1 Order Statistics

Many natural summaries of a sample depend on how the observations are *ordered*. For example, the hottest and coldest days in a year are examples of order statistics. Throughout these notes, $X_1, \dots, X_n$ is a random sample from a population with CDF F, and PMF or PDF f.

1.1 Definitions

Definition 1 (Order Statistics)**.** The *order statistics* of a sample $X_1, \dots, X_n$ are the sample values placed in increasing order,

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

In particular, $X_{(1)} = \min_{i \le n} X_i$ is the sample minimum and $X_{(n)} = \max_{i \le n} X_i$ is the sample maximum.

Several common statistics are functions of the order statistics.

Definition 2 (Sample Median, Range, and Midrange)**.** The *sample range* is $R = X_{(n)} - X_{(1)}$ and the *sample midrange* is $V = \frac{X_{(1)} + X_{(n)}}{2}$. The *sample median* is the middle observation,

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

Example 1. Suppose your commute times (in minutes) over one week in November are

$$23, \quad 18, \quad 35, \quad 21, \quad 19, \quad 95, \quad 26,$$

where the 95 minute commute happened during the first snowstorm of the year. The order statistics are

$$x_{(1)} = 18, \quad x_{(2)} = 19, \quad x_{(3)} = 21, \quad x_{(4)} = 23, \quad x_{(5)} = 26, \quad x_{(6)} = 35, \quad x_{(7)} = 95,$$

so the sample median is $x_{(4)} = 23$, the range is $95 - 18 = 77$ and the midrange is $\frac{18+95}{2} = 56.5$. The sample mean is $\bar{x} = \frac{237}{7} \approx 33.9$, which is larger than all but one of your commutes because of the single snowstorm day. If the snowstorm commute had been 200 minutes instead, the sample mean would increase to 48.9, but the sample median would still be 23. We say that the median is *robust* to outliers.

Remark 1. Even though $X_1, \dots, X_n$ are i.i.d., the order statistics are neither independent nor identically distributed. For example, $X_{(1)} \le X_{(2)}$ always, so knowing $X_{(1)}$ gives information about $X_{(2)}$, and $X_{(n)}$ tends to be larger than $X_{(1)}$.

1.2 The Minimum and the Maximum

The simplest order statistics are the extremes. Their distributions follow directly from independence.

Proposition 1 (*Distribution of the Minimum and Maximum*)

For every $x \in \mathbb{R}$,

$$F_{X_{(n)}}(x) = F(x)^n \qquad \text{and} \qquad F_{X_{(1)}}(x) = 1 - (1 - F(x))^n.$$

If the population is continuous with PDF f, then $X_{(n)}$ and $X_{(1)}$ have PDFs

$$f_{X_{(n)}}(x) = n f(x) F(x)^{n-1} \qquad \text{and} \qquad f_{X_{(1)}}(x) = n f(x)(1 - F(x))^{n-1}.$$

Example 2. If $X_1, \dots, X_n$ are i.i.d. Exp(θ) with mean θ, then $1 - F(x) = e^{-x/\theta}$ for $x \ge 0$, so

$$\mathbb{P}(X_{(1)} > x) = (e^{-x/\theta})^n = e^{-x/(\theta/n)}.$$

Therefore, $X_{(1)} \sim \text{Exp}(\frac{\theta}{n})$.

1.3 The Distribution of a General Order Statistic

To find the distribution of $X_{(j)}$, we use a simple counting observation. The jth smallest observation is at most x if and only if at least j of the observations are at most x. The number of observations that are at most x is

$$N(x) = \sum_{i=1}^{n} \mathbb{1}(X_i \le x) \sim \text{Bin}(n, F(x)),$$

since the indicators are independent $\text{Ber}(F(x))$ random variables. This gives us the following result, which holds for every population.

Theorem 1 (*CDF of an Order Statistic*)

For every $j = 1, \dots, n$ and $x \in \mathbb{R}$,

$$\mathbb{P}(X_{(j)} \le x) = \mathbb{P}(N(x) \ge j) = \sum_{k=j}^{n} \binom{n}{k} F(x)^k \big(1 - F(x)\big)^{n-k}.$$

For discrete populations, we can recover the PMF from the CDF by taking differences.

