Midterm Project 677

Ruicheng Zhang

2024-03-25

Order statistics

Uniform Distribution

1. **Scenario**: Drawing n independent samples from a uniform distribution $U(a, b)$ to find the distribution of the kth order statistic, denoted as $X_{(k)}$.
2. The PDF is defined as $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$.
3. **Distribution of Order Statistics**: The PDF of the kth order statistic is given by:

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} \left(\frac{x-a}{b-a}\right)^{k-1} \left(1 - \frac{x-a}{b-a}\right)^{n-k} \frac{1}{b-a}$$

Minimum

The distribution of the minimum value $X_{(1)}$ from a uniform distribution $U(a, b)$ is:

$$F_{X_{(1)}}(x) = 1 - (1 - \frac{x-a}{b-a})^n, \quad a \leq x \leq b$$

Maximum

The distribution of the maximum value $X_{(n)}$ is:

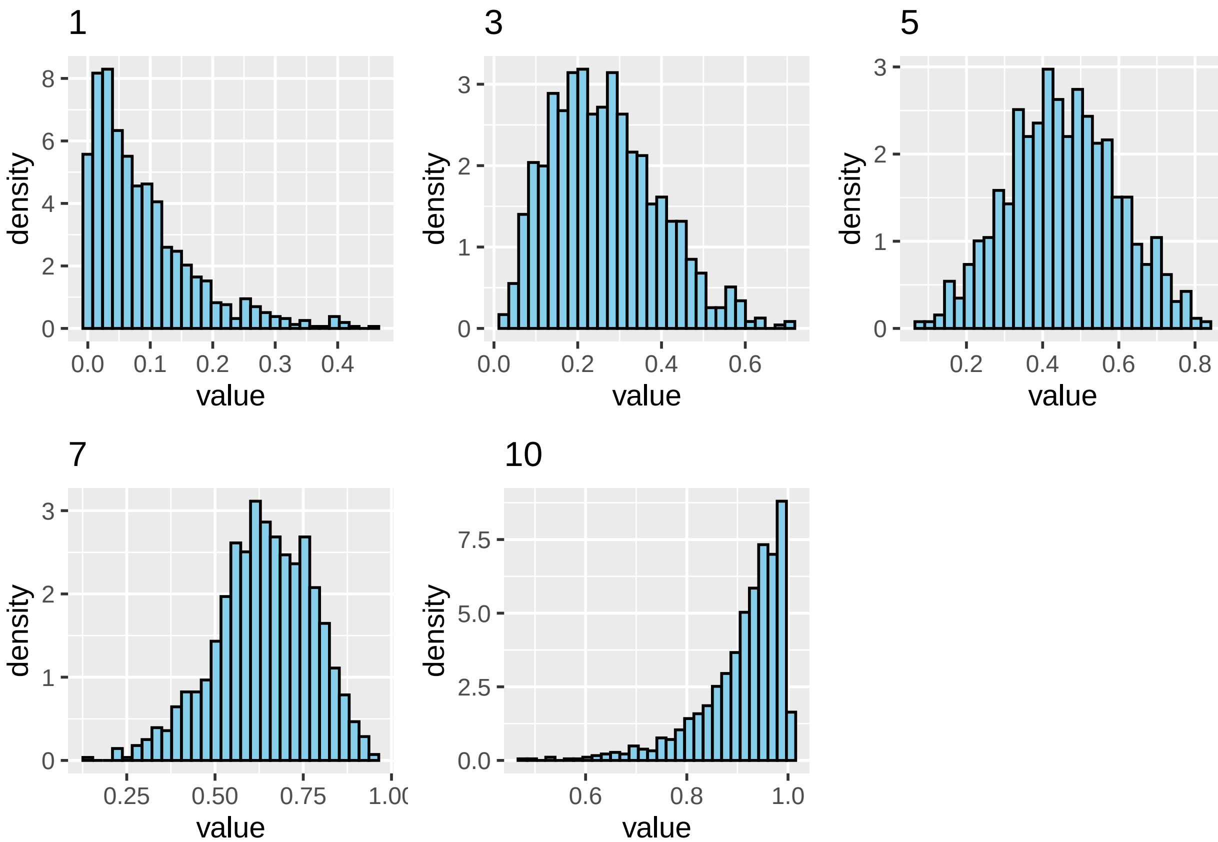
$$F_{X_{(n)}}(x) = (\frac{x-a}{b-a})^n, \quad a \leq x \leq b$$

Median, Quartiles, Percentiles

- Median: $\frac{a+b}{2}$
- First Quartile: $a + \frac{b-a}{4}$
- Third Quartile: $a + 3\frac{b-a}{4}$

Simulations

Generated 10,000 samples from a uniform distribution U(0,1), each with a sample size of 10, then calculated the values of the 1st, 3rd, 5th, 7th, and 10th order statistics, and plotted the distribution for each order statistic.



