GroupA_HW1

D. . . . , S. Cattonar, L.Ricatti, .. Valle

2024-11-05

Contents

CS - Chapter 1

Ex 1.1

Exponential random variable, $X \geq 0$, has p.d.f. $f(x) = \lambda \exp(-\lambda x)$.

1. Find the c.d.f. and the quantile function for X.
2. Find $Pr(X < \lambda)$ and the median of X.
3. Find the mean and variance of X.

Solution

1. C.D.F. and Quantile Function:

   The cumulative distribution function (c.d.f.) F(x) is:

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, \quad x \geq 0$$

   The quantile function Q(p) is the inverse of F(x):

$$Q(p) = -\frac{1}{\lambda} \ln(1 - p), \quad 0 \leq p < 1$$

2. Pr(X < λ) and Median:

   $\Pr(X < \lambda) = F(\lambda) = 1 - e^{-\lambda \lambda} = 1 - e^{-1} \approx 0.6321$

   For the median, we solve F(x) = 0.5:

$$1 - e^{-\lambda x} = 0.5$$
$$x = -\frac{1}{\lambda} \ln(0.5) = \frac{\ln(2)}{\lambda}$$

3. Mean and Variance:

Mean: $E[X] = \int_0^\infty x\lambda e^{-\lambda x}dx = \frac{1}{\lambda}$

Variance: $Var(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$

Comments on the solution:

1. The exponential distribution is characterized by its rate parameter λ, which determines both its shape and scale.

2. Interestingly, $\Pr(X < \lambda)$ is always approximately 0.6321, regardless of the value of λ. This is a unique property of the exponential distribution.

3. The median of the distribution is $\frac{\ln(2)}{\lambda}$, which is always less than the mean ($\frac{1}{\lambda}$) due to the distribution's right-skewness.

4. The mean and variance are both functions of λ. As λ increases, both the mean and variance decrease, indicating that larger values of λ result in the distribution being more concentrated near zero.

5. The standard deviation of the distribution is equal to its mean, which is a distinctive feature of the exponential distribution.

Ex 1.6

Let X and Y be non-independent random variables, such that $var(X) = \sigma_x^2$, $var(Y) = \sigma_y^2$ and $cov(X,Y) = \sigma_{xy}^2$. Using the result from Section 1.6.2, find $var(X + Y)$ and $var(X - Y)$.

Solution

Using the formula for linear transformations of random vectors from Section 1.6.2:

$$\mathbf{AX} \sim N(\mathbf{A\mu}, \mathbf{A\Sigma A^T})$$

1. Define a random vector and its covariance matrix:

$$\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix}$$

2. For $var(X + Y)$: Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \end{pmatrix}$

$$var(X + Y) = \mathbf{A\Sigma A^T} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 + \sigma_{xy}^2 \\ \sigma_{xy}^2 + \sigma_y^2 \end{pmatrix} = \sigma_x^2 + 2\sigma_{xy}^2 + \sigma_y^2$$

3. For $var(X - Y)$: Let $\mathbf{A} = \begin{pmatrix} 1 & -1 \end{pmatrix}$

$$var(X - Y) = \mathbf{A\Sigma A^T} = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 - \sigma_{xy}^2 \\ -(\sigma_{xy}^2 - \sigma_y^2) \end{pmatrix} = \sigma_x^2 - 2\sigma_{xy}^2 + \sigma_y^2$$

Therefore:

$$\mathrm{var}(X + Y) = \sigma_x^2 + 2\sigma_{xy}^2 + \sigma_y^2$$

$$\mathrm{var}(X - Y) = \sigma_x^2 - 2\sigma_{xy}^2 + \sigma_y^2$$

Comments on the solution:

1. For the sum $(X + Y)$, the covariance term is added twice, potentially increasing the overall variance if X and Y are positively correlated.

2. For the difference $(X - Y)$, the covariance term is subtracted twice, potentially decreasing the overall variance if X and Y are positively correlated.

