

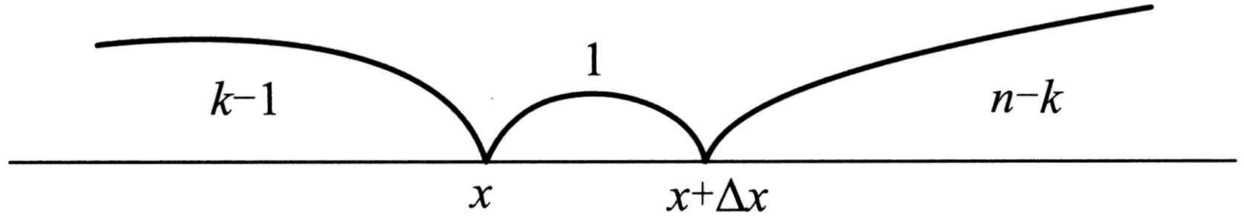
MA677 Midterm Project

Fengyuan (Vincent) Shen

2024-03-28

1. Order Statistics

First of all, let me discuss the derivation of the probability density function for the k^{th} order statistic $X_{(k)}$ of a set of n independent and identically distributed continuous random variables X_1, X_2, \dots, X_n with cumulative distribution function $F(x)$ and probability density function $p(x)$.



As is shown in the plot above, the probability density function $p_k(x)$ for the k^{th} order statistic is derived by considering the probability that $k - 1$ observations fall below x , exactly one observation falls in the small interval $[x, x + \Delta x]$ (where Δx is a very small positive number), and the remaining $n - k$ observations fall above $x + \Delta x$. The probability of $X_{(k)}$ falling between x and $x + \Delta x$ can be approximately represented as:

$$F_k(x + \Delta x) - F_k(x) \approx \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [F(x + \Delta x) - F(x)] [1 - F(x + \Delta x)]^{n-k}$$

This expression considers one observation in the interval $[x, x + \Delta x]$ which is represented as $F(x + \Delta x) - F(x)$, while $F(x)$ to the power of $k - 1$ represents the $k - 1$ observations that fall below x , and $[1 - F(x)]$ to the power of $n - k$ represents the $n - k$ observations above $x + \Delta x$.

To find the density function $p_k(x)$ of the k^{th} order statistic, we divide both sides by Δx and take the limit as Δx approaches zero:

$$p_k(x) = \lim_{\Delta x \rightarrow 0} \frac{F_k(x + \Delta x) - F_k(x)}{\Delta x} = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} p(x)$$

This function $p_k(x)$ is the probability density function of the k^{th} order statistic for any continuous distribution. Furthermore, we can derive the density functions for the specific cases of the minimum and maximum order statistics within a sample of size n . The minimum order statistic (first order statistic) $p_1(x)$ and the maximum order statistic (nth order statistic) $p_n(x)$ have the density functions:

$$p_1(x) = n \cdot [1 - F(x)]^{n-1} \cdot p(x)$$

$$p_n(x) = n \cdot [F(x)]^{n-1} \cdot p(x)$$

Uniform Distribution

The PDF and CDF for a uniform distribution $U(0, 1)$ are given by:

$$p(x) = \begin{cases} 1, & \text{for } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } 0 \leq x \leq 1 \\ 1, & \text{for } x > 1 \end{cases}$$

Therefore, the PDF for the k^{th} order statistic $X_{(k)}$ from a uniform distribution is given by:

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k}, \text{ for } 0 \leq x \leq 1$$

This is equivalent to a Beta distribution with parameters $\alpha = k$ and $\beta = n - k + 1$, that is:

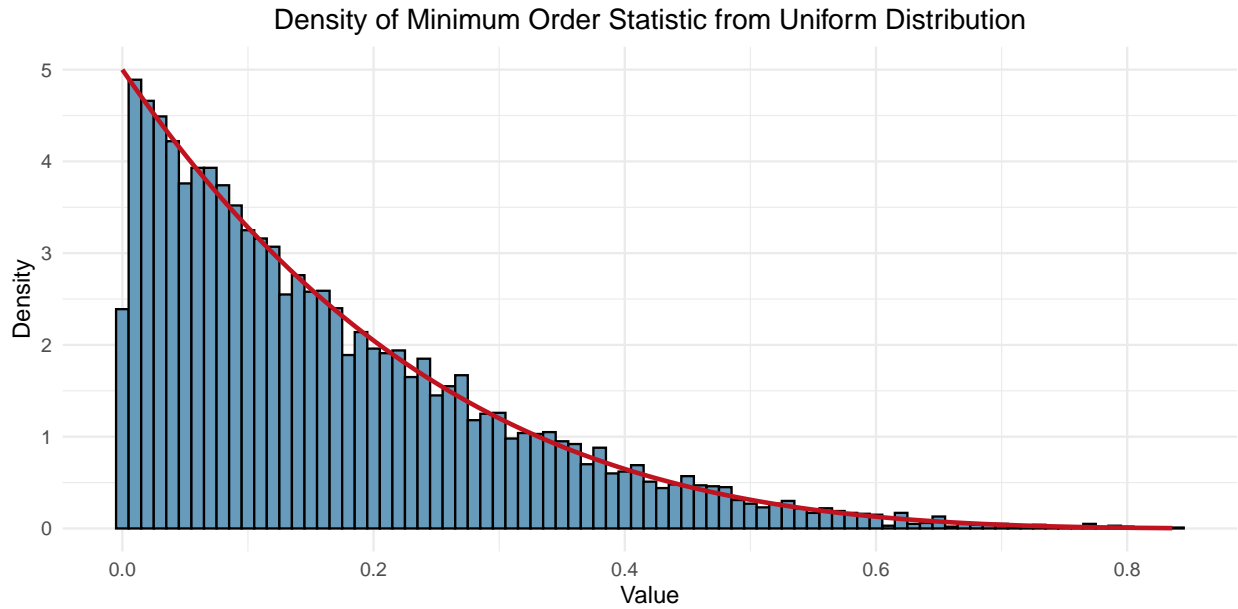
$$Beta(k, n - k + 1)$$

$$E(x_{(k)}) = \frac{k}{n+1}$$

Minimum Order Statistic

The PDF for the minimum order statistic $X_{(1)}$:

