Estimation (contd…)

Session 2

05/03/2021

Parameter Space

- Consider a random variable (r.v) with p.d.f. $f(x, \theta)$.
- In some cases, we assume the population distribution is known except for the value of the unknown parameter(s) θ which may take any value on a set Θ. This is expressed by writing the p.d.f in the form $f(x, \theta)$, $\theta \in \Theta$.
- The set Θ which is the set of all possible values of θ is called the parameter space.
- Such a situation gives rise not to one probability distribution, but a family of probability distributions which we write as $\{f(x, \theta), \theta \in \Theta\}$.

- For example if X ~ N(μ, σ^2), then the parameter space is $\Theta = \{(\mu, \sigma^2): -\infty < \mu < \infty; 0 < \sigma < \infty\}$.

- In particular, if σ^2=1, the family of probability distributions is given by

  $\{N(\mu, 1); \mu \in \Theta\}$, where $\Theta = \{\mu: -\infty < \mu < \infty\}$.

- In the following discussion, we shall consider a general family of distributions:

  $\{f(x; \theta_1, \theta_2, \dots \theta_k): \theta_i \in \Theta, i = 1,2,3 \dots k\}$

- Consider a random sample $x_1, x_2, \ldots. x_n$ of size n from a population with probability function $f(x; \theta_1, \theta_2, \ldots \theta_k)$ where $\theta_1, \theta_2, \ldots \theta_k$ are unknown population parameters.
- There will always be an infinite number of functions of sample values , called statistic, which may be proposed as estimates of one or more of the parameters.
- Evidently the best estimate would be the one that falls nearest to the true value of the parameter to be estimated.
- In other words, the statistic whose distribution concentrates as closely as possible near the true value of the parameter may be regarded as the best estimate.

Basic problem of estimation formulation

- We wish to determine the functions of the sample observations:

$t_1 = \widehat{\theta_1}(x_1, x_2, \ldots . x_n), t_2 = \widehat{\theta_2}(x_1, x_2, \ldots . x_n), \ldots$
$t_k = \widehat{\theta_k}(x_1, x_2, \ldots . x_n),$

such that their distribution is concentrated as closely as possible near the true value of the parameter. The estimating functions are then referred to as *estimators.*

Definition

- Any function of the random sample $x_1, x_2, \ldots .x_n$ that are being observed , say $t_n(x_1, x_2, \ldots .x_n)$ is called a statistic.
- Clearly statistic is a random variable.
- If it is used to estimate an unknown parameter θ of the distribution, it is called an estimator.
- A particular value of the estimator , say $t_n(x_1, x_2, \ldots .x_n)$ is called an estimate of θ.

Characteristics of a good Estimator

- The following are some of the criteria that should be satisfied by a good estimator.

➢ Unbiasedness

➢ Consistency

➢ Efficiency

➢ Sufficiency.

Unbiasedness

- An estimator $t_n = t(x_1, x_2, \ldots .x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if

  $E(t_n)$=$\gamma(\theta)$, for all $\theta \in \Theta$.
- In other words, suppose we have an estimator $\hat{\theta}$ for θ. Then $\hat{\theta}$ is said to be an unbiased estimator of θ, if $E\left(\hat{\theta}\right)$ = θ for all $\theta \in \Theta$.
- That is the mean of sampling distribution of $\hat{\theta}$ is θ. Otherwise it is said to be biased.

Sampling distribution of a Statistic

- If we draw a sample of n from a given population of size N, then the total number of possible samples is $N_{C_n} = \frac{N!}{n!(N-n)!} = \mathrm{k(say)}$
- For each of these k samples, we can compute some statistic $t=t(x_1,x_2,\ldots x_n)$, in particular the sample mean $\bar{x}$, the variance s^2 etc.as given below.

Sample Number	t	$\bar{x}$	s^2
1	t_1	$\bar{x}_1$	${s_1}^2$
2	t_2	$\bar{x}_2$	${s_2}^2$
…	…	…	…
…	…	…	…
k	t_n	$\bar{x}_k$	${s_k}^2$