3. If X and Y are independent (i.e., $\sigma_{xy}^2 = 0$), the variances of their sum and difference would both simplify to $\sigma_x^2 + \sigma_y^2$.

Ex 3.5

Solution

Comments on the solution:

Ex 3.6

Solution

Comments on the solution:

FSDS - Chapter 2

Ex 2.16

Each day a hospital records the number of people who come to the emergency room for treatment.

(a) In the first week, the observations from Sunday to Saturday are 10, 8, 14, 7, 21, 44, 60. Do you think that the Poisson distribution might describe the random variability of this phenomenon adequately. Why or why not?

Solution To assess whether the Poisson distribution might adequately describe the random variability of emergency room visits, we'll examine the data and compare it to properties of the Poisson distribution.

```r
# Data
er_visits <- c(10, 8, 14, 7, 21, 44, 60)
days <- c("Sun", "Mon", "Tue", "Wed", "Thu", "Fri", "Sat")

# Basic statistics
mean_visits <- mean(er_visits)
var_visits <- var(er_visits)

# Print results
cat("Mean of visits:", round(mean_visits, 2), "\n")
```

a)

```
## Mean of visits: 23.43
```

```r
cat("Variance of visits:", round(var_visits, 2), "\n")
```
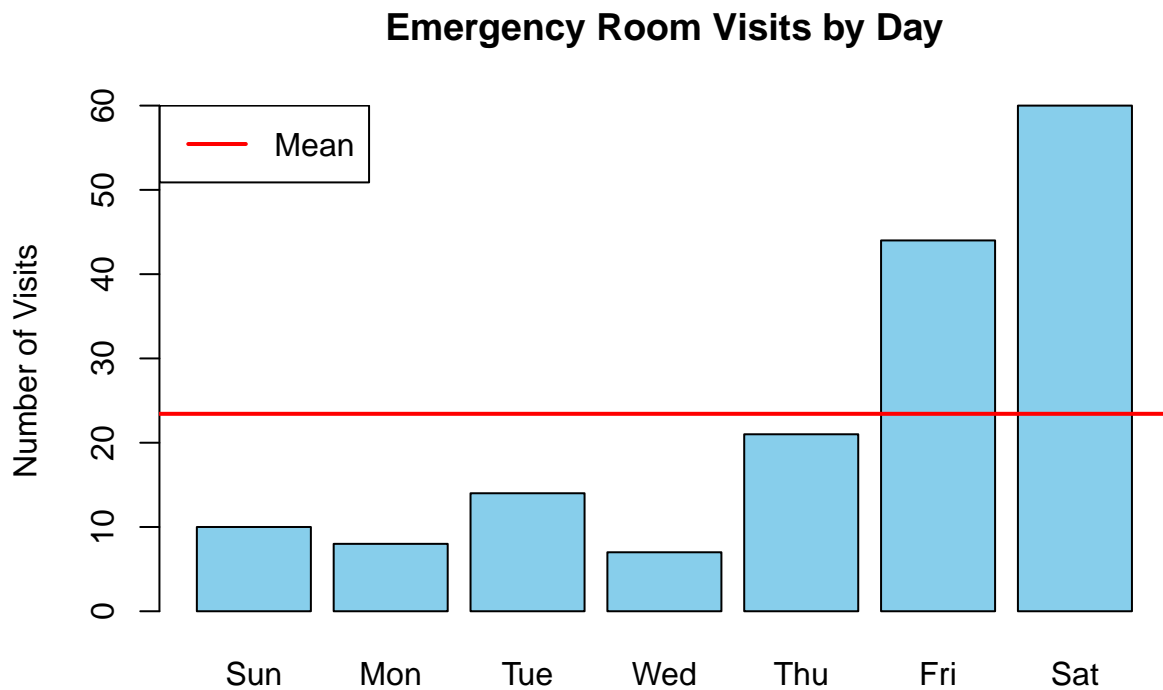
```
## Variance of visits: 423.95
```

```r
# Plot the data
barplot(er_visits, names.arg = days, main = "Emergency Room Visits by Day",
        ylab = "Number of Visits", col = "skyblue")

# Add a line for the mean
abline(h = mean_visits, col = "red", lwd = 2)
legend("topleft", legend = "Mean", col = "red", lwd = 2)
```

