Modelling Clinical Trials Enrolment

Clementine Surya

2023-07-03

Introduction

In this report, we investigate the characteristics of the data related to the number of patients enrolled in different hospitals for clinical trials by utilizing the Poisson-Gamma and Zero-Truncated Poisson-Gamma models.

1. Assumption of Independence and Identical Distribution

The number of patients enrolled in each hospital is assumed to follow a Poisson-Gamma distribution. The Poisson distribution describes the number of events (patient enrollments) occurring in a fixed interval, while the Gamma distribution represents the variability in the Poisson parameter. Based on the following observations, we can assume that the data points are independent and identically distributed for modeling enrollment to clinical trials:

- The data points represent the number of patients enrolled in different hospitals over the same period of time. It is reasonable to assume that the patient enrollments in different hospitals are independent of each other, as they are likely influenced by different factors and recruitment strategies.

- The data points are collected over the same period of time, which suggests that they are generated under similar conditions and can be considered identically distributed.

Considering these factors, it is reasonable to assume that the data points are independent and identically distributed for the purpose of modeling enrollment to clinical trials using the Poisson-Gamma model.

2. Analysis of Sample Mean and Sample Variance

To further understand the potential count model for the data, we analyze the sample mean and sample variance. The sample mean represents the average number of patients enrolled across different hospitals, providing insights into the central tendency of the data. The sample variance, on the other hand, measures the dispersion or variability of the enrollment counts.

To perform this analysis, we first load the data containing the enrollment counts from different hospitals:

```
# Load the data
x <- readRDS("~/Desktop/Github/R-Projects/Stochastic Modelling - Modeling Clinical Trial Enrolment/Enrolment
Counts.Rds")
```

Next, we calculate the mean and variance of the enrollment counts:

```
# Calculate the mean of the data
mean <- mean(x)
cat("mean:", mean, "\n")
```

```
## mean: 4.79
```

```
# Calculate the variance of the data
var <- var(x)
cat("variance:", var, "\n")
```
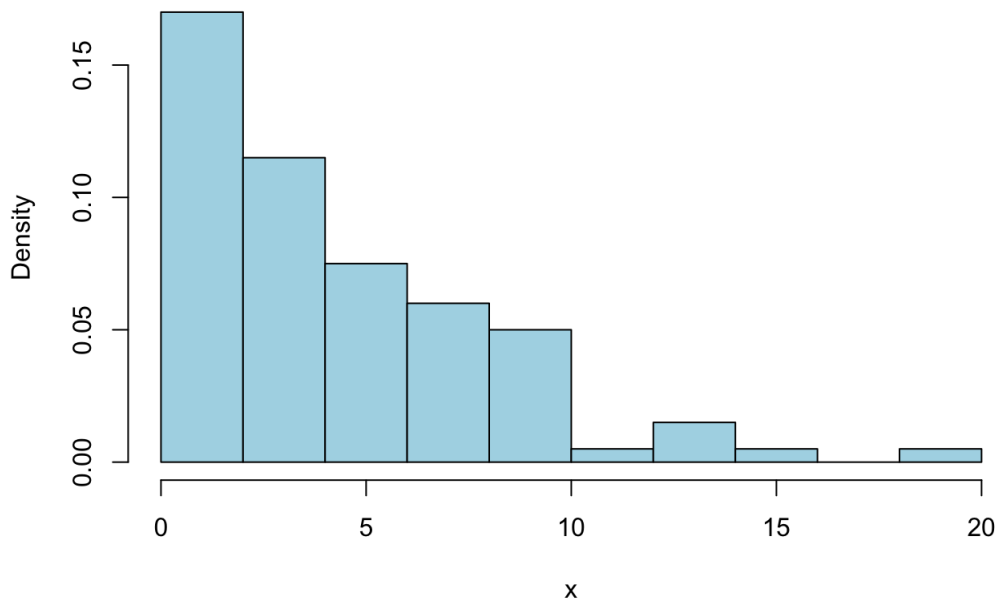
```
## variance: 13.4403
```

```
# Print summary statistics
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    2.00    4.00    4.79    7.00   20.00
```

```
# Create the histogram
hist(x, freq = FALSE, col = "lightblue", main = "Histogram of x")
```

Histogram of x

Upon analyzing the data, we observe that the sample mean is 4.79, indicating that, on average, around 4.79 patients are enrolled across the hospitals participating in the clinical trials and the sample variance is approximately 13.44.

