R Coding Assignment 4: Probability Distributions Analysis

San Diego State University - STAT550

Aiden Jajo

Overview

This assignment explores multiple discrete probability distributions and their applications to real-world scenarios. We work with Poisson, Negative Binomial, Geometric, Hypergeometric, Exponential, and Multinomial distributions.

Key Distributions Covered: - **Poisson**: Events occurring at a constant average rate - **Negative Binomial**: Number of trials needed to achieve a fixed number of successes - **Geometric**: Number of failures before the first success - **Hypergeometric**: Sampling without replacement from finite populations - **Exponential**: Time between events in a Poisson process - **Multinomial**: Extension of binomial to more than two categories

---

Question 1: Traffic Accidents - Poisson Distribution

Scenario: Traffic accidents at a particular intersection follow Poisson distribution with an average rate of 1.4 per week.

```
# a) What is the probability that the next week is accident-free?
# P(X = 0) where X follows Poisson(lambda = 1.4)
ppois(0,
      lambda = 1.4)
```

```
## [1] 0.246597
```

```
# Interpretation: ~24.7% chance of zero accidents

# b) What is the probability that there will be exactly 3 accidents next week?
# P(X = 3) - using dpois for exact probability
dpois(3,
      lambda = 1.4)
```

```
## [1] 0.112777
```

```
# Interpretation: ~11.3% chance of exactly 3 accidents

# c) What is the probability that there will be at most 2 accidents next week?
# P(X <= 2) - cumulative probability
ppois(2,
      lambda = 1.4)
```

```
## [1] 0.8334977
```

```r
# Interpretation: ~83.3% chance of 2 or fewer accidents

# d) What is the probability that there will be at least 2 accidents during the
# next two weeks?
# For 2 weeks, lambda = 2 * 1.4 = 2.8
# P(X >= 2) = 1 - P(X <= 1)
1 - ppois(1,
          lambda = 2 * 1.4)
```

```
## [1] 0.7689218
```

```r
# Interpretation: ~76.9% chance of 2+ accidents in 2 weeks

# e) What is the probability that there will be exactly 5 accidents during the
# next four weeks?
# For 4 weeks, lambda = 4 * 1.4 = 5.6
dpois(5,
      lambda = 4 * 1.4)
```

```
## [1] 0.1697109
```

```r
# Interpretation: ~17.0% chance of exactly 5 accidents in 4 weeks

# f) What is the probability that there will be exactly 2 accidents tomorrow?
# Daily rate: lambda = 1.4 / 7 = 0.2 per day
dpois(2,
      lambda = 1.4 / 7)
```

```
## [1] 0.01637462
```

```r
# Interpretation: ~1.6% chance of 2 accidents in one day

# g) What is the probability that the next accident will not occur for three days?
# Exponential distribution for time between events
# Rate = 1.4/7 per day, so P(T > 3 days)
1 - pexp(3, rate = 1.4 / 7)
```

```
## [1] 0.5488116
```

```r
# Interpretation: ~54.9% chance of waiting more than 3 days
```

Question 2: Home Run Derby - Negative Binomial Distribution

Scenario: In Major League Baseball's Home Run Derby, each contestant is allowed to keep swinging the bat until they have made 10 "outs". (An "out" is anything that is not a home run.) If Barry Bonds has a 70% chance of hitting a home run on any given swing, what is the probability that he hits at least 10 home runs before his turn is up?

```r
# Negative binomial: number of successes before r failures
# We want P(X >= 10) where X = home runs before 10 outs
# P(X >= 10) = 1 - P(X <= 9)
1 - pnbinom(9,
            size = 10,     # number of failures (outs)
            prob = 0.7)    # probability of success (home run)
```

[1] 0.03255336

```r
# Interpretation: ~3.3% chance of hitting 10+ home runs before 10 outs
# This is relatively low despite the high success rate per swing
```

Question 3: Quarterback Passes - Geometric Distribution

Scenario: A quarterback completes 44% of his passes. We want to observe this quarterback during one game to see how many pass attempts he makes before completing one pass.

```r
# A) What is the probability that the quarterback throws 3 incomplete passes
# before he has a completion?
# Geometric distribution: number of failures before first success
# P(X = 3) where X follows Geometric(p = 0.44)
dgeom(3,
      prob = 0.44)
```

[1] 0.07727104

```r
# Interpretation: ~7.7% chance of 3 incomplete passes before first completion

# B) Determine the probability that it takes more than 5 attempts before he
# completes a pass.
# P(X > 5) = 1 - P(X <= 5)
1 - pgeom(5, prob = 0.44)
```

