# Task B

Zixia Zeng

2024-11-05

## B.1

### (1)

According to the background of B.1, the probability density function $p_\lambda(x)$ of a random variable $X$ is:

$$p_\lambda(x) = \begin{cases} ae^{-\lambda(x-b)} & \text{if } x \geq b, \\ 0 & \text{if } x < b, \end{cases}$$

where: $b > 0$ is a known constant, $\lambda > 0$ is a parameter of the distribution, \$ a \$ is a constant to be determined in terms of $\lambda$ and $b$.

According to the definition of probability density function, $p_\lambda(x)$ must integrate to 1 over its domain, so an equation can be written:

$$\int_{-\infty}^{\infty} p_\lambda(x)\, dx = 1$$

When $x < b$, $p_\lambda(x) = 0$, so we only need to calculate the integral from $x = b$ to $x = \infty$. To set up the integral, the equation can be written:

$$\int_{b}^{\infty} ae^{-\lambda(x-b)}\, dx = 1$$

$a$ is a constant number, so it can be factored out and just calculate the remaining part:

$$a\int_{b}^{\infty} e^{-\lambda(x-b)}\, dx = 1$$

Let $\mu = x - b$, so domain changes to $\{0, \infty\}$ and the euqation should be transformed:

$$a\int_{0}^{\infty} e^{-\lambda\mu}\, d\mu = 1$$

Solve the integral:

$$\begin{aligned} a\int_0^\infty e^{-\lambda\mu} d\mu &= -\frac{1}{\lambda}\cdot e^{-\lambda\mu}|_0^\infty \\ &= a\cdot[\lim_{u\to\infty}(-\frac{1}{\lambda}\cdot e^{-\lambda\mu}) - \lim_{u\to 0}(-\frac{1}{\lambda}\cdot e^{-\lambda\mu})] \\ &= a\cdot[0-(-\frac{1}{\lambda})] \\ &= a\cdot\frac{1}{\lambda} = 1 \end{aligned}$$

Therefore, it is obvious that $a = \lambda$

The answer of question(1) is $a = \lambda$

**(2)**

**Mean**:

The equation of mean of $X$ is:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot p_\lambda(x) dx$$

Since $p_\lambda(x) = 0$, when $b < 0$, so the integral can be simplified to:

$$\mathbb{E}[X] = \int_{b}^{\infty} x \cdot p_\lambda(x) dx$$

Then transform $x$ to make the integral easier to solve. Let: $\mu = x - b \Rightarrow x = u + b$, then: $dx = d\mu$ and when $x = b, \mu = 0$, when $x \to \infty, \mu \to \infty$.

Substituting, and can get:

$$\mathbb{E}[X] = \int_{0}^{\infty} (\mu + b) \cdot \lambda e^{-\lambda u} d\mu$$

Expand $(u + b)$ and calculate two integrals separately:

$$\mathbb{E}[X] = \int_{0}^{\infty} \lambda u e^{-\lambda u} du + \int_{0}^{\infty} \lambda b e^{-\lambda u} du.$$

For the first integral, use intergation by parts:

$$\begin{aligned}
\int_{0}^{\infty} \lambda u e^{-\lambda u} du &= -\mu^2 e^{-\lambda \mu}|_0^\infty + \int_{0}^{\infty} e^{-\lambda u} du \\
&= \lim_{u \to \infty} (-u e^{-\lambda u}) - \lim_{u \to 0} (-u e^{-\lambda u}) + (-\frac{1}{\lambda} e^{-\lambda \mu})|_0^\infty \\
&= 0 + \lim_{u \to \infty} (-\frac{1}{\lambda} e^{-\lambda \mu}) - \lim_{u \to 0} (-\frac{1}{\lambda} e^{-\lambda \mu}) \\
&= \frac{1}{\lambda}
\end{aligned}$$

For the second integral, this is similar to $\int_0^\infty e^{-\lambda u} du$, so:

$$\int_{0}^{\infty} \lambda b e^{-\lambda u} du = \lambda b \int_{0}^{\infty} e^{-\lambda u} du = \lambda b \cdot \frac{1}{\lambda} = b$$

Combine the first integral and the second integral, the mean is:

$$\mathbb{E}(X) = \frac{1}{\lambda} + b$$

**Standard Deviation**:

To calculate the standard deviation, calculate the variance first. The variance of $X$ is:

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$\mathbb{E}(X) = \frac{1}{\lambda} + b$ is already known, so this time just calculate $\mathbb{E}(X^2)$.