Corollary 1 (*Discrete Order Statistics*)

Suppose the population takes values $x_1 < x_2 < \cdots$ (with $x_0 = -\infty$). Then

$$\begin{aligned}\mathbb{P}(X_{(j)} = x_i) &= \mathbb{P}(X_{(j)} \le x_i) - \mathbb{P}(X_{(j)} \le x_{i-1}) \\ &= \sum_{k=j}^{n} \binom{n}{k} \Big[F(x_i)^k (1 - F(x_i))^{n-k} - F(x_{i-1})^k (1 - F(x_{i-1}))^{n-k}\Big].\end{aligned}$$

Example 3. Roll 5 fair dice and let $M = X_{(3)}$ be the median roll. Since $F(3) = \frac{1}{2}$,

$$\mathbb{P}(M \le 3) = \mathbb{P}\left(\text{Bin}\big(5, \tfrac{1}{2}\big) \ge 3\right) = \frac{1}{2}$$

by symmetry. Since $F(2) = \frac{1}{3}$,

$$\mathbb{P}(M \le 2) = \mathbb{P}\left(\text{Bin}\big(5, \tfrac{1}{3}\big) \ge 3\right) = \binom{5}{3}\frac{2^2}{3^5} + \binom{5}{4}\frac{2}{3^5} + \frac{1}{3^5} = \frac{51}{243} \approx 0.210,$$

so $\mathbb{P}(M = 3) = \frac{1}{2} - \frac{51}{243} \approx 0.290$.

For continuous populations, we can differentiate the CDF, but it is more illuminating to compute the PDF directly from the joint distribution of all of the order statistics.

Theorem 2 (*Joint PDF of the Order Statistics*)

If the population is continuous with PDF f, then $(X_{(1)}, \dots, X_{(n)})$ has joint PDF

$$f_{X_{(1)},\dots,X_{(n)}}(y_1, \dots, y_n) = \begin{cases} n!\, f(y_1) f(y_2) \cdots f(y_n) & y_1 < y_2 < \cdots < y_n, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 2. The joint PDF of the sample $\prod_{i=1}^{n} f(x_i)$ is spread over all of $\mathbb{R}^n$. Sorting the sample folds the $n!$ possible orderings onto the single region $\{y_1 < \cdots < y_n\}$, which is where the factor $n!$ comes from.

To compute marginal distributions, we need to integrate the joint PDF over ordered regions. The following lemma makes this easy.

Lemma 1

Let f be a PDF with CDF F and let $-\infty \le a < b \le \infty$. For every positive integer k,

$$\int \cdots \int_{a<x_1<x_2<\cdots<x_k<b} f(x_1)f(x_2)\cdots f(x_k)\,dx_1\cdots dx_k = \frac{(F(b)-F(a))^k}{k!}.$$

Theorem 3 (*PDF of an Order Statistic*)

If the population is continuous with PDF f, then for $j = 1, \ldots, n$, $X_{(j)}$ has PDF

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

Furthermore, for $1 \le i < j \le n$, the joint PDF of $(X_{(i)}, X_{(j)})$ is

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(u)f(v)F(u)^{i-1}\big(F(v)-F(u)\big)^{j-i-1}\big(1-F(v)\big)^{n-j}$$

for $u < v$, and 0 otherwise.

Remark 3. There is a nice heuristic way to remember these formulas. For $X_{(j)}$ to be near x, we need $j-1$ observations below x, one observation near x, and $n-j$ observations above x. These three events have probabilities $F(x)$, $f(x)\,dx$, and $1-F(x)$ for each observation, and the number of ways to choose which observations fall in each group is the multinomial coefficient $\frac{n!}{(j-1)!\,1!\,(n-j)!}$. The formula for the joint PDF of $(X_{(i)}, X_{(j)})$ comes from the same reasoning with five groups.

