Methods of Estimation

BIOS 6611

CU Anschutz

Week 2

Methods of Estimation

Motivation

Once we have identified or assumed a certain probability distribution (e.g., normal, binomial, Poisson, etc.) for our population, we will need to identify methods to estimate the parameters from a sample.

Many different approaches exist, but in BIOS 6611 we will encounter and focus on three:

1. Methods of Moments
2. Maximum Likelihood Estimation
3. Least Squares Estimation

Other methods include M-estimators, maximum a posteriori (MAP), restricted maximum likelihoods, and more.

Our goal in this lecture is not to learn how to necessarily derive each estimation method, but to be aware of the terminology and general approach.

Methods of Moments

Methods of Moments

Methods of Moments (MoM) are based on the powers of our random variable, X_i.

The sample mean is a function of X_i^1: $\bar{X} = \sum_{i=1}^{n} \frac{X_i^1}{n}$.

The sample variance is a function of X_i^2: $s^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n}$

The sample skewness is a function of X_i^3: $\sum_{i=1}^{n} \frac{(X_i - \bar{X})^3}{n}$.

MoM estimators are consistent, but often biased. MoMs offer a more straightforward approach and compared to maximum likelihood estimation, but may result in estimates that are outside the parameter space.

Maximum Likelihood Estimation

Terminology

Before reviewing maximum likelihood estimation, a quick review of terminology is helpful.

If $X_1, X_2, ..., X_n$ follow the same distribution $f_X(x;\theta)$, then the **likelihood function** for a sample of n independent and identically distributed (IID) observations, $x_1, x_2, ..., x_n$ is

$$L(\theta) \propto \prod_{i=1}^{n} f_X(x_i;\theta)$$

θ represents the population parameter(s) that define the distribution (e.g., μ and σ^2 for a normal distribution, λ for a Poisson, etc.).

The distribution $f_X(x;\theta)$ represents our probability mass/density function for discrete/continuous random variables (X_i). The fixed values of θ are unknown.

Maximum Likelihood Estimation

Once we take a sample of $x_1, ..., x_n$, then we have our joint density and can estimate the unknown θ parameter(s) from our likelihood function.

By maximizing the likelihood function with respect to our unknown θ we obtain $\hat{\theta}$ ("theta-hat"), the value of the population parameter that makes the data most likely to have been observed.

The steps to generally take are:

1. Take the first derivative with respect to θ and set it equal to 0.
2. Solve using numerical methods, sometimes closed forms are possible to derive (i.e., we could do math by hand to derive the estimate).
3. Check the solution is a maximum by taking the second derivative with respect to θ.

$\hat{\theta}$ is known as our **maximum likelihood estimator** (MLE) of θ.

Maximum Likelihood Estimation

The normal distribution has MLEs of $\hat{\mu} = \sum_{i=1}^{n} \frac{X_i}{n}$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \hat{\mu})^2}{n}$. These match our MoM estimators for the sample mean and variance.

MLEs are sometimes not unbiased, but are usually consistent. They also converge more quickly to the true population parameter faster than most MoM estimators.

MLEs also have the smallest variance compared with other estimators, including MoM estimators.

Least Squares Estimation

Ordinary Least Squares

The **ordinary least squares** (OLS) estimation is used for linear regression models. It works to minimize the sums of squares of the differences between observed outcomes and those predicted by the linear regression model. We will do a deep dive into these methods later this semester.

For regression models that are not linear, we can use MLE and weighted variants of OLS to achieve better properties. These will include special topics that we touch on later in the semester.