Statistics for Engineers

Introduction to Statistical inference

Introduction to Statistical inference

Contents

- Population, sample and random sampling.
- Point estimation of parameters
  - Definitions
  - Method of moments
  - Maximum likelihood
- Fundamental sampling distributions

Keywords

- Population: the complete set of numerical information on a particular quantity in which an investigator is interested.
  - We identify the concept of the population with that of the **random variable** X.
  - The **law or the distribution of the population** is the distribution of X, F_X.
- Sample: an observed subset (say, of size n) of the population values.
  - Represented by a collection of n random variables $X_1, X_2, \ldots, X_n$, typically $\boxed{\textbf{iid (independent identically distributed)}}$.
- Parameter: a constant characterizing X or F_X.

keywords (ii)

- **Statistical inference**: the process of drawing conclusions about a population on the basis of measurements or observations made on a sample of individuals from the population.
- Statistic: a random variable obtained as a function of a random sample, $X_1, X_2, \ldots, X_n$
- Estimator of a parameter: a random variable obtained as a function, say T, of a random sample, $X_1, X_2, \ldots, X_n$, used to estimate the unknown population parameter.
- Estimate: a specific realization of that random variable, i.e., T evaluated at the observed sample, $x_1, x_2, \ldots, x_n$, that provides an approximation to that unknown parameter.

Statistical inference: example

$$
\begin{array}{ccccc}
\text{We want to know} & & \text{We have } n \text{ copies} & & \text{We have } n \\
\mu_X = \mathrm{E}[X] & & \text{of } X & & \text{observed values of} \\
 & & & & X_1, X_2, \ldots, X_n \\
 & & & & \\
 & & X_1, X_2, \ldots, X_n \sim F & & x_1, x_2, \ldots, x_n \\
X \sim F & \Longrightarrow & \text{Sample} & \Longrightarrow & \text{Observed sample} \\
 & & & & \\
\Downarrow & & \Downarrow & & \Downarrow \\
 & & & & \\
 & & \text{Estimator of } \mu_X \text{ (r. v.)} & & \text{Estimate of } \mu_X \text{ (number)} \\
\mu_X = \mathrm{E}[X] & \Longleftarrow & \bar{X} & \Longleftarrow & \bar{x} \\
\text{Expected value of } X & & \text{Sample mean} & & \text{Sample mean}
\end{array}
$$

Point estimators: introduction

- A **point estimator** of a population parameter is a function, call it T, of the sample information $\underline{X}_n = (X_1, \ldots, X_n)$ that yields a single number.
- Examples of population parameters, estimators and estimates:

Population parameter	$T(\underline{X}_n)$	Estimator: notation	Estimate: notation
Pop. mean μ_X	sample mean $\frac{X_1+\ldots+X_n}{n}$	$\bar{X} = \hat{\mu}_X$	$\bar{x}$
Pop. prop. p_X	sample prop.	$\hat{p}_X$	$\hat{p}_x$
Pop. var. σ_X^2	sample var. $\frac{\sum_i X_i^2 - n(\bar{X})^2}{n}$	$\hat{\sigma}_X^2$	$\hat{\sigma}_x^2$
Pop. var. σ_X^2	sample quasi var. $\frac{\sum_i X_i^2 - n(\bar{X})^2}{n-1}$	s_X^2	s_x^2
...	...	...	...
In general, θ_X	...	$\hat{\theta}_X$	$\hat{\theta}_x$

Point estimators: properties (i)

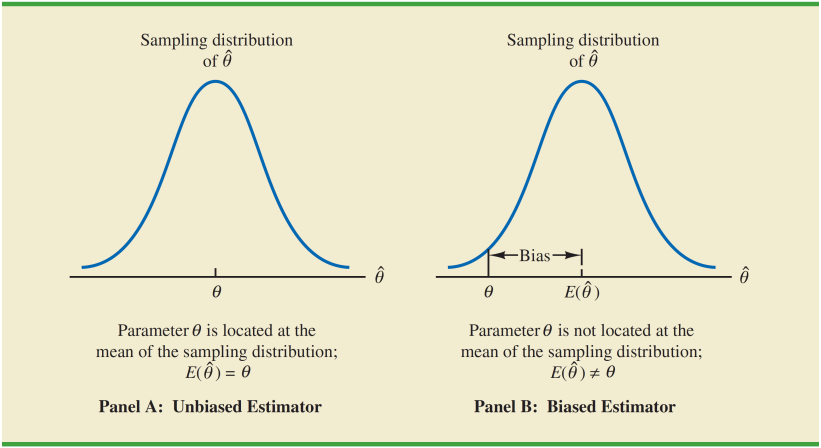
What are desirable characteristics of the estimators?