First, write the equation(already know that when $x < b$, $p_\lambda(x) = 0$):

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 p_\lambda(x) dx = \int_{b}^{\infty} \lambda x^2 e^{-\lambda(x-b)} dx$$

Let $\mu = x - b$, then re-write the equation(the transformation steps are the same as the previous proof):

$$\mathbb{E}(X^2) = \lambda \int_{0}^{\infty} (\mu + b)^2 e^{-\lambda \mu} d\mu$$

By expanding the $(\mu + b)^2$, the equation can be separated into three integrals:

$$\mathbb{E}(X^2) = \lambda \int_0^\infty u^2 e^{-\lambda u} du + 2b\lambda \int_0^\infty u e^{-\lambda u} du + b^2 \lambda \int_0^\infty e^{-\lambda u} du.$$

The second integral: $\lambda \int_0^\infty u e^{-\lambda u} du = \frac{1}{\lambda}$ and the third integral:$\lambda \int_0^\infty e^{-\lambda u} du = 1$ are solved in previous proof, so we can substitute these conclusions later.

Now we can only focus on the first integral: $\lambda \int_0^\infty u^2 e^{-\lambda u} du$.

To solve:

$$\lambda \int_0^\infty u^2 e^{-\lambda u} du$$

we can solve this integral by parts:

$$\begin{aligned}
\lambda \int_0^\infty u^2 e^{-\lambda u} du &= -\int_0^\infty u^2 de^{-\lambda u} \\
&= -(u^2 e^{-\lambda u}|_0^\infty - \int_0^\infty e^{-\lambda u} du^2) \\
&= -(\lim_{u\to\infty} u^2 e^{-\lambda u} - \lim_{u\to 0} u^2 e^{-\lambda u} - 2\int_0^\infty u e^{-\lambda u} du)
\end{aligned}$$

Since we have calculated $\mathbb{E}(x)$ before, so $\int_0^\infty u e^{-\lambda u} du) = \frac{1}{\lambda^2}$.

Substituting:

$$\begin{aligned}
\lambda \int_0^\infty u^2 e^{-\lambda u} du &= -(\lim_{u\to\infty} u^2 e^{-\lambda u} - \lim_{u\to 0} u^2 e^{-\lambda u} - 2\int_0^\infty u e^{-\lambda u} du) \\
&= -(0 - 0 - 2 \cdot \frac{1}{\lambda^2}) \\
&= \frac{2}{\lambda^2}
\end{aligned}$$

To evaluate this expression, let's break it down step by step. The expression given is:

$$-\left(u^2 e^{-\lambda u}\Big|_0^\infty\right) - \int_0^\infty e^{-\lambda u}\, du^2$$

Let's simplify each part.

1. **Evaluating** $u^2 e^{-\lambda u}\Big|_0^\infty$:

   $$\lim_{u\to\infty} u^2 e^{-\lambda u} - \lim_{u\to 0} u^2 e^{-\lambda u}$$

   As $u \to \infty$, $u^2 e^{-\lambda u} \to 0$ (since $e^{-\lambda u}$ decays to zero faster than $u^2$ grows). At $u = 0$, $u^2 e^{-\lambda u} = 0$.

   So, $u^2 e^{-\lambda u}\Big|_0^\infty = 0 - 0 = 0$.

Substitute these results back, we can get:

$$\mathbb{E}(X^2) = \frac{2}{\lambda^2} + 2b \cdot \frac{1}{\lambda} + b^2$$

Now we can calculate the variance. Substituting the results:

$$\begin{aligned} Var(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\ &= \frac{2}{\lambda^2} + 2b \cdot \frac{1}{\lambda} + b^2 - \left(\frac{1}{\lambda} + b\right)^2 \\ &= \frac{2}{\lambda^2} + 2b \cdot \frac{1}{\lambda} + b^2 - \left(\frac{1}{\lambda^2} + \frac{2b}{\lambda} + b^2\right) \\ &= \frac{1}{\lambda^2} \end{aligned}$$

The standard deviation is: $\sigma_X = \sqrt{\mathrm{Var}(X)} = \frac{1}{\lambda}$

In conclusion, the results are: **Mean**: $\mathbb{E}[X] = \frac{1}{\lambda} + b$.