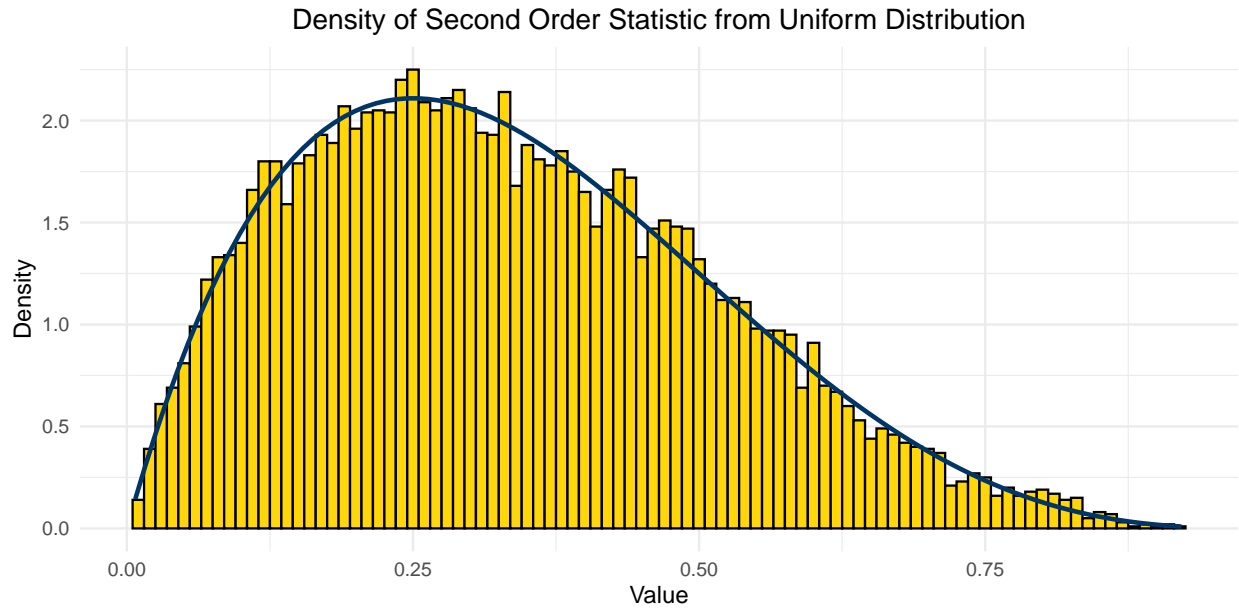
$$p_1(x) = 5(1-x)^4$$



Second Order Statistic

The PDF for the second order statistic $X_{(2)}$:

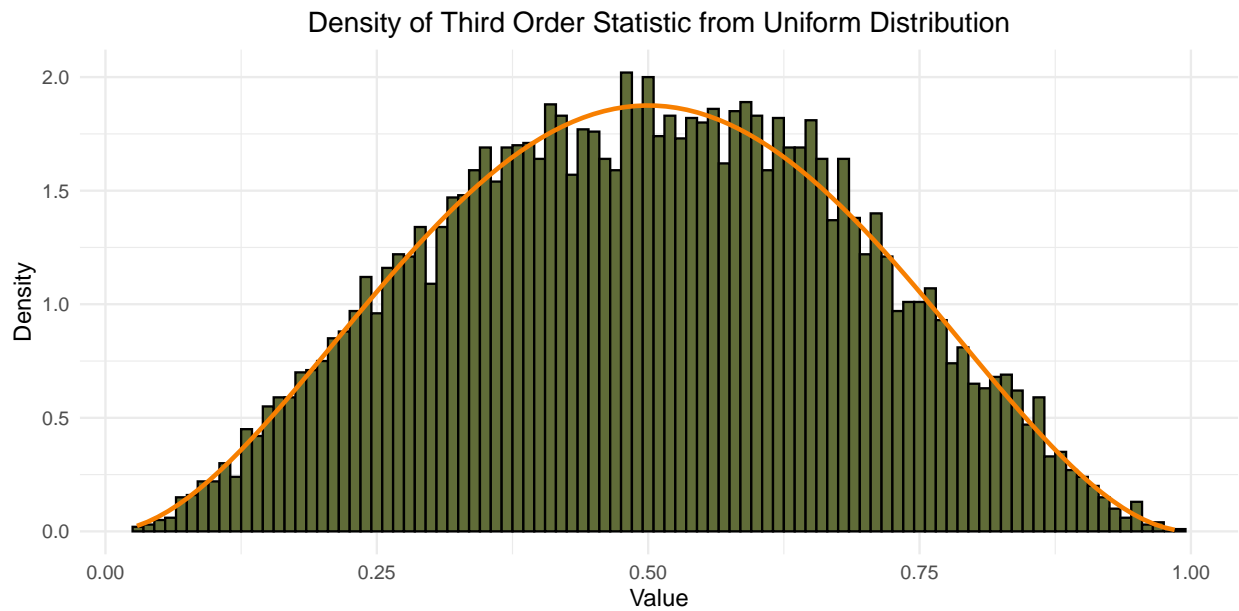
$$p_2(x) = 20x(1 - x)^3$$



Third Order Statistic

The PDF for the third order statistic $X_{(3)}$:

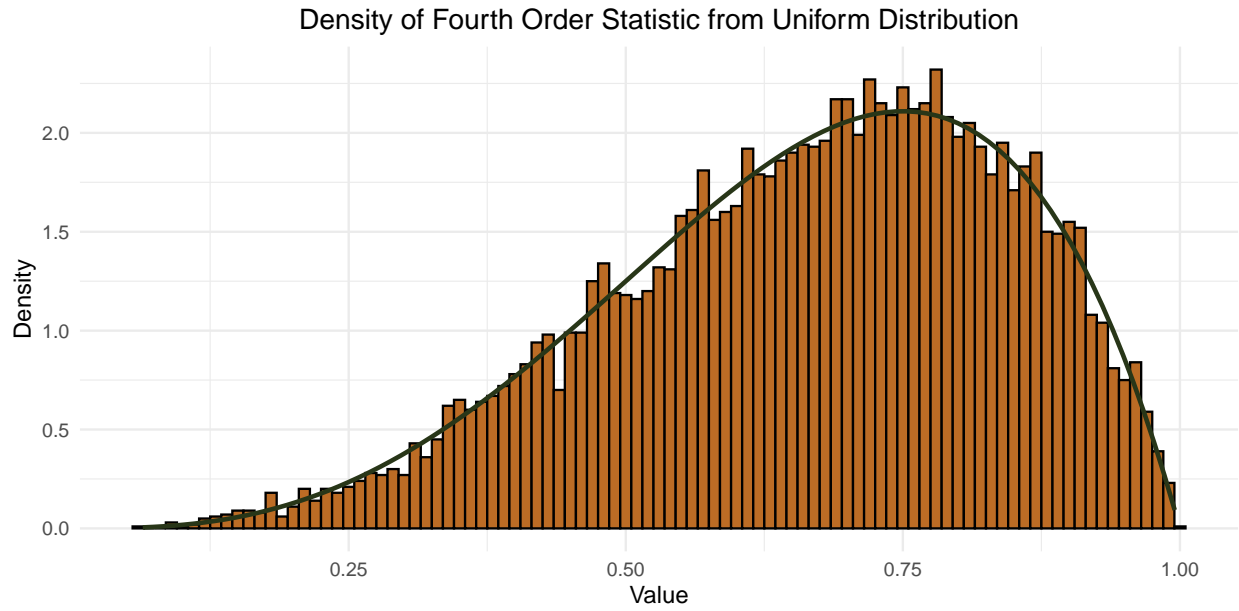
$$p_3(x) = 30x^2(1 - x)^2$$



Fourth Order Statistic

The PDF for the fourth order statistic $X_{(4)}$:

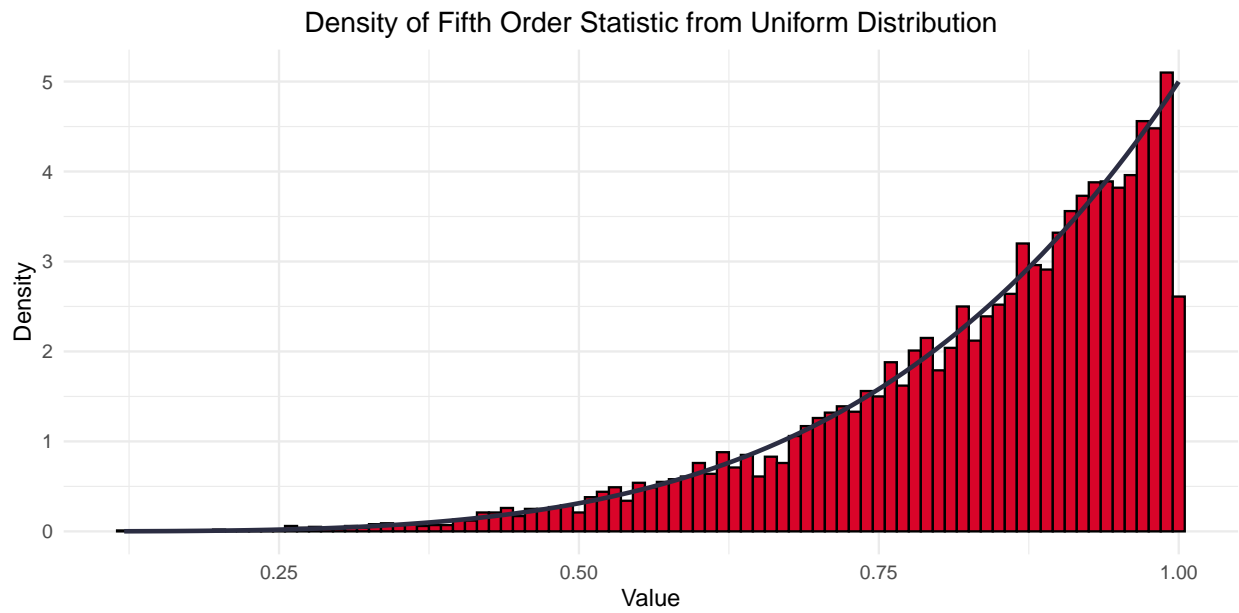
$$p_4(x) = 20x^3(1-x)$$



Fifth Order Statistic

The PDF for the maximum order statistic $X_{(5)}$:

$$p_5(x) = 5x^4$$



From the plots provided, each histogram corresponds to one of the five order statistics derived from a uniform distribution $U(0, 1)$, with a sample size of $n = 5$. The curve on each plot represents the theoretical Beta distribution that matches the order statistic being depicted.

To be more specific, the PDF for the minimum order statistic $X_{(1)}$ is equivalent to $Beta(1, 5)$. The PDF for the second order statistic $X_{(2)}$ is equivalent to $Beta(2, 4)$. The PDF for the third order statistic $X_{(3)}$ is equivalent to $Beta(3, 3)$. The PDF for the fourth order statistic $X_{(4)}$ is equivalent to $Beta(4, 2)$. The PDF for the maximum order statistic $X_{(5)}$ is equivalent to $Beta(5, 1)$.

The match between the histograms (empirical distribution) and the overlaid Beta distribution (theoretical distribution) appears to be quite good.

Exponential Distribution

For an exponential distribution with rate λ , the PDF and CDF are:

$$p(x) = \lambda e^{-\lambda x}, \text{ for } x \geq 0$$

$$F(x) = 1 - e^{-\lambda x}, \text{ for } x \geq 0$$

The PDF for the k^{th} order statistic from an exponential distribution is:

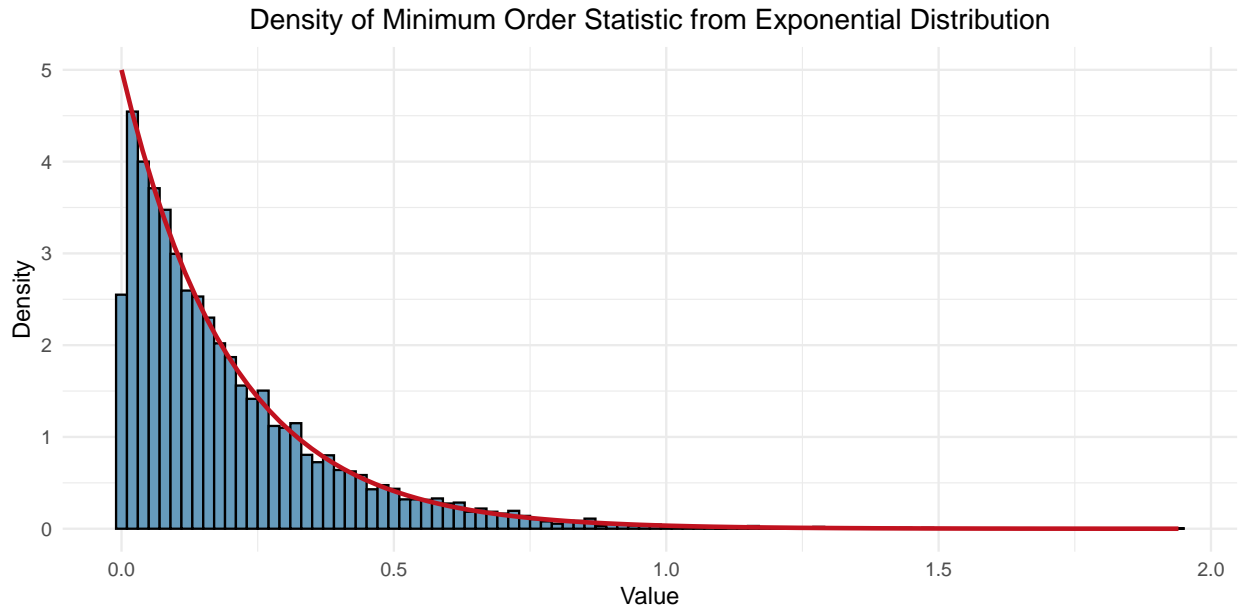
$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} (\lambda e^{-\lambda x}) (1 - e^{-\lambda x})^{k-1} (e^{-\lambda x})^{n-k}$$

Minimum Order Statistic

Here, I choose $\lambda = 1$ for better demonstration.

The PDF for the minimum order statistic $X_{(1)}$:

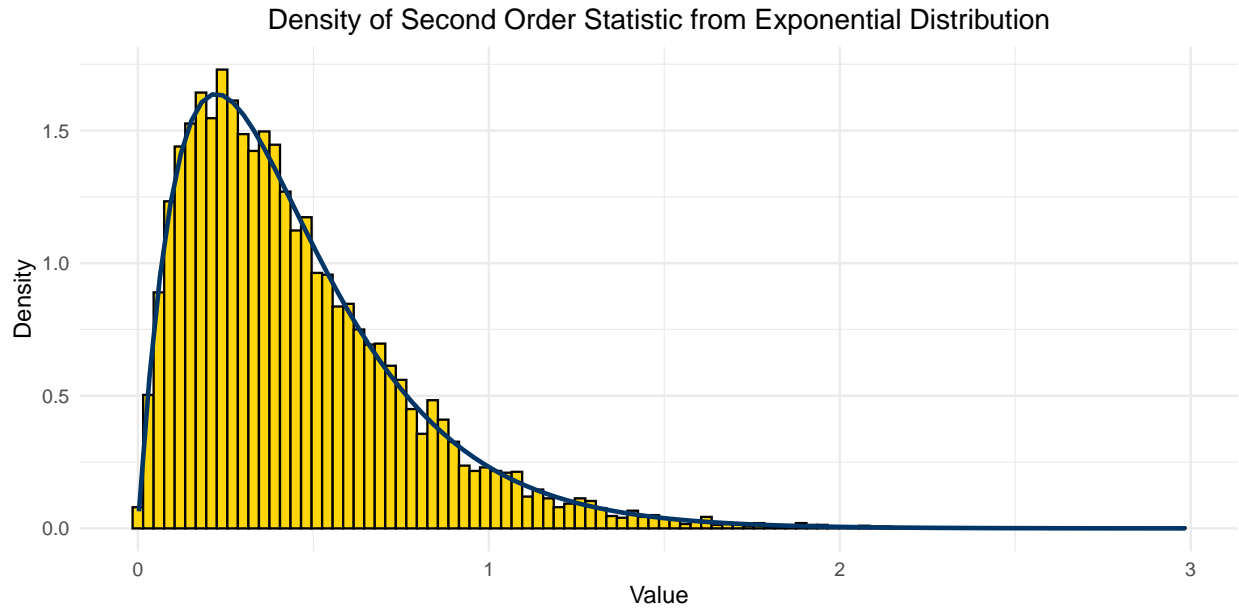
$$p_1(x) = 5e^{-5x}$$



Second Order Statistic

The PDF for the second order statistic $X_{(2)}$:

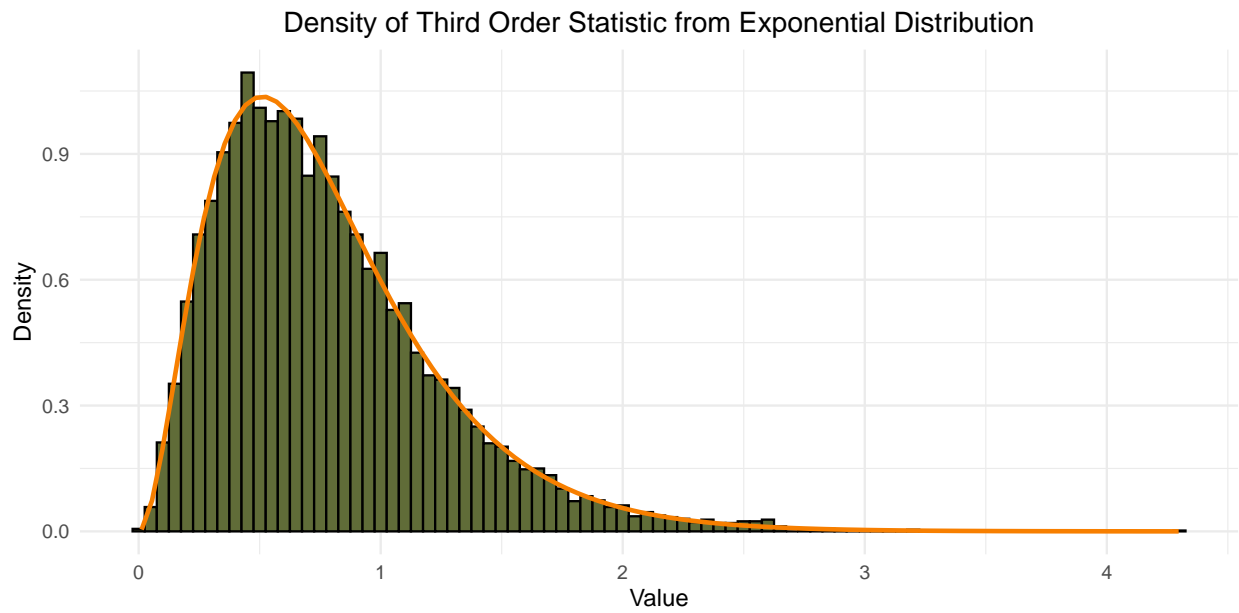
$$p_2(x) = 20(1 - e^{-x})e^{-4x}$$



Third Order Statistic

The PDF for the third order statistic $X_{(3)}$:

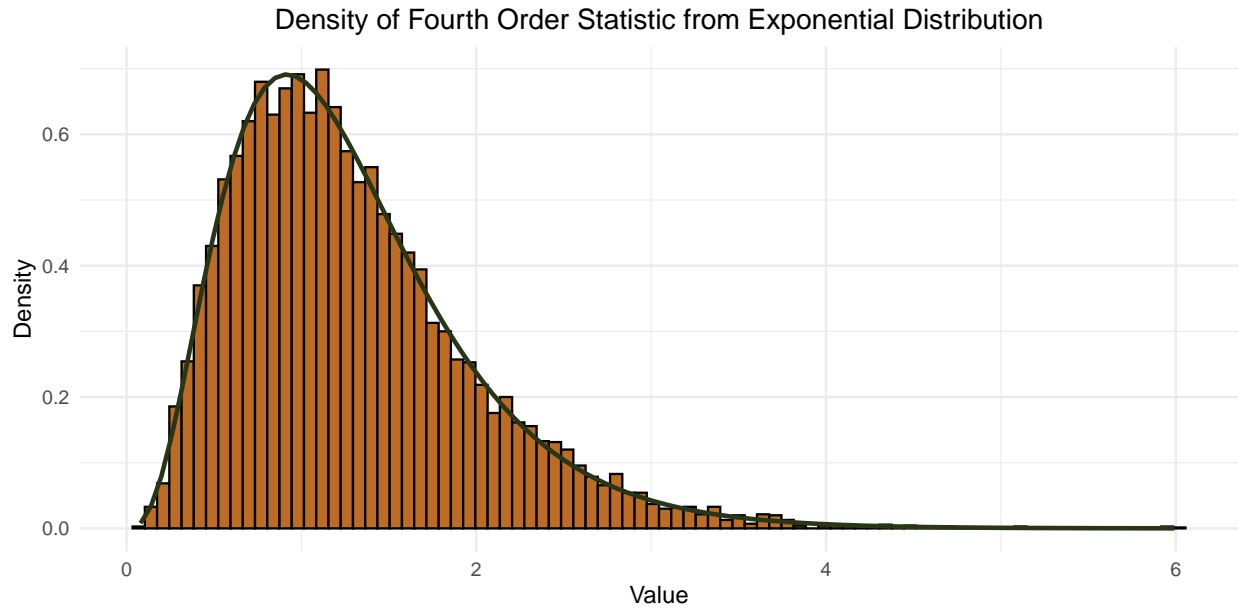
$$p_3(x) = 30(1 - e^{-x})^2e^{-3x}$$



Fourth Order Statistic

The PDF for the fourth order statistic $X_{(4)}$:

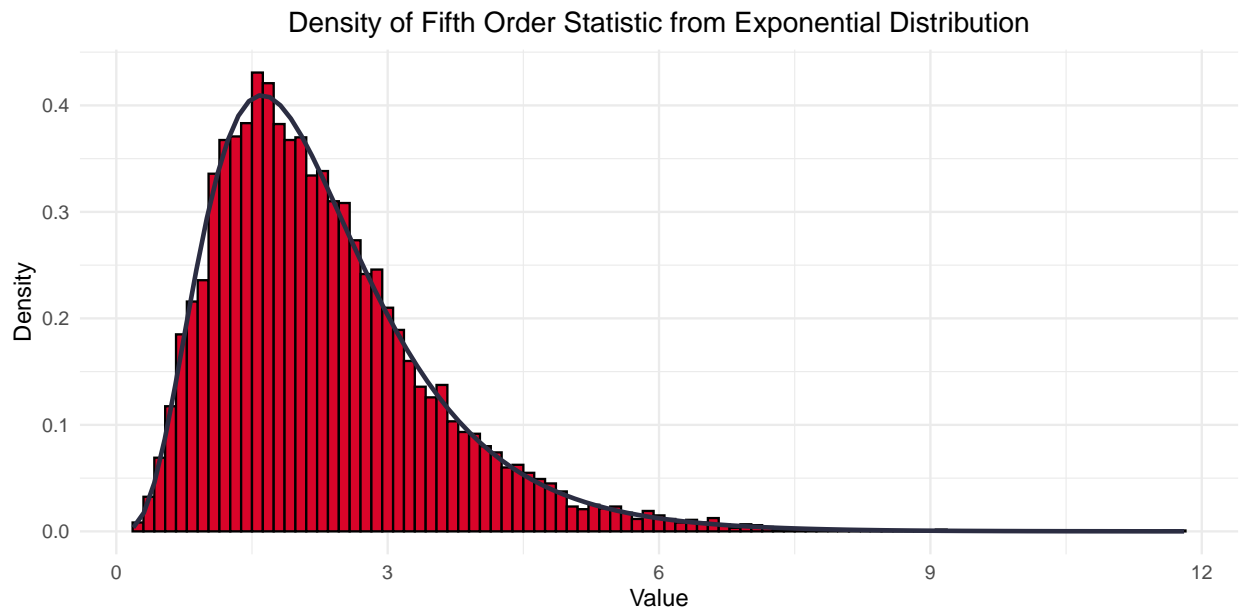
$$p_4(x) = 20(1 - e^{-x})^3 e^{-2x}$$



Fifth Order Statistic

The PDF for the maximum order statistic $X_{(5)}$:

$$p_5(x) = 5(1 - e^{-x})^4 e^{-x}$$



The minimum order statistic from $Exp(1)$ has the highest density near zero and the density decreases rapidly as the value increases, which is characteristic of the exponential distribution. As we move to higher order statistics, the peak of the distribution shifts rightward and the distribution becomes more spread out, indicating a higher variability and a greater likelihood of larger values. The maximum order statistic shows the greatest spread and the peak of the distribution occurs much farther from zero compared to the minimum order statistic.

Overall, the histograms for the order statistics from an exponential distribution exhibit a clear pattern that is consistent with the curve on each plot indicating underlying theoretical distribution.

Normal Distributions

For the k^{th} order statistic from a normal distribution $N(\mu, \sigma^2)$, the PDF can be approximated by:

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} \left[\Phi\left(\frac{x-\mu}{\sigma}\right) \right]^{k-1} \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right) \right]^{n-k} \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$$

where ϕ is the standard normal PDF, and Φ is the standard normal CDF.

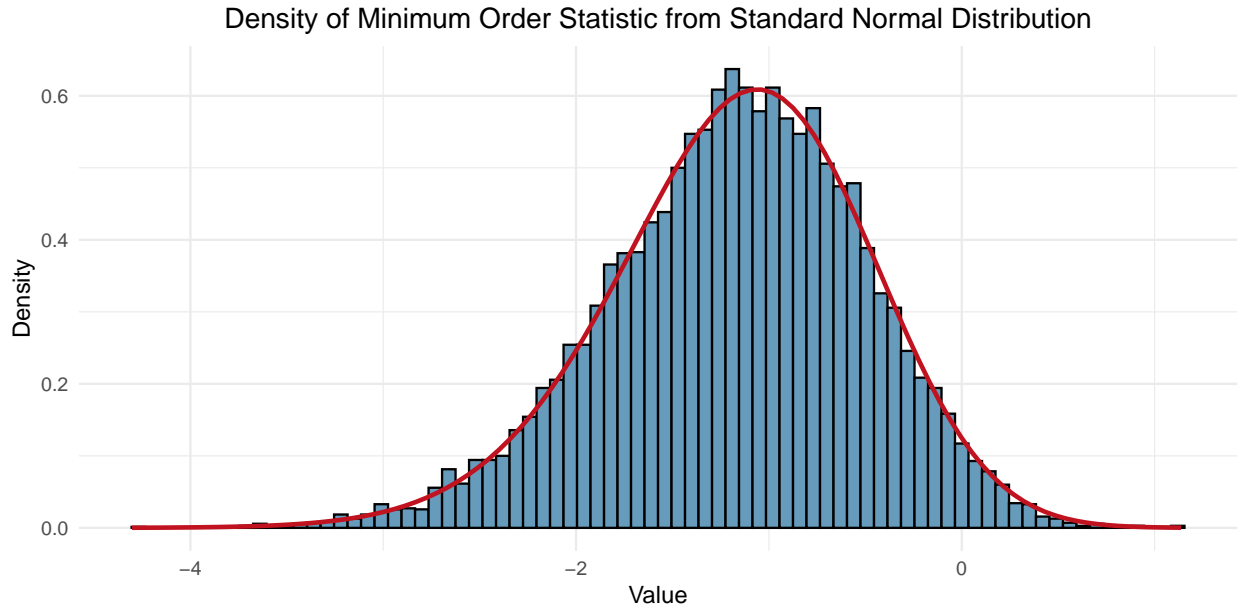
Minimum Order Statistic

Here, I choose $\mu = 0$ and $\sigma = 1$ for better demonstration. So the normal distribution becomes standard normal distribution $N(0, 1)$, and the PDF of k^{th} order statistic becomes:

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} [\Phi(x)]^{k-1} [1 - \Phi(x)]^{n-k} \phi(x)$$

The PDF for the minimum order statistic $X_{(1)}$:

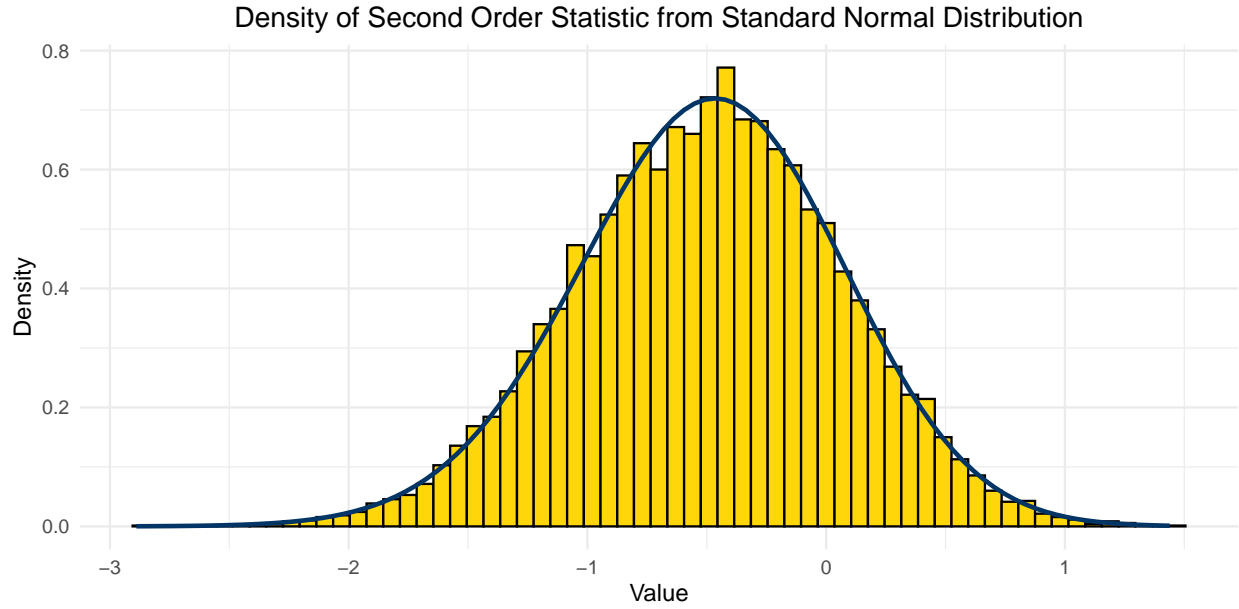
$$p_1(x) = 5[1 - \Phi(x)]^4 \phi(x)$$



Second Order Statistic

The PDF for the second order statistic $X_{(2)}$:

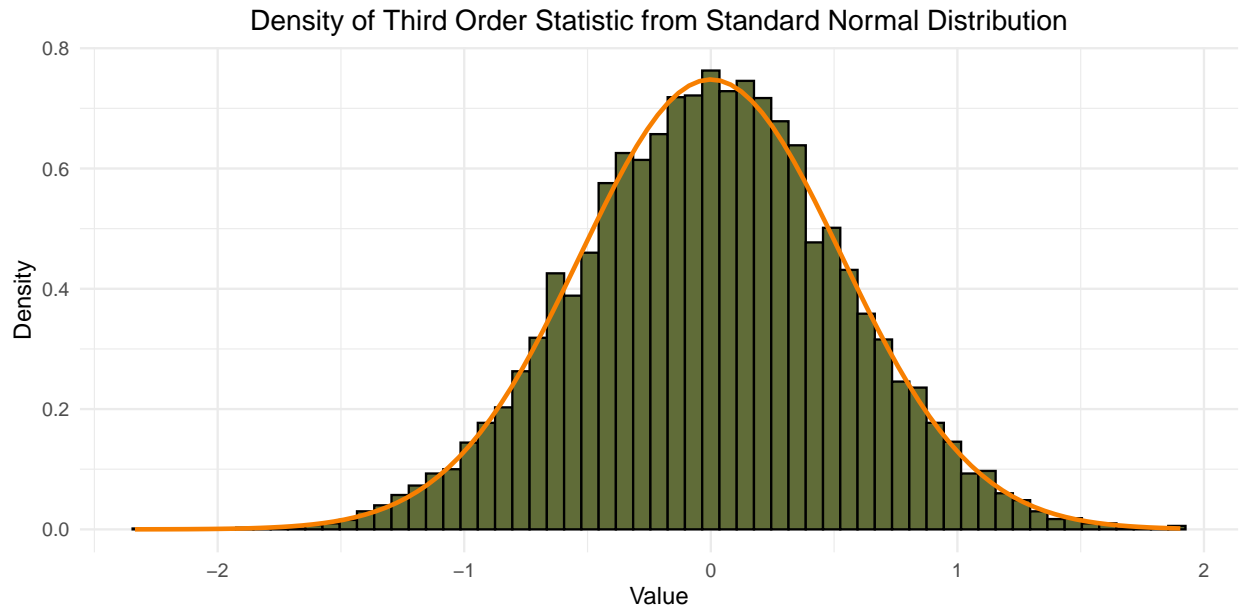
$$p_2(x) = 20[\Phi(x)][1 - \Phi(x)]^3\phi(x)$$



Third Order Statistic

The PDF for the third order statistic $X_{(3)}$:

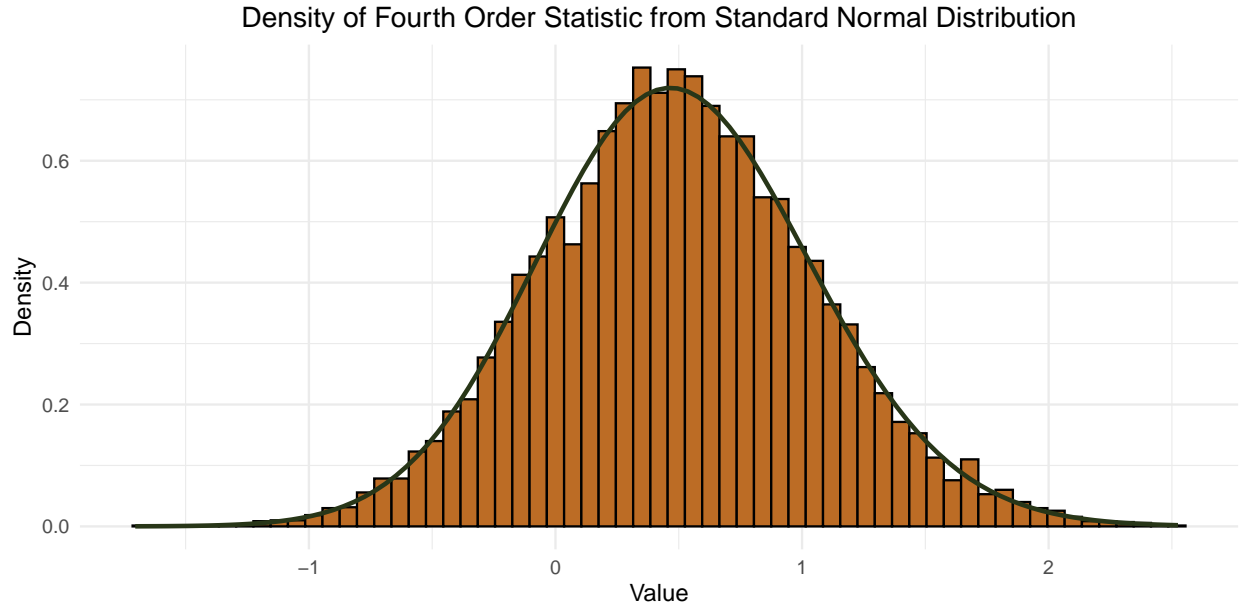
$$p_3(x) = 30[\Phi(x)]^2[1 - \Phi(x)]^2\phi(x)$$



Fourth Order Statistic

The PDF for the fourth order statistic $X_{(4)}$:

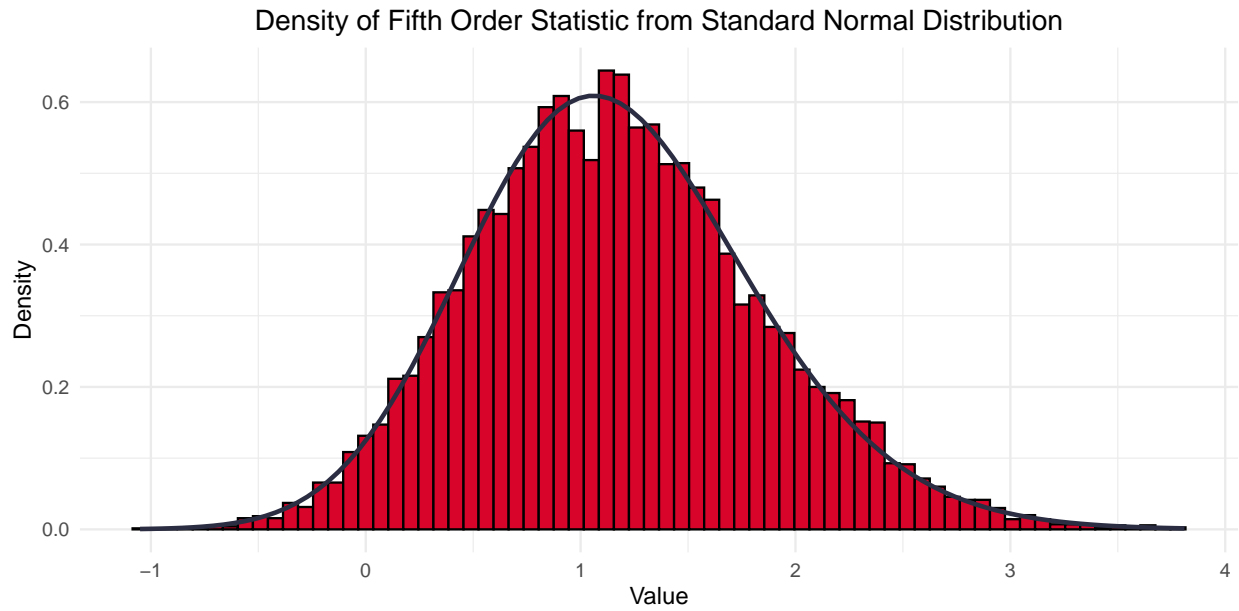
$$p_4(x) = 20[\Phi(x)]^3[1 - \Phi(x)]\phi(x)$$