Example 4. Let $U_1, \ldots, U_n$ be i.i.d. Unif$(0,1)$, so $f(x) = 1$ and $F(x) = x$ on $(0,1)$. By Theorem 3,

$$f_{U_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-j}, \qquad 0 < x < 1.$$

Since $\Gamma(j) = (j-1)!$, this is the Beta$(j, n-j+1)$ PDF, so

$$U_{(j)} \sim \text{Beta}(j, n-j+1) \qquad \text{and} \qquad \mathbb{E}[U_{(j)}] = \frac{j}{n+1}.$$

On average, the n order statistics split $[0,1]$ into $n+1$ pieces of equal length.

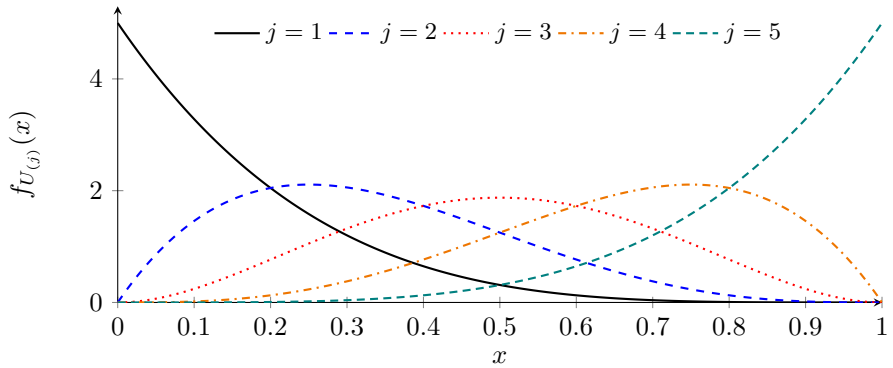


Figure: The PDFs of the order statistics $U_{(1)}, \ldots, U_{(5)}$ of a sample of size $n = 5$ from Unif$(0,1)$.

1.4 Order Statistics of the Exponential Distribution

The exponential distribution has a beautiful structure that makes its order statistics especially simple. Suppose that n light bulbs with i.i.d. $\text{Exp}(\theta)$ lifetimes are turned on at the same time. The first bulb burns out at time $X_{(1)} \sim \text{Exp}(\frac{\theta}{n})$. By the memoryless property, the remaining $n-1$ bulbs are as good as new at that time, so the additional waiting time until the next bulb burns out is the minimum of $n-1$ fresh exponentials, which is $\text{Exp}(\frac{\theta}{n-1})$, and so on. This heuristic is made precise by the following result.

Theorem 4 (*Exponential Spacings*)

Let $X_1, \dots, X_n$ be i.i.d. $\text{Exp}(\theta)$ and set $X_{(0)} = 0$. Then the spacings

$$D_k = X_{(k)} - X_{(k-1)}, \qquad k = 1, \dots, n,$$

are independent, and $D_k \sim \text{Exp}\big(\frac{\theta}{n-k+1}\big)$.

Corollary 2

Let $X_1, \dots, X_n$ be i.i.d. $\text{Exp}(\theta)$. Then

$$\mathbb{E}[X_{(j)}] = \theta \sum_{k=1}^{j} \frac{1}{n-k+1} \qquad \text{and} \qquad \text{Var}(X_{(j)}) = \theta^2 \sum_{k=1}^{j} \frac{1}{(n-k+1)^2}.$$

In particular, $\mathbb{E}[X_{(n)}] = \theta(1 + \frac{1}{2} + \cdots + \frac{1}{n})$ and $\text{Var}(X_{(n)}) = \theta^2(1 + \frac{1}{4} + \cdots + \frac{1}{n^2})$.

1.5 Example Problems

Problem 1.1. In the 17th century, the Chevalier de Méré bet that he would roll at least one six in four rolls of a fair die. Let $X_{(4)}$ denote the largest of four independent rolls.