- **Unbiasedness**. This means that the bias of the estimator is zero. What's bias? **Bias** equals the expected value of the estimator minus the target parameter

$$\mathrm{Bias}[\hat{\theta}_X] = \mathrm{E}[\hat{\theta}_X] - \theta_X$$

Population parameter	Estimator $T(\underline{X}_n)$	Bias	Unbiased?	Minimum Variance Unbiased Estimator?
Pop. mean μ_X	$\overline{X}$	$\mathrm{E}[\bar{X}] - \mu_X = 0$	Yes	Yes, if X normal
Pop. prop. p_X	$\hat{p}_X$	$\mathrm{E}[\hat{p}_X] - p_X = 0$	Yes	Yes
Pop. var. σ_X^2	$\hat{\sigma}_X^2$	$\mathrm{E}[\hat{\sigma}_X^2] - \sigma_X^2 \neq 0$	No	No
Pop. var. σ_X^2	s_X^2	$\mathrm{E}[s_X^2] - \sigma_X^2 = 0$	Yes	Yes, if X normal
In general, θ_X	$\hat{\theta}_X$	$\mathrm{E}[\hat{\theta}_X] - \theta_X$	Often	Rarely

Point estimators: properties (i)



Point estimators: properties (ii)

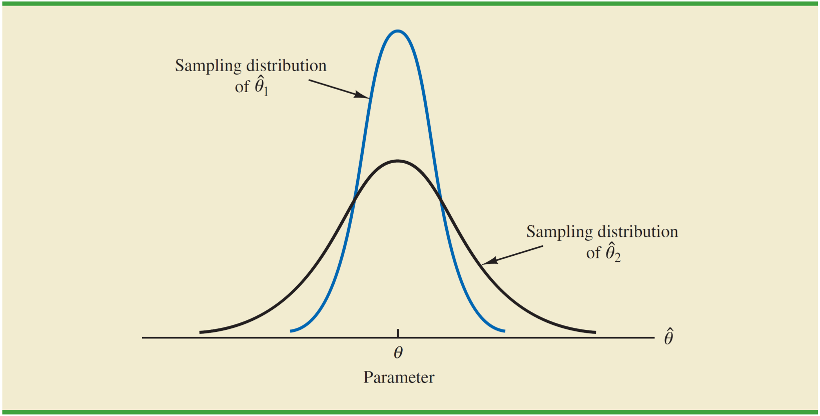
- **Efficiency**. Measured by the estimator's variance. Estimators with smaller variance are more efficient. The standard error is defined as $se \equiv \sigma_{\hat{\theta}} = \sqrt{Var[\hat{\theta}]}$.
- **Relative efficiency** of two unbiased estimators $\hat{\theta}_{X,1}$ and $\hat{\theta}_{X,2}$ of a parameter θ_X is

$$\text{Relative efficiency}(\hat{\theta}_{X,1}, \hat{\theta}_{X,2}) = \frac{\text{Var}[\hat{\theta}_{X,1}]}{\text{Var}[\hat{\theta}_{X,2}]}$$

  Note:

  - sometimes the inverse is used as a definition
  - in any case, an estimator with **smaller** variance is more efficient

Point estimators: properties (ii)



Point estimators: properties (iii)

- A more general criterion to select estimators (among unbiased and biased ones) is the **mean squared error** defined as

  $$\mathsf{MSE}[\hat{\theta}_X] = \mathsf{E}[(\hat{\theta}_X - \theta_X)^2] = \mathsf{Var}[\hat{\theta}_X] + (\mathsf{Bias}[\hat{\theta}_X])^2$$

  Note:
  - the mean squared error of an unbiased estimator equals its variance
  - an estimator with **smaller** MSE is better
  - the minimum variance unbiased estimator has the smallest variance/MSE among all estimators
- How do we come up with the definition of the estimator T?
  - In some situations, there exists an optimal estimator called **minimum variance unbiased estimator**.
  - If that's not the case, there are various alternative methods that yield reasonable estimators, for example:
    - Maximum likelihood estimation
    - Method of moments