Fifth Order Statistic

The PDF for the maximum order statistic $X_{(5)}$:

$$p_5(x) = 5[\Phi(x)]^4\phi(x)$$



The histograms for the order statistics from a standard normal distribution exhibit a clear progression in their peak locations, with the minimum being the farthest left and each subsequent statistic moving rightward towards the maximum. This shift corresponds to increasing values within the distribution.

The spread of each histogram also widens from the minimum to the maximum statistic, reflecting the greater variability as one moves through the order statistics.

The empirical data closely match the overlaid theoretical curves, underscoring the consistency of the simulation with the expected behavior of order statistics for the standard normal distribution.

2. Exponential Distribution

Derive CDF and MGF

The Probability Density Function (PDF) of the exponential distribution is given by:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0$$

where λ is the rate parameter of the distribution.

The CDF is derived by integrating the PDF from 0 to x . So:

$$F(x; \lambda) = \int_0^x \lambda e^{-\lambda t} dt = (-e^{-\lambda t}) \Big|_0^x = -(e^{-\lambda x} - e^{-\lambda \cdot 0}) = 1 - e^{-\lambda x}$$

Thus, the CDF is:

$$F(x; \lambda) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0$$

The MGF is defined by the expectation of e^{tx} over the distribution. The MGF of the exponential distribution is given by:

$$M(t) = \mathbb{E}[e^{tx}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{x(t-\lambda)} dx$$

Assuming $t < \lambda$ to ensure the integral converges:

$$M(t) = \lambda \left(\frac{e^{x(t-\lambda)}}{t-\lambda} \right) \Big|_0^\infty = \frac{\lambda}{\lambda - t},$$

Therefore, the MGF is:

$$M(t) = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda$$

Verify Mean and Variance with MGF

Given the MGF of the exponential distribution:

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

The first derivative of the MGF with respect to t gives the first moment of the distribution:

$$\frac{dM(t)}{dt} = \frac{\lambda}{(\lambda - t)^2}$$

Evaluating this derivative at $t = 0$ gives the mean:

$$E(x) = \mu = \left. \frac{\lambda}{(\lambda - t)^2} \right|_{t=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

The second derivative of the MGF with respect to t gives the second moment of the distribution:

$$\frac{d^2M(t)}{dt^2} = \frac{2\lambda}{(\lambda - t)^3}$$

Evaluating this derivative at $t = 0$ gives the second moment about the origin:

$$E(x^2) = \left. \frac{2\lambda}{(\lambda - t)^3} \right|_{t=0} = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

The variance can be calculated by subtracting the square of the mean from the second moment:

$$Var(x) = \sigma^2 = E(x^2) - [E(x)]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Central Moments About the Mean

Third Moment About the Mean (μ_3)

Given the MGF and mean of the exponential distribution:

$$M(t) = \frac{\lambda}{\lambda - t}, E(x) = \mu = \frac{1}{\lambda}$$

The third moment about the mean is:

$$\mu_3 = E[(X - \mu)^3] = \int_0^\infty (x - \mu)^3 f(x; \lambda) dx = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^3 \lambda e^{-\lambda x} dx$$

Expand it:

$$\int_0^\infty \left(x - \frac{1}{\lambda}\right)^3 \lambda e^{-\lambda x} dx = \int_0^\infty \left[x^3 - 3x^2\left(\frac{1}{\lambda}\right) + 3x\left(\frac{1}{\lambda}\right)^2 - \left(\frac{1}{\lambda}\right)^3\right] \lambda e^{-\lambda x} dx$$

Thus, I need to integrate each term of:

$$\lambda x^3 e^{-\lambda x} - 3\lambda \left(\frac{1}{\lambda}\right) x^2 e^{-\lambda x} + 3\lambda \left(\frac{1}{\lambda}\right)^2 x e^{-\lambda x} - \lambda \left(\frac{1}{\lambda}\right)^3 e^{-\lambda x}$$

Given that:

$$\int_0^\infty x^n e^{-\lambda x} dx = \frac{n!}{\lambda^{n+1}}$$

Therefore,

$$\mu_3 = E[(X - \mu)^3] = \int_0^\infty (x - \frac{1}{\lambda})^3 \lambda e^{-\lambda x} dx = \frac{6}{\lambda^3} - \frac{6}{\lambda^3} + \frac{3}{\lambda^3} - \frac{1}{\lambda^3} = \frac{2}{\lambda^2}$$

The third moment about the mean (μ_3) represents the skewness of the distribution, which measures the asymmetry of the distribution around its mean. For the exponential distribution, the skewness is positive ($\mu_3 > 0$), indicating that the distribution is right-skewed.

Fourth Moment About the Mean (μ_4)

The fourth moment about the mean is:

$$\mu_4 = E[(X - \mu)^4] = \int_0^\infty (x - \mu)^4 f(x; \lambda) dx = \int_0^\infty (x - \frac{1}{\lambda})^4 \lambda e^{-\lambda x} dx$$

Expand it:

$$\int_0^\infty (x - \frac{1}{\lambda})^4 \lambda e^{-\lambda x} dx = \int_0^\infty [x^4 - 4x^3 \left(\frac{1}{\lambda}\right) + 6x^2 \left(\frac{1}{\lambda}\right)^2 - 4x \left(\frac{1}{\lambda}\right)^3 + \left(\frac{1}{\lambda}\right)^4] \lambda e^{-\lambda x} dx$$

Thus, I need to integrate each term of:

$$x^4 \lambda e^{-\lambda x} - 4x^3 \left(\frac{1}{\lambda}\right) \lambda e^{-\lambda x} + 6x^2 \left(\frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} - 4x \left(\frac{1}{\lambda}\right)^3 \lambda e^{-\lambda x} + \left(\frac{1}{\lambda}\right)^4 \lambda e^{-\lambda x}$$

Given that:

$$\int_0^\infty x^n e^{-\lambda x} dx = \frac{n!}{\lambda^{n+1}}$$

Same as above,

$$\mu_4 = E[(X - \mu)^4] = \int_0^\infty (x - \frac{1}{\lambda})^4 \lambda e^{-\lambda x} dx = \frac{24}{\lambda^4} - \frac{24}{\lambda^4} + \frac{12}{\lambda^4} - \frac{4}{\lambda^4} + \frac{1}{\lambda^4} = \frac{9}{\lambda^4}$$

The fourth moment about the mean (μ_4) relates to the kurtosis of the distribution, which measures the “tailedness” or the peak of the distribution. For the exponential distribution, the kurtosis is higher than that of a normal distribution, indicating a more pronounced peak and heavier tails.