**Standard Deviation**: $\sigma_X = \sqrt{\mathrm{Var}(X)} = \frac{1}{\lambda}$.

**(3)**

**cumulative distribution function(CDF)**:

Assume CDF of X is:

$$F_\lambda(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p_\lambda(t)dt$$

Case $x < b$: $P(X \leq x) = 0$ because the probability density function $p_\lambda(x) = 0$.

Therefore, we can get:

$$F_\lambda(x) = 0, \quad when\ x < b.$$

Case $x \geq b$: we need to integrate $p_\lambda(t)$ from $t = b$ to $t = x$:

$$F_\lambda(x) = \int_b^x \lambda e^{-\lambda(t-b)} dt.$$

Let $u = t - b$, so $t = u + b$ and $dt = du$ so, When $t = b$, $u = 0$, When $t = x$, $u = x - b$.

Substituting into the integral, we get:

$$F_\lambda(x) = \lambda \int_0^{x-b} e^{-\lambda u}\, du.$$

Now the integral can be solved:

$$\begin{aligned} F_\lambda(x) &= \lambda \cdot \left(-\frac{1}{\lambda} e^{-\lambda\mu}\right) |_0^{x-b} \\ &= \lambda \cdot \left(-\frac{1}{\lambda} e^{-\lambda(x-b)} + \frac{1}{\lambda} e^0\right) \\ &= 1 - e^{-\lambda(x-b)} \end{aligned}$$

Combining the two cases, the cumulative distribution function is:

$$F_\lambda(x) = \begin{cases} 0 & \text{if } x < b, \\ 1 - e^{-\lambda(x-b)} & \text{if } x \geq b. \end{cases}$$

**Quantile Function**:

Quantile function is the inverse of cumulative distribution function(CDF), so for $x \geq b$, the CDF is:

$$F_\lambda(x) = 1 - e^{-\lambda(x-b)}$$

Set $F_{\lambda(x)} = p$ and find the relationship for $x$ to $p$:

$$\begin{aligned} e^{-\lambda(x-b)} &= 1-p \\ -\lambda(x-b) &= \ln(1-p) \\ x &= b - \frac{1}{\lambda} \cdot \ln(1-p) \end{aligned}$$

Therefore, the quantile function of X is:

$$x = b - \frac{1}{\lambda} \cdot \ln(1-p),\ for\ p \in [0,1)$$

In conlusion:

**Cumulative distribution function**: $F_\lambda(x) = \begin{cases} 0 & \text{if } x < b, \\ 1 - e^{-\lambda(x-b)} & \text{if } x \geq b. \end{cases}$

**Quantile function**:$x = b - \frac{1}{\lambda} \cdot \ln(1-p),\ for\ p \in [0,1)$

**(4)**

The likelihood function is defined as:

$$L(\lambda) = \prod_{i=1}^{n} p_\lambda(X_i)$$

Since $\because p_\lambda(X_i) = 0$, when$X_i < b$, so for all $x < b$ the likelihood is zero. Thus, for $X \geq b$:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda(X_i - b)}$$

Expanding the product, we can get:

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n}(X_i - b)}$$

Then transfer it to log-likelihood function:

$$\ell(\lambda) = \ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n}(X_i - b)$$

To find the maximum likelihood estimator for $\lambda$, we need to differentiate the log-likelihood function and set it equals to zero.

$$\frac{d\ell(\lambda)}{d\lambda} = 0$$

Substituting the equation:

$$\frac{d(n \ln(\lambda) - \lambda \sum_{i=1}^{n}(X_i - b))}{d\lambda} = 0n \cdot \frac{1}{\lambda} - \sum_{i=1}^{n}(X_i - b) = 0\lambda = \frac{n}{\sum_{i=1}^{n}(X_i - b)}$$

The maximum likelihood estimate (MLE) for $\lambda$ is:

$$\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n}(X_i - b)}.$$

**(5)**

Firstly, load the packages for data wrangling and visualization and the data set.

```r
# Load packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(ggplot2)

# Load data set
market_df = read.csv("supermarket_data_2024(1).csv")
```

