k)}$ is called k^{th} order statistic

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)} \leq \dots \leq X_{(n)} \quad (1)$$

- For a sample $\{X_1, X_2, \dots, X_n\}$ of size n :

$$X_{(k)} = \min\{\max\{X_j : j \in J\} : J \subset \{1, 2, \dots, n\} \mid |J| = k\} \quad (2)$$

Introduction to Order statistics

Range and Median

- For a sample $\{X_1, X_2, \dots, X_n\}$, the range is the distance between the smallest and largest observations and is denoted by R

$$R = X_{(n)} - X_{(1)} \quad (3)$$

- Median is as the middle number of a sorted sample it is denoted by M it is defined using order statistics of a sample as

$$M = \begin{cases} X_{((n+1)/2)}, & \text{if } n \text{ is odd,} \\ \frac{X_{(n/2)} + X_{(n/2+1)}}{2}, & \text{if } n \text{ is even,} \end{cases} \quad (4)$$

Introduction to Order statistics

Example Problem

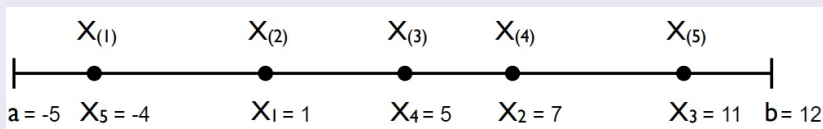
For the given data sample $\{1, 7, 11, 5, -4\}$ discuss the order statistics and find the value of 2nd order statistic, range and median

$$X_{(1)} = -4, X_{(2)} = 1, X_{(3)} = 5, X_{(4)} = 7, X_{(5)} = 11 \quad (5)$$

$$2^{\text{nd}} \text{ Order statistic} = 1 \quad (6)$$

$$\text{Range} = R = X_{(5)} - X_{(1)} = 15 \quad (7)$$

$$\text{Median} = M = X_{(3)} = 5 \quad (8)$$



CDF of k^{th} order statistic

Theorem (1)

Let $\{X_1, X_2, \dots, X_n\}$ be n i.i.d random variables with common CDF $= F(x)$ and common PDF $= f(x)$, then the marginal probability distribution of k^{th} order statistic (CDF) is denoted by $F_{(k,n)}(x)$ and it is given by

$$CDF = F_{(k,n)}(x) = \sum_{j=k}^n {}^nC_j \times (F(x))^j \times (1 - F(x))^{n-j} \quad (9)$$

Extreme cases

Equation (9) gives $X_{(1)}$ and $X_{(n)}$ as:

$$\text{Distribution of minimum} = F_{(1,n)}(x) = 1 - (1 - F(x))^n \quad (10)$$

$$\text{Distribution of maximum} = F_{(n,n)}(x) = (F(x))^n \quad (11)$$

Proof of theorem (1):

Theorem (1) Proof.

Assume after order statistics, the sample becomes $\{X_{(1)}, X_{(2)} \cdots X_{(n)}\}$ we need to find $F_{(k,n)}(x) = \Pr(X_{(k)} \leq x)$

$$X_{(i)} \leq X_{(k)} \leq x \quad \forall i \in \{1, \cdots k-1\} \quad (12)$$

$$X_{(i)} \leq x \text{ or } X_{(i)} > x \quad \forall i \in \{k+1, \cdots n\} \quad (13)$$

From the above equation (12), at least k elements in $\{X_1, X_2 \cdots X_n\}$ should be $\leq x$ and remaining $(n-k)$ elements can be $\leq x$ or $> x$

$$F_{(k,n)}(x) = \Pr(\text{At least } k \text{ elements have value } \leq x) \quad (14)$$

$$F_{(k,n)}(x) = \Pr(X_i \leq x : i \in I : I \subset \{1, \cdots n\} \mid I| \geq k) \quad (15)$$

Proof of theorem (1):

Theorem (1) Proof contd.