1. Find the PMF of $X_{(4)}$.
2. What is the probability that de Méré wins his bet?
3. Compute $\mathbb{E}[X_{(4)}]$.

Solution 1.1. **Part 1:** By Proposition 1, $\mathbb{P}(X_{(4)} \le k) = F(k)^4 = (\frac{k}{6})^4$ for $k = 0, 1, \dots, 6$. Taking differences,

$$\mathbb{P}(X_{(4)} = k) = \left(\frac{k}{6}\right)^4 - \left(\frac{k-1}{6}\right)^4 = \frac{k^4 - (k-1)^4}{1296}, \qquad k = 1, \dots, 6,$$

which gives us the table

k	1	2	3	4	5	6
$\mathbb{P}(X_{(4)} = k)$	$\frac{1}{1296}$	$\frac{15}{1296}$	$\frac{65}{1296}$	$\frac{175}{1296}$	$\frac{369}{1296}$	$\frac{671}{1296}$

Part 2: De Méré wins if the largest roll is a six, so

$$\mathbb{P}(X_{(4)} = 6) = 1 - \left(\frac{5}{6}\right)^4 = \frac{671}{1296} \approx 0.518.$$

The bet is slightly in his favor.

Part 3: Since $X_{(4)}$ is non-negative and integer valued, we can use the tail sum formula $\mathbb{E}[X_{(4)}] = \sum_{k=0}^{5} \mathbb{P}(X_{(4)} > k)$,

$$\mathbb{E}[X_{(4)}] = \sum_{k=0}^{5} \left(1 - \left(\frac{k}{6}\right)^4\right) = 6 - \frac{0 + 1 + 16 + 81 + 256 + 625}{1296} = 6 - \frac{979}{1296} \approx 5.245.$$

Problem 1.2. A server stores data on 4 hard drives whose lifetimes are i.i.d. Exp(6) (in years).

1. In a *series* configuration, the server fails as soon as any drive fails. Find the expected lifetime of the server and the probability that it lasts more than 1 year.
2. In a fully redundant *parallel* configuration, the server only fails when every drive has failed. Find the expected lifetime of the server and the probability that it lasts more than 6 years.

Solution 1.2. **Part 1:** The lifetime of the server is $X_{(1)} \sim \text{Exp}(\frac{6}{4})$ by Proposition 1. Therefore, $\mathbb{E}[X_{(1)}] = 1.5$ years and

$$\mathbb{P}(X_{(1)} > 1) = e^{-1/1.5} = e^{-2/3} \approx 0.513.$$

Part 2: The lifetime of the server is $X_{(4)}$. By Corollary 2,

$$\mathbb{E}[X_{(4)}] = 6\Big(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\Big) = 6 \cdot \frac{25}{12} = 12.5$$

years, and by Proposition 1,

$$\mathbb{P}(X_{(4)} > 6) = 1 - F(6)^4 = 1 - (1 - e^{-1})^4 \approx 0.840.$$

Redundancy increases the expected lifetime from 1.5 to 12.5 years, even though each drive only lasts 6 years on average.

Problem 1.3. Five friends each roll a fair die at the same time, once per round, until someone rolls a six. Let Y denote the number of rounds without a six before the first six appears.

1. Let $X_1, \ldots, X_n$ be i.i.d. Geo(p), where Geo(p) counts the number of failures before the first success. Show that $X_{(1)} \sim \text{Geo}(1 - (1-p)^n)$.
2. Find the distribution of Y and compute $\mathbb{E}[Y]$.

Solution 1.3. **Part 1:** If $X \sim \text{Geo}(p)$, then $X \geq x$ if and only if the first x trials are failures, so $\mathbb{P}(X \geq x) = (1-p)^x$ for $x = 0, 1, 2, \ldots$. The minimum is at least x if and only if every observation is at least x, so by independence,

$$\mathbb{P}(X_{(1)} \geq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \geq x) = (1-p)^{nx} = \big((1-p)^n\big)^x.$$

This is the tail probability of a geometric random variable with success probability $q = 1 - (1-p)^n$, so $X_{(1)} \sim \text{Geo}(1 - (1-p)^n)$.

