MM 225 – AI and Data Science

Day 16: Estimation

Instructors: Hina Gokhale, MP Gururajan, N. Vishwanathan

3 SEPTEMBER 2024

Statistical Inference

- From sample to draw conclusions about the population
- Example:
  1. An alloy is supplied to make a airframe component. The maker asks “what is the yield strength of the material?”
  2. Manufacturer has to choose from two processes A and B to produce an alloy. The engineers have to show that one process is better than the other. How will they do that?
- Example 1 refers to the case of Parameter Estimation
- Example 2 refers to the case of Hypothesis Testing.

Some Definitions

- $X_1, X_2, \dots, X_n$ is called a **"random sample"**, if X_i 's are mutually independent of each other and they come from the same probability distribution.
- Let $X_1, X_2, \dots, X_n$ denote a random sample from a population. Then any function of the sample is called **"Statistic".**
- Examples:
  1. Sample Average = $\bar{X} = \frac{1}{n}\sum X_i$ is a statistic.
  2. Sample variance = $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ is a statistic.
  3. Min$\{X_1, X_2, \dots, X_n\}$ = m is a statistic
- Statistic is a random variable.
- Therefore, statistic has a probability distribution. It is called **"Sampling Distribution"**

Point Estimator and Point Estimate

- Example:
- Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ be a random sample
- $\bar{X} = \frac{1}{n}\sum X_i$ is a point estimator of μ – This is a statistic
- If the observed sample is 10, 9, 7, 11 then
  - $\bar{x} = \frac{1}{4}(10 + 9 + 7 + 11) = 9.25$ is the point estimate of μ
- $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ is a point estimator of σ^2
- And $s^2 = 2.92$ is the point estimate of σ^2

Mean and Variance of linear combination of two or more RVs

$X_1, X_2, \ldots, X_p$ are p RVs and $c_1, c_2, \ldots, c_p$ are p constants then,

$$E\left(\sum_{i=1}^{p} c_i X_i\right) = \sum_{i=1}^{p} c_i E(X_i)$$

$$Var\left(\sum_{i=1}^{p} c_i X_i\right) = \sum_{i=1}^{p} c_1^2 Var(X_i) + 2 \sum_{i<j} \sum_{j} c_i c_j Cov(X_i, X_j)$$

Point Estimator of population mean

- Let $X_1, X_2, \dots, X_n$ be a random sample from populaton with mean μ and variance σ^2
- Hence, $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for i = 1, 2, …, n
- Then,

$$E(\bar{X}) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}\sum \mu = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2}\sum Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- It is true for random sample from any distribution with mean μ and variance σ^2

Sampling Distribution of $\bar{X}$: when population is normal

- Note: if X ~ N(μ,σ^2) and Y~ N(ξ, ω^2) are independent then

$$X + Y \sim N(\mu + \xi,\ \sigma^2 + \omega^2)$$

$$aX \sim N(a\mu, a^2\sigma^2)$$

- Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ be a random sample
- Then,

$$E(\bar{X}) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}\sum \mu = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2}\sum Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Hence $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Sampling Distribution of $\bar{X}$: for any population

If $X_1, X_2, \cdots X_n$ are n independent and identically distributed random variables with finite mean μ and finite variance σ^2, then as $\boldsymbol{n \to \infty}$

$$P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < t\right] \approx P[Z \; < t]$$

where, Z ~ N(0,1)

This implies that the random sample may come from any distribution with finite mean and standard deviation: Average of large sample is approximately distributed as Normal with distribution mean μ and distribution variance $\frac{\sigma^2}{n}$.

Standard Error of Point Estimator

- Let $\widehat{\Theta}$ denote a point estimator of θ
- Then standard Deviation of the estimator $\widehat{\Theta}$ is called standard error.

$$SE(\widehat{\Theta}) = \sqrt{Var(\widehat{\Theta})}$$

- Example:
- $\bar{X}$ is an unbiased estimator of μ
- Standard error of $\bar{X}$ is $\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

Bootstrap Standard Error

- What happens if the point estimator is complicated?
- Examples:
  1. Point estimator S of population standard deviation σ.
  2. Point estimator of scale parameter (δ) of Weibull distribution

- Way forward is to use computer-intensive technique called **Bootstrap**.