$$\begin{aligned} F_{(k,n)}(x) = & {}^nC_k \Pr(X \leq x)^k \Pr(X > x)^{n-k} + \\ & {}^nC_{k+1} \Pr(X \leq x)^{k+1} \Pr(X > x)^{n-(k+1)} + \\ & \dots\dots\dots {}^nC_n \Pr(X \leq x)^n \Pr(X > x)^0 \end{aligned} \quad (16)$$

$$F_{(k,n)}(x) = \sum_{j=k}^n {}^nC_j \Pr(X \leq x)^j \Pr(X > x)^{(n-j)} \quad (17)$$

$$\therefore F_{(k,n)}(x) = \sum_{j=k}^n {}^nC_j \times (F(x))^j \times (1 - F(x))^{n-j} \quad (18)$$



PDF of k^{th} order statistic

Theorem (2)

Let $\{X_1, X_2, \dots, X_n\}$ be n i.i.d random variables with common CDF $= F(x)$ and common PDF $= f(x)$, then the marginal probability density of k^{th} order statistic (PDF) is denoted by $f_{(k,n)}(x)$ and it is given by

$$PDF = f_{(k,n)}(x) = n {}^{n-1}C_{k-1} f(x) (F(x))^{k-1} (1 - F(x))^{n-k} \quad (19)$$

Extreme cases

Equation (19) gives $X_{(1)}$ and $X_{(n)}$ as:

$$\text{Density of minimum} = f_{(1,n)}(x) = n f(x) (1 - F(x))^{n-1} \quad (20)$$

$$\text{Density of maximum} = f_{(n,n)}(x) = n f(x) (F(x))^{n-1} \quad (21)$$

Proof of theorem (2) by differentiating CDF :

Theorem (2) Proof.

PDF of k^{th} order statistic by differentiating $F_{(k,n)}(x)$ w.r.t x

$$\frac{d}{dx} F_{(k,n)}(x) = \frac{d}{dx} \left(\sum_{j=k}^n {}^nC_j (1 - F(x))^{n-j} F(x)^j \right) \quad (22)$$

$$\begin{aligned} f_{(k,n)}(x) &= \sum_{j=k}^n {}^nC_j (j) (1 - F(x))^{n-j} F(x)^{j-1} f(x) \\ &\quad - \sum_{j=k}^n {}^nC_j (n - j) (1 - F(x))^{n-j-1} F(x)^j f(x) \quad (23) \end{aligned}$$

Let S_1, S_2 be summations such that $f_{(k,n)}(x) = S_1 - S_2$

Proof of theorem (2) by differentiating CDF :

Theorem (2) Proof contd.

$$S_1 = \sum_{j=k}^n \frac{n!}{(n-j)!(j-1)!} (1-F(x))^{n-j} F(x)^{j-1} f(x) \quad (24)$$

$$S_2 = \sum_{j=k}^n \frac{n!}{(n-j-1)!j!} (1-F(x))^{n-j-1} F(x)^j f(x) \quad (25)$$

let $i = j + 1$ change the limits for the summation in equation (25)

$$S_2 = \sum_{i=k+1}^n \frac{n!}{(n-i)!(i-1)!} (1-F(x))^{n-i} F(x)^{i-1} f(x) \quad (26)$$

$$S_1 - S_2 = \frac{n!}{(n-k)!(k-1)!} (1-F(x))^{n-k} F(x)^{k-1} f(x) \quad (27)$$

Proof of theorem (2) by differentiating CDF :

Theorem (2) Proof contd.

$$f_{(k,n)}(x) = S_1 - S_2 \quad (28)$$

$$f_{(k,n)}(x) = n^{n-1} C_{k-1} (1 - F(x))^{n-k} F(x)^{k-1} f(x) \quad (29)$$

$$\therefore \text{PDF of } k^{\text{th}} \text{ order statistic} = n^{n-1} C_{k-1} (1 - F(x))^{n-k} F(x)^{k-1} f(x)$$



Proof of theorem (2) in a simpler way:

Theorem (2) Proof.

