Group SBS, Sheet 03, Exercise 02

November 19, 2021

Contributors

Sreyas Sony (sony@uni-potsdam.de), Binoy Chacko (chacko@uni-potsdam.de), Dinesh Kumar (kumar@uni-potsdam.de) Sanika Nair (nair@uni-potsdam.de)

Solution

1. when $X_1, ..., X_n$ are independent and identical distributed random variables following the Bernoulli distribution.

θ is the unknown parameter associated with the probability of heads. The pmf of the Bernoulli distribution is thus given by,

$$f(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

The Likelihood function is given by,

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\
&= \theta^{x_1}(1-\theta)^{1-x_1} \times \theta^{x_2}(1-\theta)^{1-x_2} \times ... \times \theta^{x_n}(1-\theta)^{1-x_n} \\
&= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}
\end{aligned}$$

Taking the log of $L(\theta)$,

$$\log(L(\theta)) = \log(\theta)\sum_{i=1}^{n} x_i + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-\theta)$$

Differentiating w.r.t. θ,

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

Equating $\dfrac{\partial}{\partial\theta}\log(L(\theta))$ to 0,

$$\begin{aligned}
\frac{\sum_{i=1}^n x_i}{\theta} - \frac{n-\sum_{i=1}^n x_i}{1-\theta} &= 0 \\
\sum_{i=1}^n x_i - \theta\sum_{i=1}^n x_i - \theta n + \theta\sum_{i=1}^n x_i &= 0 \\
\Rightarrow \widehat{\theta} &= \frac{1}{n}\sum_{i=1}^n x_i
\end{aligned}$$

To check if $\widehat{\theta}$ maximizes the likelihood, taking the 2nd derivative of $\log(L(\theta))$ w.r.t. θ,

$$\frac{\partial^2}{\partial\theta^2}\log(L(\theta)) = -\frac{\sum_{i=1}^n x_i}{\theta^2} - \frac{n-\sum_{i=1}^n x_i}{(1-\theta)^2}$$

Substituting the value of $\widehat{\theta}$ in θ in the above equation,

$$\begin{aligned}
&= -\frac{n^2}{\sum_{i=1}^n x_i} - \frac{n^2}{n-\sum_{i=1}^n x_i} \\
&= \frac{-n^3 + n^2\sum_{i=1}^n x_i - n^2\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i - (\sum_{i=1}^n x_i)^2} \\
&= \frac{-n^3}{n\sum_{i=1}^n x_i - (\sum_{i=1}^n x_i)^2} < 0
\end{aligned}$$

Thus, $\widehat{\theta} = \dfrac{1}{n}\sum_{i=1}^n x_i$ maximizes the likelihood.

2. when $X_1, ..., X_n$ are independent and identical distributed random variables following the Binomial distribution.

θ is the unknown parameter associated with the probability of heads. The pmf of the Binomial distribution is thus given by,

$$\begin{aligned}
f(x_i;\theta) &= \binom{n}{x_i}\theta^{x_i}(1-\theta)^{n-x_i} \\
L(\theta) &= \prod_{i=1}^n f(x_i;\theta) \\
&= \prod_{i=1}^n \binom{n}{x_i} \times \left[\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n^2-\sum_{i=1}^n x_i}\right]
\end{aligned}$$

Taking the log of $L(\theta)$,

$$\log(L(\theta)) = \log\left(\prod_{i=1}^{n}\binom{n}{x_i}\right) + \log(\theta)\sum_{i=1}^{n} x_i + (n^2 - \sum_{i=1}^{n} x_i)\log(1-\theta)$$

Differentiating w.r.t. θ,

$$\frac{\partial}{\partial\theta}\log(L(\theta)) = 0 + \frac{1}{\theta}\sum_{i=1}^{n} x_i - \frac{1}{1-\theta}\left(n^2 - \sum_{i=1}^{n} x_i\right)$$

Equating $\frac{\partial}{\partial\theta}\log(L(\theta))$ to 0,

$$\begin{aligned}\frac{\sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i - \theta n^2 + \theta\sum_{i=1}^{n} x_i}{\theta(1-\theta)} &= 0\\ \sum_{i=1}^{n} x_i - \theta n^2 &= 0\\ \Rightarrow \widehat{\theta} &= \frac{1}{n^2}\sum_{i=1}^{n} x_i\end{aligned}$$