Bootstrap Algorithm

- Let X ~ f(θ) with known probability distribution f.

- Assume that we have an observed sample $x_1, x_2, \dots, x_n$ from which we have estimated θ as $\hat{\theta}$.
- Now generate bootstrap samples randomly from probability distribution $f(\hat{\theta})$. And repeat this process m times. So we have

$$\text{Bootstrap sample 1: } x_1^1, x_2^1, \dots, x_n^1 \ \text{ with estimate } \widehat{\theta_1}^B$$

$$\text{Bootstrap sample 2: } x_1^2, x_2^2, \dots, x_n^2 \ \text{ with estimate } \widehat{\theta_2}^B$$

$$\vdots$$

$$\text{Bootstrap sample m: } x_1^m, x_2^m, \dots, x_n^m \ \text{ with estimate } \widehat{\theta_m}^B$$

Typically m = 100 or 200.

$$\text{Bootstrap sample 1: } x_1^1, x_2^1, \dots, x_n^1 \ \text{ with estimate } \widehat{\theta_1}^B$$

$$\text{Bootstrap sample 2: } x_1^2, x_2^2, \dots, x_n^2 \ \text{ with estimate } \widehat{\theta_2}^B$$

$$\vdots$$

$$\text{Bootstrap sample m: } x_1^m, x_2^m, \dots, x_n^m \ \text{ with estimate } \widehat{\theta_m}^B$$

Typically m = 100 or 200.

Then sample bootstrap mean = $\bar{\theta}^B = \frac{1}{m}\sum_{i=1}^{m} \widehat{\theta_i}^B$

And bootstrap standard error of $\hat{\theta} = SE_B(\hat{\theta}) = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left[\widehat{\theta_i}^B - \bar{\theta}^B\right]^2}$

sometimes (m-1) in the denominator is replaced by m

Maximum Likelihood Estimator

Let $X_1, X_2, \ldots, X_n$ be a Random sample of size n from F(θ)

Thus $X_1, X_2, \ldots, X_n$ are independently and identically distributed as F(θ).

Let $f_{X_i}(x_i; \theta)$ be pdf of X_i , $i = 1, 2, \cdots . n$

Hence, the joint pdf of $X_1, X_2, \ldots, X_n$ can be given by

$$f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

Here $x_1, x_2, \cdots, x_n$ are the realisation of random sample $X_1, X_2, \ldots, X_n$.

It is important to realise that the joint pdf contains all available information on the parameter θ.

For the observed values $x_1, x_2, \cdots, x_n$, the joint density function $f(x_1, x_2, \cdots, x_n \mid \theta)$ is a function of unknown parameter θ.

As this function contains all available information on the parameter θ, it is called as Likelihood Function of parameter θ and is denoted by L(θ; $x_1, x_2, \cdots, x_n$)

$$L(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

Maximum Likelihood Estimator of θ is defined as that estimator of θ which maximises likelihood function L(θ; $x_1, x_2, \cdots, x_n$).

L(θ; $x_1, x_2, \cdots, x_n$) and log(L(θ; $x_1, x_2, \cdots, x_n$)) have their maximum at the same value, it is common to maximize log-likelihood function to obtain ML estimate.

ML estimate of parameter θ is denoted by $\hat{\theta}$

Example: MLE for Bernoulli parameter

Consider n independent Bernoulli trials $X_1, X_2, \ldots, X_n$, with p = probability of success, and

$$X_i = \begin{cases} 1 \; if \; trial \; i \; is \; success \\ 0 \qquad\qquad Otherwise \end{cases}$$

Thus $P[X_i = x] = p^x(1-p)^{1-x}$, for x = 0, 1 : p is the unknown parameter. Want to find MLE for p.

For data $x_1, x_2, \cdots, x_n$ likelihood function of p is

$$L(p;\; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$L(p; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Taking log of both sides

$$log\big(L(p; x_1, x_2, \cdots, x_n)\big) = \sum_{i=1}^{n} x_i \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p)$$

Taking the derivative with respect to p and equating to 0 we get

$$\frac{\sum_{i=1}^{n} x_i}{\hat{p}} = \frac{(n - \sum_{i=1}^{n} x_i)}{(1-\hat{p})}$$

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Thank you