Exponential Distribution

1. **Scenario**: Drawing n independent samples from an exponential distribution with rate parameter λ.

2.The PDF is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

3. **Distribution of Order Statistics**: The PDF of the kth order statistic is:

$$f_{X_{(k)}}(x) = \frac{\lambda(\lambda x)^{k-1} e^{-\lambda x}}{(k-1)!} \left(\sum_{j=0}^{n-k} \frac{(\lambda x)^j}{j!} \right)$$

Minimum

The minimum $X_{(1)}$ follows an exponential distribution, and its probability density function (PDF) is:

$$f_{X_{(1)}}(x) = n\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{n-1}$$

Maximum

The distribution of the maximum is more complex, often approached via numerical methods.

Median, Quartiles, Percentiles

- The general formula for percentiles $T(q)$ is:

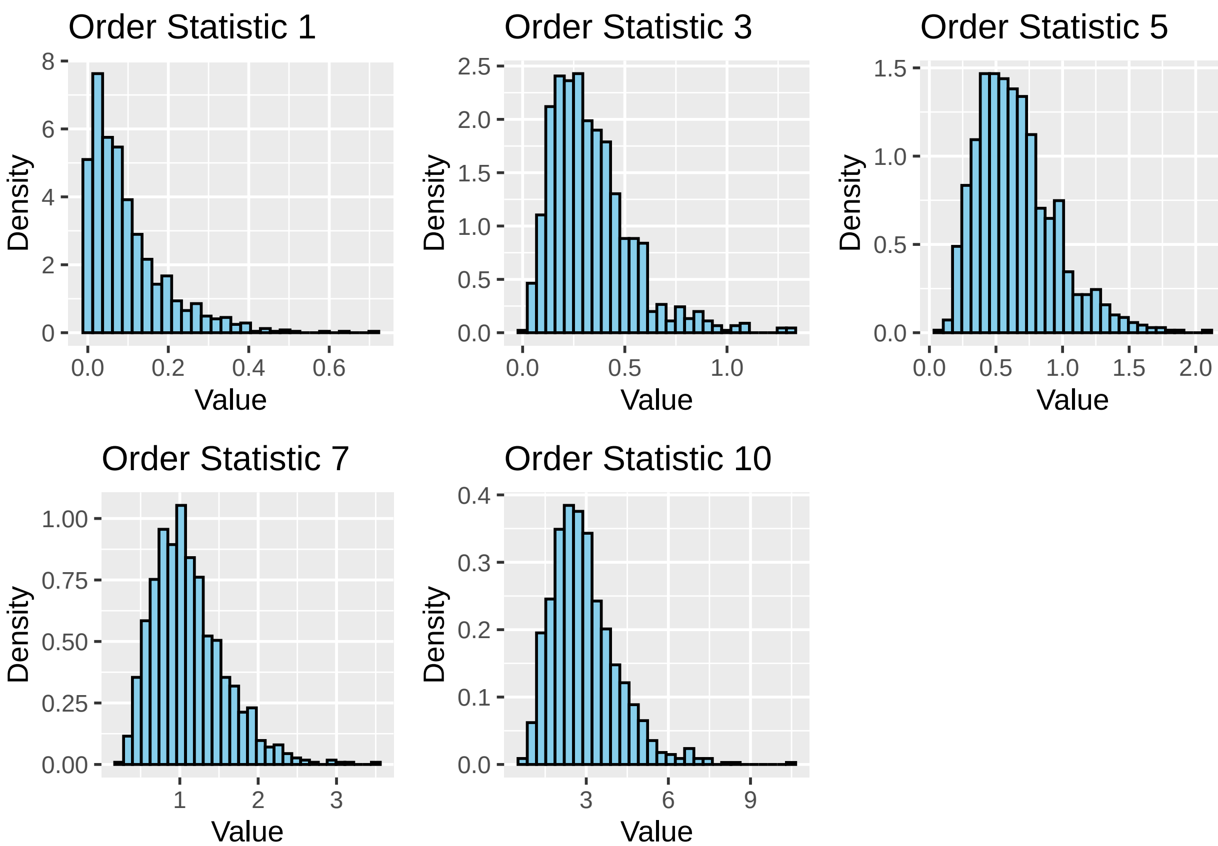
$$Q(q) = -\frac{\log(1-q)}{\lambda}$$

- The median $Q(0.5)$ is:

$$Q(0.5) = -\frac{\log(0.5)}{\lambda}$$

Simulations

Generated 10,000 samples from an exponential distribution with rate parameter $\lambda = 1$, each with a sample size of 10, then calculated the values of the 1st, 3rd, 5th, 7th, and 10th order statistics, and plotted the distribution for each order statistic.



