AID-521 Mathematics for Data Science

Module: Statistics | Lecture: 3

POINT ESTIMATION & PROPERTIES

overview, method of moments, method of maximum likelihood

Introduction -- Parametric Estimation

The Situation

→ The form of the population distribution is known (binomial, normal, etc.)

→ The parameters of the distribution (p for binomial; μ, σ^2 for normal, etc.) are unknown

→ Our objective is to obtain (best possible) estimates of these parameters using the data from our random sample

Introduction -- Parametric Estimation

General Approach

- $X_1, \ldots, X_n$ is a random sample with pdf $f(x \mid \theta_1, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are unknown population parameters.
- Determine statistics $g_i(X_1, \ldots, X_n)$, $i = 1, \ldots, m$ that can be used to estimate the parameters.
- Such statistics are called estimators for the corresponding parameters.
- Evaluation of an estimator for a given set of sample values is called estimate of the parameter.
- Note that each estimator is a random variable (function of sample r.v.s), and therefore has a (sampling) distribution.

Introduction -- Parametric Estimation

Standard Methods Available

- → Method of moments
- → Method of maximum likelihood
- → Bayes' method
- → Generalized method of moments

Some important criteria for choosing a desired point estimator are its bias, consistency, efficiency, and sufficiency.

Method of Moments

Foundational intuition

Sample moments can possibly provide good estimates of corresponding population moments, which are usually functions of population parameters.

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad \approx \quad \mu_k = E[X^k]$$

Method of Moments -- Procedure

There are n population parameters to be estimated.

$\theta = (\theta_1, ..., \theta_n)$

- → Find n population moments μ_k $(k = 1, ..., n)$, each containing one or more parameters $\theta_1, ..., \theta_n$.
- → Write the corresponding sample moments m_k $(k = 1, ..., n)$.
- → We have n moment equations containing n population parameters.
- → Solve the system of equations for parameters θ to obtain estimators $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_n)$.

Method of Moments -- Example

Try yourself

A random sample of size *n*, $Y_1, Y2, ..., Y_n$, is selected from a population having uniform probability density function over the interval $(0, \theta)$, where θ is unknown. Estimate the parameter θ.

Method of Moments -- Example

Try yourself – solution

The first moment for the uniform distribution is

$$\mu_1 = \frac{0+\theta}{2} = \theta/2.$$

The corresponding first sample moment is

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}.$$

Using the method of moments (one equation, one unknown parameter), we have

$$\frac{\theta}{2} = \bar{Y}, \quad \text{or} \quad \theta = 2\bar{Y}.$$

The method-of-moments estimator for θ is $\hat{\theta} = 2\bar{Y}$, which is a function of sample r.v.s (hence a statistic). Replacing $\bar{Y}$ with values computed from a sample $\bar{y}$ then gives an estimate of $\hat{\theta}$ as $2\bar{y}$.

Method of Moments -- Properties

Advantages

- → Often provides estimators when other methods fail to do so (or are harder to obtain).
- → Easy to compute
- → Have some desirable properties

Dis-advantages

- → Not the "best" estimators
- → Need not be unique, and sometimes may even be meaningless

Method of Maximum Likelihood

Foundational intuition

Find the values of the true (or population) parameters that would have most likely produced the data (or sample) that we have observed.

- $\rightarrow$ A sample of outcomes from unbiased coin tosses are observed.
- $\rightarrow$ If $p = 0.8$ (probability of heads), it is unlikely that we would have seen the above sample.
- $\rightarrow$ If $p = 0.5$, it is more likely than $p = 0.8$ to have produced the above sample.
- $\rightarrow$ In fact, among all potential values of p, $p = 0.5$ seems to be the value with the highest likelihood to produce the above sample.
- $\rightarrow$ So, our estimated value of p is 0.5.

Method of Maximum Likelihood

Let $f(x_1, \dots, x_n \mid \theta),\ \theta \in \mathbb{R}^k$, be the joint pdf of r.v.s $X_1, \dots, X_n$ with sample values $x_1, \dots, x_n$.

→ The likelihood function of the observed sample, a function of θ, is given by

$$L(\theta) = L(\theta \mid x_1, \dots, x_n) = f(x_1, \dots, x_n \mid \theta).$$

→ The maximum likelihood estimators (MLEs) are those values of the parameters that maximize the likelihood function with respect to the parameter θ.

$$L(\hat{\theta} \mid x_1, \dots, x_n) == \max_{\theta} L(\theta \mid x_1, \dots, x_n)$$

Method of MLE -- Procedure

- $\rightarrow$ Define the likelihood function $L(\theta)$.
  - $\rightarrow$ Often it is easier to take the natural logarithm (ln) of $L(\theta)$.
- $\rightarrow$ Solve $\frac{d\ L(\theta)}{d\theta} = 0$ or $\frac{d\ \ln(L(\theta))}{d\theta} = 0$ for θ.
- $\rightarrow$ The form of the above solution of θ is the estimator $\hat{\theta}$.

Method of MLE -- Example

Try yourself

$X_1, \ldots, X_n$ is a sample from a normal population $\mathcal{N}(\mu, \sigma^2)$.

μ is unknown and σ^2 is known to be σ_0^2.

Method of MLE -- Example

Try yourself – solution

Let $\theta = \sigma^2$, for ease of notation.

$$L(\mu, \theta = \sigma^2) \ = \ \frac{1}{(2\pi\theta)^{n/2}} \cdot \exp\left(\frac{-\ \sum_{i=1}^{n}(x_i - \mu)^2}{2\theta} \right)$$

$$\ln\ L(\mu, \theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\theta}$$

Only one parameter μ is needed to be estimated. Setting

$$\frac{\partial}{\partial \mu}\ln\ L(\mu, \theta_0 = \sigma_0^2) = \frac{2\sum_{i=1}^{n}(x_i - \mu)}{2\theta_0} \quad = \quad 0,$$

we get $\sum_{i=1}^{n}(x_i - \mu) = 0$, or $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$. Hence, $\hat{\mu} = \bar{X}$.

Method of MLE -- Properties

→ For most cases of practical interest, the performance of MLEs is optimal for large enough data

→ One of the most versatile methods for fitting parametric statistical models to data

→ Have “good” desirable properties

Dis-advantages

→ Computation of optimization can be difficult

→ Need not be unique, and sometimes may even be meaningless

Desirable Properties of Point Estimators

.. in next lecture ..