Packet 3: Point Estimation

In this section, we assume the function form of the p.d.f. or p.m.f. is known but the distribution depends on an unknown parameter theta that may have any value in the parameter space.

Learning objects:

- Understand several criteria to evaluate point estimators.
- Learn and use the method of moments.
- Learn and use the maximum likelihood method.

Point estimator / Point estimate

In estimation, we take a random sample $\{X_1, X_2, \ldots, X_n\}$ with the observed values $x = \{x_1, x_2, \ldots, x_n\}$ to infer the unknown parameter. n is called sample size.

We define a statistic as a function of the random sample.

If we use a statistic $g(X_1, X_2, \ldots, X_n)$ to estimate a parameter θ, then the statistic is called a point estimator of θ, denoted by

$$\hat{\theta} = g(X_1, X_2, \ldots, X_n).$$

the statistic will change if we draw another random sample. so its randomness comes from the sampling process.

Some facts of a statistics:

- A statistic $g(X_1, X_2, \ldots, X_n)$ is a function of $X_1, X_2, \ldots, X_n$, but does not depend on θ; itself is a random variable and has a distribution.

- A point estimator (a statistic) is designed to describe and make inference about the features of a population, i.e. θ.
  Unfortunately, populations are often too large to measure all individuals. So, we draw a random sample from the population and then use the measurements (data) taken on the sample to draw conclusions about the population feature.

  θ : population summary

- A statistic is a summary of the random sample but not the population, so if the sample changes, the value of the statistic will change.

  after observing data

- A point estimate $\hat{\theta} = \hat{\theta}(x) = g(x_1, x_2, \ldots, x_n)$ is computed from the observed data.

Example: For normal distribution, $X_i \sim N(\mu, \sigma^2)$, we would like to estimate the mean parameter μ, and we draw a random sample with sample size =6. Suppose we observe a set of data = 5, 3, 4, 7, 8, 6. Which of the following statements are true? $X_1, X_2 \cdots X_6$

$x_1 \; x_2 \; \cdots \; x_6$

- $\bar{X}$ is a point estimator for μ and 5.5 is the estimate. Yes

$\bar{X} = \frac{1}{6}\sum_{i=1}^{6} X_i$ is a function of $X_1 \cdots X_6$, and is used to estimate μ, so it is an estimator.

- The fourth smallest sample $X_{(4)}$ is a point estimator for μ and 6 is the estimate. Yes

Order statistic is also a function of $X_1 \cdots X_6$, so it is an estimator.

- The first observed sampleX_1 is a point estimator for μ and 5 is the estimate. Yes

X_1 is a function $g(X_1 \cdots X_6)$ it does not neccesary to involve all X_i's.

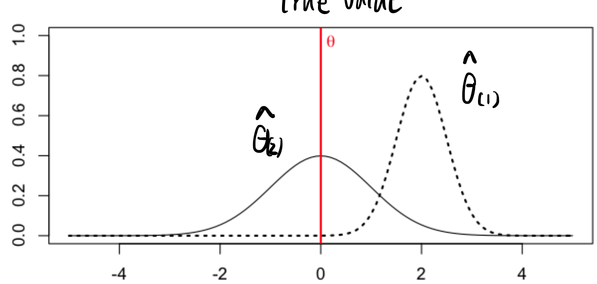
$E(X_1) = \mu$ is unbiased.

Several criteria to evaluate point estimators

Which estimator is better? How good is this estimate? What makes an estimate good?

Unbiasedness: An estimator $\hat{\theta}$ of a parameter θ is *unbiased* if $E(\hat{\theta}) = \theta$, i.e., if the mean of the (sampling) distribution of $\hat{\theta}$ is θ (for all possible values of θ in the parameter space Ω).

true value



$E(\hat{\theta}_{(1)}) > \theta$ biased

$E(\hat{\theta}_{(2)}) = \theta$ unbiased

We also define the bias of $\hat{\theta}$ as

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

If the bias is positive, then $\hat{\theta}$ tends to overestimate θ on average. On the other hand, if the bias is negative, then then $\hat{\theta}$ tends to underestimate θ on average.

$E(\bar{X}) = \mu$ $\quad E(X_1) = \mu$

both unbiased

Consistency: An estimator is a consistent estimator if it converges to θ as the sample size n increases, i.e., $\hat{\theta} \to \theta$ (in probability) as n increases.

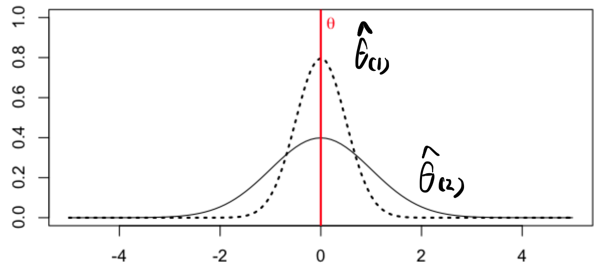
$\bar{X} \sim N(\mu, \sigma^2/n)$ $\quad \bar{X} \to \mu$ as $n \to \infty$ so $\bar{X}$ is consistent $\quad X_1$ is not

Sufficient statistic: The statistic $g(X_1, X_2, \ldots, X_n)$ carries all the information about θ, no other statistic that can provide any additional information as to the value of the θ.

