Discrete Response Model
Lecture 1

datascience@berkeley

Maximum Likelihood Estimation (2)

Finding MLE in General

In general, the MLE can be found as follows:

1. Find the natural log of the likelihood function, $\log\left[L(\pi \mid y_1,\ldots,y_n)\right]$
2. Take the derivative of $\log\left[L(\pi \mid y_1,\ldots,y_n)\right]$ with respect to π.
3. Set the derivative equal to 0 and solve for π to find the maximum likelihood estimate. Note that the solution is the maximum of $L(\pi \mid y_1,\ldots,y_n)$ provided certain "regularity" conditions hold (see Mood, Graybill, Boes, 1974).

For the field goal example:

$$\log\left[L(\pi \mid y_1,\ldots,y_n)\right] = \log\left[\pi^{\sum_{i=1}^{n} y_i}(1-\pi)^{n-\sum_{i=1}^{n} y_i}\right]$$

$$= \sum_{i=1}^{n} y_i \log(\pi) + \left(n - \sum_{i=1}^{n} y_i\right)\log(1-\pi)$$

where log means natural log.

$$\frac{\partial \log\left[L(\pi \mid y_1,\ldots,y_n)\right]}{\partial \pi} = \frac{\sum_{i=1}^{n} y_i}{\pi} - \frac{n - \sum_{i=1}^{n} y_i}{1-\pi} = 0$$

Finding MLE in General

$$\Rightarrow \frac{\sum_{i=1}^{n} y_i}{\pi} = \frac{n - \sum_{i=1}^{n} y_i}{1-\pi}$$

$$\Leftrightarrow \frac{1-\pi}{\pi} = \frac{n - \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i}$$

$$\Leftrightarrow \frac{1}{\pi} = \frac{n - \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i}$$

$$\Rightarrow \boxed{\pi = \frac{\sum_{i=1}^{n} y_i}{n}}$$

Therefore, the maximum likelihood estimator of π is the proportion of field goals made. To avoid confusion between a parameter and a statistic, we will denote the estimator as $\hat{\pi} = \sum_{i=1}^{n} y_i / n$.

Again, Maximum likelihood estimation will prove to be extremely important in this class, as it is used in pretty much all of the statistical models we will study.

Properties of MLE

Why are we interested in MLE?

It comes with many desirable statistical properties.

$\hat{\pi}$ will vary from sample to sample. We can mathematically quantify this variation for maximum likelihood estimators in general as follows:

- Asymptotic normality: For a large sample, maximum likelihood estimators can be treated as normal random variables.
- For a large sample, the variance of the maximum likelihood estimator can be computed from the second derivative of the log likelihood function.

Properties of MLE

Thus, in general for a maximum likelihood estimator $\hat{\theta}$ for θ, we can say that

$$\hat{\theta} \sim N\left(\theta, Var(\hat{\theta})\right)$$

for a large sample $Y_1, \ldots, Y_n$, where

$$Var(\hat{\theta}) = -\left[E\left(\frac{\partial^2 \log[L(\theta \mid Y_1, \ldots, Y_n)]}{\partial \theta^2}\right)\right]^{-1}\Bigg|_{\theta=\hat{\theta}}$$

The use of "for a large sample" can also be replaced with the word "asymptotically." You will often hear these results talked about using the phrase "asymptotic normality of maximum likelihood estimators."

Properties of MLE

Thus, in general for a maximum likelihood estimator $\hat{\theta}$ for θ, we can say that

$$\hat{\theta} \sim N\left(\theta, Var(\hat{\theta})\right)$$

for a large sample Y_1, …, Y_n, where

$$Var(\hat{\theta}) = -\left[E\left(\frac{\partial^2 \log[L(\theta \mid Y_1,\ldots,Y_n)]}{\partial\theta^2}\right)\right]^{-1}\Bigg|_{\theta=\hat{\theta}}$$

The use of "for a large sample" can also be replaced with the word "asymptotically." You will often hear these results talked about using the phrase "asymptotic normality of maximum likelihood estimators."

Berkeley SCHOOL OF INFORMATION