Emergency Room Visits by Day

1. The Poisson distribution has the property that its mean and variance are equal. In our data:

   - Mean: 23.4285714
   - Variance: 423.952381 The large difference between these values suggests that the Poisson distribution may not be appropriate.

2. The bar plot shows a clear increasing trend throughout the week, with a sharp increase on Friday and Saturday. This pattern is not consistent with the Poisson distribution, which assumes a constant rate of events.

3. The Poisson distribution assumes:

   - Events occur independently
   - The average rate of occurrences is constant

   In this case, the number of ER visits doesn't satisfy these assumptions:

   - There may be dependencies (e.g., a local event affecting multiple people)
   - The rate clearly varies by day of the week

4. The data shows overdispersion (variance much larger than the mean), which is not characteristic of the Poisson distribution.

Given these observations, we can conclude that the Poisson distribution does not adequately describe the random variability of emergency room visits in this hospital. A more complex model that accounts for day-of-week effects and overdispersion (such as a negative binomial distribution or a time series model) would likely be more appropriate.

b) Solution

Yes, we would expect the Poisson distribution to better describe the number of weekly admissions to the hospital for a rare disease.

1. **Rare events**: The Poisson distribution is particularly well-suited for modeling rare events. A rare disease, by definition, occurs infrequently.

2. **Independence**: Admissions for a rare disease are more likely to be independent of each other, especially if the disease is not contagious.

3. **Constant rate**: The occurrence of a rare disease is less likely to be affected by day-of-week patterns or other cyclical factors that we observed in general ER admissions.

4. **No simultaneous occurrences**: With rare diseases, the probability of two or more admissions occurring simultaneously is extremely low, which aligns with another assumption of the Poisson distribution.

5. **Lower variance**: Rare events typically have a lower variance, which is more likely to be closer to the mean.

6. **Small numbers**: The Poisson distribution is often used to model count data when the counts are small, which is likely the case for weekly admissions of a rare disease.