Then calculate maximum likelihood estimator. According to (4), the equation of MLE of $\lambda$ is $\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n}(X_i - b)}$

```r
# Instant b is given
b = 300

# Number of observations
time_lengths = market_df$TimeLength
n = length(time_lengths)

# Calculate the MLE
lambda_mle = n / sum(time_lengths - b)

# Print the MLE for lambda
cat("The MLE for lambda is:", lambda_mle, "\n")
```

```
## The MLE for lambda is: 0.01988426
```

**(6)**

```r
# Use boot package to compute the 95% level confidence interval
# Load package
library(boot)

# Set random seed
set.seed(123)

# Define a function which computes the lambda of a chosen column
compute_lambda_MLE = function(df,indicies,col_name){
  # Extract sub-sample
  sub_sample = slice(df,indicies) %>%
```

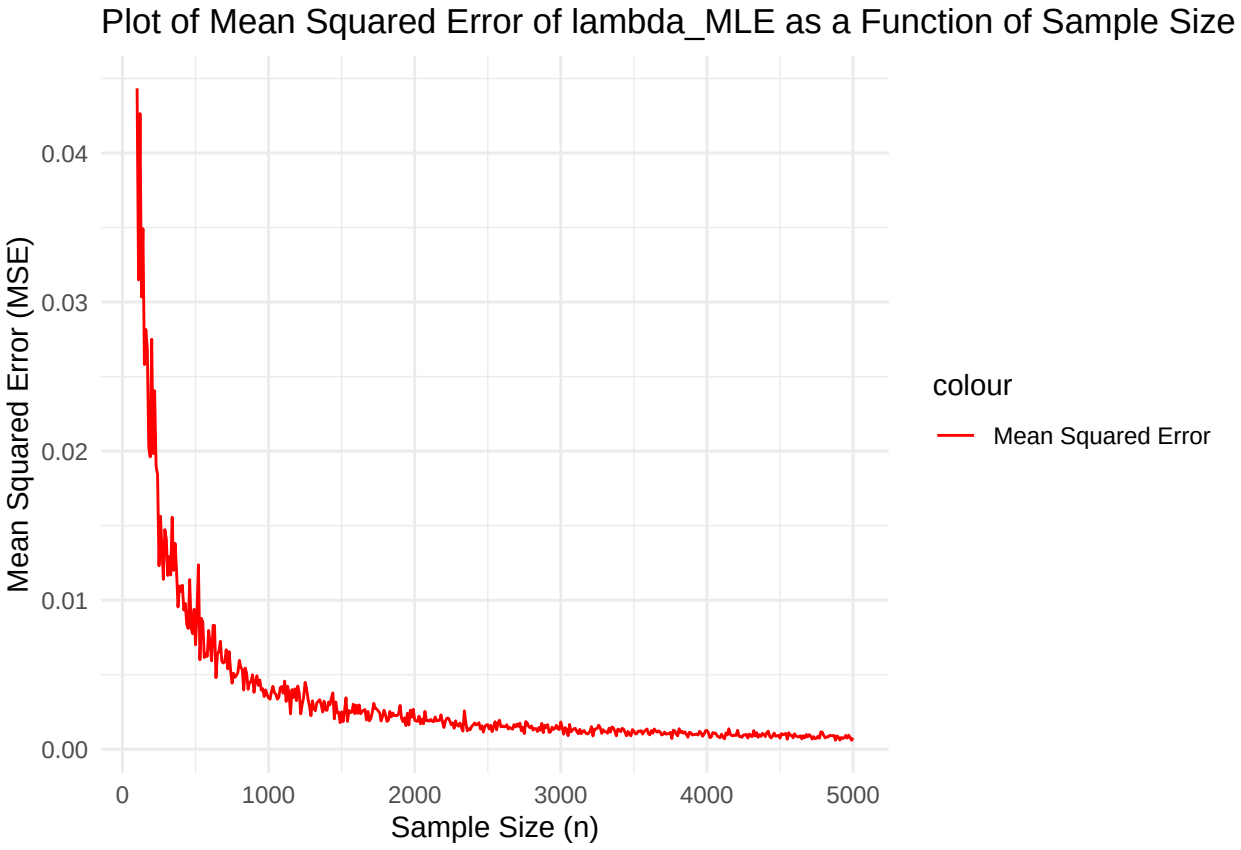
```
    pull(all_of(col_name))
  # Return lambda_MLE
  return(length(sub_sample)/sum(sub_sample-b))
}

# Use boot function to generate the bootstrap statistics
res = boot(data = market_df, statistic = compute_lambda_MLE, col_name = "TimeLength", R = 10000)

# Compute the 95% level confidence interval for lambda
boot.ci(boot.out = res, type = "basic", conf = 0.95)
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = res, conf = 0.95, type = "basic")
##
## Intervals :
## Level      Basic
## 95%   ( 0.0191,  0.0207 )
## Calculations and Intervals on Original Scale
```

