EXPONENTIAL DISTRIBUTION

TEAM MEMBERS

- Abhishek Sonone (MBA202224-009)
- Akash Talokar (MBA202224-012)
- Ashish Kumar (MBA202224-029)
- Ashutosh Sahoo (MBA202224-031)
- Diptarshi Maiti (MBA202224-051)
- Prateek Shukla (MBA202224-109)

ABOUT

- The exponential distribution is a continuous probability distribution that describes the time between independent, random events occurring at a constant rate. It has one parameter, called the rate parameter, denoted by λ, which represents the expected number of events per unit of time.

- P.D.F = $f(x) = \lambda e^{(-\lambda x)}$, for $x \geq 0$
- where x is the time between two successive event

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

- The mean of the distribution is $1/\lambda$
- The variance of the distribution is $1/\lambda^2$
- Moment Generating Function is $\lambda/(1-\lambda)$

Insurance Sector

- The exponential distribution is often used in the insurance industry to model the time until a specific event occurs, such as the time between losses or claims. Insurance companies use this distribution to estimate the probability of a loss or claim occurring within a given time frame, which helps them determine premiums and policy limits.
- Suppose an insurance company sells a policy that covers losses due to theft for a certain type of merchandise. Historical data shows that the average time between theft claims for this type of merchandise is 6 months, and the insurer wants to use this information to calculate the probability of a theft claim in the next 3 months.
- Using the exponential distribution, we can calculate the probability of a theft claim occurring within the next 3 months as follows:
- The rate parameter, λ, can be calculated by taking the reciprocal of the average time between theft claims, which is 1/6. Thus, $\lambda = 1/6$.

- We want to find the probability of a theft claim occurring within the next 3 months, which is equivalent to finding the probability that the time until the next theft claim is less than 3 months. Therefore, we need to calculate the cumulative distribution function (CDF) of the exponential distribution at $x=3$:

- $P(x<=3) = 1 - e^{(-\lambda x)} = 1 - e^{(-1/2)} \approx 0.393$

- This means that the probability of a theft claim occurring within the next 3 months is approximately 0.393, or 39.3%.

- The insurer can use this information to set appropriate premiums for the policyholders, taking into account the expected frequency and severity of theft claims. By using the exponential distribution to model the time between theft claims, the insurer can better estimate the risk of losses due to theft and make more informed decisions about pricing and risk management.

PROPERTIES

- **Memorylessness:** The probability of an event occurring at any given time is independent of the time elapsed since the last event.
- **Lack of upper bound:** The exponential distribution has no upper bound, meaning there is a non-zero probability that the time between events can be arbitrarily large.
- **Monotonicity:** The exponential distribution is a decreasing function, meaning that the probability of longer times between events decreases as time increases.

Amazon uses exponential distribution to predict the delivery time of its products.

- The company uses various factors such as the distance between the warehouse and the customer, the number of products in stock, and the shipping mode to estimate the time it would take to deliver a product.
- This information is used to optimize the delivery process and improve customer satisfaction.
- Additionally, Amazon uses exponential distribution to model the time between customer visits to its website, which helps the company understand user behaviour and optimize its website design and functionality.
- By using exponential distribution, Amazon can estimate the probability of various time intervals, which helps it make informed decisions about its operations and customer experience.

- Let's consider an example of using exponential distribution to predict the delivery time of products for an online store. Suppose that the store has a warehouse located 20 miles away from the customer, and the average speed of delivery is 5 miles per hour. We can use the exponential distribution to estimate the probability of the time it would take to deliver the product.
- The probability density function of the exponential distribution is given by:
- $f(x; \lambda) = \lambda * e^{(-\lambda * x)}$ for $x \geq 0$
- where λ is the rate parameter, which represents the average number of events per unit of time, and x is the random variable representing the time between events.
- We can estimate the value of λ using the average speed of delivery:
- $\lambda = 1 / (\text{average speed}) = 1 / 5 = 0.2$

Now, let's say we want to estimate the probability that the delivery time is less than or equal to 4 hours:

$$P(X \leq 4) = \int_{[0,4]} \lambda * e^{(-\lambda*x)}\, dx$$

Using integration, we can calculate:

$$\begin{aligned} P(X \leq 4) &= [-e^{(-\lambda x)}] \text{ from } 0 \text{ to } 4 \\ &= -e^{(-0.24)} + e^{(-0.2*0)} \\ &= -e^{(-0.8)} + 1 \\ &\approx 0.5498 \end{aligned}$$

Therefore, the probability of delivery time being less than or equal to 4 hours is approximately 0.5498, assuming that the delivery time follows an exponential distribution with a rate parameter of 0.2.

