TOPIC 3: POINT ESTIMATION

3.1. KEY ISSUES. PROPERTIES OF POINT ESTIMATORS

KEY ISSUES

- In parametric inference we apply point estimation in order to assign a single value to the unknown parameter
- Some probabilistic models have a single parameter:

$$B(1,p) \qquad \chi^2_n \qquad t_n \qquad P(\lambda)$$

- Other have more than one:

$$B(n,p) \qquad U(a,b) \qquad N(\mu,\sigma) \qquad F_{m,n}$$

- In order to estimate parameters we use estimators
- An estimator is a statistic aimed at giving values to a parameter
- A statistic is any function of sampling data
- An estimator is a random variable before selecting the sample (a priori) and it will become a value once the sample has been obtained (a posteriori)
- The population mean and the population variance are key parameters to estimate
- By analogy, the sample mean, the sample variance and the bias-corrected sample variance are going to be candidates as point estimators
- To know their sampling distribution will be an important matter (already shown in chapter 2)
- In summary, the point estimator process will have the next

steps:

1.- We have a population with one or more unknown parameters (for instance a normal random variable with unknown mean and variance of 4).

2.- We want to assign a value to the unknown parameters (the mean in the previous example)

3.- We think in obtaining a sample $X(x_1, x_2, x_3 \dots x_n)$ based on a probabilistic method. In this subject we will apply *only* s.r.s. We would like the size *n* be big enough (however for simplicity in our example we will take n=4)

4.- Then we choose a point estimator $\hat{\theta}(x_1, x_2, x_3 \dots x_n)$ (the sample mean in our example)

5.- The simple random sample is obtained (in our case: $x_1=1$, $x_2=1$, $x_3=2$, $x_4=2$).

6.- The sample is incorporated in the estimator in order to come up with an estimate (in our example:

$$\hat{\mu}(1,1,2,2) = \bar{x} = \frac{1+1+2+2}{4} = 1.5)$$

7.- That estimate is the value to be assigned to the unknown parameter μ.

❖ However, how to choose the best estimator among the infinite statistics? We will use that estimator verifying certain properties (unbiased estimator, efficient estimator, consistency and other)

❖ Thanks God there are statistical methods to find out estimators fulfilling some or many of those properties

- Two or those methods are the Maximum Likelihood Estimation (MLE) and the Generalized Method of Moments (GMM)

Next we will approach the properties of estimators and then the methods for deriving good estimators.

PROPERTIES OF POINT ESTIMATORS

- UNBIASED ESTIMATOR
  - θ^* is an unbiased estimator of a parameter θ if it verifies, for any n the following rule:

$$E(\theta^*) = \theta$$

  - This property is referred to the behavior of the estimator on average, considering all the values the estimator may take.
  - Otherwise, we have a biased estimator. The bias $\varphi(\theta^*)$ is defined as it follows:

$$\varphi(\theta^*) = E(\theta^*) - \theta$$

  - The bias can be either positive or negative (when the bias is nule the estimator is unbiased)

Some properties:

- Noncentral sampling moments a_r are unbiased estimators of respective noncentral population moments μ_r.
- Un estimator is asymptotically unbiased if the bias approaches cero when n approaches infinitum

$$\lim_{n\to\infty} \varphi\left(\theta^*\right) = 0$$
$$\rightarrow \theta^* \; is\; asymtotically\; unbiased$$

- Let θ_1* and θ_2* be two unbiased estimators of θ. Then a convex linear combination of them will produce another unbiased estimador of θ:

$$\theta_3^* = c \cdot \theta_1^* + (1-c) \cdot \theta_2^*$$

- The following estimators of μ are unbiased:

$$\mu^* = \sum_{i=1}^{n} c_i \cdot x_i \qquad donde \; \sum_{i=1}^{n} c_i = 1$$

x_i being the elements of a s.r.s.

❖ EFFICIENCY

- Efficiency is referred to the variability of the estimator. The lower the dispersion the higher the efficiency, and viceversa, the higher the variability the lower the efficiency.
- Two concepts must be considered: the most efficient estimator and the relative efficiency.
- Among all unbiased estimators of a given parameter, the

Minimum Variance Unbiased Estimator (MVUE) will be that one with the smallest variance. This will be the most efficient estimator.

- Relative efficiency arises when comparing the variances of two estimators of a given parameter
- An estimator θ_1* is more efficient than θ_2* if, for any size *"n"*, the variance of the former is equal or lower than the variance of the latter:

$$V(\theta_1^*) \leq V(\theta_2^*)$$

❖ MEAN SQUARE ERROR (MSE)

- Let θ_1* and θ_2* be two point estimators of θ:

$$\varphi(\theta_1^*) < \varphi(\theta_2^*) \quad but \quad V(\theta_1^*) > V(\theta_2^*)$$

Which estimator must I choose?