(7)

```
# Set given true parameters
true_lambda = 2
b = 0.01

# Define the range of sample sizes (from 100 to 5000 in increments of 10)
sample_sizes = seq(100, 5000, by = 10)
num_trials = 100

# Initialize a variable to store the mean squared errors for each sample
mse_values = numeric(length(sample_sizes))

# Function to calculate the MLE of lambda
calculate_lambda_mle = function(sample, b) {
  n = length(sample)
  n / sum(sample - b)
}

# Start the simulation study
# Set random seed
set.seed(321)
for (j in seq_along(sample_sizes)) {
  # Set sample size for this trail
  n = sample_sizes[j]

  # Store the MLEs for the current sample size across trials
  mle_estimates = numeric(num_trials)
```

```r
  for (i in 1:num_trials) {
    # Generate a sample from exponential distribution with rate = true_lambda
    sample = b + rexp(n, rate = true_lambda)

    # Compute the MLE for the current sample
    mle_estimates[i] = calculate_lambda_mle(sample, b)
  }

  # Calculate the mean squared error for current sample size
  mse_values[j] = mean((mle_estimates - true_lambda)^2)
}

# Create a data frame and ready for plot
mse_data = data.frame(
  SampleSize = sample_sizes,
  MSE = mse_values
)

# Plot the Mean Squared Error as a function of sample size using ggplot2
ggplot(mse_data, aes(x = SampleSize, y = MSE,color = "Mean Squared Error" )) +
  geom_line() +
  labs(
    title = "Plot of Mean Squared Error of lambda_MLE as a Function of Sample Size",
    x = "Sample Size (n)",
    y = "Mean Squared Error (MSE)",
    color = "colour"
  ) +
  scale_color_manual(values = c("Mean Squared Error" = "red")) +
  theme_minimal()
```




## B.2

### (1)

Since $X$ equals the number of red balls minus the number of blue balls, there are three possible outcomes:

$X = 2$, when both balls drawn are red balls.

$X = 0$, when one red ball and one blue ball are drawn.

$X = -2$, when both balls drawn are blue balls.

The probability of $X = 2$ is:

$$P(X = 2) = \frac{\binom{a}{2}}{\binom{a+b}{2}}$$

Where $\binom{a}{2}$ means the number of ways to choose two red balls from $a$ red balls, and $\binom{a+b}{2}$ means the number of ways to choose 2 balls from $a + b$ balls.

The probability of $X = 0$ is:

$$P(X = 0) = \frac{\binom{a}{1}\binom{b}{1}}{\binom{a+b}{2}}$$

Where $\binom{a}{1}$ means choose one red balls from $a$ red balls, and $\binom{b}{1}$ means choose one blue balls from $b$ blue balls.(Do not need to consider the sequence)

The probability of $X = -2$ is:

$$P(X = -2) = \frac{\binom{b}{2}}{\binom{a+b}{2}}$$

Where $\binom{b}{2}$ means choose two red balls from $a$ red balls.

