

STA 3032 Lab 3: Exploring Distributions in R

Ansel Gillman

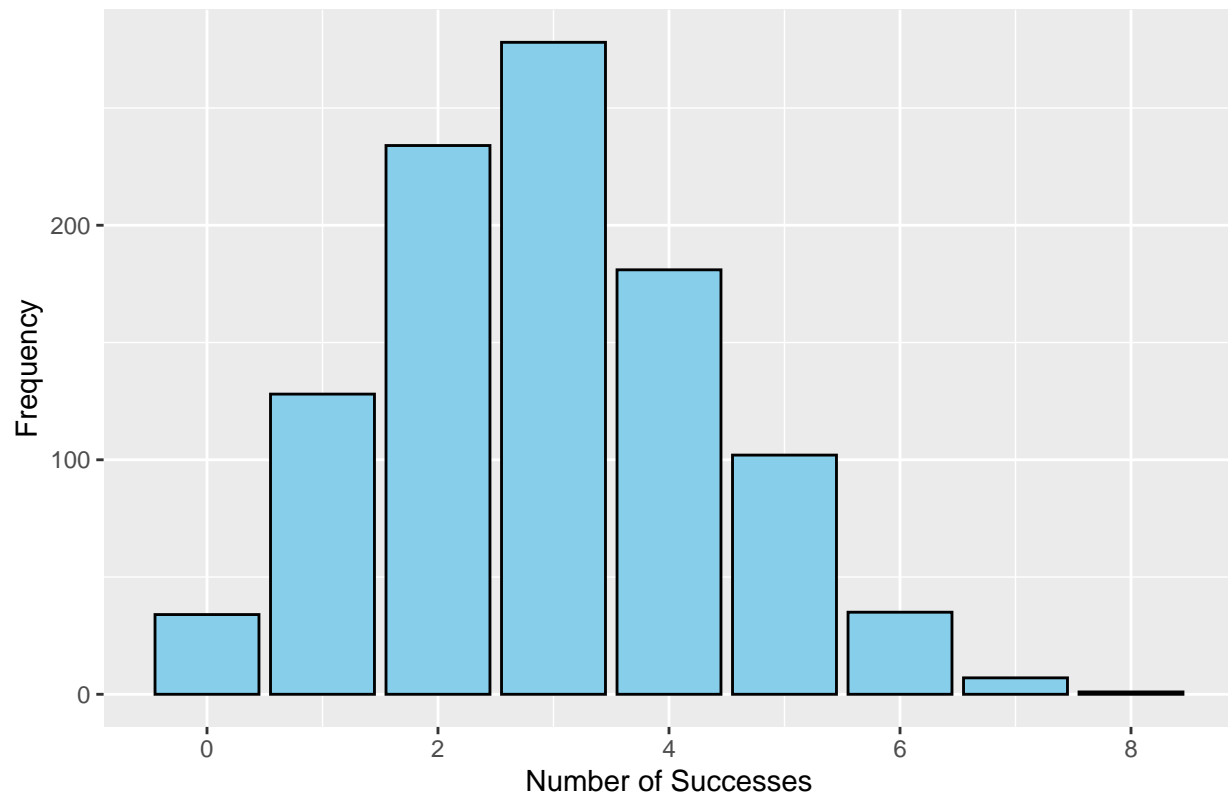
2025-10-05

```
library(tidyverse)
library(knitr)

# -----
# 1. Binomial Distribution
# -----
binom_data <- tibble(x = rbinom(1000, size = 10, prob = 0.3))
binom_mean <- mean(binom_data$x)
binom_var <- var(binom_data$x)

ggplot(binom_data, aes(x = x)) +
  geom_bar(fill = "skyblue", color = "black") +
  labs(title = "Binomial(10, 0.3) Distribution",
       x = "Number of Successes", y = "Frequency")
```

