Mathematical statistics - science of explaining an unknown phenomen (reality) with the aid of observational (sampled) data

Population - underlying set, subject of interest The complete set of individuals, objects or scores of interest Denoted by U

Sample - a subset of the population Denoted S

> random sample $x_1, x_2, \dots, x_n$ from F consists of observations on independen RVs $X_1, X_2, \dots, X_n$ each with distribution F

> convenience sample

Aim is to generilise from sample to population

Stat inference (SI) the process of drawing conclusions about a population based on information in sample

Areas of SI

- point estimation
- interval estimation (conf. interval)
- hypothesis testing

Describing population

- Distribution of X has parameters say $\theta_1, \dots, \theta_k$
- We have n objects in our sample
- We measure (observe) variable X on those n sample objects and get values $x_1, x_2, \dots, x_n$ we call these values "sample" or "realised sample".
- For simplicity take 1 variable X

Let X be normally distributed in the population, $X \sim N(\mu, \sigma^2)$.
Population μ and σ^2 are never known exactly

- We take the sample
- We find estimates of parameters μ σ^2 from the sample
- Sample characteristics $\bar{X}$ and S^2 are estimates of population char-s μ and σ^2

Describing sample

How to describe sample probabilistically?

- Sample has to be representative.
- Two req-s: independence of objects, equal probability for objects to be selected into the sample
- Assume all selected objects have the same distribution
- Mathematically this means that we have random variables $X_1, X_2, \ldots, X_n$ $X_i \sim F$, where F is some distribution and variables X_i are independent

- The set $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is called a theoretical sample
- The set $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ is called realised sample

Sample

Let $X \sim F$ describe population.
A sample $\boldsymbol{x} = (x_1, x_2 \dots x_n)$ from F consists of observations on independent random variables $\mathbf{X} = (X_1, \dots, X_n)$ each with distribution F.

Let F be the distribution of rv X. the distribution depends on the unknown param-s: $\theta_1, \dots, \theta_k$

$$F = F(\theta_1, \dots, \theta_k)$$

For simplicity consider one param θ, $F(\theta)$. Let $\boldsymbol{x} = (x_1, \dots, x_n)$ be the sample from F and we use it to estimate θ

NB! θ is abstract parameter which has a concrete interpretation in the real world

Statistic, Estimator

> A function g of theoretical sample $y = g(X_1, \dots, X_n) = g(X)$ is called a statistic

If values of a statistic can be considered as estimates of θ, the function is called an estimator of θ and denoted $\hat{\theta}(X)$.

Def

Given a sample of realised observations, the number $\hat{\theta}(x)$ is called a point estimate of θ

Def

A point estimator of θ is a function $\hat{\theta}(X)$

The distribution of an estimator

Distinction between $\hat{\theta}(x)$ and $\hat{\theta}(X)$

- Point estimate $\hat{\theta}(x)$ is a numerical value computed from the sample
- Point estimator $\hat{\theta}(X)$ is a random variable

The properties of a point estimator are described with its distribution. Finding the distribution of a point estimator is an important and sometimes difficult task. It can be solved in different ways.

Finding the distribution:

1. Analitic method. The distribution is derived by means of probability theory, exactly or approximately
2. Simulations. The sampling procedure is repeated many times and an approx. to the distribution is then obtained by tabulating all values of $\hat{\theta}(X)$ in the spirit of descriptive statistics

Point estimation: example

Let θ be unknown population param, e.g. the mean of population and there are 3 possible estimates

$$\hat{\theta}_1(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$ good estimator

$$\hat{\theta}_2(\boldsymbol{x}) = \frac{x_1 + x_n}{2}$$ might be biased

$$\hat{\theta}_3(\boldsymbol{x}) = \frac{x_{(1)} + x_{(n)}}{2}$$

smallest biggest

Bias

Point estimate $\hat{\theta}(x)$ is said to be unbiased if the corresponding estimator has expectation θ that is for each $\theta \in A$

$$E[\hat{\theta}(X)] = \theta$$

where A is a set of possible parameter values

If the expectation is different from the θ the estimate is said to be biased.

Bias denoted by B and computed

$$B = E[\hat{\theta}(X)] - \theta$$

Example (continue)

Let $X \sim N(\theta, \sigma^2)$ we had $\hat{\theta}_1(\boldsymbol{x})$ and $\hat{\theta}_2(\boldsymbol{x})$ (from prev. example)

Let's see if they are unbiased:

$$E(\hat{\theta}_1(X)) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} n E[X] = \theta$$

$$E(\hat{\theta}_2(X)) = E\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{2}[EX_1 + EX_2] = \theta$$

Two estimates are unbiased

Efficiency

if two estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased and

$$Var[\hat{\theta}_1(X)] \leq Var[\hat{\theta}_2(X)]$$

$\forall \theta \in A$, then $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$

To estimate $\hat{\theta}$ for which the estimator variance $Var[\hat{\theta}(X)]$ is lower, should be preffered

Example

Let $X \sim N(\theta, \sigma^2)$

$$Var(\hat{\theta}_1(X)) = Var\left[\frac{1}{n}\sum^n X_i\right] = \frac{1}{n^2} Var(\sum X_i) = \frac{1}{n}\sigma^2$$

$$Var(\hat{\theta}_2(X)) = Var\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{2}\sigma^2$$

$\hat{\theta}_1$ is more efficient

Consistency

if for any fixed $\theta \in A$ and for any given $\varepsilon > 0$

$$P(|\hat{\theta}(X) - \theta| > \varepsilon) \to 0$$

as the sample size $n \to \infty$, then the point estimate $\hat{\theta}(x)$ is said to be consistent.

