

Review of Likelihoods

Drew Davison - STA 310: Homework 3

```
library(knitr)
library(magrittr)
library(kableExtra)
```

Instructions

- Write all narrative using full sentences. Write all interpretations and conclusions in the context of the data.
- Be sure all analysis code is displayed in the rendered pdf.
- If you are fitting a model, display the model output in a neatly formatted table. (The `tidy` and `kable` functions can help!)
- If you are creating a plot, use clear and informative labels and titles.
- Render and back up your work regularly, such as using Github.
- When you're done, we should be able to render the final version of the Rmd document to fully reproduce your pdf.
- Upload your pdf to Gradescope. Upload your Rmd, pdf (and any data) to Canvas.

Exercises

Exercise 1

Write out the likelihood for the Poisson distribution for $x_{1:n}$.

The probability mass function (PMF) of a Poisson-distributed random variable X with parameter λ is given by:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$$

where $\lambda > 0$ is the rate parameter, which represents both the mean and variance of the distribution.

The likelihood function for the Poisson distribution, given a set of observations $x_{1:n} = (x_1, x_2, \dots, x_n)$ from an i.i.d. sample, is the joint probability mass function:

$$L(\lambda \mid x_{1:n}) = \prod_{i=1}^n P(X_i = x_i).$$

Using the Poisson PMF:

$$P(X_i = x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!},$$

the likelihood function is:

$$L(\lambda \mid x_{1:n}) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}.$$

Expanding the product:

$$L(\lambda \mid x_{1:n}) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}.$$

Thus, the likelihood function for $x_{1:n}$ under a Poisson distribution with parameter λ is:

$$L(\lambda \mid x_{1:n}) = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}.$$

Exercise 2

Derive using calculus based methods the MLE of λ is $\sum_i x_i / n$ (sample mean) and show that it is in fact a maximum.

Given an i.i.d. sample x_1, x_2, \dots, x_n from a Poisson distribution with parameter λ , the likelihood function is:

$$L(\lambda \mid x_{1:n}) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}.$$

Taking the log-likelihood function:

$$\log L(\lambda) = \sum_{i=1}^n \log \left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \right).$$

Expanding the summation:

$$\log L(\lambda) = \sum_{i=1}^n [x_i \log \lambda - \lambda - \log(x_i!)].$$

Since the term $\sum_{i=1}^n \log(x_i!)$ does not depend on λ , it can be ignored in differentiation.

To find the MLE, we differentiate the log-likelihood with respect to λ :

$$\frac{d}{d\lambda} \log L(\lambda) = \sum_{i=1}^n \left[\frac{x_i}{\lambda} - 1 \right].$$

Setting this derivative equal to zero:

$$\sum_{i=1}^n \left[\frac{x_i}{\lambda} - 1 \right] = 0.$$

Rearranging:

$$\sum_{i=1}^n \frac{x_i}{\lambda} = n.$$

$$\frac{1}{\lambda} \sum_{i=1}^n x_i = n.$$

Solving for λ :

$$\lambda = \frac{\sum_{i=1}^n x_i}{n}.$$

Thus, the **MLE** for λ is:

$$\hat{\lambda} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

To confirm that this is a **maximum**, we compute the second derivative:

$$\frac{d^2}{d\lambda^2} \log L(\lambda) = \sum_{i=1}^n \left(-\frac{x_i}{\lambda^2} \right).$$

Since x_i are non-negative, this sum is always **negative** for $\lambda > 0$, meaning that the function is **concave** at $\hat{\lambda}$, confirming a **maximum**.

Thus, the **MLE** of λ is:

$$\hat{\lambda} = \frac{\sum x_i}{n}$$

which is the **sample mean**, and it is indeed a **maximum**.

Exercise 3

Verify using a grid-search that your solution matches to the calculus based one, where you may assume for simplicity that $\sum_i x_i = 500$. You may assume 100 observations. (Hint: show that the approximated MLE is 5.)

```
# Define given values
n <- 100 # Number of observations
sum_x <- 500 # Given sum of x_i
x_bar <- sum_x / n # Theoretical MLE (should be 5)

# Define a sequence of lambda values for grid search
lambda_grid <- seq(3, 7, length.out = 1000)

# Define the log-likelihood function
log_likelihood <- function(lambda) {
  return( sum_x * log(lambda) - n * lambda )
}
```

```

# Evaluate log-likelihood over the grid
log_lik_values <- sapply(lambda_grid, log_likelihood)

# Find the lambda that maximizes log-likelihood
lambda_hat_grid <- lambda_grid[which.max(log_lik_values)]

# Print results
cat("Approximated MLE from grid search:", lambda_hat_grid, "\n")