Binomial(10, 0.3) Distribution

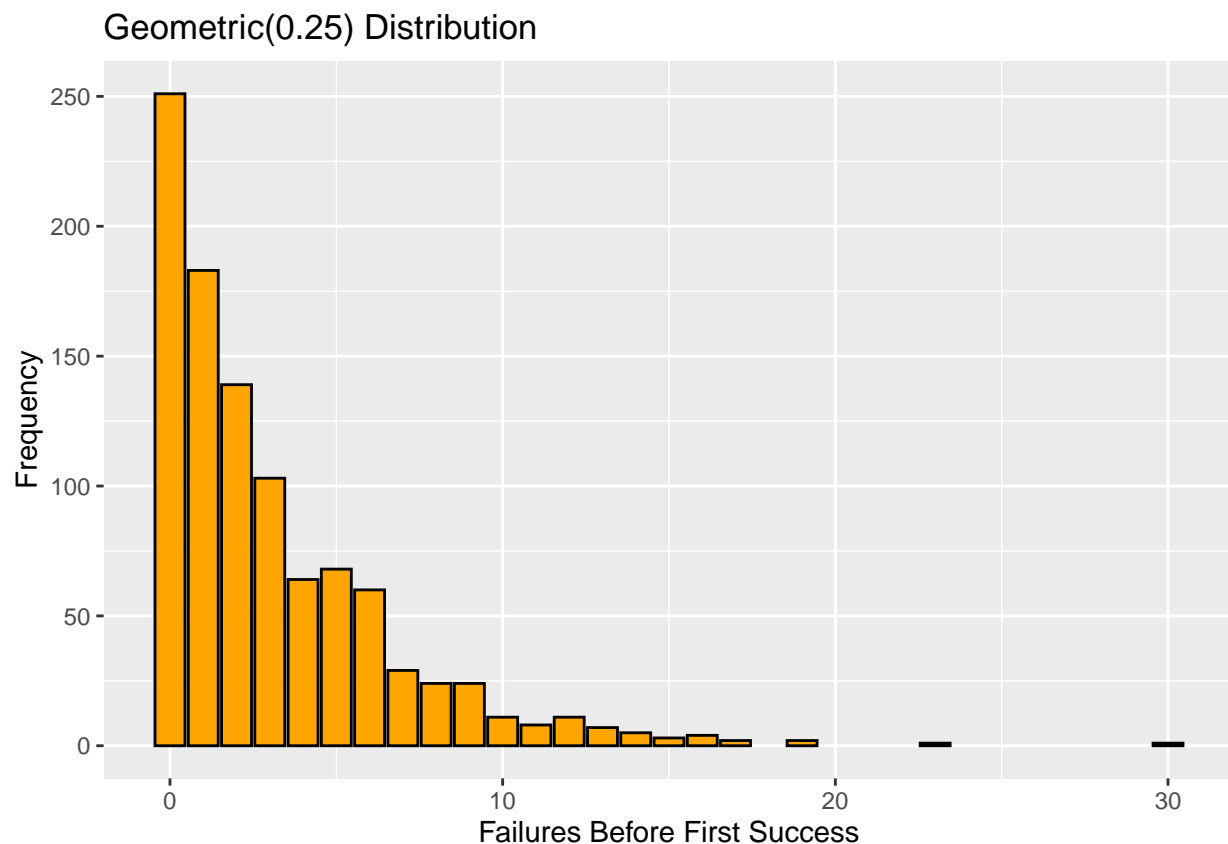


```
cat("
Binomial Discussion:
- Models the number of successes in a fixed number of independent trials (n=10) with probability p=0.3.
- The distribution is roughly symmetric since p is moderate.
- Simulated mean and variance (", round(binom_mean, 2), ", ", round(binom_var, 2), ") are close to theoretical values (3, 2.1).")
```

```
##
## Binomial Discussion:
## - Models the number of successes in a fixed number of independent trials (n=10) with probability p=0.3.
## - The distribution is roughly symmetric since p is moderate.
## - Simulated mean and variance ( 2.93 , 2.09 ) are close to theoretical values (3, 2.1).
```

```
# -----
# 2. Geometric Distribution
# -----
geom_data <- tibble(x = rgeom(1000, prob = 0.25))
geom_mean <- mean(geom_data$x)
geom_var <- var(geom_data$x)

ggplot(geom_data, aes(x = x)) +
  geom_bar(fill = "orange", color = "black") +
  labs(title = "Geometric(0.25) Distribution",
       x = "Failures Before First Success", y = "Frequency")
```