let us assume $X_{(k)} \in [x, x + dx]$ for $dx \rightarrow 0$. Solving using 4 sub -jobs

$$\text{Job 1} = \text{Choose } X_i = X_{(k)} \text{ and } \Pr(X_i \in [x, x + dx]) = {}^nC_1 f(x) \quad (30)$$

$$\text{Job 2} = \text{Choose } k - 1 \text{ elements from remaining} = {}^{n-1}C_{k-1} \quad (31)$$

$$\text{Job 3} = \Pr(X \leq x) \text{ for chosen } (k - 1) \text{ elements} = F(x)^{k-1} \quad (32)$$

$$\text{Job 4} = \Pr(X > x) \text{ for } (n - k) \text{ elements} = (1 - F(x))^{n-k} \quad (33)$$

$$\Pr(X_{(k)} \in [x, x + dx]) = f_{(k,n)}(x) = \prod_{i=1}^4 \text{Job } i \quad (34)$$

$$\therefore f_{(k,n)}(x) = n {}^{n-1}C_{k-1} f(x) (F(x))^{k-1} (1 - F(x))^{n-k} \quad (35)$$



Uniform order statistics

Introduction

Let $\{X_1, \dots, X_n\}$ be i.i.d from a uniform distribution on $[0, 1]$ such that $f(x) = 1$ and $F(x) = x$, from theorem (2), equation (19)

$$f_{(k,n)}(x) = n^{n-1} C_{k-1} f(x) (F(x))^{k-1} (1 - F(x))^{n-k} \quad (36)$$

$$f_{(k,n)}(x) = n^{n-1} C_{k-1} x^{k-1} (1 - x)^{n-k} \quad (37)$$

Since equation (37) is marginal probability density (PDF)

$$\int_0^1 n^{n-1} C_{k-1} x^{k-1} (1 - x)^{n-k} dx = 1 \quad (38)$$

$$\int_0^1 x^{k-1} (1 - x)^{n-k} dx = \frac{1}{n^{n-1} C_{k-1}} \quad (39)$$

Introduction contd

$$\int_0^1 x^{k-1} (1-x)^{n-k} dx = \frac{(k-1)! (n-k)!}{n!} \quad (40)$$

$$\int_0^1 x^{k-1} (1-x)^{n-k} dx = \frac{\Gamma(k) \Gamma(n-k+1)}{\Gamma((n-k+1) + k)} \quad (41)$$

Definition

let $r = k$ and $s = n - k + 1$ The **Beta function** is defined for $r, s > 0$

$$B(r, s) = \int_0^1 x^{r-1} (1-x)^{s-1} dx = \frac{\Gamma(r) \Gamma(s)}{\Gamma(r+s)} \quad (42)$$

Beta function

Substituting in equation (36) and (37)

By using the above definition and equation

$$f_{(k,n)}(x) = \frac{1}{B(k, n-k+1)} f(x) (F(x))^{k-1} (1-F(x))^{n-k} \quad (43)$$

$$f_{(k,n)}(x) = \frac{x^{k-1} (1-x)^{n-k}}{B(k, n-k+1)} \text{ (Uniform order statistics)} \quad (44)$$

Beta distribution

The Beta distribution is a continuous distribution defined on the range $(0, 1)$ whose PDF given by

$$f(x) = \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1} \quad (45)$$

Beta Distribution

Beta distribution contd

Let $B_x(r, s) = \int_0^x x^{r-1} (1-x)^{s-1} dx$, CDF of Beta distribution:

$$F(x) = \int_0^x f(x) dx = \int_0^x \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1} dx \quad (46)$$

$$F(x) = \frac{\int_0^x x^{r-1} (1-x)^{s-1} dx}{B(r, s)} = \frac{B_x(r, s)}{B(r, s)} \quad (47)$$

(48)

$$\therefore \text{CDF} = F(x) = \frac{B_x(r, s)}{B(r, s)} \quad (49)$$

Beta distribution

Mean value of Beta distribution

$$E(x) = \int_0^1 x \times f(x) dx = \int_0^1 \frac{x}{B(r, s)} x^{r-1} (1-x)^{s-1} dx \quad (50)$$

$$= \frac{\int_0^1 x^{(r+1)-1} (1-x)^{s-1} dx}{B(r, s)} = \frac{B(r+1, s)}{B(r, s)} \quad (51)$$

$$= \frac{\Gamma(r+1) \Gamma(s)}{\Gamma(r+s+1)} \times \frac{\Gamma(r+s)}{\Gamma(r) \Gamma(s)} \quad (52)$$