Comparing the sample mean and sample variance, we note that they are not equal. This discrepancy suggests that the data may not perfectly align with the assumptions of a simple Poisson count model. The difference between the sample mean and variance could be indicative of overdispersion, meaning that there is more variability in the data than what a Poisson model can account for.

Considering the presence of overdispersion, it may be appropriate to explore alternative count models that can accommodate the observed variability in the enrollment counts. One potential model is the Poisson-Gamma model, which incorporates the Gamma distribution as a prior for the Poisson parameter, allowing for increased flexibility in capturing overdispersion.

In addition, by examining the histogram above, we can further observe that the distribution is right-skewed where the mean (4.79) is greater than the median (4.00).

3. Estimation of Model Parameters using Method of Moments

To fit the Poisson-Gamma model to the data and estimate the model parameters, we can utilize the method of moments. The method of moments involves equating the theoretical moments of the distribution to the sample moments and solving for the parameter values.

The Poisson-Gamma model is defined by the following equations:

$$E(x) = \alpha\beta \qquad (1)$$
$$V(x) = \alpha\beta + \alpha\beta^2 = \alpha\beta(1 + \beta) \qquad (2)$$
$$\bar{x} = \alpha\beta \qquad (3)$$
$$s^2 = \alpha\beta(1 + \beta) \qquad (4)$$
$$\beta = \frac{s^2}{\bar{x}} - 1 \qquad (5)$$
$$\alpha = \frac{\bar{x}}{\beta} \qquad (6)$$

Let's calculate the estimated model parameters using the method of moments:

```
# Calculate for estimate model parameters
beta_mom <- var / mean - 1
alpha_mom <- mean / beta_mom
cat("Estimated alpha:", alpha_mom, "\n")
```

```
## Estimated alpha: 2.652404
```

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

```
## Estimated beta: 1.805909
```

By using method of moment, we obtain the estimated model parameters, which are $\hat{\alpha}$: 2.65 and $\hat{\beta}$: 1.81.

4. Maximum Likelihood Estimation and Confidence Intervals

Now, we will fit the Poisson-Gamma model to the data using maximum likelihood estimation (MLE) and obtain 95% confidence intervals for the model parameters.

To perform maximum likelihood estimation, we define the dPoissonGamma function to compute the Poisson-Gamma probability density or log density, and the loglik function to compute the log likelihood of the Poisson-Gamma model. We then use the optim function with the BFGS optimization method to estimate the parameters.

```r
# Define the dPoissonGamma function to compute the Poisson-Gamma log probability mass function or probabilit
y mass function
dPoissonGamma <- function(x,alpha,beta,log=FALSE)
{
if (log)
{
lognumer <- lgamma(x+alpha)+x*log(beta)
logdenom <- lgamma(alpha)+(alpha+x)*log(1+beta)+lgamma(x+1)
res <- lognumer-logdenom
}else
{
numer <- gamma(x+alpha)*beta^x
denom <- gamma(alpha)*(1+beta)^(alpha+x)*gamma(x+1)
res <- numer/denom
}
res
}
```

```r
# Define the loglik function to compute the log likelihood of the Poisson-Gamma model
loglik_PoissonGamma <- function(theta,x)
{
  alpha <- theta[1]
  beta <- theta[2]
  sum(dPoissonGamma(x, alpha,beta,log=TRUE))
}

# Initial values for alpha, beta
alpha0 <- 1
beta0 <- 1
theta0 <- c(alpha0, beta0)

# Estimate the parameters
fit_PoissonGamma <- optim(par = theta0, loglik_PoissonGamma, method = "BFGS", x = x, control = list(fnscale
= -1), hessian = TRUE)
```

```r
# Retrieve the estimated parameters from the fitted model
alpha_PoissonGamma <- fit_PoissonGamma$par[1]
beta_PoissonGamma <-  fit_PoissonGamma$par[2]

# Create a vector to store the estimated parameters
par_PoissonGamma <- c(alpha_PoissonGamma, beta_PoissonGamma)

# Print the estimated alpha and beta parameters
cat("Estimated alpha:", alpha_PoissonGamma, "\n")
```

```
## Estimated alpha: 3.011026
```

```r
cat("Estimated beta:", beta_PoissonGamma, "\n")
```

```
## Estimated beta: 1.59082
```

By using maximum likelihood estimation, we obtain the estimated model parameters: $\hat{\alpha}$: 3.01 and $\hat{\beta}$: 1.59. It is noteworthy that the maximum likelihood estimation provides slightly higher estimates for alpha and lower estimates for beta compared to the method of moments.

