

22692037GLM (STA4813)Assignment 02Question 1

$$a) f(x_i, \theta) = \binom{x_i + r - 1}{r - 1} \theta^r (1 - \theta)^{x_i} \quad x = 0, 1, \dots$$

$$\begin{aligned} f(x_i; \theta) &= \exp(\log(1 - \theta)^{x_i} + \log \theta^r + \log \binom{x_i + r - 1}{r - 1}) \\ &= \exp(x_i \log(1 - \theta) + r \log \theta + \log \binom{x_i + r - 1}{r - 1}) \end{aligned}$$

Where :

$$a(x) = x$$

$$b(\theta) = \log(1 - \theta)$$

$$c(\theta) = \log(\theta)$$

$$d(x) = \binom{x_i + r - 1}{r - 1}$$

Therefore  $f(x_i, \theta)$  belongs to the exponential family of Probability Mass Function.

$$\begin{aligned} b) E(x) &= -\frac{c'(\theta)}{b'(\theta)} = \left[ \frac{1}{\theta} \div \frac{-1}{(1 - \theta)} \right] \\ &= \frac{(1 - \theta)}{\theta} \end{aligned}$$

Where

$$c'(\theta) = \frac{1}{\theta}; \quad b'(\theta) = \frac{-1}{1 - \theta}$$

$$\text{Var}(x) = \frac{b''(\theta) c'(\theta) - c''(\theta) b'(\theta)}{[b'(\theta)]^3}$$

$$= \left[ \frac{1}{\theta(1-\theta)^2} + \frac{1}{\theta^2(1-\theta)} \right] \times \frac{(1-\theta)^3}{-1}$$

$$= - \frac{(1-\theta)^3}{\theta^2(1-\theta)^2}$$

$$= - \frac{(1-\theta)}{\theta^2} \quad ???$$

$$G) U = a(x) b'(\theta) + c'(\theta)$$

$$= x \left( \frac{-1}{1-\theta} \right) + \frac{1}{\theta}$$

$$= \frac{\theta(x-1) + (\theta+1)}{\theta(1-\theta)}$$

$$\approx \frac{\theta x - \theta + 1}{\theta(1-\theta)}$$

$$\approx \frac{\theta(x-2) + 1}{\theta(1-\theta)}$$

Verify :

$$E(u) = b'(\theta) E(a(y)) + c'(\theta)$$

$$= \left[ \frac{-1}{1-\theta} \right] \left[ \frac{(1-\theta)}{\theta} \right] + \frac{1}{\theta}$$

$$= -\frac{1}{\theta} + \frac{1}{\theta} = 0$$

$$\begin{aligned}
 d) J &= E(-U) = E\left[\sum_{i=1}^n U_i'\right] \\
 &= \sum_{i=1}^n \left[ \frac{b''(\theta) c'(\theta)}{b'(\theta)} - c''(\theta) \right] \\
 &= \sum_{i=1}^n \left[ \frac{\left(\frac{1}{(1-\theta)^2}\right) \left(\frac{1}{\theta}\right)}{-\frac{1}{(1-\theta)}} - \left(-\frac{1}{\theta^2}\right) \right] \\
 &= \sum_{i=1}^n \left( \frac{1-\theta}{\theta(1-\theta)^2} + \frac{1}{\theta^2} \right) \\
 &= \frac{n}{\theta^2(1-\theta)} \rightarrow
 \end{aligned}$$

$$\begin{aligned}
 e) L(\theta|x) &= \prod_{i=1}^n \binom{x_i+r-1}{r-1} \theta^r (1-\theta)^{x_i} \\
 \log(L(\theta|x)) &= \sum_{i=1}^n \log \binom{x_i+r-1}{r-1} + r \log \theta + x \log(1-\theta) \\
 &= nr \log \theta + \sum_{i=1}^n x_i \log(1-\theta)
 \end{aligned}$$

$$\frac{\partial}{\partial \theta} \log(L) = \frac{nr}{\theta} + \frac{\sum_{i=1}^n x_i (-1)}{1-\theta}$$

$$0 = \frac{nr}{\hat{\theta}} - \frac{\sum_{i=1}^n x_i}{1-\hat{\theta}}$$

$$\begin{aligned}
 (1-\hat{\theta})nr &= \hat{\theta} \sum_{i=1}^n x_i \\
 nr &= \hat{\theta} \left( \sum_{i=1}^n x_i + nr \right) \\
 \hat{\theta} &= \frac{nr}{\sum_{i=1}^n x_i + nr}
 \end{aligned}$$



$$f) W_n = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\text{Var}(\hat{\theta}_n)}}$$

$$\text{Var}(\hat{\theta}_n) = \frac{\theta(1-\theta)}{n-1}$$

$$\therefore W = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{\theta_n(1-\theta_n)}{nr}}}$$

$$g) \hat{\theta}_n \pm Z_{1-\alpha/2} \times \text{Se}(\hat{\theta}_n)$$

$$= \hat{\theta}_n \pm Z_{1-\alpha/2} \times \hat{\theta}_n$$

Question 2:

Given  $f(y, p) = p(1-p)^{y-1}$

$$a) \log(f(y, p)) = \log(p(1-p)^{y-1})$$

$$= \log p + y \log(1-p) - \log(1-p)$$

$$\therefore f(y, p) = \exp(y \log(1-p) - \log(1-p) - \log(p))$$

Where

$$a(y) = y$$

$$b(p) = \log(1-p)$$

$$c(p) = \log(p)$$

$$d(y) = 0$$

$$\begin{aligned}
 (b) \quad J &= E(-U) = E \left[ \sum_{i=1}^n U_i' \right] \\
 &= \sum_{i=1}^n \left[ \frac{b''(p) C'(p)}{b'(p)} - C''(p) \right] \\
 &= \sum_{i=1}^n \left[ \frac{\left( \frac{1}{(1-p)^2} \right) \left( \frac{1}{p} \right)}{\frac{-1}{1-p}} - \left( -\frac{1}{p^2} \right) \right] \\
 &= \sum_{i=1}^n \left( \frac{1}{p(1-p)} + \frac{1}{p^2} \right) \\
 &= \frac{n}{p^2(1-p)} \rightarrow
 \end{aligned}$$

$$c) \quad L(p|y) = \pi (1-p)^{y-1} p$$

$$\begin{aligned}
 \log(L(p|y)) &= \sum_{i=1}^n \log(p(1-p)^{y_i-1}) \\
 &= \sum_{i=1}^n (\log p + \log(1-p)^{y_i-1}) \\
 &= n \log p + \log(1-p) \sum (y_i - n)
 \end{aligned}$$

$$\frac{d}{dp} (\log(L(p|y))) = n/p + (-1/(1-p)) \sum y_i - n$$

$$\Rightarrow n/p + (-1/(1-\hat{p})) \sum (y_i - n) = 0$$

$$\Rightarrow \hat{p} \sum (y_i - n) = n(1-\hat{p})$$

$$\Rightarrow \hat{p} = \frac{n}{\sum y_i}$$

$$\therefore \hat{p} = 1/\bar{y} \approx \underline{\underline{\bar{y}^{-1}}}$$

# Assignment02\_Rmarkdown

22692037\_Tshepiso Mashiane

2024-05-24

## Question3

### Loading the data

```
library(gridExtra)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:gridExtra':
##
##      combine
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
gamma_arrivals1 <- read.csv("C:/Users/36076724/Desktop/BSC Hons Statistic/STA4813_GLM/2024/Assignment_02/gamma_arrivals2.csv", header = TRUE)
inter_arrival_times <- gamma_arrivals1$Arrivals
summary(gamma_arrivals1)
```

```
##      Arrivals
## Min.      : 0.112
## 1st Qu.: 23.410
## Median : 55.553
## Mean    : 79.935
## 3rd Qu.:110.986
## Max.     :729.135
```

```
head(gamma_arrivals1)
```

```
## Arrivals
## 1 38.02663
## 2 409.44521
## 3 272.46339
## 4 98.67453
## 5 87.85801
## 6 79.83959
```

The `gamma_arrivals` data above contains a set of gamma-ray data consisting of the times between arrivals (inter-arrival times) of 3935 photons in seconds. The inter-arrival times range from 0.112 to 729.135, with a median time of 55.553 seconds and a mean time of 79.935 seconds

## Examine the data

Negative values are invalid for a gamma distribution, hence we need to examine the data to ensure there are no negative values in the data