To illustrate this point, we can simulate weekly admissions for a hypothetical rare disease:

```r
set.seed(123)
weeks <- 52
lambda <- 1.5  # Average 1.5 admissions per week for the rare disease
rare_disease_admissions <- rpois(weeks, lambda)

# Basic statistics
mean_admissions <- mean(rare_disease_admissions)
var_admissions <- var(rare_disease_admissions)

# Print results
cat("Mean of admissions:", round(mean_admissions, 2), "\n")
```

```
## Mean of admissions: 1.56
```

```r
cat("Variance of admissions:", round(var_admissions, 2), "\n")
```
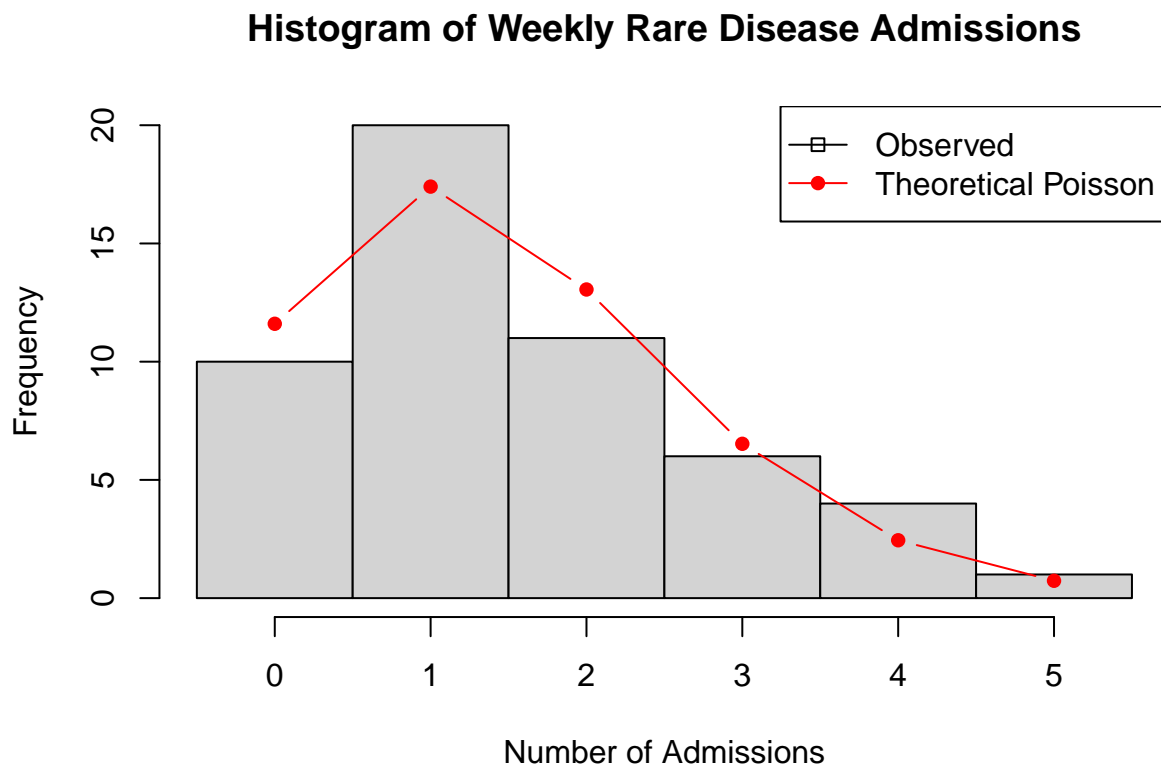
Variance of admissions: 1.58

```r
# Plot the data
hist(rare_disease_admissions, breaks = seq(-0.5, max(rare_disease_admissions) + 0.5, by = 1),
     main = "Histogram of Weekly Rare Disease Admissions",
     xlab = "Number of Admissions", ylab = "Frequency")

# Overlay Poisson distribution
x <- 0:max(rare_disease_admissions)
lines(x, dpois(x, lambda) * weeks, col = "red", type = "b", pch = 16)
legend("topright", legend = c("Observed", "Theoretical Poisson"),
       col = c("black", "red"), lty = 1, pch = c(22, 16))
```

Histogram of Weekly Rare Disease Admissions

Comments on Solution:

1. The mean (1.56) and variance (1.58) of the simulated data are much closer to each other, which is characteristic of the Poisson distribution.

2. The histogram of simulated admissions closely follows the theoretical Poisson distribution (red line), indicating a good fit.

3. The number of admissions per week is small and varies within a narrow range, which is typical for rare events and well-described by the Poisson distribution.

Ex 2.21

Plot the gamma distribution by fixing the shape parameter k = 3 and setting the scale parameter θ = 0.5, 1, 2, 3, 4, 5. What is the effect of increasing the scale parameter? (See also Exercise 2.48.)