To assess the precision of these parameter estimates, we calculate the standard errors and 95% confidence intervals.

The standard errors are computed as follows:

```
# Calculate the standard errors of the estimated parameters
inf<- -fit_PoissonGamma$hessian # Inverse of the Hessian matrix
se <- sqrt(diag(solve(inf))) # Standard errors of the parameters

# Print the standard errors
cat("alpha standard error:", se[1], "\n")
```

```
## alpha standard error: 0.6908697
```

```
cat("beta standard error:", se[2], "\n")
```

```
## beta standard error: 0.3833
```

The 95% confidence interval for alpha is calculated using the formula:

```
# Calculate the upper and lower bounds of the confidence interval for alpha
UB_alpha <- fit_PoissonGamma$par[1]+1.96*se[1] # Upper bound of the confidence interval
LB_alpha <- fit_PoissonGamma$par[1]-1.96*se[1] # Lower bound of the confidence interval

# Print the range of the confidence interval for alpha
cat("Range of alpha:", LB_alpha, "to", UB_alpha)
```

```
## Range of alpha: 1.656921 to 4.36513
```

Similarly, the 95% confidence interval for beta is computed using:

```
# Calculate the upper and lower bounds of the confidence interval for beta
UB_beta <- fit_PoissonGamma$par[2]+1.96*se[2] # Upper bound of the confidence interval
LB_beta <- fit_PoissonGamma$par[2]-1.96*se[2] # Lower bound of the confidence interval

# Print the range of the confidence interval for beta
cat("Range of beta:", LB_beta, "to", UB_beta)
```

```
## Range of beta: 0.8395523 to 2.342088
```

The 95% confidence interval for alpha is 1.66 to 4.37, and for beta is 0.84 to 2.34. These intervals provide a range of plausible values for the model parameters based on the maximum likelihood estimation.

5. Examining the Occurrence of Zero-Patient Hospitals

Using the Poisson-Gamma model fitted in Part 3, we can calculate the expected number of hospitals with zero patients in a sample of $n = 100$. This provides insights into the likelihood of observing hospitals with no enrolled patients.

```
# Calculate the probability of observing zero using the method of moments estimates
prob_zero <- dPoissonGamma(0, alpha_mom, beta_mom, log = FALSE)

# Set the sample size (number of observations)
n <- 100

# Calculate the expected number of zeros in the sample
expected_zero <- prob_zero * n

cat("The expected number of hospitals with zero patients:", round(expected_zero,1))
```

```
## The expected number of hospitals with zero patients: 6.5
```

Based on the estimated model parameters, we expect approximately 6 to 7 hospitals out of the 100 sampled hospitals to have zero enrolled patients. This suggests that a considerable proportion of hospitals may not enroll any patients in the clinical trial, indicating potential challenges in patient recruitment across different healthcare facilities.

6. Zero-Truncated Poisson-Gamma Model

The Zero-Truncated Poisson-Gamma model is obtained by conditioning the Poisson-Gamma model on the event that the count variable is greater than zero. This model is commonly used in situations where zero values are not observed or are not of interest.

The probability mass function (PMF) of the Poisson-Gamma model is derived as follows:

$$f(x) = \frac{\Gamma(x + \alpha)\beta^x}{\Gamma(\alpha)(1 + \beta)^{x+\alpha}\Gamma(x + 1)}$$

where $x = 0, 1, 2, \ldots$ and $\alpha, \beta > 0$

And the PMF for a Zero-Truncated Poisson-Gamma distribution:

$$\begin{aligned}
f(x) &= P(X = x \mid x > 0) \\
&= \frac{P(X = x, X > 0)}{P(X > 0)} \\
&= \frac{P(X = x)}{1 - P(X = 0)} \\
&= \frac{(1 + \beta)^\alpha}{(1 + \beta)^\alpha - 1} \frac{\Gamma(x + \alpha)\beta^x}{\Gamma(\alpha)(1 + \beta)^{x+\alpha}\Gamma(x + 1)} \\
&= \frac{\Gamma(x + \alpha)\beta^x(1 + \beta)^\alpha}{\Gamma(\alpha)(1 + \beta)^{x+\alpha}\Gamma(x + 1)((1 + \beta)^\alpha - 1)}
\end{aligned}$$

where $x = 1, 2, \ldots$ and $\alpha, \beta > 0$

Below is the code that defines the Probability Mass Function of Zero-Truncated Poisson-Gamma:

```
# Define dPoissonGamma_ZeroTruncated_pmf function to compute probability mass function
dPoissonGamma_ZeroTruncated_pmf <- function(x,alpha,beta)
{
numer <- gamma(x+alpha)*beta^x*(1+beta)^alpha
denom <- gamma(alpha)*(1+beta)^(alpha+x)*gamma(x+1)*((1+beta)^alpha-1)
res <- numer/denom
}
```