Normal Distribution

1. **Scenario**: Drawing n independent samples from a normal distribution $N(\mu, \sigma^2)$.
2. **Complexity of Normal Order Statistics**: The distribution of order statistics for a normal distribution does not have a simple closed form and is typically evaluated through numerical methods.
3. **Approach for Normal Order Statistics**: Due to the lack of a simple closed-form solution, the focus is usually on using software or numerical methods to estimate the properties of order statistics for the normal distribution.

Minimum and Maximum

No simple closed forms exist for the minimum or maximum; these are approached through numerical methods or simulation.

Median, Quartiles, Percentiles

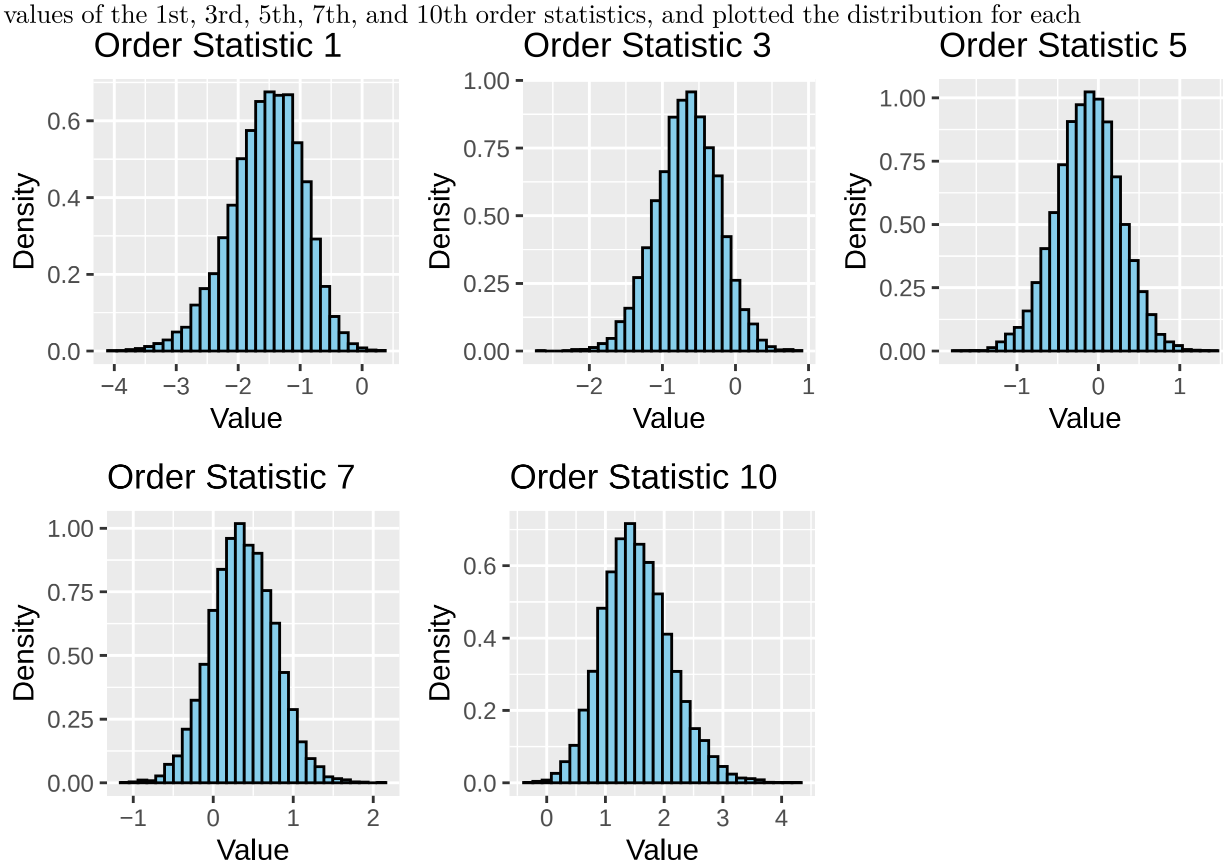
- The distribution of the median and quartiles is centered around the mean μ.
- The formula for percentiles, derived from the cumulative distribution function (CDF), is given by:

$$Q(q) = \mu + \sigma Z_q$$

  where Z_q is the z-score corresponding to the qth percentile.
- The quartiles $Q(0.25)$ and $Q(0.75)$, as well as any percentile Q, can be calculated by finding the z-score Z_q and applying the normal distribution formula.