This information can be used by the online store to optimize its delivery operations and improve customer satisfaction by providing more accurate delivery estimates.

PHARMACEUTICAL INDUSTRY

The exponential distribution is commonly used in the pharma industry to determine the **time-to-failure of drug products** and the **interarrival times between customer orders** and the **time taken for quality control checks.** Here are some important applications of exponential distribution in the pharma industry:

- **Drug Absorption and Elimination:** It helps pharmaceutical companies determine the appropriate drug dosage, and companies can estimate the time required for a drug to reach its maximum effectiveness and the time it takes to eliminate the drug.
- **Quality Control:** This helps manufacturers estimate the expected time for a quality control check and plan their production accordingly and it is also used to determine the distribution of defect rates, allowing companies to monitor and control the quality of their products.
- **Reliability Analysis:** It is widely used to find the time-to-failure of components and systems. By using these distribution, manufacturers can estimate the expected lifespan of their products and plan their maintenance and replacement strategies accordingly.

Example : A drug has a failure rate of 0.02 per hour. What is the probability that the drug will still be effective after 10 hours?

Solution:

The probability that the drug will still be effective after 10 hours can be calculated as follows:

$$P(X > 10) = e^{(-\lambda*10)}$$

Substituting $\lambda = 0.02$ and 10 for X, we get:

$$P(X > 10) = e^{(-0.02*10)} = e^{(-0.2)} = 0.8187$$

Therefore, the probability that the drug will still be effective after 10 hours is approximately **0.8187 or 81.87%.**

The exponential distribution is commonly used in the pharmaceutical industry to calculate the **half-life** of a drug, which is an essential parameter for determining the drug's appropriate dosage and administration schedule. The half-life of a drug is defined as the **time it takes for the concentration of the drug in the body to decrease to half of its initial value.**

To calculate the half-life using exponential distribution, the drug concentration in the body over time is modelled as an exponential decay process. This means that the rate of change of the drug concentration is proportional to the current concentration of the drug.

Formula = $C(t) = C(0) * e^{(-kt)}$

Where $C(0)$ = Initial concentration, k = decay rate, t = Time

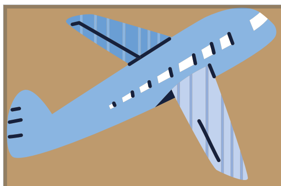
Application of exponential distribution on THE airline industry

- The exponential distribution can be applied to various aspects of the airline industry, such as flight arrival and departure times, maintenance schedules, and delay times. Here are a few examples:

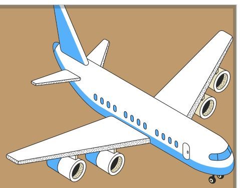
- **Flight arrival and departure times:**
- The exponential distribution can be used to model the time between consecutive arrivals or departures of flights at an airport. This can help airlines and airport operators predict the probability of delays and plan their schedules accordingly.

- **Maintenance schedules:**
- Airlines must perform regular maintenance on their aircraft to ensure they are safe and reliable. The exponential distribution can be used to model the time between maintenance events, such as engine overhauls or tire replacements. This can help airlines optimize their maintenance schedules and minimize downtime.

- **Delay times:**
- Delays are a common occurrence in the airline industry, and they can have significant financial and operational impacts. This can help airlines predict the probability of delays and develop strategies to minimize their impact. It is also important for passengers and airlines to calculate the time difference between connected flights.



Solution



The amount of delay time for a given flight is exponentially distributed with a mean of 0.5 hour. Ten passengers on this flight need to take a subsequent connecting flight. The scheduled connection time is either 1 or 2 hours depending on the final destination. Suppose 3 and 7 passengers are associated with these connection times, respectively.

(a) Suppose John is one of the 10 passengers needing a connection. What is the probability that he will miss his connection? (ans. 0.053)

(b) Suppose he met Mike on the plane, and Mike also needs to make a connection. However, Mike is going to another destination and thus has a different connection time from John's. What is the probability that both John and Mike will miss their connections? (ans. 0.018)

Solution:

Let J_1 and J_2 denote the events that John's scheduled connection time is 1 and 2 hours, respectively, where $P(J_1) = 0.3$ and $P(J_2) = 0.7$. Also, let X be the delay time of the flight in hours. Note that

$$\begin{aligned} P(X > x) &= 1 - P(X \le x) = 1 - F(x) = 1 - (1 - e^{-x/0.5}) \\ &= e^{-x/0.5} \end{aligned}$$