This probability mass function allows us to calculate the probability of observing a non-zero count value from the Zero-Truncated Poisson-Gamma model.

7. Zero-Truncated Poisson-Gamma Model Fitting

To fit the Zero-Truncated Poisson-Gamma model to the data, we use the maximum likelihood estimation method. The model parameters are estimated by maximizing the log likelihood function.

We define the dPoissonGamma_ZeroTruncated function to compute the Zero-Truncated Poisson-Gamma log probability mass function or probability mass function. This function incorporates the zero-truncation condition in the calculation.

```
# Define the dPoissonGamma_ZeroTruncated function to compute the Zero-Truncated Poisson-Gamma log probabilit
y mass function or probability mass function
dPoissonGamma_ZeroTruncated <- function(x,alpha,beta,log=FALSE)
{
if (log)
{
lognumer <- lgamma(x+alpha)+x*log(beta)+alpha*log(1+beta)
logdenom <- lgamma(alpha)+(alpha+x)*log(1+beta)+lgamma(x+1)+log((1+beta)^alpha-1)
res <- lognumer-logdenom
}else
{
numer <- gamma(x+alpha)*beta^x*(1+beta)^alpha
denom <- gamma(alpha)*(1+beta)^(alpha+x)*gamma(x+1)*((1+beta)^alpha-1)
res <- numer/denom
}
res
}
```

Next, we define the loglik_ZeroTruncated_PoissonGamma function, which computes the log likelihood of the Zero-Truncated Poisson-Gamma model. Using the initial values of alpha and beta, we estimate the parameters by optimizing the log likelihood function. The optim function is used with the BFGS method for optimization. The resulting estimates are stored in the fit_ZeroTruncated object.

```
# Define the loglik_ZeroTruncated function to compute the log likelihood of the Zero-Truncated Poisson-Gamma
model
loglik_ZeroTruncated_PoissonGamma <- function(theta,x)
{
  alpha <- theta[1]
  beta <- theta[2]
  sum(dPoissonGamma_ZeroTruncated(x, alpha,beta,log=TRUE))
}

# Initial values for alpha, beta
alpha0 <- 1
beta0 <- 1
theta0 <- c(alpha0, beta0)

# Estimate the parameters
fit_ZeroTruncated_PoissonGamma <- optim(par = theta0, loglik_ZeroTruncated_PoissonGamma, method = "BFGS", x
= x, control = list(fnscale = -1), hessian = TRUE)
```

```
# Retrieve the estimated parameters from the fitted Zero-Truncated model
alpha_ZeroTruncated_PoissonGamma <- fit_ZeroTruncated_PoissonGamma$par[1]
beta_ZeroTruncated_PoissonGamma <- fit_ZeroTruncated_PoissonGamma$par[2]

# Create a vector to store the estimated parameters
par_mle_ZeroTruncated_PoissonGamma <- c(alpha_ZeroTruncated_PoissonGamma, beta_ZeroTruncated_PoissonGamma)

# Print the estimated parameters
par_mle_ZeroTruncated_PoissonGamma
```

```
## [1] 1.614707 2.589808
```

Estimated alpha, $\hat{\alpha}$: 1.61

Estimated beta, $\hat{\beta}$: 2.59

By fitting the Zero-Truncated Poisson-Gamma model, we observe that the parameter estimates are different from the original Poisson-Gamma model. This is expected since the Zero-Truncated model accounts for the absence of zero values in the data.

8. Model Comparison using E[X] and V[X] Quantities

In order to compare the models from Parts 4 and 7, we examine the quantities $E[X] = \alpha\beta$ and $V[X] = \alpha\beta + \alpha\beta^2$

Mean and variance for Poisson-Gamma Model (Part 4):

```
# Calculate the mean of the distribution
mean_PoissonGamma <- alpha_PoissonGamma * beta_PoissonGamma

# Calculate the variance of the distribution
var_PoissonGamma <- alpha_PoissonGamma * beta_PoissonGamma + alpha_PoissonGamma * beta_PoissonGamma^2

# Print the mean and variance
mean_PoissonGamma
```

```
## [1] 4.790001
```

```
var_PoissonGamma
```

```
## [1] 12.41003
```

For the Poisson-Gamma model, we obtain an estimated expected value of 4.79 and an estimated variance of 12.41.