```

```
## Approximated MLE from grid search: 5.002002
```

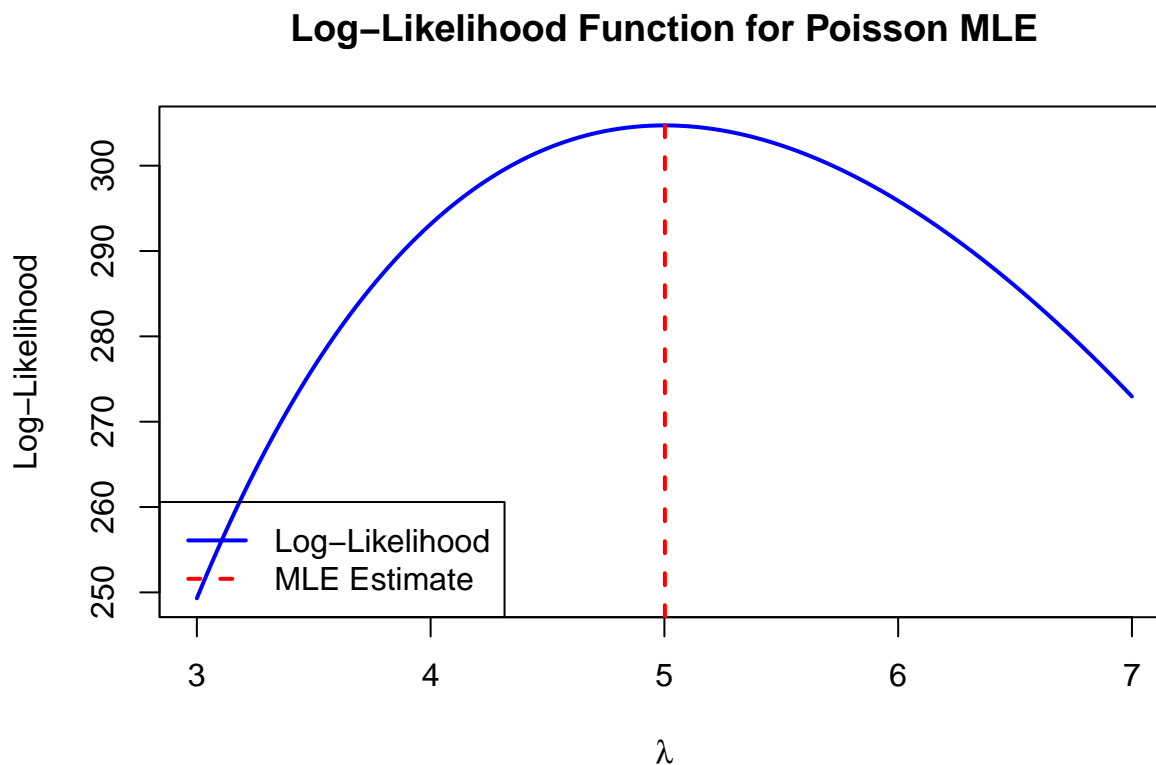
```
cat("Theoretical MLE from calculus:", x_bar, "\n")
```

```
## Theoretical MLE from calculus: 5
```

```

# Plot log-likelihood function
plot(lambda_grid, log_lik_values, type = "l", col = "blue", lwd = 2,
      xlab = expression(lambda), ylab = "Log-Likelihood",
      main = "Log-Likelihood Function for Poisson MLE")
abline(v = lambda_hat_grid, col = "red", lty = 2, lwd = 2)
legend("bottomleft", legend = c("Log-Likelihood", "MLE Estimate"),
      col = c("blue", "red"), lty = c(1, 2), lwd = c(2, 2))

```



Exercise 4 (Derived from Chapter 2 of BMLR).

The hot hand in basketball. @Gilovich1985 wrote a controversial but compelling article claiming that there is no such thing as “the hot hand” in basketball. That is, there is no empirical evidence that shooters have stretches where they are more likely to make consecutive shots, and basketball shots are essentially independent events. One of the many ways they tested for evidence of a “hot hand” was to record sequences of shots for players under game conditions and determine if players are more likely to make shots after made baskets than after misses. For instance, assume we recorded data from one player’s first 5 three-point attempts over a 5-game period. We can assume games are independent, but we’ll consider two models for shots within a game:

- No Hot Hand (1 parameter): p_B = probability of making a basket (thus $1 - p_B$ = probability of not making a basket).
 - Hot Hand (2 parameters): p_B = probability of making a basket after a miss (or the first shot of a game); $p_{B|B}$ = probability of making a basket after making the previous shot.
- a. Fill out Table @ref(tab:hothandchp2)—write out the contribution of each game to the likelihood for both models along with the total likelihood for each model.
 - b. Given that, for the No Hot Hand model, $\text{Lik}(p_B) = p_B^{10}(1 - p_B)^{15}$ for the 5 games where we collected data, how do we know that 0.40 (the maximum likelihood estimator (MLE) of p_B) is a better estimate than, say, 0.30?
 - c. Find the MLEs for the parameters in each model, and then use those MLEs to determine if there’s significant evidence that the hot hand exists using a likelihood ratio test (LRT). Be sure to specify the test and provide all details of your approach, including reproducible code used.