$m(X_1 \cdots X)$

Efficiency: Standard error of $\hat{\theta}$ is small. If there are several unbiased estimators, one with the smallest variance is preferred.

$\hat{\theta}_1$ is preferable b.c. of a small variance



The method of moments (MoM)

In probability, the k-th moment of an r.v. X is defined as

$$\mu_k = E(X^k).$$

Also, when the data $X_1, X_2, \ldots, X_n$ are i.i.d. r.v.s, the k-th sample moment is defined as

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

The MoM (Chebyshev, 1887) is a way to obtain estimators by matching moments of an r.v. with their estimators, i.e., sample moments.

First, we calculate low-order moments and express them in terms of unknown parameters.

For example, if there is only one unknown parameter, computing the first moment (μ_1) might be enough.

If there are two unknown parameters, computing the first two moments (μ_1, μ_2) might be enough.

If needed, we calculate higher-order moments until we have enough equations to solve for the parameters. Note that moments (μ_k's) are functions of parameters (θ),

$$\mu_1 = E(X^1) = g_1(\theta).$$

Example: Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. with p.d.f.

$$f(x) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0.$$

Find the MoM estimator for θ.

Sol: ① $\mu_1 = E(X) = \int_0^1 x f(x)\,dx = \int_0^1 x\,\theta\,x^{\theta-1}\,dx$

$$= \theta \int_0^1 x^{\theta}\,dx = \theta \times \frac{1}{\theta+1} x^{\theta+1} \Big|_0^1 = \frac{\theta}{\theta+1}$$

② Set up the equation

$$\mu_1 \doteq \hat{\mu}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

③ $\frac{\theta}{\theta+1} = \bar{X}$ $\theta = \theta\bar{X} + \bar{X}$

$\theta(1-\bar{X}) = \bar{X}$ $\hat{\theta} = \frac{\bar{X}}{1-\bar{X}}$ MoM estimator for θ

Example: van den Bergh [1985] considers the luminosity (a measure of the radiant power emitted by a star) for globular clusters in various galaxies. In the paper, vdB's conclusion is that the luminosity for clusters in the Milky Way is adequately described by the $N(\mu, \sigma^2)$ distribution, where μ represents the population average brightness and σ is the population standard deviation of brightness. Its p.d.f. is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{for } x \in \mathbb{R}.$$

We are interested in estimating μ and σ^2 using a random sample of n globular clusters. ① Find the first two moments.

$* \; Var(X) = E(X^2) - (E(X))^2$ Chap 5.3

$\mu_1 = E(X) = \mu$

$\mu_2 = E(X^2) = \int x^2 f(x)\,dx = Var(X) + (E(X))^2 = \sigma^2 + \mu^2$

hard to integrate

Second, we invert the equations to write the parameters in terms of moments: $\theta = g_1^{-1}(\mu_1)$.

② Let $\mu_1 = \hat{\mu}_1 = \bar{X}$ $\qquad \mu_2 = \hat{\mu}_2 = \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$\mu = \bar{X}$ $\qquad \sigma^2 + \mu^2 = \overline{X^2}$

Finally, we plug in the sample moments and obtain estimators for the parameters: $\hat{\theta} = g_1^{-1}(\hat{\mu}_1)$

③ $\hat{\mu} = \bar{X}$ $\qquad \sigma^2 + \bar{X}^2 = \overline{X^2} \Rightarrow \hat{\sigma}^2 = \overline{X^2} - \bar{X}^2$

MoM estimator for μ, σ^2

from Chap 5.5 $\quad \bar{X} \sim N(\mu, \sigma^2/n)$

Are the MoM estimators for μ and σ^2 biased?

$E(\hat{\mu}) = \mu$ so it is unbiased

$E(X_i^2) = Var(X_i) + (E(X))^2$
$= \sigma^2 + \mu^2$

$E(\bar{X}^2) = Var(\bar{X}) + (E(\bar{X}))^2$
$= \sigma^2/n + \mu^2$

$E(\hat{\sigma}^2) = E(\overline{X^2}) - E(\bar{X}^2)$

$= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \left[Var(\bar{X}) + (E(\bar{X}))^2\right]$

$= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - (\sigma^2/n + \mu^2)$

$= \sigma^2 + \mu^2 - \sigma^2/n - \mu^2 = \frac{n-1}{n}\sigma^2 < \sigma^2$

so $\hat{\sigma}^2$ is biased but is still consistent