Solution

To visualize the effect of increasing the scale parameter on the gamma distribution, we'll create a plot showing multiple gamma distributions with a fixed shape parameter and varying scale parameters.

```
# Set parameters
k <- 3   # Shape parameter
theta <- c(0.5, 1, 2, 3, 4, 5)   # Scale parameters
colors <- c("#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00")

# Create x values
x <- seq(0, 30, length.out = 1000)

# Plot
plot(x, dgamma(x, shape = k, scale = theta[1]), type = "l", col = colors[1],
     main = "Gamma Distribution with k = 3 and Varying theta",
     xlab = "x", ylab = "Density", ylim = c(0, 0.5))

# Add lines for other scale parameters
for (i in 2:length(theta)) {
  lines(x, dgamma(x, shape = k, scale = theta[i]), col = colors[i])
}

# Add legend
legend("topright", legend = paste("theta =", theta), col = colors, lty = 1, cex = 0.8)
```


Gamma Distribution with k = 3 and Varying theta
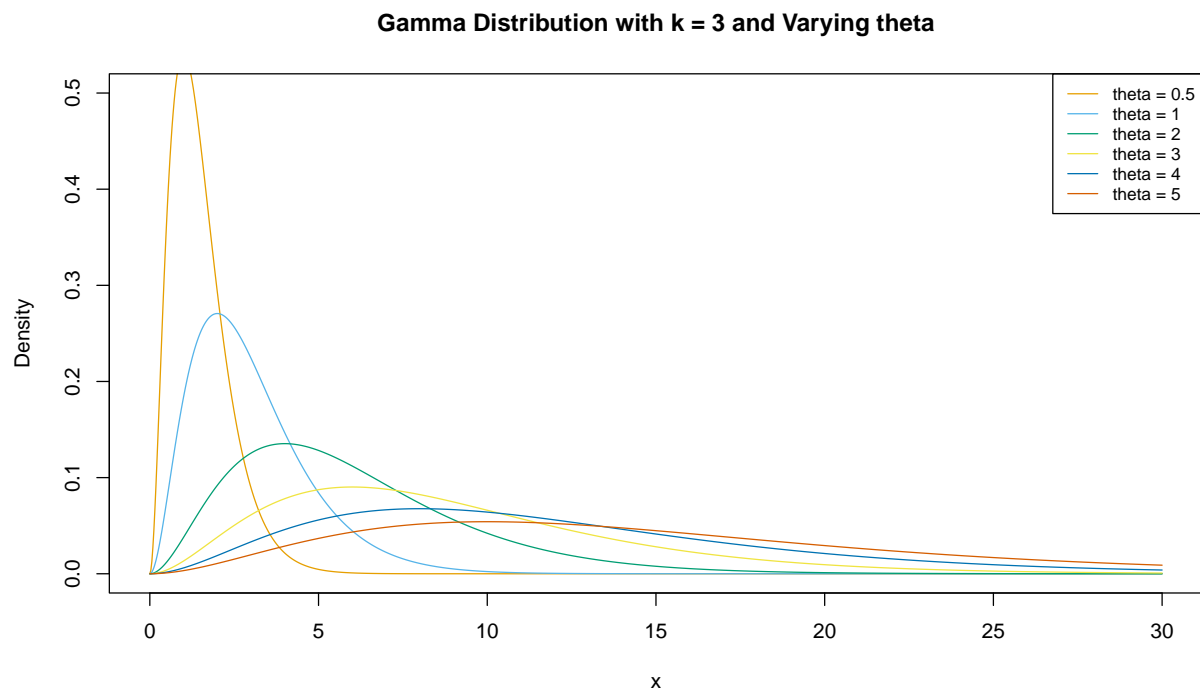
Comments on the solution:

1. **Shape of the distribution**: As we increase the scale parameter θ, we observe:

   a. The peak of the distribution shifts to the right (towards larger x values).
   b. The height of the peak decreases.
   c. The distribution becomes wider.

2. **Interpretation**:

   - A larger scale parameter θ indicates greater variability and a shift in the distribution.
   - This can be explained by the relationship between the parameters of the gamma distribution:
     - The mean of the gamma distribution is $\mu = k\theta$
     - The variance is $\sigma^2 = k\theta^2$
   - As θ increases:
     - The mean increases linearly $(k\theta)$
     - The variance increases quadratically $(k\theta^2)$
   - This quadratic increase in variance relative to the mean explains the greater spread and variability we observe with larger θ values.