50008 - Probability and Statistics - Lecture 9

Oliver Killane

08/03/22

Lecture Recording

Lecture recording is available here

Maximum Likelihood Estimate

Given some distribution with an unknown parameter θ:

$$X \sim Distribution(\dots\theta\dots)$$

And a sample taken from the distribution $\underline{X}$:

$$\underline{X} = (X_1, X_2, \dots, X_n)$$

We want to know the value of θ for which the likelihood of the sample occurring is highest.

Definition: Likelihood Function

The likelihood of some observations $x_1, x_2, \dots, X_n$ occurring given some θ are:

$$\begin{aligned} L(\theta) &= P(x_1, x_2, \dots, x_n|\theta) \\ &= \prod_{i=1}^{n} f(x_i|\theta) \end{aligned}$$

This is as f is the **probability mass function**, and as each observation is independent we can multiply their probabilities.

Definition: Log Likelihood Function

Used more often than likelihood (easier to work with, and converts decimal small values to large negative values - avoids floating point errors)

$$l(\theta) = \ln L(\theta)$$

To do this, we construct the likelihood (or log likelihood) function from the distribution and sample in term of θ.

Then we can differentiate the function to determine the value of θ for the maximum.

This value of θ is the **Maximum Likelihood Estimate** ($\hat{\theta}$).

Common Maximum Likelihood Estimates

Given a sample $\underline{x} = (x_1, x_2, \dots, x_n)$, we can use formulas for the maximum likelihood.

Exponential Distribution

$$X \sim Exp(\theta) \Rightarrow f(x) = \theta e^{-\theta x}$$

First we determine the **likelihood** in terms of θ.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\ &= \theta^n \prod_{i=1}^{n} e^{-\theta x_i} \\ &= \theta^n e^{-\theta \sum_{i=1}^{n} x_i} \end{aligned}$$

Next we can derive the **log likelihood**

$$\begin{aligned} l(\theta) &= \ln L(\theta) \\ &= \ln \left(\theta^n e^{-\theta \sum_{i=1}^{n} x_i}\right) \\ &= n \ln \theta - \theta \sum_{i=1}^{n} x_i \end{aligned}$$

Next we can differentiate and set equal to zero:

$$\begin{aligned} \frac{dl(\theta)}{d\theta} &= n\frac{1}{\theta} - \sum_{i=1}^{n} x_i = 0 \\ 0 &= \frac{n}{\theta} - \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i &= \frac{n}{\theta} \\ \theta &= \frac{n}{\sum_{i=1}^{n} x_i} \end{aligned}$$

Hence the maximum likelihood estimator is the reciprocal of the mean of the sample.

$$\hat{\theta} = {}^{1}\!/_{\bar{x}}$$

Geometric Distribution

$$X \sim Geo(\theta) \Rightarrow f(x) = \theta(1-\theta)^{x-1}$$

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

Now we find the **log likelihood**.

$$
\begin{aligned}
l(\theta) &= \ln L(\theta) \\
&= \ln\left(\theta^n (1-\theta)^{\left(\sum_{i=1}^{n} x_i\right) - n}\right) \\
&= n \ln\theta + \left(\left(\sum_{i=1}^{n} x_i\right) - n\right) \ln(1-\theta)
\end{aligned}
$$

Now we differentiate, and set equal to zero to find $\hat{\theta}$.

$$
\begin{aligned}
\frac{dl(\theta)}{d\theta} &= \frac{n}{\theta} + \left(\left(\sum_{i=1}^{n} x_i\right) - n\right) \frac{1}{\theta - 1} = 0 \\
0 &= \frac{n(\theta-1)}{\theta(\theta-1)} + \left(\left(\sum_{i=1}^{n} x_i\right) - n\right) \frac{\theta}{\theta(\theta-1)} \\
0 &= n(\theta-1) + \left(\left(\sum_{i=1}^{n} x_i\right) - n\right) \theta \\
0 &= n\theta - n + \left(\left(\sum_{i=1}^{n} x_i\right) - n\right) \theta \\
n &= \left(\sum_{i=1}^{n} x_i\right) \theta \\
\frac{n}{\sum_{i=1}^{n} x_i} &= \theta
\end{aligned}
$$

Hence the maximum likelihood estimator is the reciprocal of the mean of the sample.

$$\hat{\theta} = {}^{1}\!/_{\overline{x}}$$

Binomial Distribution

$$X \sim Binomial(m, \theta) \Rightarrow f(x) = \binom{m}{x} \theta^x (1-\theta)^{m-x}$$

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(x_i) \\
&= \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i}(1-\theta)^{m-x_i} \\
&= \prod_{i=1}^{n} \binom{m}{x_i} \times \prod_{i=1}^{n} \theta^{x_i} \times \prod_{i=1}^{n} (1-\theta)^{m-x_i} \\
&= \prod_{i=1}^{n} \binom{m}{x_i} \times \theta^{\sum_{i=1}^{n} x_i} \times (1-\theta)^{\sum_{i=1}^{n} m-x_i} \\
&= \prod_{i=1}^{n} \binom{m}{x_i} \times \theta^{\sum_{i=1}^{n} x_i} \times (1-\theta)^{mn-\sum_{i=1}^{n} x_i}
\end{aligned}
$$

Now we find the **log likelihood**.

$$
\begin{aligned}
l(\theta) &= \ln L(\theta) \\
&= \ln \left(\prod_{i=1}^{n} \binom{m}{x_i} \times \theta^{\sum_{i=1}^{n} x_i} \times (1-\theta)^{mn-\sum_{i=1}^{n} x_i} \right) \\
&= \ln \prod_{i=1}^{n} \binom{m}{x_i} + \ln \theta^{\sum_{i=1}^{n} x_i} + \ln(1-\theta)^{mn-\sum_{i=1}^{n} x_i} \\
&= \ln \prod_{i=1}^{n} \binom{m}{x_i} + \sum_{i=1}^{n} x_i \ln \theta + \left(mn - \sum_{i=1}^{n} x_i \right) \ln(1-\theta)
\end{aligned}
$$

Now we differentiate, and set equal to zero to find $\hat{\theta}$.

$$
\begin{aligned}
\frac{dl(\theta)}{d\theta} &= 0 + \sum_{i=1}^{n} x_i \frac{1}{\theta} + \left(mn - \sum_{i=1}^{n} x_i \right) \frac{1}{\theta - 1} = 0 \\
0 &= \sum_{i=1}^{n} x_i \frac{\theta - 1}{\theta(\theta - 1)} + \left(mn - \sum_{i=1}^{n} x_i \right) \frac{\theta}{\theta(\theta - 1)} \\
0 &= \sum_{i=1}^{n} x_i(\theta - 1) + \left(mn - \sum_{i=1}^{n} x_i \right) \theta \\
0 &= \theta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i + mn\theta - \theta \sum_{i=1}^{n} x_i \\
0 &= -\sum_{i=1}^{n} x_i + mn\theta \\
\frac{\sum_{i=1}^{n} x_i}{mn} &= \theta
\end{aligned}
$$

Hence the maximum likelihood estimator is the sample mean divided by the number of trials (for binomial):

$$
\hat{\theta} = \frac{\overline{x}}{m}
$$