Part 2: Let X_i be the number of rolls without a six for friend i if they kept rolling forever, so $X_i \sim \text{Geo}(\frac{1}{6})$ are independent and $Y = X_{(1)}$. By Part 1, $Y \sim \text{Geo}(q)$ with $q = 1 - (\frac{5}{6})^5 \approx 0.598$, which is the probability that at least one of the five friends rolls a six in a given round. Therefore,

$$\mathbb{E}[Y] = \frac{1-q}{q} = \frac{(5/6)^5}{1 - (5/6)^5} \approx 0.672.$$

Remark 4. This is the discrete analogue of the fact that the minimum of independent exponentials is exponential. Both results come from the memoryless property, and the formula for the tail of the minimum in Proposition 1.

Problem 1.4. Let $X_1, \dots, X_n$ be a random sample from $\text{Unif}(0, \theta)$ where $\theta > 0$ is unknown. We want to estimate θ, the largest possible value.

1. Show that $T_1 = 2\bar{X}$ satisfies $\mathbb{E}[T_1] = \theta$, and compute $\text{Var}(T_1)$.
2. Show that $T_2 = \frac{n+1}{n} X_{(n)}$ satisfies $\mathbb{E}[T_2] = \theta$, and compute $\text{Var}(T_2)$.
3. Which estimator would you prefer?

Solution 1.4.

Part 1: Since $\mathbb{E}[X_i] = \frac{\theta}{2}$ and $\text{Var}(X_i) = \frac{\theta^2}{12}$, the properties of the expected value and variance implies that

$$\mathbb{E}[T_1] = 2 \cdot \frac{\theta}{2} = \theta \qquad \text{and} \qquad \text{Var}(T_1) = 4 \cdot \frac{\theta^2}{12n} = \frac{\theta^2}{3n}.$$

Part 2: The random variables $U_i = \frac{X_i}{\theta}$ are i.i.d. $\text{Unif}(0,1)$, and $X_{(n)} = \theta U_{(n)}$ with $U_{(n)} \sim \text{Beta}(n, 1)$. The PDF of $U_{(n)}$ is nx^{n-1} on $(0,1)$, so

$$\mathbb{E}[U_{(n)}] = \int_0^1 x \cdot nx^{n-1}\, dx = \frac{n}{n+1} \qquad \text{and} \qquad \mathbb{E}[U_{(n)}^2] = \int_0^1 x^2 \cdot nx^{n-1}\, dx = \frac{n}{n+2}.$$

Therefore, $\text{Var}(U_{(n)}) = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)}$, and

$$\mathbb{E}[T_2] = \frac{n+1}{n} \cdot \theta \cdot \frac{n}{n+1} = \theta \qquad \text{and} \qquad \text{Var}(T_2) = \frac{(n+1)^2}{n^2} \cdot \theta^2 \cdot \frac{n}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}.$$

Part 3: Both estimators are correct on average, but

$$\frac{\text{Var}(T_1)}{\text{Var}(T_2)} = \frac{\theta^2/(3n)}{\theta^2/(n(n+2))} = \frac{n+2}{3},$$

so T_2 has a much smaller variance when n is large. For example, with $n = 10$ observations, the variance of T_1 is 4 times larger. The variance of T_2 decreases like $\frac{1}{n^2}$ instead of the usual $\frac{1}{n}$, because the maximum is very close to θ when n is large.

Remark 5. The discrete version of this problem is known as the *German tank problem*. During World War II, the Allies estimated German tank production from the serial numbers of captured tanks using an estimator based on the largest observed serial number, and the estimates turned out to be much more accurate than intelligence reports.

Problem 1.5. Let $U_1, \dots, U_n$ be i.i.d. $\text{Unif}(0,1)$ with $n \geq 2$. Show that the sample range $R = U_{(n)} - U_{(1)}$ has a $\text{Beta}(n-1, 2)$ distribution, and compute $\mathbb{E}[R]$.