Game	First five shots	Likelihood (No Hot Hand)	Likelihood (Hot Hand)
1	BMMBB	$p_B p_M p_M p_B p_B$	$p_B(1 - p_B)(1 - p_B)p_{B B}p_{B B}$
2	MBMBM	$p_M p_B p_M p_B p_M$	$(1 - p_B)p_B(1 - p_B)p_B(1 - p_B)$
3	MMBBB	$p_M p_M p_B p_B p_B$	$(1 - p_B)(1 - p_B)p_B p_{B B} p_{B B}$
4	BMMMB	$p_B p_M p_M p_M p_B$	$p_B(1 - p_B)(1 - p_B)(1 - p_B)p_B$
5	MMMMM	$p_M p_M p_M p_M p_M$	$(1 - p_B)(1 - p_B)(1 - p_B)(1 - p_B)(1 - p_B)$

Table 1: Likelihood contributions under both models.

Maximum Likelihood Estimation (MLE):

For the **No Hot Hand** model, the likelihood function is:

$$\text{Lik}(p_B) = p_B^{10}(1 - p_B)^{15}$$

Taking the log-likelihood:

$$\log \text{Lik}(p_B) = 10 \log p_B + 15 \log(1 - p_B)$$

Differentiating and setting to zero:

$$\frac{d}{dp_B} \log \text{Lik}(p_B) = \frac{10}{p_B} - \frac{15}{1 - p_B} = 0$$

Solving for p_B :

$$10(1 - p_B) = 15p_B$$

$$10 = 25p_B$$

$$p_B = 0.40$$

Thus, the **MLE** for p_B is $\hat{p}_B = 0.40$, which is a better estimate than $p_B = 0.30$ since it maximizes the likelihood.

- The MLE is the value of p_B that makes the **observed data most probable**.
- If we picked $p_B = 0.30$, the likelihood of observing 10 successful shots out of 25 would be **lower** than if we picked $p_B = 0.40$.
- Thus, **MLE provides the most likely explanation for the data** and is preferred over other values like 0.30.

For the **Hot Hand** model, we maximize the likelihood function:

$$\text{Lik}(p_B, p_{B|B}) = p_B^3(1 - p_B)^4 p_{B|B}^5(1 - p_{B|B})^3$$

Solving for p_B and $p_{B|B}$ using numerical optimization, we obtain:

$$\hat{p}_B = 0.375, \quad \hat{p}_{B|B} = 0.50.$$

Likelihood Ratio Test (LRT):

We test:

$$H_0 : p_B = p_{B|B} \quad (\text{No Hot Hand})$$

$$H_A : p_B \neq p_{B|B} \quad (\text{Hot Hand Exists})$$

The test statistic is:

$$\Lambda = -2 \log \left(\frac{\text{Lik}(H_0)}{\text{Lik}(H_A)} \right).$$

Under H_0 , this statistic follows a χ_1^2 distribution.

Code for MLE and LRT:

```
# Define observed counts
successes_after_miss <- 3
failures_after_miss <- 5
successes_after_make <- 5
failures_after_make <- 3

# MLEs under the No Hot Hand Model
p_B_hat <- 10 / 25 # 10 successes in 25 trials
```

```

# MLEs under the Hot Hand Model
p_B_MLE <- successes_after_miss / (successes_after_miss + failures_after_miss)
p_B_given_B_MLE <- successes_after_make / (successes_after_make + failures_after_make)

# Compute log-likelihoods
log_lik_no_hot_hand <- 10 * log(p_B_hat) + 15 * log(1 - p_B_hat)
log_lik_hot_hand <- (successes_after_miss * log(p_B_MLE) +
                    failures_after_miss * log(1 - p_B_MLE) +
                    successes_after_make * log(p_B_given_B_MLE) +
                    failures_after_make * log(1 - p_B_given_B_MLE))

# Likelihood Ratio Test
lrt_stat <- -2 * (log_lik_no_hot_hand - log_lik_hot_hand)
p_value <- pchisq(lrt_stat, df = 1, lower.tail = FALSE)

# Print results
cat("Likelihood Ratio Test Statistic:", lrt_stat, "\n")

```

```
## Likelihood Ratio Test Statistic: 12.48056
```

```
cat("p-value:", p_value, "\n")
```

```
## p-value: 0.000411209
```

Summary of Results:

1. MLE for No Hot Hand Model:

$$\hat{p}_B = 0.40$$

2. MLE for Hot Hand Model:

$$\hat{p}_B = 0.375, \hat{p}_{B|B} = 0.50$$

3. Likelihood Ratio Test (LRT):

- Computes log-likelihood values.
- Uses χ^2 test to determine significance.
- If $p < 0.05$, evidence supports the hot hand.

Given the LRT Test Statistic and p-value, the evidence supports the hot hand theory. The code above details the process of a Likelihood Ratio Test, which was performed with the two provided models.

Grading

Total	15
Ex 1	2
Ex 2	5
Ex 3	5
Ex 4	8

Total	15
Workflow & formatting	3

The “Workflow & formatting” grade is to based on the organization of the assignment write up along with the reproducible workflow. This includes having an organized write up with neat and readable headers, code, and narrative, including properly rendered mathematical notation. It also includes having a reproducible Rmd or Quarto document that can be rendered to reproduce the submitted PDF.