Therefore, the probability mass function $p_X(x)$ is:

$$p_X(x) = \begin{cases} \frac{\binom{a}{2}}{\binom{a+b}{2}}, & \text{if } x = 2, \\ \frac{\binom{a}{1}\binom{b}{1}}{\binom{a+b}{2}}, & \text{if } x = 0, \\ \frac{\binom{b}{2}}{\binom{a+b}{2}}, & \text{if } x = -2, \\ 0, & \text{otherwise.} \end{cases}$$

**(2)**

The formula of expectation $\mathbb{E}(x)$ is:

$$\mathbb{E}(X) = 2 \cdot P(X = 2) + 0 \cdot P(X = 0) + (-2) \cdot P(X = -2)$$

Substituting the probabilities we found:

$$\mathbb{E}(X) = 2 \cdot \frac{\binom{a}{2}}{\binom{a+b}{2}} + 0 \cdot \frac{\binom{a}{1}\binom{b}{1}}{\binom{a+b}{2}} + (-2) \cdot \frac{\binom{b}{2}}{\binom{a+b}{2}}$$

Simplifying:

$$\mathbb{E}(X) = 2 \cdot \frac{\frac{a(a-1)}{2}}{\frac{(a+b)(a+b-1)}{2}} - 2 \cdot \frac{\frac{b(b-1)}{2}}{\frac{(a+b)(a+b-1)}{2}}$$

The expression of the expectation of $X$ is:

$$\mathbb{E}(X) = \frac{2(a(a-1) - b(b-1))}{(a+b)(a+b-1)}$$

**(3)**

To compute the variance, first we need to compute $E(X^2)$.

$$\mathbb{E}(X^2) = 2^2 \cdot P(X = 2) + 0^2 \cdot P(X = 0) + (-2)^2 \cdot P(X = -2)$$

Substituting $P(X)$ respectively:

$$\mathbb{E}(X^2) = 4 \cdot \frac{\binom{a}{2}}{\binom{a+b}{2}} + 0 \cdot \frac{\binom{a}{1}\binom{b}{1}}{\binom{a+b}{2}} + 4 \cdot \frac{\binom{b}{2}}{\binom{a+b}{2}}$$

$$\mathbb{E}(X^2) = 4 \cdot \frac{\frac{a(a-1)}{2}}{\frac{(a+b)(a+b-1)}{2}} + 4 \cdot \frac{\frac{b(b-1)}{2}}{\frac{(a+b)(a+b-1)}{2}}$$

Now $E(X^2)$ is:

$$\mathbb{E}(X^2) = \frac{4(a(a-1)+b(b-1))}{(a+b)(a+b-1)}$$

Therefore, the variance is:

$$Var(X) = E(X^2) - (E(X))^2$$

Substituting the expressions of $E(X^2)$ and $E(X)$ to calculate the variance.

$$Var(X) = \frac{4(a(a-1)+b(b-1))}{(a+b)(a+b-1)} - \left(\frac{2(a(a-1)-b(b-1))}{(a+b)(a+b-1)}\right)^2$$

**(4)**

```
# Function to compute the expectation E(X)
compute_expectation_X = function(a, b) {
  # Expectation E(X)
  expectation = (2 * (a * (a - 1) - b * (b - 1))) / ((a + b) * (a + b - 1))
  return(expectation)
}

# Function to compute the variance Var(X)
compute_variance_X = function(a, b) {
  # Compute E(X^2)
  E_X2 = 4 * (a*(a-1)+b*(b-1)) / ((a+b)*(a+b-1))

  # Compute E(X)
  E_X = compute_expectation_X(a, b)

  # Variance Var(X) = E(X^2) - (E(X))^2
  variance = E_X2 - (E_X)^2
  return(variance)
}
```

**(5)**

Since $\overline{X}$ is defined as the sample mean of $X$:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Since $X_1, X_2, \ldots, X_n$ are i.i.d. random variables and they all have the same expectation, so we have $\mathbb{E}(X_i) = \mathbb{E}(X)$ for all $i$. Therefore, the expectation of the sample mean can be written as:

$$\mathbb{E}(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X)$$

Simplify:

$$\mathbb{E}(\overline{X}) = \frac{1}{n}\cdot n\cdot\mathbb{E}(X) = \mathbb{E}(X)$$

Therefore, the expectation of $\overline{X}$ is:

$$\mathbb{E}(\overline{X}) = \mathbb{E}(X) = \frac{2(a(a-1)-b(b-1))}{(a+b)(a+b-1)}$$