Solution 1.5. By Theorem 3 with $i = 1$ and $j = n$, the joint PDF of $(U_{(1)}, U_{(n)})$ is

$$f_{U_{(1)},U_{(n)}}(u, v) = \frac{n!}{0!\,(n-2)!\,0!}(v-u)^{n-2} = n(n-1)(v-u)^{n-2}, \qquad 0 < u < v < 1.$$

Consider the transformation $(R, U_{(1)})$ with inverse $u = u$ and $v = r + u$, whose Jacobian is 1. The constraints $0 < u < v < 1$ become $0 < r < 1$ and $0 < u < 1 - r$, so

$$f_{R,U_{(1)}}(r, u) = n(n-1)r^{n-2}, \qquad 0 < r < 1, \quad 0 < u < 1 - r.$$

Integrating out u,

$$f_R(r) = \int_0^{1-r} n(n-1)r^{n-2}\,du = n(n-1)r^{n-2}(1-r), \qquad 0 < r < 1.$$

Since $\frac{\Gamma(n+1)}{\Gamma(n-1)\Gamma(2)} = n(n-1)$, this is the Beta$(n-1, 2)$ PDF. Therefore,

$$\mathbb{E}[R] = \frac{n-1}{(n-1)+2} = \frac{n-1}{n+1},$$

which agrees with $\mathbb{E}[U_{(n)}] - \mathbb{E}[U_{(1)}] = \frac{n}{n+1} - \frac{1}{n+1}$.

Problem 1.6. Let U_1, U_2 be i.i.d. Unif$(0, 1)$. Show that $\mathrm{Cov}(U_{(1)}, U_{(2)}) = \frac{1}{36}$.

Solution 1.6. The trick is that the product of the order statistics is the same as the product of the original sample, since sorting does not change the product: $U_{(1)}U_{(2)} = U_1U_2$. Therefore, by independence,

$$\mathbb{E}[U_{(1)}U_{(2)}] = \mathbb{E}[U_1U_2] = \mathbb{E}[U_1]\,\mathbb{E}[U_2] = \frac{1}{4}.$$

By the example in Section 1.3, $\mathbb{E}[U_{(1)}] = \frac{1}{3}$ and $\mathbb{E}[U_{(2)}] = \frac{2}{3}$, so

$$\mathrm{Cov}(U_{(1)}, U_{(2)}) = \frac{1}{4} - \frac{1}{3}\cdot\frac{2}{3} = \frac{9-8}{36} = \frac{1}{36}.$$

The covariance is positive: a large minimum forces the maximum to be large as well. In particular, $U_{(1)}$ and $U_{(2)}$ are not independent.

Remark 6. The same trick works for any symmetric function of the sample. For example, $\sum_i X_{(i)} = \sum_i X_i$, so $\sum_{i=1}^n \mathbb{E}[X_{(i)}] = n\,\mathbb{E}[X_1]$.

1.6 Proofs of Key Results

Problem 1.7. Prove Proposition 1.

Solution 1.7. The maximum is at most x if and only if every observation is at most x, so by independence,

$$F_{X_{(n)}}(x) = \mathbb{P}(X_1 \le x, \ldots, X_n \le x) = \prod_{i=1}^n \mathbb{P}(X_i \le x) = F(x)^n.$$

Similarly, the minimum is larger than x if and only if every observation is larger than x, so

$$1 - F_{X_{(1)}}(x) = \mathbb{P}(X_1 > x, \ldots, X_n > x) = \prod_{i=1}^n \mathbb{P}(X_i > x) = (1 - F(x))^n.$$

If the population is continuous, differentiating with the chain rule gives the PDFs.

Problem 1.8. Prove Theorem 1 and Corollary 1.

Solution 1.8.