$$= \frac{r!}{(r+s)!} \times \frac{(r+s-1)!}{(r-1)!} = \frac{r}{r+s} \quad (53)$$

$$\therefore \text{Mean value of } X (E(x)) = \frac{r}{r+s} \quad (54)$$

Beta distribution

Variance of Beta distribution

$$\text{Var}(x) = E(x^2) - (E(x))^2 \quad (55)$$

$$= \frac{\int_0^1 x^{(r+2)-1} (1-x)^{s-1} dx}{B(r, s)} - \left(\frac{r}{r+s} \right)^2 \quad (56)$$

$$= \frac{B(r+2, s)}{B(r, s)} - \frac{r^2}{(r+s)^2} \quad (57)$$

$$= \frac{\Gamma(r+2)\Gamma(s)}{\Gamma(r+s+2)} \times \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} - \frac{r^2}{(r+s)^2} \quad (58)$$

$$= \frac{(r+1)!}{(r+s+1)!} \times \frac{(r+s-1)!}{(r-1)!} - \frac{r^2}{(r+s)^2} \quad (59)$$

Variance of Beta distribution contd

$$= \frac{r(r+1)}{(r+s)(r+s+1)} - \frac{r^2}{(r+s)^2} \quad (60)$$

$$= \frac{r(r+1)(r+s) - r^2(r+s+1)}{(r+s)^2(r+s+1)} \quad (61)$$

$$= \frac{(r^3 + r^2s + r^2 + rs) - (r^3 + r^2s + r^2)}{(r+s)^2(r+s+1)} \quad (62)$$

$$\text{Var}(x) = \frac{rs}{(r+s)^2(r+s+1)} \quad (63)$$

$$\therefore \text{Variance of } X \text{ (Var}(x)) = \frac{rs}{(r+s)^2(r+s+1)} \quad (64)$$

Beta Distribution

Beta distribution summary

PDF, CDF, Expectation value and Variance of Beta distribution

$$f(x) = \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1} \quad (65)$$

$$F(x) = \frac{B_x(r, s)}{B(r, s)} \quad (66)$$

$$E(x) = \frac{r}{r+s} \quad (67)$$

$$Var(x) = \frac{rs}{(r+s)^2 (r+s+1)} \quad (68)$$

In Uniform order statistics on $[0,1]$ the PDF of k^{th} order statistic follows Beta distribution with $r = k$, $s = n - k + 1$ and PDF $f_{(k,n)}(x)$ from the above equation (44)

GATE 2021 (ST), Q.17 (STATISTICS SECTION)

If the marginal probability density function of the k^{th} order statistic of a random sample of size 8 from a uniform distribution on $[0, 2]$ is

$$f(x) = \begin{cases} \frac{7}{32} x^6 (2 - x), & 0 < x < 2, \\ 0, & \text{otherwise,} \end{cases}$$

then k equals _____

Solution (Using Theorem(2))

Method 1:

The problem involves the uniform order statistics on $[0, 2]$, sample size 8, we begin the solution by finding the PDF of the given sample:

$$\int_0^2 f(x) dx = 1 \quad (69)$$

$$f(x) = \frac{1}{2} \quad (70)$$

CDF of the given sample:

$$F(x) = \int_0^x f(x) dx = \frac{x}{2} \quad (71)$$

Solution (Using Theorem(2))

Using Theorem (2) PDF of k^{th} order statistic is given by

$$f_{(k,n)}(x) = n^{n-1} C_{k-1} f(x) (F(x))^{k-1} (1 - F(x))^{n-k} \quad (72)$$

$$f_{(k,n)}(x) = n^{n-1} C_{k-1} \frac{1}{2} \left(\frac{x}{2}\right)^{k-1} \left(1 - \frac{x}{2}\right)^{n-k} \quad (73)$$

$$f_{(k,8)}(x) = \frac{8}{2(1+(k-1)+(8-k))} \times {}^7C_{k-1} x^{k-1} (2-x)^{8-k} \quad (74)$$