(a) Let M denote the event that John misses his connection, i.e. the flight delay time exceeded his scheduled time for connection. Using the theorem of total probability,

$$\begin{aligned} P(M) &= P(M \mid J_1)P(J_1) + P(M \mid J_2)P(J_2) \\ &= P(X > 1)\times 0.3 + P(X > 2)\times 0.7 \\ &= e^{-1/0.5} \times 0.3 + e^{-2/0.5} \times 0.7 \\ &= 0.135335283\times 0.3 + 0.018315639\times 0.7 \cong \mathbf{0.053} \end{aligned}$$

(b) Regardless of whether John has a connection time of 1 hour and Mike has 2, or the opposite, for them to both miss their connections the flight must experience a delay of more than two hours, and the probability of such an event is

$$P(X > 2) = e^{-2/0.5} \cong \mathbf{0.018}$$

Epidemiology:

- Epidemiology is the study of how diseases spread and affect populations.
- Allows to make predictions about the spread and impact of diseases based on the rate of occurrence of the event.
- Used to model the distribution of the time between infections or the time until recovery or death.
- Public health officials can make informed decisions about how to allocate resources and respond to outbreaks.
- The exponential distribution is also used in epidemiological models, such as the ***SIR*** **(Susceptible-Infectious-Recovered)** model, which is a mathematical model used to study the spread of infectious diseases within a population.
- In this model, the exponential distribution is used to model the time between successive infections, which allows us to estimate the basic **reproduction number (*RO*)** of the disease and make predictions about the future trajectory of the epidemic.
- Here are some examples of the exponential distribution used in epidemiology:
    - **Infectious disease modelling**
    - **Survival analysis**
    - **Vaccination modelling**

- Let's consider the time between successive disease cases in a population.
- The average time between two successive disease cases is 10 days.
- The probability density function of the exponential distribution is given by:
- **f(x) = λ * e^(-λx)**
- λ is the rate parameter. λ is 1/10, i.e. the time between successive cases is **10 days**.
- Suppose we want to calculate the probability that the time between two successive cases is **less than or equal to 5 days**. We can use the cumulative distribution function **(*CDF*)** of the exponential distribution, which is given by:
- **F(x) = 1 - e^(-λx)**
- Substituting the values, we get:
- **F (5) = 1 - e^ (-1/2) = 0.393**
- This means that the probability of the time between two successive cases being less than or equal to 5 days is ***0.393*** or ***39.3%.***

Electronics Industry

- The exponential distribution is commonly used in reliability engineering to model the time between failures of electronic components. The exponential distribution is a suitable model for many electronic components because they tend to have a constant failure rate over time, meaning that the probability of a failure occurring at any given time is independent of how long the component has been in use.

- To use the exponential distribution to find the failure of an electronic component, we need to know the failure rate, which is also known as the hazard rate or lambda (λ). The failure rate represents the probability of the component failing per unit of time, and it can be estimated based on historical failure data or from manufacturer specifications.

- Suppose we have a particular electronic component with a failure rate of λ = 0.001 failures per hour, and we want to find the probability that the component will fail within the next 100 hours of operation. Using the exponential distribution formula, we can calculate the probability density function (PDF) as follows:

$$f(t) = \lambda * e^{(-\lambda t)}$$

Where t is the time in hours since the last failure or since the component was put into operation.

Substituting the values, we get:

$f(100) = 0.001 * e^{(-0.001 * 100)} \approx 0.0009$

This means that the probability of the component failing within the next 100 hours of operation is approximately 0.0009 or 0.09%.

Using the same formula, we can also find the probability of the component not failing within the next 100 hours, which is given by the cumulative distribution function (CDF):

$$F(t) = 1 - e^{(-\lambda t)}$$

Substituting the values, we get:

$F(100) = 1 - e^{(-0.001 * 100)} \approx 0.0993$

This means that the probability of the component not failing within the next 100 hours of operation is approximately 0.0993 or 9.93%.

- Suppose we know that the failure rate of a certain type of electronic component is 0.1 per minute. We want to calculate the probability that the component will fail within the next 10 minutes.

We can solve such problems of exponential distribution with the help of python and R.