Proof of Theorem 1: If at least j of the observations are at most x, then the jth smallest observation is at most x. Conversely, if $X_{(j)} \le x$, then $X_{(1)}, \dots, X_{(j)}$ are all at most x, so at least j of the observations are at most x. Therefore, $\{X_{(j)} \le x\} = \{N(x) \ge j\}$. The indicators $\mathbb{1}(X_i \le x)$ are independent (since they are functions of independent random variables) and each is $\text{Ber}(F(x))$, so $N(x) \sim \text{Bin}(n, F(x))$ and

$$\mathbb{P}(X_{(j)} \le x) = \mathbb{P}(N(x) \ge j) = \sum_{k=j}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}.$$

Proof of Corollary 1: If the population only takes the values $x_1 < x_2 < \cdots$, then so does $X_{(j)}$, so $\{X_{(j)} \le x_i\}$ is the disjoint union of $\{X_{(j)} \le x_{i-1}\}$ and $\{X_{(j)} = x_i\}$. Therefore, applying Theorem 1, we have that

$$\begin{aligned}\mathbb{P}(X_{(j)} = x_i) &= \mathbb{P}(X_{(j)} \le x_i) - \mathbb{P}(X_{(j)} \le x_{i-1}) \\ &= \sum_{k=j}^{n} \binom{n}{k} \Big[F(x_i)^k (1 - F(x_i))^{n-k} - F(x_{i-1})^k (1 - F(x_{i-1}))^{n-k}\Big].\end{aligned}$$

Problem 1.9. Prove Theorem 2.

Solution 1.9. Let $Y = (X_{(1)}, \dots, X_{(n)})$ and let $O = \{y \in \mathbb{R}^n : y_1 < y_2 < \cdots < y_n\}$. Since the population is continuous, $\mathbb{P}(X_i = X_j) = 0$ for $i \ne j$, so with probability one the observations are distinct and $Y \in O$. In this case, there is exactly one permutation σ of $\{1, \dots, n\}$ such that

$$X_{\sigma(1)} < X_{\sigma(2)} < \cdots < X_{\sigma(n)},$$

and then $Y = (X_{\sigma(1)}, \dots, X_{\sigma(n)})$. Let $A \subseteq O$. The event $\{Y \in A\}$ is the disjoint union over all $n!$ permutations σ of the events $\{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in A\}$, since a permuted sample can only land in $A \subseteq O$ if σ is the permutation that sorts the sample. Since the X_i are i.i.d., every permuted vector $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ has the same joint PDF $\prod_{i=1}^n f(y_i)$ as $(X_1, \dots, X_n)$. Therefore,

$$\mathbb{P}(Y \in A) = \sum_{\sigma} \mathbb{P}\left((X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in A\right) = \sum_{\sigma} \int_A \prod_{i=1}^n f(y_i)\, dy = \int_A n! \prod_{i=1}^n f(y_i)\, dy.$$

Since this holds for all $A \subseteq O$ and $\mathbb{P}(Y \notin O) = 0$, the joint PDF of Y is $n! \prod_{i=1}^n f(y_i)$ on O and 0 elsewhere.

Problem 1.10. Prove Lemma 1.

Solution 1.10. The cube $(a, b)^k$ is the union of the $k!$ regions $\{a < x_{\sigma(1)} < \cdots < x_{\sigma(k)} < b\}$ over all permutations σ. We can assume at the inequalities are strict because the probability that two coordinates are equal is zero. The integrand $\prod_{l=1}^k f(x_l)$ is symmetric in $x_1, \dots, x_k$, so relabeling the variables shows that its integral over each of the $k!$ regions is the same. Therefore, by Fubini's theorem,

$$k! \int \cdots \int_{a < x_1 < \cdots < x_k < b} \prod_{l=1}^k f(x_l)\, dx = \int_{(a,b)^k} \prod_{l=1}^k f(x_l)\, dx = \prod_{l=1}^k \int_a^b f(x_l)\, dx_l = (F(b) - F(a))^k,$$

and dividing by $k!$ proves the lemma.

Problem 1.11. Prove Theorem 3.

Solution 1.11.