$$f_{(k,8)}(x) = {}^7C_{k-1} \frac{1}{32} x^{k-1} (2-x)^{8-k} \quad (75)$$

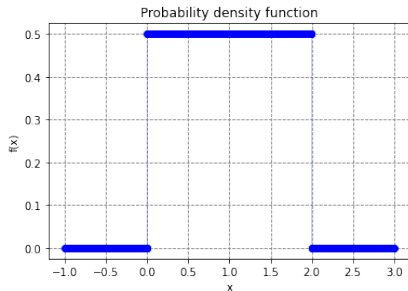
Comparing equation (75) with the given PDF of k^{th} order statistic

$$\frac{{}^7C_{k-1}}{32} x^{k-1} (2-x)^{n-k} = \frac{7}{32} x^6 (2-x) \quad (76)$$

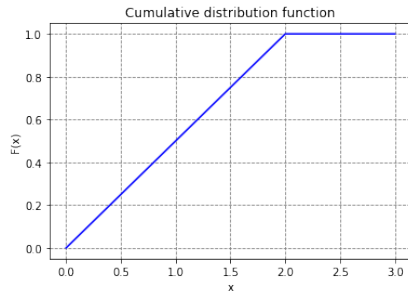
$$\therefore k = 7 \quad (77)$$

PDF and CDF of Uniform Distribution

The plots of PDF, CDF of uniform statistics in equations (70) and (71):



(a) PDF $f(x)$



(b) CDF $F(x)$

Figure: PDF and CDF of given uniform statistics

Solution(Using Beta distribution)

Method 2:

Since the k^{th} order statistic of a uniform distribution on $[0, 1]$ follows Beta distribution, convert the random variables, given PDF in $[0, 1]$ range

$$\int_0^2 f(x) dx = \int_0^2 \frac{7}{32} x^6 (2 - x) dx \quad (78)$$

$$\int_0^2 f(x) dx = \int_0^2 56 \left(\frac{x}{2}\right)^6 \left(1 - \frac{x}{2}\right) d\left(\frac{x}{2}\right) \quad (79)$$

$$\text{Let, } f_{(k,8)}(t) = 56 t^6 (1 - t) \quad (80)$$

Let new random variable be t such that $t = x/2$, New sample be $\{T_1, \dots, T_8\}$ such that $T_i = X_i/2$.

Solution(Using Beta distribution)

The Uniform distribution of new random sample is on $[0, 1]$ such that $f(t) = 1$ and $F(t) = t$

$$\int_0^2 f(x) dx = \int_0^1 56 t^6 (1 - t) dt = \int_0^1 f(t) dt \quad (81)$$

Given k^{th} order statistic (after conversion)

$$f_{(k,8)}(t) = \begin{cases} 56 t^6 (1 - t), & 0 < t < 1, \\ 0, & \text{otherwise,} \end{cases} \quad (82)$$

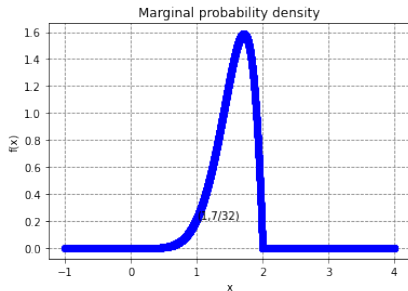
Since equation (82) is a Beta distribution with $r = k$, $s = n - k + 1$

$$k - 1 = 6 \quad (83)$$

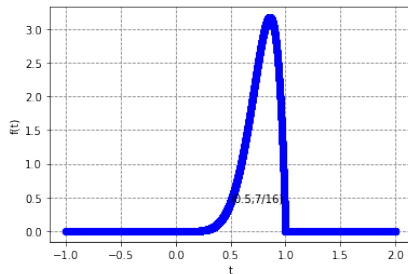
$$\therefore k = 7 \quad (84)$$

Plots of PDF of k^{th} order statistic

The plots for given PDF and PDF in equation (82) are shown below:



(a) PDF of $f_{(7,8)}(x)$



(b) PDF of $f_{(7,8)}(t)$

Figure: Plots of k^{th} order statistic

THANK YOU

Assignment link for reference:

[https://github.com/Suraj11050/Assignments-AI1103/tree/main/
Assignment4](https://github.com/Suraj11050/Assignments-AI1103/tree/main/Assignment4)