

Review of Likelihoods

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.

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

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.

First, start by getting the log-likelihood:

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

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

$$\ell(\lambda) = \sum_{i=1}^n (x_i \log \lambda - \lambda - \log(x_i!))$$

Since we will be deriving with respect to λ , we can ignore the $-\log(x_i!)$ term.

$$\ell(\lambda) = \sum_{i=1}^n (x_i \log \lambda - \lambda)$$

Next, we take the derivative and set it equal to zero to find maximum:

$$\frac{d}{d\lambda} \ell(\lambda) = \sum_{i=1}^n \frac{x_i}{\lambda} - n$$

$$\sum_{i=1}^n \frac{x_i}{\lambda} - n = 0$$

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

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

This has shown us that the maximum likelihood estimate for λ is the sample mean:

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

To verify that this is a maximum, we can take the second derivative:

$$\frac{d^2}{d\lambda^2} \ell(\lambda) = - \sum_{i=1}^n \frac{x_i}{\lambda^2}$$

Since λ will always be > 0 , this second derivative will always be negative, meaning that the MLE is, in fact, 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.)

```
poisson_lik_log <- function(lambda) {
  x_sum <- 500
  n <- 100
  log_likelihood <- x_sum * log(lambda) - n * lambda
  return(log_likelihood)
}

nGrid = 1000
lambda_vals <- seq(0, 10, length = nGrid)
log_likelihood_vals <- sapply(lambda_vals, poisson_lik_log)

lambda_vals[log_likelihood_vals == max(log_likelihood_vals)]

## [1] 5.005005
```

This grid-search using the log-likelihood has returned a MLE of approximately 5, which aligns with our calculus-based solution that the MLE is the sample mean (in this case, we get $500/100 = 5$).

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?

We can justify this by plugging in 0.40 and 0.30 into the likelihood function, and seeing which gives us a greater likelihood of observing the data from these 5 games:

```
Lik_0.40 <- (0.40)^10 * (1-0.40)^15
Lik_0.30 <- (0.30)^10 * (1-0.30)^15

Lik_0.40

## [1] 4.930247e-08
```

Lik_0.30

[1] 2.803388e-08

As we can see, the likelihood is greater (therefore a better estimate) for 0.40 than for 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.

Model 1 (unconditional): p_B (probability of a make)

Model 2 (conditional): p_B (probability of make after miss or first basket of game), $p_{B|B}$ (probability of a make after a make)

Model 1 is nested in Model 2. The parameters of Model 2 ($p_B, p_{B|B}$) can be set equal to p_B (from Model 1) to get Model 1. Knowing this, we can generate the following hypothesis test:

H_0 : p_B (conditional) = $p_{B|B}$ (conditional) = p_B (unconditional)

H_a : At least of $p_B, p_{B|B}$ differs from the others

Next, we will find the MLEs for each parameter in each model:

Model 1:

$$\log \mathcal{L}(p_B) = \log(p_B^{10}(1-p_B)^{15})$$

$$\ell(p_B) = 10 \log(p_B) + 15 \log(1-p_B)$$

$$\frac{d}{dp_B} \ell(p_B) = \frac{d}{dp_B} (10 \log(p_B) + 15 \log(1-p_B))$$

$$\frac{d}{dp_B} \ell(p_B) = \frac{10}{p_B} - \frac{15}{1-p_B}$$

$$\frac{10}{p_B} - \frac{15}{1-p_B} = 0$$

$$p_B = \frac{10}{25} = \frac{2}{5}$$

To confirm the critical point as a maximum, check second derivative:

$$\frac{d^2}{dp_B^2} \ell(p_B) = -\frac{10}{p_B^2} - \frac{15}{(1-p_B)^2}$$

This will be negative when $p_B = 0.4$, so this confirms a maximum.

Model 2:

$$\log \mathcal{L}(p_B, p_{B|B}) = \log(p_B^7(1-p_B)^{11}p_{B|B}^3(1-p_{B|B})^4)$$

$$\ell(p_B, p_{B|B}) = 7 \log(p_B) + 11 \log(1-p_B) + 3 \log(p_{B|B}) + 4 \log(1-p_{B|B})$$

We now have to take partial derivatives with respect to both p_B and $p_{B|B}$:

$$\frac{\partial}{\partial p_B} \ell(p_B, p_{B|B}) = \frac{7}{p_B} - \frac{11}{1 - p_B}$$

$$\frac{\partial}{\partial p_{B|B}} \ell(p_B, p_{B|B}) = \frac{3}{p_{B|B}} - \frac{4}{1 - p_{B|B}}$$

Now we can set both equal to zero to find the critical points:

$$\frac{7}{p_B} - \frac{11}{1 - p_B} = 0$$

$$\frac{3}{p_{B|B}} - \frac{4}{1 - p_{B|B}} = 0$$

Now we can solve for the MLE of each parameter:

$$p_B = \frac{7}{18}$$

$$p_{B|B} = \frac{3}{7}$$

We now take the second derivatives to test these critical points as maxima:

$$\frac{\partial^2}{\partial p_B^2} \ell(p_B, p_{B|B}) = -\frac{7}{p_B^2} - \frac{11}{(1 - p_B)^2}$$

$$\frac{\partial^2}{\partial p_{B|B}^2} \ell(p_B, p_{B|B}) = -\frac{3}{p_{B|B}^2} - \frac{4}{(1 - p_{B|B})^2}$$

Plugging in the critical points, both second derivatives are negative, confirming these points as maxima.

Now that we have the MLEs for all parameters in both models, we can plug them back into log-likelihood functions to output the maximum log-likelihoods for each model and find the difference between these two likelihoods:

```
model1_lik <- function(pb) {
  10*log(pb) + 15*log(1 - pb)
}
model1_lik(10/25)

## [1] -16.82529

model2_lik <- function(pb, pbb) {
  7*log(pb) + 11*log(1 - pb) + 3*log(pbb) + 4*log(1 - pbb)
}
model2_lik(7/18, 3/7)

## [1] -16.80883

diff <- model2_lik(7/18, 3/7) - model1_lik(10/25)
diff
```

```
## [1] 0.01646234
```

Next, we will run a likelihood ratio test (LRT) to determine whether or not this difference between the two models is significant (whether or not to reject our null hypothesis):

```
LRT <- 2 * diff  
pchisq(LRT, 1, lower.tail = FALSE)
```

```
## [1] 0.8560131
```

p-value = 0.8560131

The p-value is very large, so we fail to reject H_0 . We do not have convincing evidence that the conditional model (using conditional p_B and $p_{B|B}$) is an improvement over the unconditional model (only unconditional p_B).

d. Table from Question 4.a

Game	First five shots	Likelihood (no hot hand)	Likelihood (hot hand)
1	BMMBB	$p_B^3(1-p_B)^2$	$p_B^2(1-p_B)p_{B B}(1-p_{B B})$
2	MBMBM	$p_B^2(1-p_B)^3$	$p_B^2(1-p_B)(1-p_{B B})^2$
3	MMBBB	$p_B^3(1-p_B)^2$	$p_B(1-p_B)^2p_{B B}^2$
4	BMMMB	$p_B^2(1-p_B)^3$	$p_B^2(1-p_B)^2(1-p_{B B})$
5	MMMMM	$(1-p_B)^5$	$(1-p_B)^5$
	TOTAL	$p_B^{10}(1-p_B)^{15}$	$p_B^7(1-p_B)^{11}p_{B B}^3(1-p_{B B})^4$
	LIKELIHOOD		

Grading

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