Mean and variance for Zero-Truncated Poisson-Gamma Model (Part 7):

```
# Calculate the mean of the distribution
mean_ZeroTruncated_PoissonGamma <- alpha_ZeroTruncated_PoissonGamma * beta_ZeroTruncated_PoissonGamma

# Calculate the variance of the distribution
var_ZeroTruncated_PoissonGamma <- alpha_ZeroTruncated_PoissonGamma * beta_ZeroTruncated_PoissonGamma + alpha
_ZeroTruncated_PoissonGamma * beta_ZeroTruncated_PoissonGamma^2

# Print the mean and variance
mean_ZeroTruncated_PoissonGamma
```

```
## [1] 4.18178
```

```
var_ZeroTruncated_PoissonGamma
```

```
## [1] 15.01179
```

For the Zero-Truncated Poisson-Gamma model, we find an estimated expected value of 4.18 and an estimated variance of 15.01.

We observe that the Poisson-Gamma model yields a slightly higher expected value (4.79) compared to the Zero-Truncated Poisson-Gamma model (4.18). This suggests that, on average, the Poisson-Gamma model predicts a slightly larger number of patients occurring in the given interval.

It's important to note that the variance of the Zero-Truncated Poisson-Gamma model (15.01) is higher than the variance of the Poisson-Gamma model (12.41). This indicates that the Zero-Truncated Poisson-Gamma model accounts for a greater level of variability in the data.

9. Model Comparison using AIC and BIC

To determine which model is better for the data, we compare the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for both the Poisson-Gamma model and the Zero-Truncated Poisson-Gamma model.

Poisson-Gamma Model:

```
# Number of parameter
p <- 2

# Compute the log likelihood of Poisson-Gamma model using the estimated parameters
log_likelihood_PoissonGamma <- loglik_PoissonGamma(par_PoissonGamma, x)

# Calculate AIC for Poisson-Gamma model
AIC_PoissonGamma <- 2*log_likelihood_PoissonGamma - 2*p
AIC_PoissonGamma_round <- round(AIC_PoissonGamma, 2)

# Calculate BIC for Poisson-Gamma model
BIC_PoissonGamma <- 2*log_likelihood_PoissonGamma - log(100)*p
BIC_PoissonGamma_round <- round(BIC_PoissonGamma, 2)
```

Zero-Truncated Poisson-Gamma Model:

```
# Compute the log likelihood of Zero-Truncated Poisson-Gamma model using the estimated parameters
log_likelihood_ZeroTruncated_PoissonGamma <- loglik_ZeroTruncated_PoissonGamma(par_mle_ZeroTruncated_Poisson
Gamma, x)

# Calculate AIC for Zero-Truncated Poisson-Gamma model
AIC_ZeroTruncated_PoissonGamma <- 2*log_likelihood_ZeroTruncated_PoissonGamma - 2*p
AIC_ZeroTruncated_PoissonGamma_round <- round(AIC_ZeroTruncated_PoissonGamma, 2)

# Calculate BIC for Zero-Truncated Poisson-Gamma model
BIC_ZeroTruncated_PoissonGamma <- 2*log_likelihood_ZeroTruncated_PoissonGamma - log(100)*p
BIC_ZeroTruncated_PoissonGamma_round <- round(BIC_ZeroTruncated_PoissonGamma, 2)
```

