R Coding Assignment 5: Continuous Probability Distributions

San Diego State University - STAT550

Aiden Jajo

Overview

This assignment explores continuous probability distributions including Normal, Exponential, and Uniform distributions. We apply these distributions to real-world scenarios in healthcare, economics, and quality control.

Key Distributions Covered: - **Normal Distribution**: Bell-shaped, symmetric distribution for continuous data - **Exponential Distribution**: Time between events in a Poisson process - **Uniform Distribution**: All values equally likely within an interval

Key R Functions: - `pnorm()`, `qnorm()`: Normal distribution probabilities and quantiles - `pexp()`: Exponential distribution probabilities - `punif()`: Uniform distribution probabilities

---

Question 1: Birth Weight Analysis - Normal Distribution

Scenario: Birth weights of babies in the United States can be modeled by a normal distribution with mean 3250 grams and standard deviation 550 grams. Those weighing less than 2500 grams are considered to be of low birth weight.

```r
# A) What proportion of babies would the normal distribution predict as weighing
# more than 10 pounds (4536 grams) at birth?
# P(X > 4536) = 1 - P(X <= 4536)
1 - pnorm(4536, mean = 3250, sd = 550)
```

```
## [1] 0.009688909
```

```r
# Interpretation: ~0.97% of babies weigh more than 10 pounds at birth
# This is a relatively rare occurrence

# B) Determine the probability that a randomly selected baby weighs between 3000
# and 4000 grams at birth
# P(3000 < X < 4000) = P(X < 4000) - P(X < 3000)
pnorm(4000, mean = 3250, sd = 550) - pnorm(3000, mean = 3250, sd = 550)
```

```
## [1] 0.5889408
```

```r
# Interpretation: ~58.9% of babies fall within this weight range
# This represents the majority of newborns

# C) How little would a baby have to weigh to be among the lightest 2.5% of all
# newborns?
# Find the 2.5th percentile using qnorm (quantile function)
qnorm(0.025, mean = 3250, sd = 550)
```

```
## [1] 2172.02
```

```r
# Interpretation: Babies weighing less than ~2172 grams are in the lightest 2.5%
# These would be considered very low birth weight

# D) How much would a baby have to weigh to be among the heaviest 10% of all
# newborns?
# Find the 90th percentile (top 10% starts at 90th percentile)
qnorm(0.90, mean = 3250, sd = 550)
```

```
## [1] 3954.853
```

```r
# Interpretation: Babies weighing more than ~3955 grams are in the heaviest 10%
```

Question 2: Teacher Salaries - Normal Distribution

Scenario: The weekly salaries of teachers in one state are normally distributed with a mean of $490 and a standard deviation of $45. What is the probability that a randomly selected teacher earns more than $525 a week?

```r
# P(X > 525) = 1 - P(X <= 525)
# X follows Normal(mean = 490, sd = 45)
1 - pnorm(525, mean = 490, sd = 45)
```

```
## [1] 0.21835
```

```r
# Interpretation: ~21.8% of teachers earn more than $525 per week
# This represents roughly 1 in 5 teachers
```

Question 3: Airline Ticket Purchases - Exponential Distribution

Scenario: The number of days ahead travelers purchase their airline tickets can be modeled by an exponential distribution with the average amount of time equal to 15 days. Find the probability that a traveler will purchase a ticket fewer than ten days in advance.

```r
# Exponential distribution with mean = 15, so rate = 1/15
# P(X < 10) where X follows Exponential(rate = 1/15)
pexp(10, rate = 1 / 15)
```

```
## [1] 0.4865829
```

```r
# Interpretation: ~48.7% of travelers purchase tickets less than 10 days ahead
# Nearly half of all ticket purchases occur within this window
```

Question 4: Package Weight - Uniform Distribution

Scenario: Packages have a nominal net weight of 1000 g. However their actual net weights have a uniform distribution over the interval 980 g to 1030 g.

```r
# A) Find the probability that the net weight of a package is less than 1000 g
# P(X < 1000) where X follows Uniform(980, 1030)
punif(1000, min = 980, max = 1030)
```

```
## [1] 0.4
```

```r
# Interpretation: 40% of packages weigh less than the nominal 1000g
# Since 1000 is at the 40% mark of the interval [980, 1030]

# B) Find the probability that the net weight of a package is less than w g,
# where 990 < w < 1030
# General formula: P(X < w) = (w - 980) / (1030 - 980) = (w - 980) / 50
# This is the CDF formula for uniform distribution

# Example: For w = 1010
w <- 1010
punif(w, min = 980, max = 1030)
```

```
## [1] 0.6
```

```r
# Interpretation: For w = 1010, P(X < 1010) = 0.6 or 60%

# The general answer is: P(X < w) = (w - 980) / 50 for 990 < w < 1030
# This represents a linear relationship where probability increases uniformly
```

Key Takeaways

Distribution Characteristics

1. **Normal Distribution**

- Symmetric, bell-shaped curve
    - Completely defined by mean and standard deviation
    - Use `pnorm()` for probabilities, `qnorm()` for percentiles
    - Approximately 68% of data within 1 SD, 95% within 2 SD

2. **Exponential Distribution**
    - Models time between events
    - Memoryless property
    - Parameter: rate = 1/mean
    - Right-skewed distribution

3. **Uniform Distribution**
    - All values equally likely in interval
    - Rectangular probability density
    - CDF is linear: P(X < x) = (x - a) / (b - a)

R Function Usage

- **pnorm(x, mean, sd)**: P(X <= x) for normal distribution
- **qnorm(p, mean, sd)**: Value x where P(X <= x) = p (inverse CDF)
- **pexp(x, rate)**: P(X <= x) for exponential distribution
- **punif(x, min, max)**: P(X <= x) for uniform distribution

Practical Applications

- **Healthcare**: Birth weight analysis for identifying at-risk newborns
- **Economics**: Salary distributions and workforce planning
- **Business**: Customer behavior modeling (ticket purchase timing)
- **Quality Control**: Manufacturing tolerance and acceptance criteria