Run Debug Stop Share Save {} Beautify

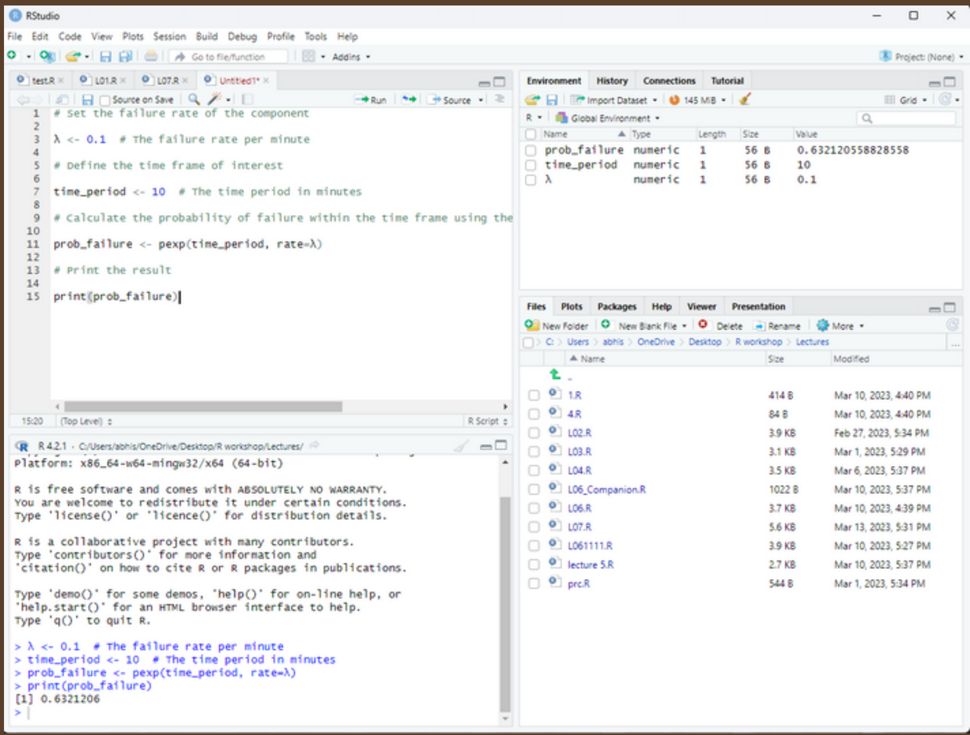
main.py

```python
import numpy as np

# Set the parameters of the distribution
λ = 0.1  # The failure rate per minute
size = 10000  # The number of samples to draw from the distribution

# Draw samples from the exponential distribution
samples = np.random.exponential(scale=1/λ, size=size)

# Calculate the proportion of samples that fall within the time frame of 10 minutes
prob = np.sum(samples <= 10) / size

print("The probability of the component failing within the next 10 minutes is:", prob)

```

input

```
The probability of the component failing within the next 10 minutes is: 0.6264


...Program finished with exit code 0
Press ENTER to exit console.
```

Using R


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
test.R
L01.R
L07.R
Untitled1*
Source on Save
Run
Source
Set the failure rate of the component
λ <- 0.1 # The failure rate per minute
# Define the time frame of interest
time_period <- 10 # The time period in minutes
# Calculate the probability of failure within the time frame using the
prob_failure <- pexp(time_period, rate=λ)
# Print the result
print(prob_failure)
Environment History Connections Tutorial
Import Dataset
Global Environment
Name Type Length Size Value
prob_failure numeric 1 56 B 0.632120558828558
time_period numeric 1 56 B 10
λ numeric 1 56 B 0.1
Files Plots Packages Help Viewer Presentation
New Folder New Blank File Delete Rename More
C: > Users > abhis > OneDrive > Desktop > R workshop > Lectures
Name Size Modified
1.R 414 B Mar 10, 2023, 4:40 PM
4.R 84 B Mar 10, 2023, 4:40 PM
L02.R 3.9 KB Feb 27, 2023, 5:34 PM
L03.R 3.1 KB Mar 1, 2023, 5:29 PM
L04.R 3.5 KB Mar 6, 2023, 5:37 PM
L06_Companion.R 1022 B Mar 10, 2023, 5:37 PM
L06.R 3.7 KB Mar 10, 2023, 4:39 PM
L07.R 5.6 KB Mar 13, 2023, 5:31 PM
L061111.R 3.9 KB Mar 10, 2023, 5:27 PM
lecture 5.R 2.7 KB Mar 10, 2023, 5:37 PM
prc.R 544 B Mar 1, 2023, 5:34 PM
R 4.2.1 · C:/Users/abhis/OneDrive/Desktop/R workshop/Lectures/
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> λ <- 0.1 # The failure rate per minute
> time_period <- 10 # The time period in minutes
> prob_failure <- pexp(time_period, rate=λ)
> print(prob_failure)
[1] 0.6321206

THANK YOU