### (6) For the variance of the sample mean, the formula is:

$$Var(\overline{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

Since $X_1, X_2, \ldots, X_n$ are i.i.d., each $X_i$ has the same variance as $X$, so we have:

$$\sum_{i=1}^{n} Var(X_i) = n \cdot Var(X)$$

Therefore, $Var(\overline{X})$ is:

$$Var(\overline{X}) = \frac{1}{n^2} \cdot n \cdot Var(X) = \frac{1}{n} \cdot Var(X)$$

$$Var(\overline{X}) = \frac{1}{n}\left(\frac{4(a(a-1)+b(b-1))}{(a+b)(a+b-1)} - \left(\frac{2(a(a-1)-b(b-1))}{(a+b)(a+b-1)}\right)^2\right)$$

### (7)

```r
# Function to generate a sample of X1, X2, ..., Xn of independent copies of X
sample_Xs = function(a, b, n) {
  # Define the probability of 2 red balls
  p_2r = (a * (a - 1)) / ((a + b) * (a + b - 1))

  # Define the probability of 1 red ball and 1 blue ball
  p_1r1b = (2 * a * b) / ((a + b) * (a + b - 1))

  # Define the probability of 2 blue balls
  p_2b = (b * (b - 1)) / ((a + b) * (a + b - 1))

  # Generate the sample of n independent copies of X
  sample <- sample(c(2, 0, -2), size = n, replace = TRUE, prob = c(p_2r, p_1r1b, p_2b))

  return(sample)
}
```

### (8)

```r
# Set default parameters
a = 3
b = 5
n = 100000

# Compute numerical value of E(X) and Var(X)
E_X <- compute_expectation_X(a, b)
Var_X <- compute_variance_X(a, b)

# Generate a sample X1, X2, ..., Xn of X
# Set random seed
set.seed(228)
sample_X <- sample_Xs(a, b, n)

# Compute the sample mean and sample variance
```

```r
sample_mean <- mean(sample_X)
sample_variance <- var(sample_X)

# Print the results
cat("Population E(X):", E_X, "\n")
```

```
## Population E(X): -0.5
```

```r
cat("Population Var(X):", Var_X, "\n")
```

```
## Population Var(X): 1.607143
```

```r
cat("Sample Mean of X:", sample_mean, "\n")
```

```
## Sample Mean of X: -0.50926
```

```r
cat("Sample Variance of X:", sample_variance, "\n")
```

```
## Sample Variance of X: 1.60663
```

```r
# Calculate the difference
cat("The difference between sample mean to E(X):",abs(E_X - sample_mean),"\n")
```

```
## The difference between sample mean to E(X): 0.00926
```

```r
cat("The difference between sample variance to Var(X):",abs(Var_X - sample_variance),"\n")
```

```
## The difference between sample variance to Var(X): 0.0005125384
```

**Explanation**

According to the result, the sample mean is very close to the population mean($\mathbb{E}(X)$) and the sample variance is very close to the population variance($Var(X)$).

For sample mean and population mean, according to the Central Limit Theorem, as the sample size increases, the sample mean becomes closer to the population mean. If the sample size becomes larger than 100000, the differences between sample mean and population mean will be even smaller.

For sample variance and population variance, according to the Central Limit Theorem, as the sample size increases, the sample variance will coverage to the population variance, and become more accurate.