Here, I create table comparison of AIC and BIC values for both models:

```r
# Create a data frame to store the AIC and BIC values
AIC_BIC_comparison <- data.frame(
  Criterion = c("AIC", "BIC"),
  Poisson_Gamma = c(AIC_PoissonGamma_round, BIC_PoissonGamma_round),
  Zero_Truncated_Poisson_Gamma = c(AIC_ZeroTruncated_PoissonGamma_round, BIC_ZeroTruncated_PoissonGamma_roun
d)
)

# Print the comparison table
print(AIC_BIC_comparison)
```

```
##   Criterion Poisson_Gamma Zero_Truncated_Poisson_Gamma
## 1       AIC       -509.40                      -492.81
## 2       BIC       -514.61                      -498.02
```

```r
# Create the histogram
hist(x, freq = FALSE, col = "lightblue", main = "Histogram of x")

xseq <- seq(0, max(x), length = 101)

# Plot the density function of the Poisson-Gamma distribution on the same plot
points(xseq, dPoissonGamma(xseq, alpha_PoissonGamma, beta_PoissonGamma), col = "red", type = "l")

# Plot density function Zero-Truncated Poisson-Gamma distribution on the same plot
points(xseq, dPoissonGamma_ZeroTruncated(xseq, alpha_ZeroTruncated_PoissonGamma, beta_ZeroTruncated_PoissonG
amma), col = "blue", type = "l")

# Add legend
legend("topright", legend = c("Poisson-Gamma model", "Zero-Truncated Poisson-Gamma model"),
       col = c("red", "blue"), lty = 1, lwd = 1)
```
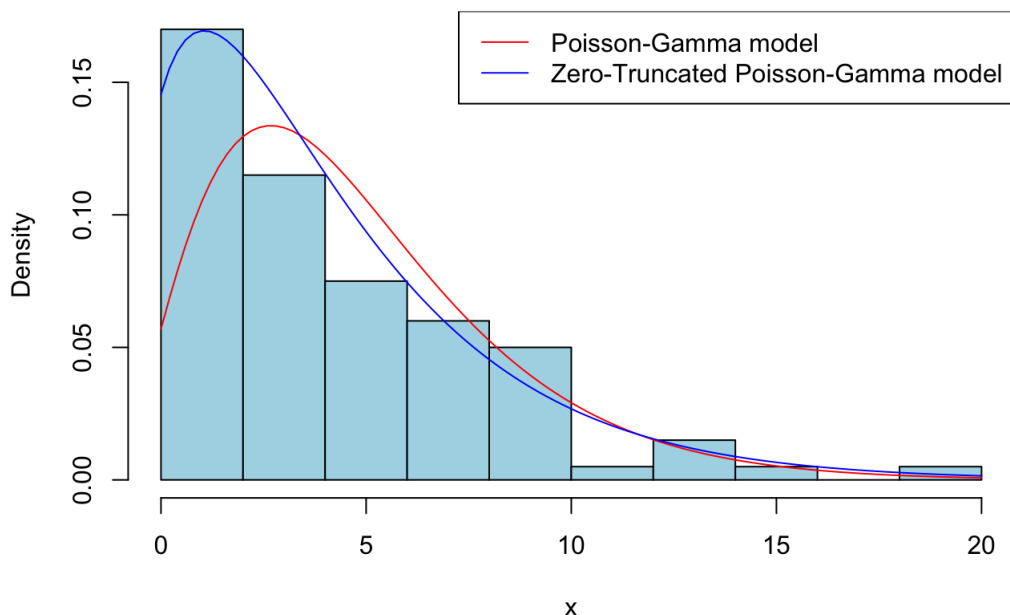

Histogram of x

After calculating the AIC and BIC values, we find that the Zero-Truncated Poisson-Gamma model has a bigger AIC (-492.81) and BIC (-498.02) compared to Poisson-Gamma model, which has an AIC of -509.40 and BIC of -514.61. The bigger AIC and BIC values suggest that the Zero-Truncated Poisson-Gamma model is more favorable in explaining the observed data.

Moreover, based on visual analysis of the plot, the Zero-Truncated Poisson-Gamma distribution appears to provide a better fit to the data compared to the Poisson-Gamma distribution.

Conclusion

In this analysis, we explored the use of the Poisson-Gamma model and the Zero-Truncated Poisson-Gamma model to describe the distribution of patient counts in a sample of hospitals. We compared the two models based on their parameter estimates, expected number of hospitals with zero patients, moments of the distribution, and model fit using AIC and BIC.

By using maximum likelihood, the Poisson-Gamma model provided parameter estimates with alpha estimated at 3.01 and beta estimated at 1.59. The Zero-Truncated Poisson-Gamma model also produced parameter estimates, with alpha estimated at 1.61 and beta estimated at 2.59.

Upon comparing the two models, we discovered that the Poisson-Gamma model exhibited a slightly higher average patient count in comparison to the Zero-Truncated Poisson-Gamma model. However, the Zero-Truncated Poisson-Gamma model demonstrated greater variability.

The AIC and BIC values indicated that the Zero-Truncated Poisson-Gamma model provided a better fit to the data.