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 θ 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

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

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.

- 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.

- 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?

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

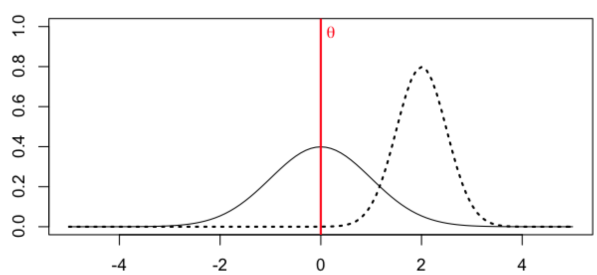
- The fourth smallest sample $X_{(4)}$ is a point estimator for μ and 6 is the estimate.

- The first observed sample X_1 is a point estimator for μ and 5 is the estimate.

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).



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

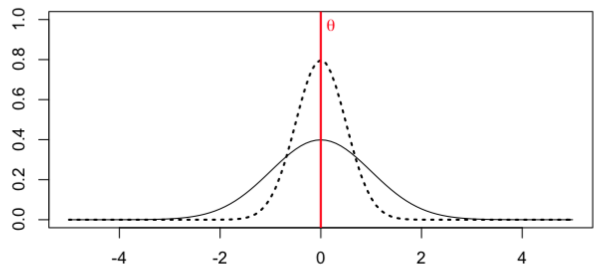
$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

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

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.

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.

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



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.

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. Find the first two moments.

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

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

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