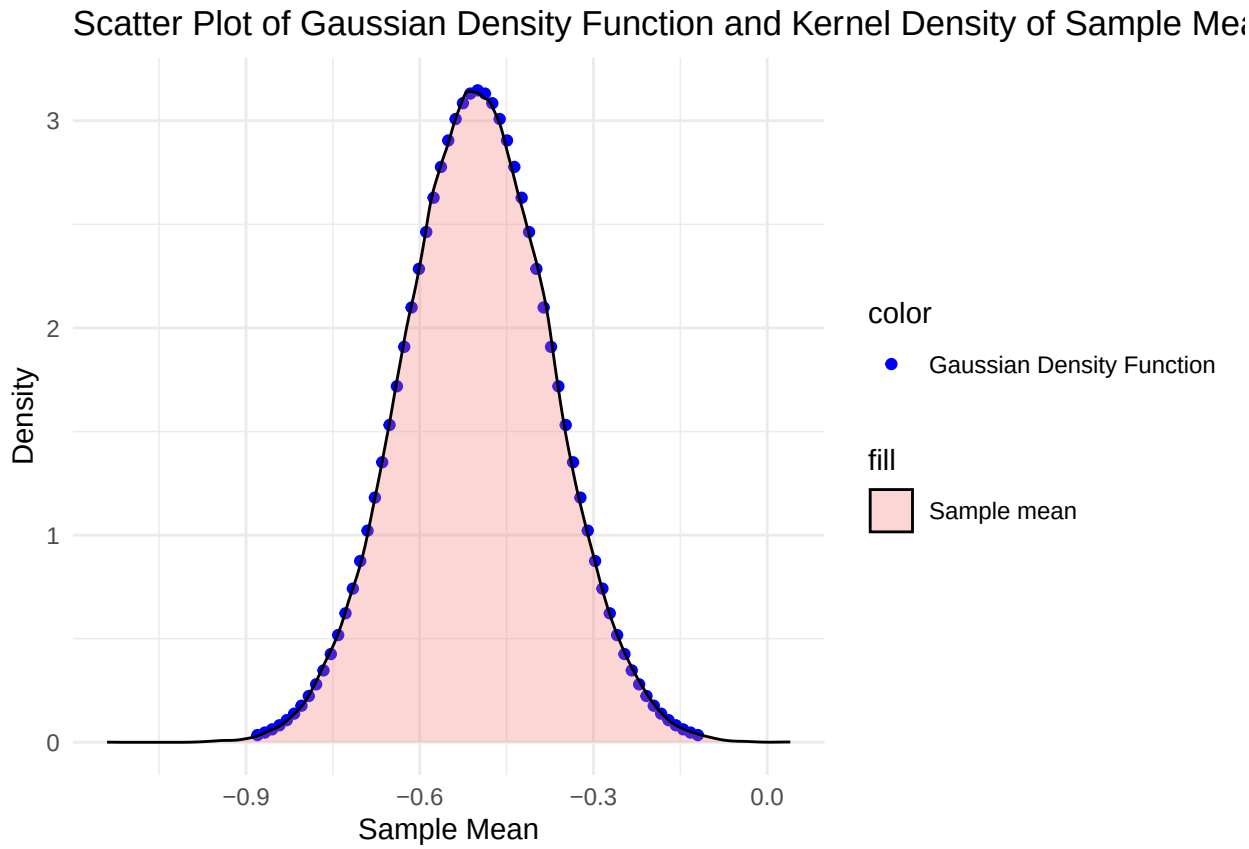
**(9)**

```r
# Set default variables
a <- 3
b <- 5
n <- 100
trials <- 50000
```

```r
# Define a vector to store sample mean for each trial
sample_means = numeric(trials)

# Generate sample_means
for(i in 1:trials){
  x = sample_Xs(a,b,n)
  sample_means[i] = mean(x)
}
```

**(10)**

```r
# Load ggplot package
library(ggplot2)

# Set mu and sigma
mu = E_X
sigma = sqrt(Var_X/n)

# Create sequence of Xi
x_seq = seq(mu - 3 * sigma,mu + 3 * sigma, by = 0.1 * sigma)

# Create Gaussian density function
f_g = dnorm(x_seq, mean = mu, sd = sigma)

# Create a data frame for the plot
f_g_data <- data.frame(x = x_seq, f_g = f_g)

# Create the plot for the density plot and the scatter plot
ggplot() +
  # Scatter plot of Gaussian points
  geom_point(data = f_g_data, aes(x = x, y = f_g, color = "Gaussian Density Function")) +
  # Kernel density of sample means
  geom_density(aes(x = sample_means, fill = "Sample mean"), alpha = 0.3) +
  labs(title = "Scatter Plot of Gaussian Density Function and Kernel Density of Sample Mean",
       x = "Sample Mean",
       y = "Density",
       color = "color") +
  scale_color_manual(values = c("Gaussian Density Function" = "blue", "Kernel" = "red")) +
  theme_minimal()
```




**(11)**

From the plot, it is easy to find that the red density curve (kernel density of simulated sample means $\overline{X}$) aligns closely with the blue scatter points (the Gaussian function $f_{\mu,\sigma}(x)$). This shows that the sample mean distribution is similar to a normal distribution.

The relationship between the density of the sample mean $\overline{X}$ and the Gaussian function $f_{\mu,\sigma}(x)$ can be explained by the **Central Limit Theorem(CLT)**: as the sample size increases, the distribution of the sample mean converges to a normal distribution with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This close match also confirms CLT's conclusion that for large sample size $n$, the sample mean will be approximately normal.