To check if $\widehat{\theta}$ maximizes the likelihood, taking the 2nd derivative of $\log(L(\theta))$ w.r.t. θ,

$$\frac{\partial^2}{\partial\theta^2}\log(L(\theta)) = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} - \frac{n^2 - \sum_{i=1}^{n} x_i}{(1-\theta)^2}$$

Substituting the value of $\widehat{\theta}$ in θ in the above equation,

$$\begin{aligned}&= -\frac{n^4}{\sum_{i=1}^{n} x_i} - \frac{n^4}{n^2 - \sum_{i=1}^{n} x_i}\\ &= \frac{-n^6 + n^4\sum_{i=1}^{n} x_i - n^4\sum_{i=1}^{n} x_i}{n^2\sum_{i=1}^{n} x_i - (\sum_{i=1}^{n} x_i)^2}\\ &= \frac{-n^6}{n^2\sum_{i=1}^{n} x_i - (\sum_{i=1}^{n} x_i)^2} < 0\end{aligned}$$

Thus, $\widehat{\theta} = \frac{1}{n^2}\sum_{i=1}^{n} x_i$ maximizes the likelihood.

Additional: When $X = X_1 + ... + X_n$ is Binomial

In this case, each X_i is an iid Bernoulli trial, then X is the total sum of successes from each trial.

$$f(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

The Likelihood function is given by,

$$\begin{aligned} L(X|\theta) &= f(x;\theta) \\ &= \binom{n}{x}\theta^x(1-\theta)^{n-x} \end{aligned}$$

Taking the log of $L(\theta)$,

$$\log(L(\theta)) = \log\left(\binom{n}{x}\right) + x\log(\theta) + (n-x)\log(1-\theta)$$

Differentiating w.r.t. θ,

$$\frac{\partial}{\partial\theta}\log(L(\theta)) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

Equating $\frac{\partial}{\partial\theta}\log(L(\theta))$ to 0,

$$\begin{aligned} \frac{x}{\theta} - \frac{n-x}{1-\theta} &= 0 \\ x - x\theta - n\theta + x\theta &= 0 \\ \Rightarrow \widehat{\theta} &= \frac{x}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i \text{ because } X = X_1 + X_2 + ... + X_n \end{aligned}$$

As proved in the 1st part, taking the second derivative proves that $\widehat{\theta}$ maximizes the likelihood.

We can see that, the MLE based on n independent Bernoulli random variables and the MLE based on a single binomial random variable are the same, since binomial is the result of n independent Bernoulli trials anyway. In general, whenever we have repeated independent Bernoulli trials with the same probability of success θ for each trial, the MLE will always be the sample proportion of successes. This is true regardless of whether we know the outcomes of the individual trials $X_1, X_2, ...X_n$ or just the total number of successes for all trials $X = X_1 + X_2 + ... + X_n$.

Python Implemenation

November 19, 2021

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

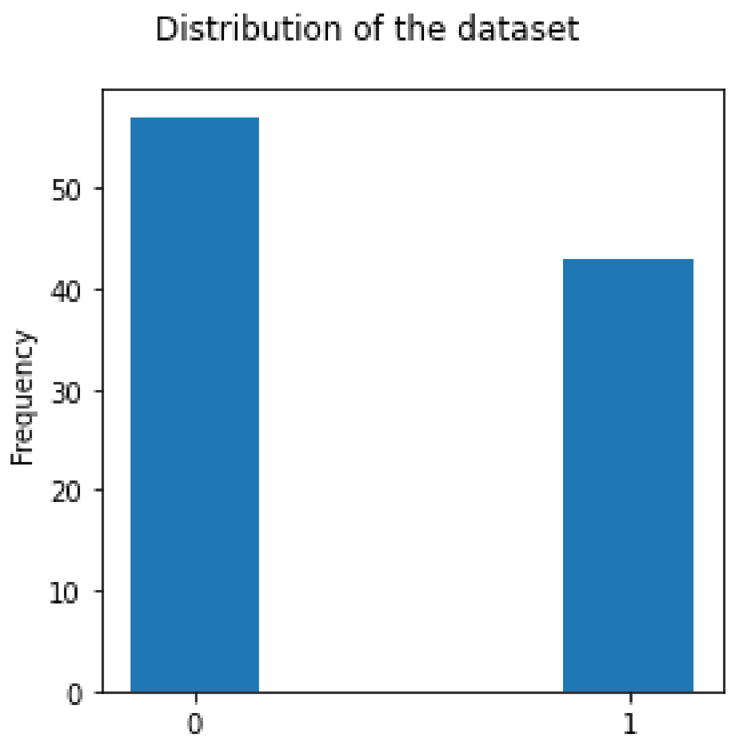
[2]:
```
# Loading the data
data = pd.read_csv('sampleset.txt', header=None)[0].tolist()
print(data)
```

```
[1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1,
0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0,
1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1]
```

