STAT 231 — Lecture 9

Bartosz Antczak *Instructor: Michael Wallace* *September 27, 2017*

9.1 Maximum Likelihood Estimate

Let's refer back to our example from the last lecture. Our probability function is

$$P(Y = 10; \theta) = \binom{25}{10} \theta^{10}(1 - \theta)^{15}$$

We see that the maximum probability will occur when $\widehat{\theta} = 0.4$. The approach we used to estimate θ is called the **method of maximum likelihood**. This is the most widely used method of estimation among statisticians.

9.1.1 Likelihood Function

The **likelihood function** for θ is defined as

$$L(\theta) = L(\theta; y) = P(Y = y; \theta)$$

We interpret the likelihood function as the probability we observe the data y as a function of θ. We determine how plausible an estimate for θ is based on one criteria: values of θ which make the observed data probable are considered to be more reasonable, and this is determined by finding values of θ that maximize $L(\theta)$.

9.1.2 Relative Likelihood

$$R(\theta) = \frac{L(\theta)}{L(\widehat{\theta})}$$

For our recurring example we refer to, we determined that $\widehat{\theta} = 0.4$ for $y = 10$ and $n = 25$. This means our relative function is

$$R = \frac{\theta^{10}(1 - \theta^{15})}{0.4^{10}0.6^{15}}$$

(Note: constants don't affect the relative probabilities and so they are cancelled out in the fraction).

9.1.3 Log Likelihood Function

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

The log likelihood function is maximized for the same value of θ as the regular likelihood function, but often times taking the logarithm makes our algebra much easier.
The graph of our log likelihood function is typically quadratic.

9.1.4 Summary

Let $\widehat{\theta}$ denote the maximum likelihood estimate of a population parameter θ for a particular dataset. Then:

- $\widehat{\theta}$ maximizes $L(\theta)$
- $\widehat{\theta}$ maximizes $R(\theta)$ and $R(\widehat{\theta}) = 1$
- $\widehat{\theta}$ maximizes $\ell(\theta)$