Methods of Point Estimation: Method of Moments

- Let $X_1, X_2, ..., X_n$ be a random sample from a probability density function (continuous case) or probability mass function (discrete case) $f(X)$. The kth population moment (or distribution moment) is $E[X^k]$, $k = 1, 2, \ldots$. The corresponding kth sample moment is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.
- Let $X_1, X_2, ..., X_n$ be a random sample from a probability density function (continuous case) or probability mass function (discrete case) with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The moment estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

Methods of Point Estimation: Maximum Likelihood Estimation

Given independent observations $x_1, x_2, ..., x_n$ from a probability density function (continuous case) or probability mass function (discrete case) $f(x;\theta)$, the maximum likelihood estimator $\hat{\Theta}$ is that θ which maximizes the likelihood function:

$$L(\theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$$

- Note that the variable of the likelihood function is θ.
- Quite often it is convenient to work with the natural *log* of the likelihood function in finding the maximum of that function.
- The maximum likelihood estimator is unbiased asymptotically or in the limit.
- The variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator.
- $\hat{\Theta}$ has an approximate normal distribution.

Sampling Distributions

- Since a statistic is a random variable that depends only on the observed sample, it must have a probability distribution.
- The probability distribution of a statistic is called a sampling distribution.
- The sampling distribution of a statistic depends on the distribution of the population, the size of the samples, and the method of choosing the samples.

Sampling Distributions: Sample mean $\overline{X}$

- The sample mean

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

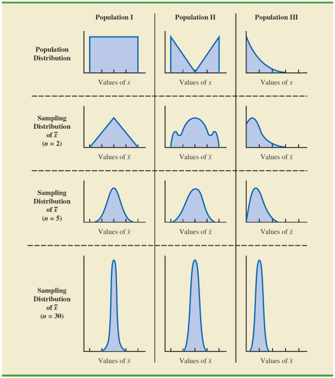
  is a natural estimator of the population mean μ. It is unbiased with variance σ^2/n, where σ is the standard deviation of X.

- In accordance with the CLT, for any distribution of X, whenever the sample size n is sufficiently large

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx N(0, 1)$$

Sampling Distributions: Sample mean $\overline{X}$

- **If the population has a normal distribution**, the sampling distribution of $\bar{x}$ is normally distributed.
- **If the population does not have a normal distribution**, using the Central Limit Theorem we can approximate the sampling distribution of $\bar{x}$ by a normal distribution as the sample size becomes large.



Sampling Distributions: Sample proportion $\hat{p}$

- We denote by p the population proportion of individuals with a certain characteristic. The r.v. X that assumes value 1 on individuals with the characteristic and 0 on the remaining individuals follows a $Be(p)$ and $Bin(1, p)$ distributions.

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

where

$$E[\hat{p}] = p \qquad V[\hat{p}] = \frac{p(1-p)}{n}$$

if $n \geq 30$ and $np(1-p) > 5$, we can apply the CLT. Some statisticians use the condition $np > 5$ and $n(1-p) > 5$.

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0, 1)$$

Addendum: Chi-square distribution χ^2

- Given n independent standard Normal random variables $X_1, X_2, \ldots, X_n$, the random variable $Y = X_1^2 + X_2^2 + \cdots + X_n^2$ follows a Chi-square distribution with n degrees of freedom:

$$Y \sim \chi_n^2$$

where

$$E[Y] = n \qquad V[Y] = 2n$$

Sampling Distributions: Sample variance S^2

- The sample variance is an unbiased estimator of the population variance (σ^2):

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$$

- When the sample is taken from a normal population,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

we then have

$$V[S^2] = 2\frac{\sigma^4}{(n-1)}$$

Is the sample variance is an unbiased estimator of the population variance (σ^2)

- Let assume we have a random variable X with the following mean and variance:

$$\begin{aligned}\mathrm{E}[X] &= \mu \\ \mathrm{Var}[X] &= \sigma^2 = \mathrm{E}\left[X^2\right] - \mu^2\end{aligned}$$

- Let assume we have sample $X_1, \ldots, X_i, \ldots, X_n$ coming from distribution X:

$$\begin{aligned}\mathrm{E}[X_i] &= \mu \\ \mathrm{Var}[X_i] &= \sigma^2 = \mathrm{E}\left[X_i^2\right] - \mu^2\end{aligned}$$

- The sample mean is a random variable with the following mean and variance:

$$\begin{aligned}\mathrm{E}\left[\overline{X}\right] &= \mu \\ \mathrm{Var}\left[\overline{X}\right] &= \frac{\sigma^2}{n} = \mathrm{E}\left[\overline{X}^2\right] - \mu^2\end{aligned}$$