Simulations

Generated 10,000 samples from a standard normal distribution $N(0, 1)$, each with a sample size of 10, then calculated the values of the 1st, 3rd, 5th, 7th, and 10th order statistics, and plotted the distribution for each



order statistic.

Conclusion

For the Uniform Distribution U(0,1): The order statistics from a uniform distribution exhibit a nearly linear increase from the minimum to the maximum. The distribution of the k-th order statistic is evenly spread between 0 and 1, with early order statistics (e.g., minimum values) concentrated towards the lower end, and later order statistics (e.g., maximum values) concentrated towards the upper end.The distributions of order statistics are symmetric around the midpoint (e.g., the 5th order statistic in a sample of 10), reflecting the uniform distribution's symmetry. For the Exponential Distribution : The exponential distribution is skewed to the right, and this property influences its order statistics. Early order statistics tend to be closely clustered near zero, reflecting the exponential distribution's high density near zero. For the Normal Distribution:The normal distribution is symmetric and bell-shaped, and its order statistics reflect this property. Middle order statistics (e.g., the median) from samples of a normal distribution tend to be normally distributed around the population mean, with reduced variance.

The distribution of order statistics from uniform samples is quite distinct and easily recognized due to its linear progression from the minimum to the maximum values within the bounded range of the distribution. This characteristic is unique compared to the skewed distributions of exponential order statistics and the symmetric, bell-shaped distributions of normal order statistics.

Exponential distribution

Start

The PDF of the Exponential Distribution is given by:

$$f(x, \lambda) = \lambda e^{-\lambda x}$$

for $x \geq 0$, and $f(x, \lambda) = 0$ for $x < 0$, where $\lambda > 0$ is the rate parameter of the distribution.

CDF

To find the CDF, we integrate the PDF from $-\infty$ to x:

$$F(x; \lambda) = \int_{-\infty}^{x} \lambda e^{-\lambda t} dt$$

Since $f(x; \lambda) = 0$ for $x < 0$, we actually integrate from 0 to x:

$$F(x; \lambda) = \int_{0}^{x} \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}$$

MGF

The MGF, $M(t)$, of a random variable X is given by:

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

For the exponential distribution:

$$M(t) = \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \int_{0}^{\infty} \lambda e^{x(t-\lambda)} dx$$

Solving this integral:

$$M(t) = \frac{\lambda}{\lambda - t}$$

for $t < \lambda$.

Mean and Variance

The mean (μ) and variance (σ^2) of a distribution can be found using the MGF as follows:

- Mean: $\mu = M'(0)$
- Variance: $\sigma^2 = M''(0) - [M'(0)]^2$

For our MGF:

- $M'(t) = \frac{\lambda}{(\lambda-t)^2}$
- $M''(t) = \frac{2\lambda}{(\lambda-t)^3}$

At $t = 0$:

- $\mu = M'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$
- $\sigma^2 = M''(0) - [M'(0)]^2 = \frac{2\lambda}{\lambda^3} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

the Third, Fourth, and Fifth Moments

The nth moment about the mean can be found using the derivatives of the MGF. Specifically:

- Third Moment: $\mu_3 = M'''(0)$
- Fourth Moment: $\mu_4 = M''''(0)$
- Fifth Moment: $\mu_5 = M'''''(0)$

For the exponential distribution, these higher moments involve higher powers of $\frac{1}{\lambda}$, with factorials in the numerator due to the pattern observed in derivatives of $M(t)$.

The moments of a distribution, like the mean, variance, and higher moments, show about the shape, spread, and tail behavior of the distribution. The exponential distribution is closely related to the Poisson distribution, where it describes the time between events in a Poisson process.

Distribution of the sum of exponential

The MGF of an exponential distribution with rate parameter λ is:

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

When considering the sum of independent exponential variates, say $X_1, X_2, ..., X_n$ each with the same rate parameter λ, the MGF of the sum $S = X_1 + X_2 + ... + X_n$ is the product of the individual MGFs, thanks to the independence:

$$M_S(t) = M_{X_1}(t) \times M_{X_2}(t) \times ... \times M_{X_n}(t)$$

Given the MGF of each exponential variate:

$$M_S(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

This MGF, $\left(\frac{\lambda}{\lambda-t}\right)^n$, is recognizable as the MGF of a Gamma distribution with shape parameter n and rate parameter λ. This conclusion comes from the fact that the Gamma distribution generalizes the exponential distribution, where the exponential distribution is a special case of the Gamma distribution with the shape parameter $n = 1$.

