Likelihoods

Rebecca C. Steorts

This material loosely follows Chapter 2 of Roback and Legler.

Reading

BMLR: Chapter 2:
https://bookdown.org/roback/bookdown-BeyondMLR/

Topics

- What is a likelihood
- Principle of a maximum likelihood estimator
- How to obtain a maximum likelihood estimator

Notation

Consider observed data $x_{1:n}$ and a fixed, but unknown parameter θ.

Our data can come from any type of distribution. Let's consider a situation where the data is not normally distributed.

Example

Assume we observe one coin flip (observed data) and a success is the coin landing on heads.

What is the distribution of our data?

Distribution of our data

Each trial can be summarized as a Bernoulli coin flip with unknown parameter θ.

The probability mass function is given by

$$P(X = x \mid \theta) = \theta^x (1 - \theta)^{1-x}.$$

Likelihood

A likelihood is function that tells us how likely we are to observe our data for a given parameter value θ.

Likelihood Function

The likelihood function for the Bernoulli distribution becomes

$$L(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

Example

Suppose we observe 3 heads and 7 tails. We do not know θ.

The likelihood for this example is as follows:

$$L(\theta) = \theta^3(1-\theta)^7 \tag{1}$$

Graphical Maximum Likelihood Estimate

To graphically approximate the Maximum Likelihood Estimator (MLE) of θ for the Bernoulli distribution, we can visually identify the value of θ where the likelihood function reaches its maximum.

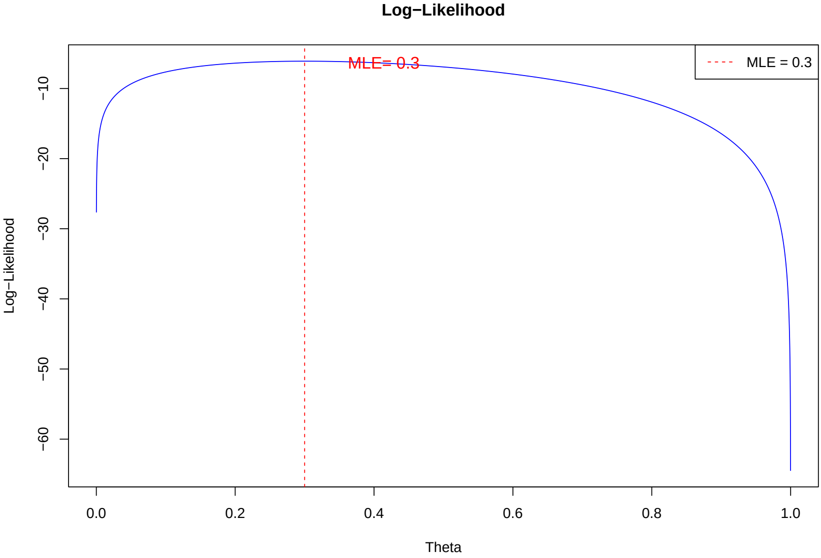
Since the Bernoulli distribution's likelihood function is unimodal (it has a single peak), the MLE corresponds to the value of θ that maximizes the likelihood function.

Steps

1. Set up a grid of θ values. We will create a sequence of values between 0 and 1 to evaluate the log-likelihood function at each point.
2. Evaluate the log-likelihood function at each point: For each θ value in the grid, we will compute the log-likelihood for the observed data.
3. Find the value of θ that maximizes the log-likelihood. The value of θ that gives the highest log-likelihood is the MLE.

Grid Search

```
## MLE for theta: 0.30003
```



Finding the MLE using Calculus

A more general way to find the MLE is using calculus, which provides us with a generalized solution.

Perhaps think about why we would not want to perform a grid-search in practice? Think about potentially computational issues!

Finding the MLE

Recall that

$$L(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$$

Consider the log-likelihood

$$\begin{aligned} \ell(\theta) &= \log\left[\theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}\right] && (2) \\ &= \sum_i x_i \log\theta + (n - \sum_i x_i)\log(1-\theta) && (3) \end{aligned}$$

Finding the MLE

Recall that

$$\ell(\theta) = \sum_i x_i \log\theta + (n - \sum_i x_i) \log(1-\theta) \tag{4}$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\sum_i x_i}{\theta} - \frac{(n - \sum_i x_i)}{(1-\theta)} =: 0$$

Now, we solve for θ.

Finding the MLE

$$
\begin{aligned}
\frac{\sum_i x_i}{\theta} &= \frac{(n - \sum_i x_i)}{(1-\theta)} && (5)\\
\sum_i x_i(1-\theta) &= (n - \sum_i x_i)\theta && (6)\\
\sum_i x_i - \theta \sum_i x_i &= n\theta - \theta \sum_i x_i && (7)\\
\sum_i x_i &= n\theta - \theta \sum_i x_i + \theta \sum_i x_i && (8)\\
\sum_i x_i &= n\theta && (9)\\
\theta &= \frac{1}{n} \sum_i x_i && (10)
\end{aligned}
$$

Second derivative check

The second derivative of $\ell(\theta)$ is

$$-\frac{\sum_i x_i}{\theta^2} - \frac{n - \sum_i x_i}{(1-\theta)^2}$$

Because θ is between 0 and 1, both terms are negative, so the second derivative is negative, confirming that the critical point corresponds to a maximum.

Circuling back to our example

Thus, for our particular example, we can calculate the MLE directly (and can do so for any problem moving forward instead of performing a grid search).

Practice with the Poisson Distribution

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

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

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

Exercise (goes with Homework 3)

1. Write out the likelihood for the Poisson distribution for $x_{1:n}$.
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.
3. Verify using a grid-search that your solution matches to the calculus based one. Assume $\sum_{i=1}^{n} x_i = 500.$ and $n = 100.$ In this example, you should find that $\hat{\lambda}_{MLE} \approx 5.$

Discrete Likelihoods

Please make sure to read through Chapter 2 regarding a different treatment of likelihoods and you may find these slides helpful.