[3]:
```
sample_average = np.mean(data)
print(f'The sample average is {sample_average}')
```

```
The sample average is 0.43
```

[4]:
```
# Plotting the distribution of the given dataset
height = np.histogram(data, bins=2)
plt.figure(figsize=(4,4))
plt.bar(x=['0', '1'], height=height[0], width=0.3)
plt.ylabel("Frequency")
plt.suptitle("Distribution of the dataset")
plt.tight_layout()
```



0.1 Parameter estimation using Bernoulli distribution

0.1.1 Theroretical method

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

This is equal to the sample average

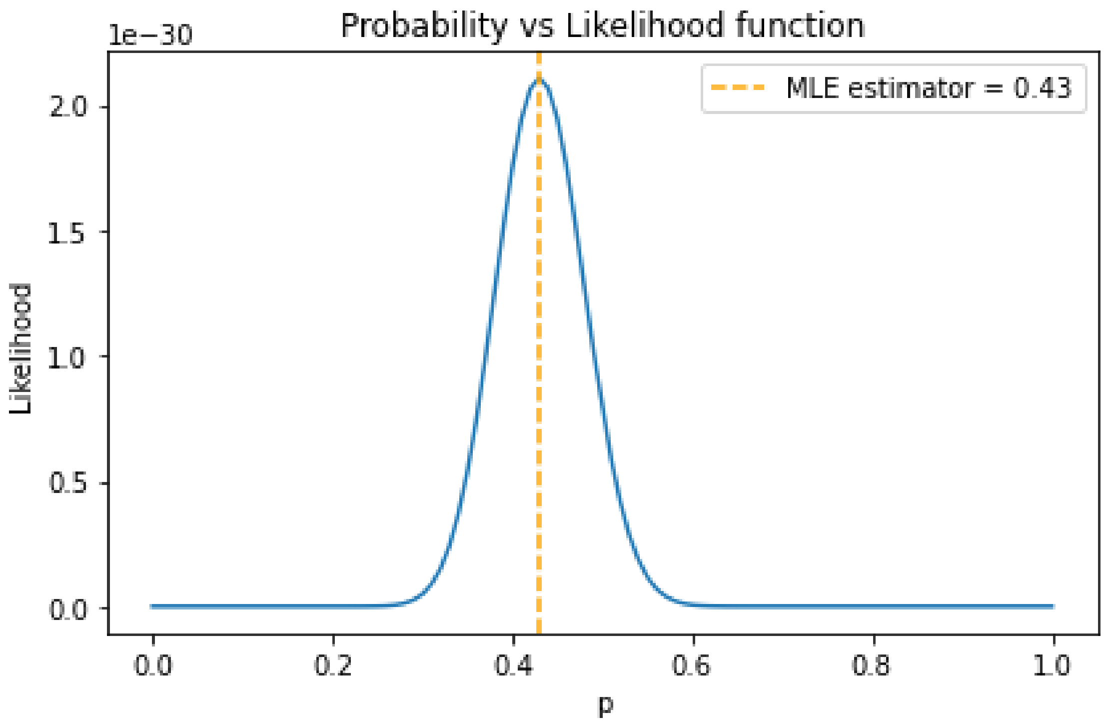
[5]:
```
theta_estimate_Ber = np.mean(data)
print(f'The estimate for p is {theta_estimate_Ber}')
```

```
The estimate for p is 0.43
```

0.1.2 Graphical method

[6]:
```
def likelihood_Ber(counts, n=100):
    """
    counts: total number of success
    n: number of trials
    """
    p = np.round(np.linspace(0,1,100), 2)
    #likelihood function
    L = p**counts * (1-p)**(n-counts)
    return p, L
```

[7]:
```python
# Total number of success in the given dataset
# number of trials
n = 100
success = np.sum(data)
x, y =  likelihood_Ber(counts=success, n=100)
# getting the index value where likelihood is maximum
idx = np.argmax(y)
plt.plot(x, y)
plt.axvline(x[idx], ls='--', c='orange', label= f'MLE estimator = {x[idx]}')
plt.xlabel("p")
plt.ylabel("Likelihood")
plt.title("Probability vs Likelihood function")
plt.legend()
plt.tight_layout()
```



0.2 Parameter estimation using Binomial distribution

0.2.1 Theroretical method

Considering number of success is the sum of n i.i.d. Bernoulli trials, which is then a single Binomial random variable.