In simpler terms, the sum of n independent exponential variates with the same rate λ follows a Gamma distribution with parameters n (shape) and λ (rate). This tells us about the time you'd wait for n events to happen in a process where events occur continuously and independently at a constant average rate.

Irrigation – Those green circles

The aim is to calculate the 90% confidence interval for the average speed of the rotating arm at the outer wheels of an irrigation system, based on provided rotation times.

Calculation

Mean Rotation Time: The mean rotation time is calculated by averaging all the provided rotation times. This gives us an overall average that represents the central tendency of the rotation times dataset.

$$\text{Mean Rotation Time} = \frac{\sum \text{Rotation Times}}{N}$$

Standard Deviation: Where N is the number of observations (rotation times). For our dataset, the mean rotation time was found to be 22.836184 hours.

$$\text{Standard Deviation} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}}$$

Where x_i is each individual rotation time, $\bar{x}$ is the mean rotation time, and $N-1$ represents the degrees of freedom (used for sample standard deviation). The calculated standard deviation for our dataset is 1.227864. The standard error of the mean rotation time is approximately 0.21 hours.The Z-score for a 90% confidence interval is 1.645.

First, we calculate the mean rotation time, which represents the average time it takes for the system to complete one full rotation.

```
rotation_times <- c(
    21.80086, 23.74087, 24.6675, 22.1376, 21.4186,
    23.80423, 23.11184, 24.23174, 24.826, 21.44181,
    22.09314, 22.96205, 22.27362, 23.23669, 22.05037,
    21.8075, 22.5501, 24.55148, 23.21969, 24.36872,
    24.56083, 23.8828, 21.84536, 21.90287, 21.55993,
    22.91966, 22.74965, 24.86386, 21.56766, 24.81992,
    22.77892, 21.23745, 22.1006, 21.12459, 21.05793
)
radius_feet <- 1320
rotation_speed <-  2*pi*radius_feet /rotation_times

mean_rotation_speed <- mean(rotation_speed)
std_deviation <- sd(rotation_speed)
n <- length(rotation_speed)
df <- n - 1
t_critical <- qt(0.95, df)
```

```r
margin_of_error <- t_critical * (std_deviation / sqrt(n))


ci_lower_speed <- mean_rotation_speed - margin_of_error
ci_upper_speed <- mean_rotation_speed+ margin_of_error

cat("Mean Rotation Speed: ", mean_rotation_speed, "\n")
```

```
## Mean Rotation Speed:  364.195
```

```r
cat("Standard Deviation: ", std_deviation, "\n")
```

```
## Standard Deviation:  19.31709
```

```r
cat(" Confidence Interval Lower Bound: ", ci_lower_speed, "\n")
```

```
##  Confidence Interval Lower Bound:  358.6738
```

```r
cat(" Confidence Interval Upper Bound: ", ci_upper_speed , "\n")
```

```
##  Confidence Interval Upper Bound:  369.7162
```

The 90% confidence interval for the mean rotation speed of the rotating arm at the outer wheels, based on the given rotation times,is between 358.6738 feet per hour and 369.7162 feet per hour. The mean rotation time for the irrigation system is approximately 364.195 feet per hour, with a standard deviation of about 19.31709 feet per hour. Based on the 90% confidence interval calculation, we can be 90% confident that the speed of the rotating arm at the outer wheels is between 358.6738 feet per hour and 369.7162 feet per hour.

Explanation

The confidence interval for speed requires you to adjust the margin of error from the mean rotation time to a margin of error for speed. Since the relationship between time and speed is inversely proportional, the calculations involve using the square of the mean rotation time.With our analysis, we've determined that the mean rotation speed of the irrigation system's arm is 364.195 feet per hour. Accounting for variability in our data, signified by a standard deviation of 19.31709, we can confidently assert that the typical speed will generally fall between 358.6738 feet per hour and 369.7162 feet per hour. This range is calculated with 90% confidence, which serves as a reliable guide for both farmers monitoring their irrigation systems and data scientists seeking to model agricultural processes. This research discovery holds significant implications. The identified mean rotation speed and its confidence interval provide a practical benchmark for monitoring the efficiency of their irrigation systems. We can compare their actual pivot speeds against this range to gauge if their equipment is operating within expected parameters, or if it requires maintenance or adjustment.