Estimate is consistent if it will get more precise (is closer to the actual value θ, has lower variance) as the sample size n grows

Proving concistency

$\hat{\theta}$ is consistent iff

(1) $\lim_{n \to \infty} E(\hat{\theta}(X)) = \theta$

(2) $\lim_{n \to \infty} Var(\hat{\theta}(X)) = 0$

Mean Square Error (MSE)

MSE of an estimator $\hat{\theta}(X)$ for estimating θ

$$MSE(\hat{\theta}) = E(\hat{\theta}(X) - \theta)^2 =$$

$$= Var(\hat{\theta}(X)) + B^2$$

Note.

- The above statement is also called bias and variance decomposition of MSE
- Prove

$$E[\hat{\theta}(X) - \theta]^2 = E[(\hat{\theta}(X) - E(\hat{\theta}(X))) + (E(\hat{\theta}(X)) - \theta)]^2$$

$$= E[(\hat{\theta}(X) - E[\hat{\theta}(X)])^2 + 2(\hat{\theta}(X) - E(\hat{\theta}(X)))(E(\hat{\theta}(X)) - \theta) + (E - \theta)^2]$$

$$= \underbrace{E[\hat{\theta}(X) - E(\hat{\theta}(X))]^2}_{Var} + 2E[\underbrace{\cancel{\hat{\theta}(X) - E(\hat{\theta}(X))}}_{0} \cdot (\quad)] + \underbrace{E[E\hat{\theta}(X) - \theta]^2}_{B^2}$$

$$E([\hat{\theta}(X) - \theta]^2) = E[((\hat{\theta}(X) - E(\hat{\theta}(X))) + (E(\hat{\theta}(X)) - \theta))^2]$$

$$= E[(\hat{\theta}(X) - E(\hat{\theta}(X)))^2] + 2E[(\quad)(\quad)] + \underbrace{E[(\quad)^2]}_{B^2}$$

$$= Var\,\hat{\theta}(X)$$

Descriptive statistics

Let $x_1, x_2, \dots, x_n$ be the realised values of a random variables X from a sample size n.

The sample mean is defined:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

the sample variance S^2

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \overline{X})^2$$

standard deviation of a sample

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \overline{X})^2}$$

$$E[S^2] = E\left[\frac{1}{n-1}\sum^{n}(X_i - \overline{X})^2\right] =$$

$$= \frac{1}{n-1} E\left[\sum^{n}\left((X_i - m) - (\overline{X} - m)\right)^2\right] =$$

$$= \frac{1}{n-1} E\left[\sum\left((X_i - m)^2 - 2(\overline{X} - m)\underbrace{(X_i - m)}_{n(\overline{X}-m)} + n(\overline{X} - m)\right)\right] =$$

$$= \frac{1}{n-1} E\left[\sum(X_i - m)^2 - n(\overline{X} - m)^2\right] = \frac{1}{n-1}\left[nE\left[(X - m)^2\right] - nE\left[(\overline{X} - m)^2\right]\right] =$$

$$= \frac{1}{n-1}\left[n \operatorname{Var} X - n \underbrace{\operatorname{Var} \overline{X}}_{\text{there's a theorem for RV } \overline{X}}\right] = \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] =$$

$$= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

Standard Error (SE)

Since s is an estimate of σ, an estimate of $\sigma/\sqrt{n}$ is

$$SE_{\bar{X}} = \frac{s}{\sqrt{n}}$$ SE of the sample mean

- Standard deviation describes the variability of the data
- SE - is measure of the precision of the sample mean

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum^{n} X_i\right) = \frac{n}{n^2} Var X =$$

$$= \frac{Var X}{n}$$

Example tossing coin page 209

$$Var(p^*) = E\left[\left(\frac{X}{n} - E\left[\frac{X}{n}\right]\right)^2\right] =$$

$$= \frac{1}{n^2} E\left[X^2 - 2X \cdot EX + (EX)^2\right] =$$

$$= \frac{1}{n^2}\left[E[X^2] - (EX)^2\right] = \frac{1}{n^2} Var X = \frac{1}{n^2} p(1-p)$$

$$D(p^*) = \sqrt{Var(p^*)} = \sqrt{\frac{p(1-p)}{n^2}}$$

Standard deviation of estimat

Now construct an estimate of SD of estimate

$$d(p^*) = \sqrt{\frac{p^*(1-p^*)}{n^2}}$$