$X = X_1 + X_2 + X_3 + \ldots + X_n$

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

the MLE based on n i.i.d. Bernoulli random variables and the MLE based on a single binomial random variable are the same, since the binomial is the result of n independent Bernoulli trials.

0.2.2 Parameter estimation using Binomial distribution

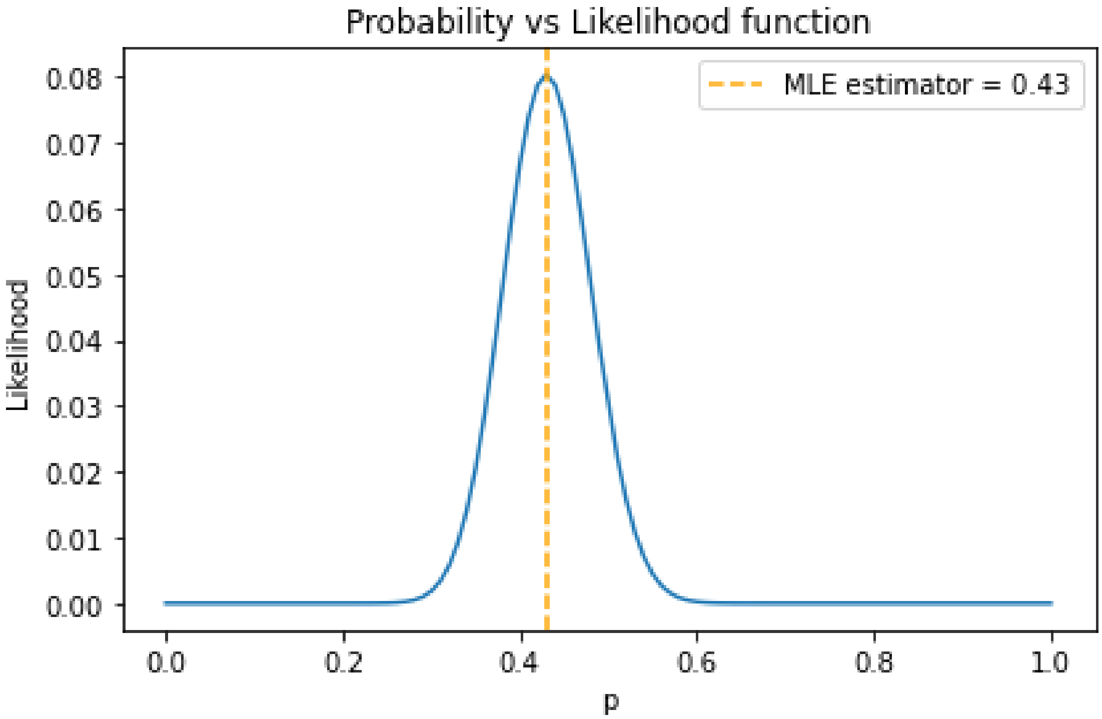
[8]:
```
theta_estimate_Binom = np.mean(data)
print(f'The estimate for p is {theta_estimate_Binom}')
```

```
The estimate for p is 0.43
```

Since the likelihood function is same for b

[9]:
```
from math import factorial
def likelihood_Binom(counts, n=100):
    """
    counts: total number of success
    n: number of trials
    """
    p = np.round(np.linspace(0,1,100), 2)
    #likelihood function
    nCk = factorial(n)/(factorial(counts)*factorial(n-counts))
    L = nCk*p**counts * (1-p)**(n-counts)
    return p, L
```

[10]:
```
# Total number of success in the given dataset
# number of trials
n = 100
success = np.sum(data)
x, y =  likelihood_Binom(counts=success, n=n)
# getting the index value where likelihood is maximum
idx = np.argmax(y)
plt.plot(x, y)
plt.axvline(x[idx], ls='--', c='orange', label= f'MLE estimator = {x[idx]}')
plt.xlabel("p")
plt.ylabel("Likelihood")
plt.title("Probability vs Likelihood function")
plt.legend()
plt.tight_layout()
```



The MLE based on n i.i.d. Bernoulli random variables and the MLE based on a single binomial random variable are the same, since the binomial is the result of n i.i.d. Bernoulli trials. The difference between the likelihood function graph for Bernoulli and Binomial is only in the value of likelihood, which is due to the constant term (nCk) in the Binomial distribution.

[ ]:

```
```