Fifth Moment About the Mean (μ_5)

The fifth moment about the mean is:

$$\mu_5 = E[(X - \mu)^5] = \int_0^\infty (x - \mu)^5 f(x; \lambda) dx = \int_0^\infty (x - \frac{1}{\lambda})^5 \lambda e^{-\lambda x} dx$$

Expand it:

$$\int_0^\infty (x - \frac{1}{\lambda})^5 \lambda e^{-\lambda x} dx = \int_0^\infty [x^5 - 5x^4 \left(\frac{1}{\lambda}\right) + 10x^3 \left(\frac{1}{\lambda}\right)^2 - 10x^2 \left(\frac{1}{\lambda}\right)^3 + 5x \left(\frac{1}{\lambda}\right)^4 - \left(\frac{1}{\lambda}\right)^5] \lambda e^{-\lambda x} dx$$

Thus, I need to integrate each term of:

$$x^5 \lambda e^{-\lambda x} - 5x^4 \left(\frac{1}{\lambda}\right) \lambda e^{-\lambda x} + 10x^3 \left(\frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} - 10x^2 \left(\frac{1}{\lambda}\right)^3 \lambda e^{-\lambda x} + 5x \left(\frac{1}{\lambda}\right)^4 \lambda e^{-\lambda x} - \left(\frac{1}{\lambda}\right)^5 \lambda e^{-\lambda x}$$

Given that:

$$\int_0^\infty x^n e^{-\lambda x} dx = \frac{n!}{\lambda^{n+1}}$$

Same as above,

$$\mu_5 = E[(X - \mu)^5] = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^5 \lambda e^{-\lambda x} dx = \frac{120}{\lambda^5} - \frac{120}{\lambda^5} + \frac{60}{\lambda^5} - \frac{20}{\lambda^5} + \frac{5}{\lambda^5} - \frac{1}{\lambda^5} = \frac{44}{\lambda^5}$$

The fifth moment about the mean (μ_5) provide additional shape characteristics of the distribution, especially regarding its asymmetry and the behavior of its tails.

Relation to Other Distributions

The exponential distribution is closely related to the gamma distribution. To be more specific, when considering the sum of independent exponential variables, the resulting distribution is a gamma distribution. As a matter of fact, the exponential distribution is a special case of the gamma distribution with shape parameter $k = 1$.

$$Gamma(1, \lambda) = Exp(\lambda)$$

To prove that, considering the MGF of an exponential distribution with rate parameter λ and the MGF of a gamma distribution with shape k and rate λ (or scale $\theta = 1/\lambda$):

$$M_{\text{exp}}(t) = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda$$

$$M_{\text{gamma}}(t) = \mathbb{E}[e^{tx}] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{xt} x^{\alpha-1} e^{-\lambda x} dx = (1 - \theta t)^{-k} = \left(\frac{\lambda}{\lambda - t}\right)^k, \quad \text{for } t < \lambda$$

For k independent exponential random variables X_1, X_2, \dots, X_k each with rate λ , the MGF of their sum is the product of their each MGF:

$$M_{\text{sum}}(t) = (M_{\text{exp}}(t))^k = \left(\frac{\lambda}{\lambda - t}\right)^k = M_{\text{gamma}}(t)$$

This MGF corresponds to that of a gamma distribution with parameters k and λ , demonstrating that the sum of k independent exponential random variables with rate λ follows a gamma distribution:

$$\sum_{i=1}^k Exp(\lambda) = Exp(\lambda) + Exp(\lambda) + \dots + Exp(\lambda) = Gamma(k, \lambda)$$

3. Irrigation - Those Green Circles

Read data:

```
##                               V1
## 1    21.80086 23.74087 24.6675 22.1376 21.4186
## 2    23.80423 23.11184 24.23174 24.826 21.44181
## 3    22.09314 22.96205 22.27362 23.23669 22.05037
## 4    21.8075 22.5501 24.55148 23.21969 24.36872
## 5    24.56083 23.8828 21.84536 21.90287 21.55993
## 6    22.91966 22.74965 24.86386 21.56766 24.81992

## Rows: 7
## Columns: 1
## $ V1 <chr> "21.80086 23.74087 24.6675 22.1376 21.4186", "23.80423 23.11184 24.~
```

To calculate a 90% confidence interval for the speed of the rotating arm at the outer wheels, we should perform the following steps:

1. Convert rotation time to rotation speed.
2. Calculate the mean and the standard deviation of the rotation speed.
3. Use t-distribution to find the appropriate multiplier for the 90% confidence level.
4. Calculate the 90% confidence interval for the rotation speed.

Let me do it step-by-step.

Convert Rotation Time to Rotation Speed

Rotation speed v is inversely proportional to rotation time t . Given the circumference of the path traced by the outer wheels, we can calculate the speed:

$$v = \frac{\text{distance}}{\text{time}}$$

The distance is the circumference of the circle with a radius that includes the length of the arm and the extension of the end-gun, which is $1320 + 100 = 1420$ feet. Therefore,

$$\text{distance} = 2 \cdot \pi \cdot \text{radius} = 2 \cdot \pi \cdot 1420 = 2840 \cdot \pi$$

The rotation speed is:

$$v = \frac{\text{distance}}{\text{time}} = \frac{2840 \cdot \pi}{\text{time}}$$

By calculation, I can derive the speed of the rotating arm at the outer wheels:

```
## Rotation Speed (feet per hour):
## 409.2556 375.8128 361.6955 403.0303 416.5596 374.8125 386.0412 368.1999 359.3863 416.1087 403.8413
```

Calculate Mean and Standard Deviation

For calculating mean of rotation speed:

$$\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$$

For calculating standard deviation of rotation speed:

$$\text{Standard Deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2}$$

By calculation, I can derive the mean and standard deviation of rotation speed:

```
## Average Rotation Time: 22.83618 hours
## Average Speed: 391.7855 feet per hour
## Standard Deviation of Speed: 20.78051 feet per hour
```

Calculate t-distribution

Then let me calculate the t_score to determine how far the confidence interval should extend around the mean:

$$t_{\alpha/2, n-1} = QT\left(\frac{1 + \text{confidence level}}{2}, n-1\right)$$

```
## t-score For a 90% Confidence Interval: 1.690924
```

Calculate 90% Confidence Interval

Finally, using the following formula, I derive the 90% confidence interval for the speed of the rotating arm at the outer wheels:

$$\text{Standard Error} = \frac{\text{Standard Deviation}}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (v_i - \bar{v})^2}$$

$$90\% \text{ Confidence Interval} = [\bar{v} - t_score \cdot \text{Standard Error}, \bar{v} + t_score \cdot \text{Standard Error}]$$

Therefore,

```
## 90% Confidence Interval: [ 385.8461 , 397.725 ] feet per hour
```

Explanation for Data scientists and Farmers

A center pivot irrigation system is a marvel of modern agriculture, allowing for efficient watering of crops over large areas. To understand how quickly it moves, I analyzed the rotation speed of one such system, discovering that on average, it completes a rotation in about **22.83618** hours. Given the size of the circle it covers, this translates to an average speed of **391.7855** feet per hour for the outer edge of the system.

Understanding the system's speed helps in planning and optimizing water and chemical application, ensuring crops receive the right amount of resources at the right time. For my analysis, I am not only interested in the average speed but also in the range within which the true average speed likely falls. I calculated a 90% confidence interval, which means I am 90% confident that the true average speed lies between **385.8461** and **397.725** feet per hour. This range is helpful for making informed decisions about irrigation schedules and system maintenance, ensuring that the system operates efficiently and effectively.