GENERAL ASSEMBLY

MAXIMUM LIKELIHOOD ESTIMATION

Matt Brems

Data Science Immersive, GA DC

LEARNING OBJECTIVES

- Define point estimation.
- Define likelihood and common distributions associated with likelihood.
- Understand the mathematical underpinnings of maximum likelihood estimation.
- Find MLEs through simulation.

OPENING: POINT ESTIMATION

- An estimator is a function of a sample. (This can also be called a statistic.)
  - Ex: an estimator of the population mean μ is the sample mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

- An estimate is the realized value of an estimator given a particular sample.
  - Ex: an estimate of the population mean μ is the realized value $\bar{x} = 6$.

- "Point estimation" describes the process by which we develop estimators (statistics) to estimate parameters.

OPENING: METHODS OF POINT ESTIMATION

- Sometimes we have really intuitive estimators.
  - Population mean can be estimated by the sample mean.
  - Population standard deviation can be estimated by the sample standard deviation.
  - Population β_0 for a regression model can be estimated by $\hat{\beta}_0$.

- Sometimes we don't. What then?
  - Method of Moments
    - Moments: $\mu_1 = \int_{-\infty}^{\infty} x f(x) dx\,; \mu_n = \int_{-\infty}^{\infty} (x - \mu_1)^n f(x) dx$
    - Process: Find sample moments, find population moments, equate them, solve equations.
  - Bayes Estimators
    - Find posterior distribution of θ, use $E[\theta|y]$ (or something similar).
  - Maximum Likelihood Estimation
  - EM Algorithm (Expectation-Maximization)
    - Replaces MLE method with easier maximizations whose limit is the true answer.

MAXIMUM LIKELIHOOD

LIKELIHOOD

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} = \frac{L(\theta|y)f(\theta)}{f(y)} \propto L(\theta|y)f(\theta)$$

- Recall that the likelihood $L(\theta|y)$ refers to the likelihood of the parameter θ given our data y.

- By definition, $L(\theta|y) = f(y|\theta)$, where $f(y|\theta)$ is the probability density function associated with a random variable Y and where θ is the parameter that defines the pdf f.

- Our goal is to identify the value of θ that maximizes the likelihood.

LIKELIHOOD

- Our goal is to identify the value of θ that maximizes the likelihood.

- Find Maximum Likelihood Estimator $\hat{\theta}$ of θ.
  - Step 1: Write out likelihood $L(\theta|y)$.
  - Step 2: Take derivative with respect to θ.
  - Step 3: Set equal to zero.
  - Step 4: Solve for θ.

PROPERTIES OF MLEs

- MLEs are “consistent.”
  - As $n \to \infty$, $\hat{\theta} \to \theta$.
- MLEs are “asymptotically normal.”
  - As $n \to \infty$, $\hat{\theta} \sim N\big(\theta, I(\theta)\big)$, where $I(\theta) =$ Fisher information matrix.
- MLEs are “efficient.”
  - No consistent estimator has a lower asymptotic MSE than the MLE.
- If $\hat{\theta}$ is the MLE of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$.