Classical and Bayesian inference

AMS 132

January 29, 2018

Introduction

- Here, we will develop a relatively simple method of constructing an estimator without having to specify a loss function and a prior distribution.
- The maximum likelihood method was introduced by R. A. Fisher in 1912.

Some Terminology:

- Here, we are going to say that $X_1, \ldots, X_n$ form a random sample from a distribution with p.d.f. or p.f. $f(x \mid \theta)$, where θ is unknown (not random!)

Definition of the Maximum Likelihood Estimator

- Let $X_1, \ldots, X_n$ form a random sample from a distribution with p.f. or p.d.f. $f(x \mid \theta)$. Recall that when $\boldsymbol{x}$ is observed, the likelihood function (the joint p.f. or p.d.f.) is $f_n(\boldsymbol{x} \mid \theta)$.
- For each possible observed vector $\boldsymbol{x}$, we will consider a value of θ for which the likelihood function is a maximum and use this value as an estimate of θ.

Definition (Maximum Likelihood Estimator / Estimate)

For each possible observed vector $\boldsymbol{x}$, let $\delta(\boldsymbol{x}) \in \Omega$ denote a value of $\theta \in \Omega$ for which the likelihood function $f_n(\boldsymbol{x} \mid \theta)$ is a maximum, and let $\widehat{\theta} = \delta(\boldsymbol{X})$ be the estimator of θ defined in this way. The estimator $\widehat{\theta}$ is called a *maximum likelihood estimator* of θ. After $\boldsymbol{X} = \boldsymbol{x}$ is observed, the value $\delta(\boldsymbol{x})$ is called a *maximum likelihood estimate* of θ.

Definition of the Maximum Likelihood Estimator

- The expressions *maximum likelihood estimator* and *maximum likelihood estimate* are abbreviated M.L.E..
- The M.L.E. is required to be an element of the parameter space.
- The M.L.E. of θ might not exist.
- The M.L.E. of θ might not be unique.

Examples

Example (Lifetime of a component)

Suppose that X_1, X_2, and X_3 form a random sample from the exponential distribution with rate θ. Suppose that $X_1 = 3$, $X_2 = 1.5$, and $X_3 = 2.1$ are observed.

- Plot the likelihood function and its logarithm (See R).
- Find the M.L.E. for θ.

Example (Bernoulli sampling, discrete Ω)

Suppose that X follows Bernoulli distribution with probability of success θ. Suppose $\Omega = \{0.1, 0.9\}$. Find the M.L.E. of θ.

Examples

Example (Uniform sampling)

Suppose that $X_1, \ldots, X_n$ form a random sample from a uniform distribution on the interval $[0, \theta]$, where θ in unknown. Find the M.L.E. of θ.

Example (Normal sampling with known mean)

Suppose that $X_1, \ldots, X_n$ form a random sample from a Normal distribution with unknown mean θ and known variance σ^2. Find the M.L.E. of θ.

Properties of the Maximum Likelihood Estimator

Invariance

- Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution that has p.f. or p.d.f. $f(x \mid \theta)$ and we find that the M.L.E for θ is $\widehat{\theta}$. If we are interested in the parameter $\Psi = g(\theta)$, is there a relationship between the M.L.E of θ and the M.L.E. of Ψ?

Theorem (Invariance Property of the M.L.E.)

If $\widehat{\theta}$ is the maximum likelihood estimator of θ and if g is a one-to-one function, then $g(\widehat{\theta})$ is the maximum likelihood estimator of $g(\theta)$.

Proof:

Example (Lifetime of a component)

Suppose that X_1, X_2, and X_3 form a random sample from the exponential distribution with rate θ. Find the M.L.E. of the mean of the random variables.

Properties of the Maximum Likelihood Estimator

Invariance

Theorem (Invariance Property of the M.L.E.)

Let $\widehat{\theta}$ be the maximum likelihood estimator of θ and let $g(\theta)$ be a function of θ, then a maximum likelihood estimator of $g(\theta)$ is $g(\widehat{\theta})$.

Example

Suppose that X_1, X_2, and X_3 form a random sample from the Bernoulli distribution with probability of success θ. Find the M.L.E. of the standard deviation of X.

Properties of the Maximum Likelihood Estimator

Consistency

- Consider an estimation problem in which a random sample is to be taken from a distribution involving a parameter θ.
- Suppose that for every sufficiently large sample size, n, there exists a unique M.L.E. of θ.
- Then, under certain conditions, the sequence of M.L.E. converges in probability to the unknown value of θ, as $n \to \infty$.
- So, under certain general conditions the sequence of Bayes estimators and the sequence of M.L.E. will be very close to each other, and very close to the unknown value of θ.