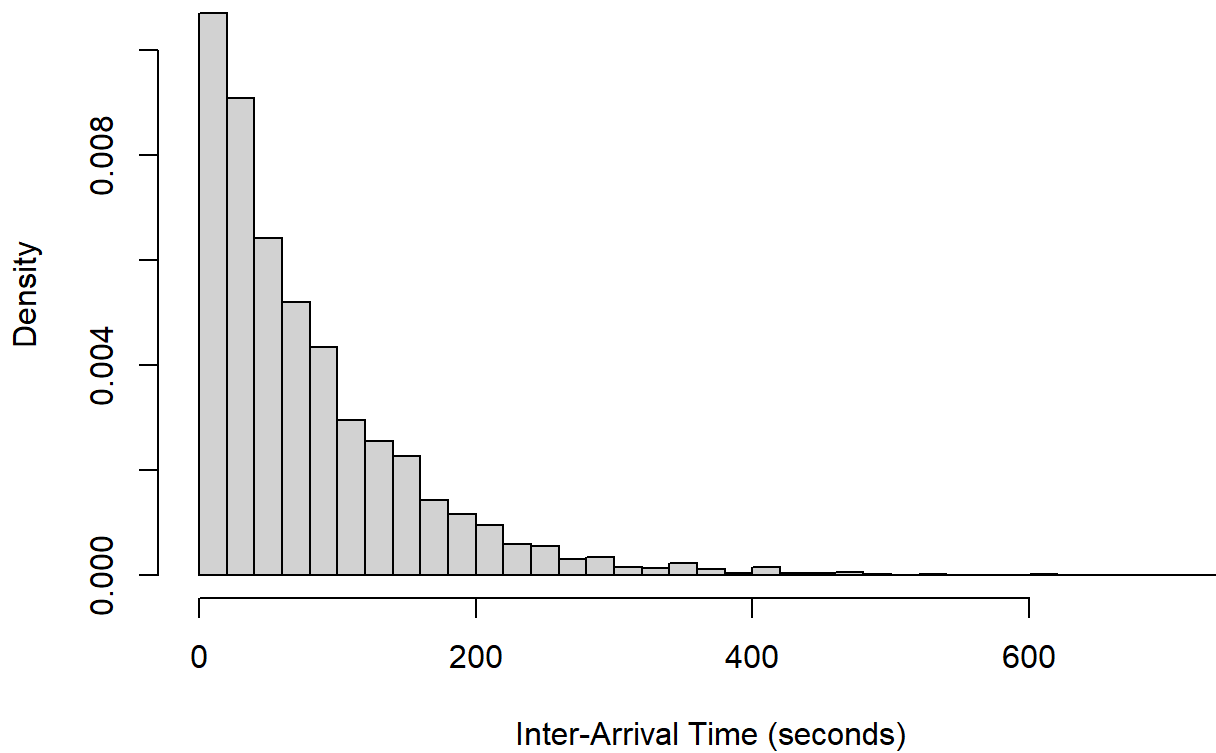
```
if (any(inter_arrival_times <= 0)) {
  stop("Negative value detected")
}
```

Since nothing is returned this means there is no negative value detected. Therefore we can continue with our analysis.

### a. Histogram of the inter-arrival times

```
hist(inter_arrival_times, breaks = 50, probability = TRUE,
     main = "Histogram of Inter-Arrival Times",
     xlab = "Inter-Arrival Time (seconds)", ylab = "Density")
```

## Histogram of Inter-Arrival Times



Since the data is concentrated around lower arrival time values and has a long tail of higher values, this means the inter-arrival times are right-skewed, which is a characteristic of a gamma distribution. This suggests that the gamma distribution is a plausible model for the data.

b. Estimate the model parameters using the method of maximum likelihood estimation



```

# Define the negative log-likelihood function for the Gamma distribution
neg_log_likelihood <- function(params, data) {
  alpha <- params[1]
  beta <- params[2]

  # Ensure parameters are positive
  if (alpha <= 0 || beta <= 0) return(Inf)

  n <- length(data)

  # Calculate the log likelihood
  log_likelihood <- n * (alpha * log(beta) - lgamma(alpha)) +
    (alpha - 1) * sum(log(data)) - beta * sum(data)

  # Return the negative log likelihood
  return(-log_likelihood)
}

# Initial parameter guesses
initial_params <- c(1, 1)

# Optimize the parameters using nlm
mle_result <- optim(initial_params, neg_log_likelihood, data = inter_arrival_times,
  method = "L-BFGS-B", lower = c(1e-8, 1e-8))

# Extract the estimated parameters
alpha_hat <- mle_result$par[1]
beta_hat <- mle_result$par[2]

# Print the estimated parameters
cat("Estimated alpha:", alpha_hat, "\n")