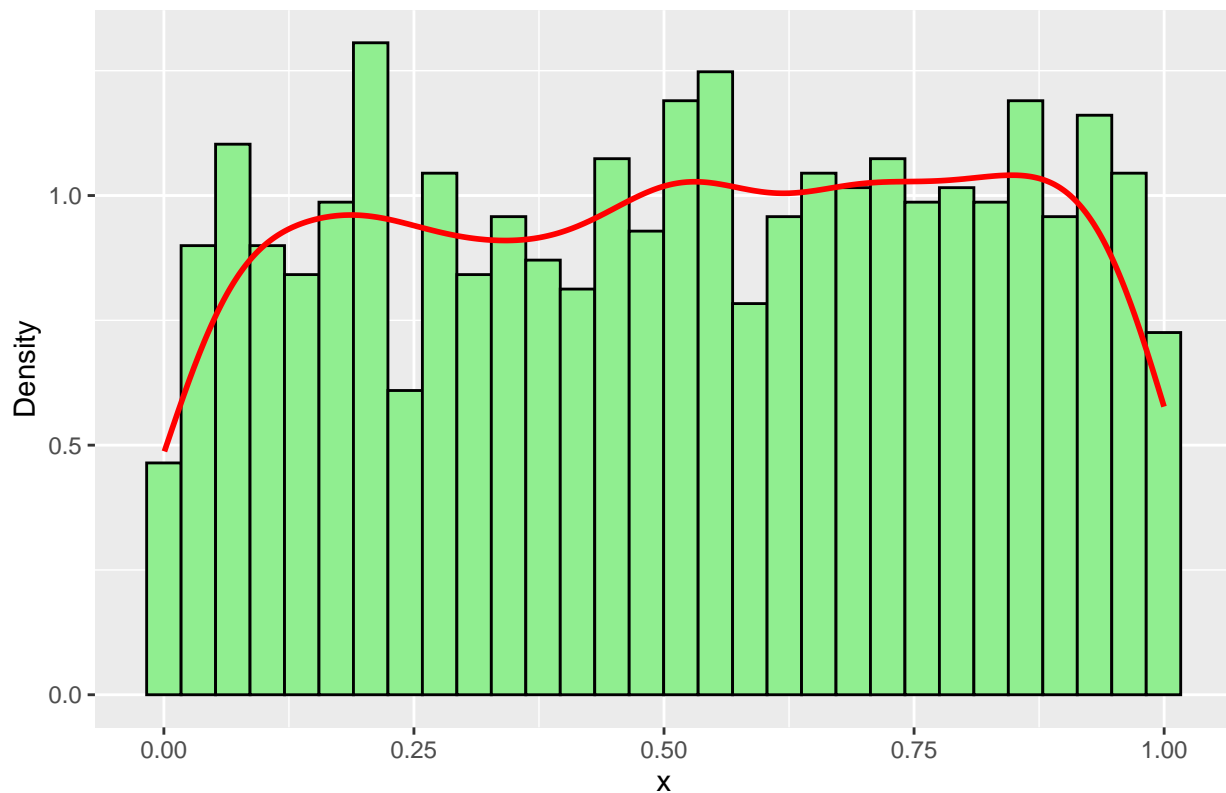
```
cat("
Geometric Discussion:
- Models the number of failures before the first success; right-skewed.
- Theoretical mean = 3, variance = 12.
- Simulated mean and variance (", round(geom_mean, 2), ", ", round(geom_var, 2), ") are close to theoretical.
")
```

```
##
## Geometric Discussion:
## - Models the number of failures before the first success; right-skewed.
## - Theoretical mean = 3, variance = 12.
## - Simulated mean and variance ( 3.06 , 12.33 ) are close to theoretical.
```

```
# -----
# 3. Uniform Distribution
# -----
uniform_data <- tibble(x = runif(1000))
uniform_mean <- mean(uniform_data$x)
uniform_var <- var(uniform_data$x)

ggplot(uniform_data, aes(x = x)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "lightgreen", color = "black") +
  geom_density(color = "red", size = 1) +
  labs(title = "Uniform(0,1) Distribution", x = "x", y = "Density")
```