Is the sample variance is an unbiased estimator of the population variance (σ^2)

- Let assume we have a random variable X with the following mean and variance:

$$
\begin{aligned}
\mathrm{E}\left[S^2\right] &= \mathrm{E}\left[\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}\right] = \mathrm{E}\left[\frac{\sum_{i=1}^{n}(X_i^2+\overline{X}^2-2X_i\overline{X})}{n-1}\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}X_i^2+n\overline{X}^2-2\overline{XX}n\right] = \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right] \\
&= \frac{1}{n-1}\left(n\mathrm{E}\left[X_i^2\right]-n\mathrm{E}\left[\overline{X}^2\right]\right) = \frac{1}{n-1}\left(n\sigma^2+\cancel{n\mu^2}-n\left(\frac{\sigma^2}{n}+\cancel{\mu^2}\right)\right) \\
&= \frac{n\sigma^2-\sigma^2}{n-1} = \sigma^2,
\end{aligned}
$$

which proves that S^2 is an unbiased estimator of the population variance.

Addendum: Student's t distribution

- Given a standard normal random variable X and Y independent of X following a chi-square distribution (χ^2_n) with n degrees of freedom, the random variable $\frac{X}{\sqrt{Y/n}}$ follows a distribution t with n degrees of freedom:

$$Z = \frac{X}{\sqrt{Y/n}} \sim t_n$$

where

$$E[Z] = 0 \text{ if } n \geq 0 \qquad V[Z] = \frac{n}{n-2} \text{ if } n \geq 3$$

Sampling Distributions: Sample mean $\overline{X}$ with unknown variance

- When the sample is taken from a normal population with unknown variance σ^2, we replace it by the sample variance S^2 to obtain:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \approx t_{n-1}$$

Addendum: Fisher's F distribution

- Given two independent chi-squared random variables X_1 and X_2 such that $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$, the random variable $\frac{X_1/n_1}{X_2/n_2}$ follows a F distribution with n_1 and n_2 degrees of freedom:

$$Z = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$$

- **Property:** if $Z \sim F_{n_1,n_2}$ then $Z^{-1} \sim F_{n_2,n_1}$

Sampling Distributions: Variance ratio.

- When two independent samples (sample from X_1 and sample from X_2) of respective sizes n_1 and n_2 are taken from two normal populations, the ratio of their sample variances follows an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom

$$\frac{S^2_{X_1}/\sigma^2_{X_1}}{S^2_{X_2}/\sigma^2_{X_2}} \sim F_{n_1-1,n_2-1}$$

Appendix: Bootstraping

There are situations in which the sampling distribution of an estimator $\hat{\theta}$ (statistic) is unknown or difficult to derive.

- Bootstrap $\rightarrow$ Computer-intensive technique to obtain simulated values of the population by only using values from the sample.
- Suppose that we are sampling from a population that can be modeled by the probability distribution $f(x, \theta)$. The random sample results in data values $x_1, x_2, \dots, x_N$ and we obtain $\hat{\theta}$ as the point estimate of θ. We would now use a computer to obtain *bootstrap samples* of size $n < N$ from the original sample, and for each of these subsamples we calculate the bootstrap estimate $\hat{\theta}^*$ of θ. This results in:

Bootstrap Sample	Observations	Bootstrap Estimate
1	$x_1^*, x_2^*, \dots, x_n^*$	$\hat{\theta}_1^*$
2	$x_1^*, x_2^*, \dots, x_n^*$	$\hat{\theta}_2^*$
$\vdots$	$\vdots$	$\vdots$
B	$x_1^*, x_2^*, \dots, x_n^*$	$\hat{\theta}_B^*$

Appendix: Bootstraping

- Usually $B = 100$ or 200 of these bootstrap samples are taken.
- We can then consider

$$\overline{\theta}^* = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i^*$$

  to approximate the sample mean of the bootstrap estimator $\hat{\theta}^*$.
- Similarly, the standard error of $\hat{\theta}$ can be approximated as:

$$se_{\hat{\theta}} = \sqrt{\frac{\sum_{i=1}^{B}\left(\hat{\theta}_i^* - \overline{\theta}^*\right)^2}{B-1}}$$

- The empirical distribution of $\hat{\theta}^*$ approximates its true distribution.