```

```
## Estimated alpha: 1.029497
```

```
cat("Estimated beta:", beta_hat, "\n")
```

```
## Estimated beta: 0.01290321
```

```

# Define the Gamma density function
gamma_density <- function(x, alpha, beta) {
  return((beta^alpha / gamma(alpha)) * x^(alpha - 1) * exp(-beta * x))
}

```

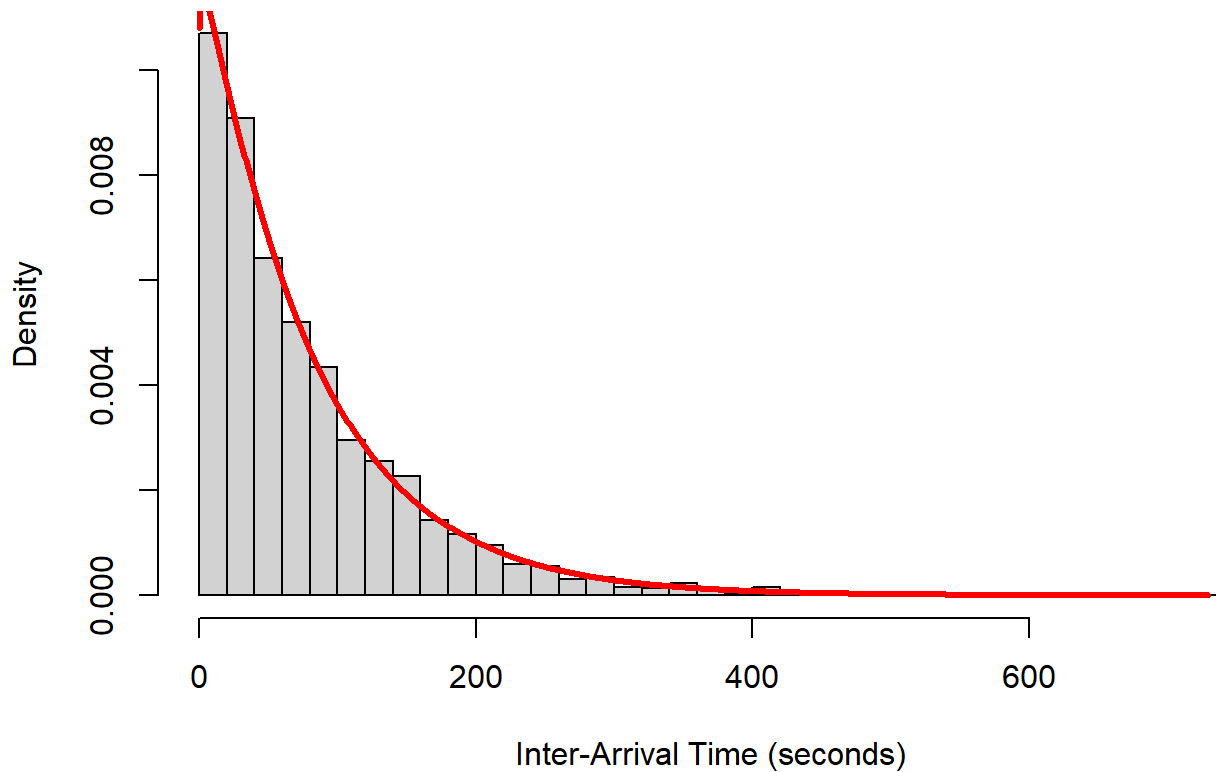
### c. Fitted gamma densities

```
# Plot the histogram with the fitted Gamma densities
hist(inter_arrival_times, breaks = 50, probability = TRUE,
     main = "Histogram of Inter-Arrival Times with Fitted Gamma Densities",
     xlab = "Inter-Arrival Time (seconds)", ylab = "Density")

# Generate a sequence of x values for the density plot
x_values <- seq(min(inter_arrival_times), max(inter_arrival_times), length.out = 1000)

# Plot the fitted Gamma density
lines(x_values, gamma_density(x_values, alpha_hat, beta_hat), col = "red", lwd = 3)
```

## Histogram of Inter-Arrival Times with Fitted Gamma Densities



The fitted gamma density appears to be a reasonable model for the data.

The gamma distribution captures the right-skewed shape of the histogram, and the peak of the fitted density aligns well with the mode of the histogram. While the fit may not be perfect, the gamma distribution seems to be a plausible model for the inter-arrival times of the gamma-rays data. Therefore the fitted gamma density appears to be a reasonable model for the data.