Uniform(0,1) Distribution



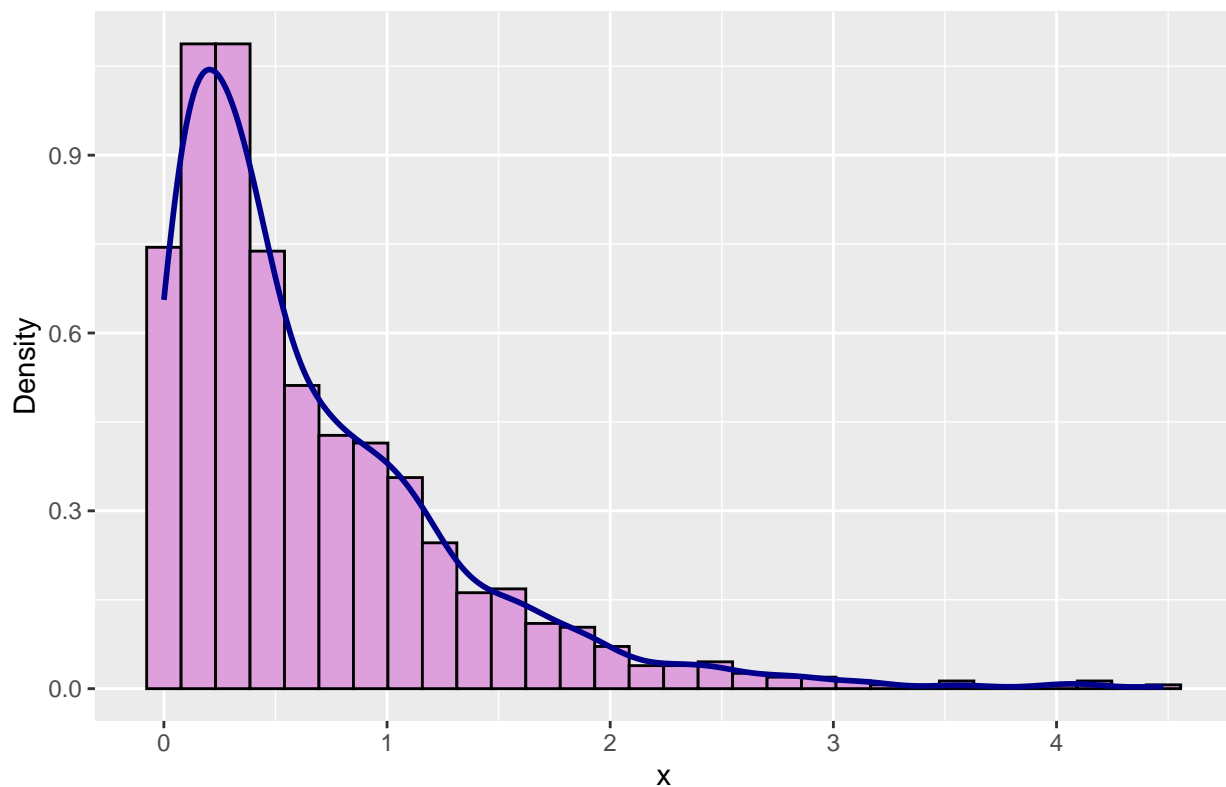
```
cat("
Uniform Discussion:
- Continuous distribution; all values equally likely between 0 and 1.
- Flat and symmetric shape.
- Simulated mean and variance (", round(uniform_mean, 2), ", ", round(uniform_var, 4), ") match theoretical
")
```

```
##
## Uniform Discussion:
## - Continuous distribution; all values equally likely between 0 and 1.
## - Flat and symmetric shape.
## - Simulated mean and variance ( 0.52 , 0.0853 ) match theoretical (0.5, 0.0833).
```

```
# -----
# 4. Exponential Distribution
# -----
exp_data <- tibble(x = rexp(1000, rate = 1.5))
exp_mean <- mean(exp_data$x)
exp_var <- var(exp_data$x)

ggplot(exp_data, aes(x = x)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "plum", color = "black") +
  geom_density(color = "darkblue", size = 1) +
  labs(title = expression(paste("Exponential(", lambda, " = 1.5) Distribution")),
       x = "x", y = "Density")
```

Exponential($\lambda = 1.5$) Distribution



```

cat("
Exponential Discussion:
- Continuous distribution modeling waiting times between events; right-skewed like geometric.
- Theoretical mean = 0.667, variance = 0.444.
- Simulated mean and variance (", round(exp_mean, 3), ", ", round(exp_var, 3), ") are close to theoretical.
")

##
## Exponential Discussion:
## - Continuous distribution modeling waiting times between events; right-skewed like geometric.
## - Theoretical mean = 0.667, variance = 0.444.
## - Simulated mean and variance ( 0.658 , 0.42 ) are close to theoretical.

# -----
# 5. Summary Table
# -----
summary_table <- tibble(
  Distribution = c("Binomial(10,0.3)", "Geometric(0.25)", "Uniform(0,1)", "Exponential(1.5)"),
  Simulated_Mean = c(binom_mean, geom_mean, uniform_mean, exp_mean),
  Simulated_Variance = c(binom_var, geom_var, uniform_var, exp_var),
  Theoretical_Mean = c(3, 3, 0.5, 1/1.5),
  Theoretical_Variance = c(2.1, 12, 1/12, 1/1.5^2)
)

kable(summary_table, digits = 4, caption = "Simulated vs Theoretical Summary")

```

Table 1: Simulated vs Theoretical Summary

Distribution	Simulated_Mean	Simulated_Variance	Theoretical_Mean	Theoretical_Variance
Binomial(10,0.3)	2.9310	2.0903	3.0000	2.1000
Geometric(0.25)	3.0620	12.3305	3.0000	12.0000
Uniform(0,1)	0.5152	0.0853	0.5000	0.0833
Exponential(1.5)	0.6583	0.4203	0.6667	0.4444