PDF of $X_{(j)}$: We integrate out all of the variables except $y_j = x$ in the joint PDF from Theorem 2. For fixed x, the region $\{y_1 < \cdots < y_{j-1} < x < y_{j+1} < \cdots < y_n\}$ is a product of the region $\{y_1 < \cdots < y_{j-1} < x\}$ for the variables below x and the region $\{x < y_{j+1} < \cdots < y_n\}$ for the variables above x. By Fubini's theorem and Lemma 1 (with $(a, b) = (-\infty, x)$ and $(a, b) = (x, \infty)$ respectively),

$$\begin{aligned} f_{X_{(j)}}(x) &= n! f(x) \left(\int_{y_1 < \cdots < y_{j-1} < x} \prod_{l=1}^{j-1} f(y_l)\, dy \right) \left(\int_{x < y_{j+1} < \cdots < y_n} \prod_{l=j+1}^{n} f(y_l)\, dy \right) \\ &= n! f(x) \frac{F(x)^{j-1}}{(j-1)!} \frac{(1 - F(x))^{n-j}}{(n-j)!}. \end{aligned}$$

Joint PDF of $(X_{(i)}, X_{(j)})$: The same argument works with two fixed variables $y_i = u$ and $y_j = v$ for $u < v$. The remaining variables split into three groups: $i-1$ variables in $(-\infty, u)$, $j-i-1$ variables in (u, v), and $n-j$ variables in (v, ∞), each ordered within its group. Applying Lemma 1 to each group,

$$f_{X_{(i)}, X_{(j)}}(u, v) = n! f(u) f(v) \frac{F(u)^{i-1}}{(i-1)!} \frac{(F(v) - F(u))^{j-i-1}}{(j-i-1)!} \frac{(1 - F(v))^{n-j}}{(n-j)!}.$$

Remark 7. Alternatively, the PDF of $X_{(j)}$ can be found by differentiating the CDF in Theorem 1. The derivative of the sum telescopes, and only one term survives. This is the approach taken in the textbook Casella–Berger.

Problem 1.12. Prove Theorem 4 and Corollary 2.

Solution 1.12.

Proof of Theorem 4: By Theorem 2, the joint PDF of $Y = (X_{(1)}, \ldots, X_{(n)})$ is

$$f_Y(y) = n! \prod_{i=1}^{n} \frac{1}{\theta} e^{-y_i/\theta} = \frac{n!}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^{n} y_i}, \qquad 0 < y_1 < \cdots < y_n.$$

The spacings are a linear transformation of Y with inverse $y_k = d_1 + d_2 + \cdots + d_k$. The matrix of this inverse transformation is lower triangular with ones on the diagonal, so its Jacobian is 1, and the region $0 < y_1 < \cdots < y_n$ corresponds to $d_1, \ldots, d_n > 0$. Each spacing d_k appears in $y_k, y_{k+1}, \ldots, y_n$, so

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \sum_{k=1}^{i} d_k = \sum_{k=1}^{n} (n - k + 1) d_k.$$

Since $n! = \prod_{k=1}^{n} (n - k + 1)$, the joint PDF of $D = (D_1, \ldots, D_n)$ is

$$f_D(d) = \frac{n!}{\theta^n} \prod_{k=1}^{n} e^{-\frac{(n-k+1)d_k}{\theta}} = \prod_{k=1}^{n} \frac{n-k+1}{\theta} e^{-\frac{(n-k+1)d_k}{\theta}}, \qquad d_1, \ldots, d_n > 0.$$

This is a product of $\mathrm{Exp}(\frac{\theta}{n-k+1})$ PDFs, so by the factorization theorem (Week 5, Proposition 1), the spacings are independent with $D_k \sim \mathrm{Exp}(\frac{\theta}{n-k+1})$.

Proof of Corollary 2: Since $X_{(j)} = D_1 + \cdots + D_j$, linearity of expectation and the independence of the spacings imply that

$$\mathbb{E}[X_{(j)}] = \sum_{k=1}^{j} \frac{\theta}{n-k+1} \qquad \text{and} \qquad \operatorname{Var}(X_{(j)}) = \sum_{k=1}^{j} \frac{\theta^2}{(n-k+1)^2},$$

using that an $\operatorname{Exp}(\beta)$ random variable has mean β and variance β^2. Taking $j = n$ and reindexing with $i = n - k + 1$ gives the formulas for the maximum.