[1] 0.03084098

```r
# Interpretation: ~3.1% chance of more than 5 attempts needed
```

Question 4: Telemarketer Sales - Negative Binomial Distribution

Scenario: Suppose a telemarketer has a 15% chance of making a sale on any given phone call. He is required to make 10 successful sales before leaving for the day. What is the probability that he needs to make more than 40 calls?

```
# Negative binomial: trials needed to achieve r successes
# If more than 40 calls total are needed, that means more than 30 failures
# before getting 10 successes
# P(failures > 30) = 1 - P(failures <= 30)
1 - pnbinom(30,
            size = 10,      # required successes
            prob = 0.15)    # probability of success per call
```

```
## [1] 0.9327797
```

```
# Interpretation: ~93.3% chance of needing more than 40 calls
# With only 15% success rate, this is highly likely
```

Question 5: Microprocessor Quality Control - Hypergeometric Distribution

Scenario: A company supplies microprocessors to a manufacturer in batches of 50. The consumer regards a batch as acceptable provided that there are not more than 5 defective microprocessors in the batch. Rather than test all microprocessors, 10 are selected at random and tested.

```
# A) Find the probability that out of a sample of 10, d = 0, 1, 2, 3, 4, 5 are
# defective when there are actually 5 defective microprocessors in the batch.
# Hypergeometric: sampling without replacement
# Parameters: m = 5 defective, n = 45 good, k = 10 sampled
sapply(0:5,
       function(d) dhyper(d,
                          m = 5,     # defective in population
                          n = 45,    # good in population
                          k = 10))   # sample size
```

```
## [1] 0.3105627820 0.4313371972 0.2098397176 0.0441767826 0.0039645831
## [6] 0.0001189375
```

```
# Results show probability for each possible number of defectives in sample

# B) Suppose that the consumer will accept the batch provided that not more than
# m=2 defectives are found in the sample of 10.

# i) Find the probability that the batch is accepted when there are 5 defectives
# in the batch.
# P(X <= 2) when 5 defectives exist
sum(dhyper(0:2,
           m = 5,
           n = 45,
           k = 10))
```

```
## [1] 0.9517397
```

```
# Interpretation: ~95.2% chance batch is accepted despite having 5 defects
# This shows the sampling method might not catch all defective batches

# ii) Find the probability that the batch is rejected when there are 3 defectives
# in the batch.
# P(X > 2) when only 3 defectives exist = 1 - P(X <= 2)
1 - sum(dhyper(0:2,
               m = 3,
               n = 47,
               k = 10))
```

```
## [1] 0.006122449
```

```
# Interpretation: ~0.6% chance of rejecting a batch with only 3 defects
# This is a Type I error - rejecting a good batch
```

Question 6: Voter Preferences - Multinomial Distribution

Scenario: In a certain town, 40% of the eligible voters prefer candidate A, 10% prefer candidate B, and the remaining 50% have no preference. You randomly sample 10 eligible voters. What is the probability that 4 will prefer candidate A, 1 will prefer candidate B, and the remaining 5 will have no preference?

```
# Multinomial distribution: extension of binomial to k > 2 categories
# P(X1 = 4, X2 = 1, X3 = 5) where probabilities are (0.4, 0.1, 0.5)
dmultinom(c(4, 1, 5),          # observed counts
          prob = c(0.4, 0.1, 0.5))  # category probabilities
```

```
## [1] 0.1008
```

```
# Interpretation: ~10.1% chance of this exact outcome
# This matches the expected distribution fairly closely
```

Key Takeaways

Distribution Selection Guide

1. **Poisson**: Count of events in fixed time/space with constant average rate
2. **Negative Binomial**: Number of trials to achieve r successes
3. **Geometric**: Special case of negative binomial (r = 1)
4. **Hypergeometric**: Sampling without replacement from finite population
5. **Exponential**: Time between events in Poisson process
6. **Multinomial**: Categorical outcomes with k > 2 categories

R Functions Summary

- dpois(), ppois(): Poisson probabilities
- dnbinom(), pnbinom(): Negative binomial probabilities
- dgeom(), pgeom(): Geometric probabilities
- dhyper(): Hypergeometric probabilities
- pexp(): Exponential cumulative probabilities
- dmultinom(): Multinomial probabilities

Practical Applications

- Traffic safety and infrastructure planning
- Sports analytics and performance modeling
- Quality control and acceptance sampling
- Sales forecasting and workforce planning
- Election polling and preference analysis