- Answer: that one with the lowest MSE:

$$MSE(\theta^*) = E[\theta^* - \theta]^2 = V(\theta^*) + \varphi(\theta^*)^2$$

❖ CONSISTENCY

- θ* is a consistent estimator of θ when it approaches the parameter as the sample size *"n"* increases.
- This property is verified whenever two conditions take place as sample size "*n*" approaches infinitum:

$$if\ n \to \infty \begin{cases} E(\theta^*) \to \theta \\ V(\theta^*) \to 0 \end{cases}$$

- This property must be required whenever efficient unbiased estimators are not available.

3.2. GENERALIZED METHOD OF MOMENTS (GMM)

❖ It assigns as GMM estimator of a parameter, its analogous in the sample

❖ The method is based on the fact that, in general, the noncentral population moment μ_r depends on the vector of unknown parameters:

- In a discrete random variable:

$$\mu_r(\theta_1,...\theta_k) = \sum_{i=1}^{N} X_i^r \cdot P(\xi = X_r)$$

- In a continuous random variable:

$$\mu_r(\theta_1,...\theta_k) = \int_{-\infty}^{+\infty} X^r \cdot f(X, \theta_1,...\theta_k) dX$$

Steps:

❖ First, a s.r.s. of size n is considered

❖ Then the noncentral sampling moments a_r are equalled to their corresponding noncentral population moments μ_r resulting a system of k equations with k unknowns:

$$\mu_1(\theta_1,...\theta_k) = a_1$$
$$.......$$
$$\mu_k(\theta_1,...\theta_k) = a_k$$

❖ Whose solutions give the GMM estimators of the *k* parameters:

$$\theta_1^* = \theta_1^*(a_1,...a_k)$$
$$.......$$
$$\theta_k^* = \theta_k^*(a_1,...a_k)$$

❖ Properties:

- Consistency
- In general, they are neither unbiased nor efficient estimators.

❖ The main advantage of this method is its simplicity.

❖ Disadvage: it does not use all the information in the sample, provided that it ignores the probability distribution of the population under study.

3.3. MAXIMUM LIKELIHOOD ESTIMATORS (MLE)

❖ Maximum likelihood estimation is based on the principle that usually it happens what is most likely to happen

❖ What it happens is the Likelihood Function (LF), which is the sample joint distribution for a given s.r.s. of size "*n*" fluctuating with the parameters values

❖ That estimator maximizing the likelihood function is the MLE, hence producing the highest likelihood for a given s.r.s. of size "*n*", this being what is most likely to happen.

❖ Stages:

- Let ξ be a random variable with probability distribution $P(\xi = x;\theta)$ or with density function $f(x;\theta)$ depending on ξ being either a discrete random variable or a continuous one and being θ an unknown parameter

- A s.r.s. $X(x_1, x_2, x_3 \ldots x_n)$ of size *n* is selected from the population where:

$$x_i \sim i.i.d.\, P(\xi = x;\theta) \quad or \quad x_i \sim i.i.d.\, f(x;\theta)$$

- In the discrete case, the corresponding sample joint distribution will give the probability of that sample:

$$P(X) = P(x_1 \cdots x_n) = P(\xi = x_1;\theta) \cdots P(\xi = x_n;\theta) = f(x_1;\theta) \cdots f(x_n;\theta)$$

- In the continuos case, the sample joint distribution is given by the joint density function:

$$f(X) = f(x_1 \cdots x_n) = f(x_1 \cdots x_n;\theta) = f(x_1;\theta) \cdots f(x_n;\theta)$$

  Do not confuse with the probability of the sample in this case:

$$P(X) = P(x_1 < \xi \le x_1 + dx_1; \cdots, x_n < \xi \le x_n + dx_n) = f(x_1;\theta)dx_1 \cdots f(x_n;\theta)dx_n$$
$$= f(X)dX$$

- Then giving θ different values, either in $P(X)$ or $f(X)$, we will obtain different outcomes for the sample joint distribution and this is what we call LF:

$$L(X;\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdots f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

- Notice that the sample X is fixed and the variable is θ
- In consequence, θ_1 is more likely than θ_2 if LF is higher with θ_1 than with θ_2
- Then MLE is obtained by maximizing LF:

$$\max_{\theta} L(X;\theta)$$

- In order to make calculus easier, we take natural logarithms and maximize this last function:

$$l(X;\theta) = lnL(X;\theta)$$

$$\max_{\theta} l(X;\theta) = \max_{\theta} lnL(X;\theta)$$

- Then we apply the necessary condition for a maximum:

$$\frac{\partial l(X;\theta)}{\partial \theta} = 0$$

- And finally the sufficient condition for a maximum:

$$\frac{\partial^2 l(X;\theta)}{\partial \theta^2} < 0$$

- If the function is not derivable this method cannot be used

❖ Excellent properties in large samples:

- MLEs are asymptotically normally distributed
- MLEs are asymptotically unbiased
- MLEs are asymptotically minimum variance