1 Point estimates

We begin with two examples.

Example 1 Let $x_1, \ldots, x_n$ be the results obtained from measuring a certain distance.

Suppose that the true distance is μ and that each measurement has a measurement error. If the measurement error is assumed to be normally distributed with expectation 0 and standard deviation σ then the observations can be modelled as outcomes of independent random variables $X_1, \ldots, X_n$ such that $X_i \in N(\mu, \sigma)$, $i = 1, \ldots, n$ since

$$X_i = \mu + Z_i$$

where $Z_i \in N(0, \sigma)$ denotes the measurement error for measurement number i and the Z_i's are assumed to be independent.

A common estimate of the distance μ is the mean of the measurements/observations, i.e.

$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Should you also need an estimate of σ^2 it is common to use the **sample variance**

$$(\sigma^2)^*_{obs} = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - n\bar{x}^2\right).$$

Compare this to the definition of variance $\sigma^2 = E[(X-\mu)^2]$. The standard deviation σ can be estimated by the sample standard deviation s. □

Example 2 Opinion poll. Let p be the proportion of people supporting (s) in the electorate, so in all the people allowed to vote. Interview $n = 3002$ persons and let x be the number of persons claiming to support (s), say $x = 1053$.

An estimate of p is given by

$$p^*_{obs} = p^*(x) = \frac{x}{n} = \frac{1053}{3002} = 0.3508$$

□

Definition 1 *A* **random sample** $x_1, \ldots, x_n$ *(from F) consists of observations of independent random variables $X_1, \ldots, X_n$ (with the distribution function F).*
A **point estimate** *of a parameter θ is a function of a random sample/the sample values.*

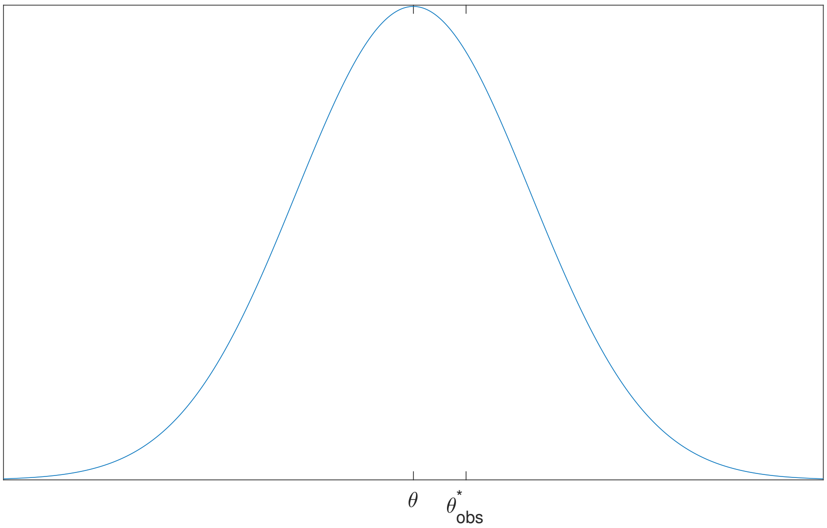
$$\theta^*_{obs} = \theta^*(x_1, \ldots, x_n).$$

The point estimate is described by the **sample variable** *(or* **statistic***)*

$$\theta^* = \theta^*(X_1, \ldots, X_n).$$

(θ^*_{obs} is an outcome of θ^*).

Remark 1 Oftentimes the random varibles $X_1, \ldots, X_n$ are identically distributed (if the random sample is a sample from F), but they do not have to be.



Figur 1: Example of a density function of a sample variable θ^* (a sample distribution).

Which properties are desirable of a point estimate?

Definition 2 *A point estimate* $\theta^*_{obs} = \theta^*(x_1, \ldots, x_n)$ *of a parameter* θ *is said to be*

- **consistent** *if*
$$P(|\theta^* - \theta| > \epsilon) \to 0 \qquad \textit{as } n \to \infty,$$
- **unbiased** *if*
$$E[\theta^*] = \theta.$$
- *If both* θ^*_{obs} *and* $\hat{\theta}_{obs}$ *are unbiased point estimates of* θ *then* θ^*_{obs} *is said to be* **more efficient** *if*
$$V(\theta^*) \leq V(\hat{\theta}).$$

Example 3 Continuing with Example 1 we see that the estimate

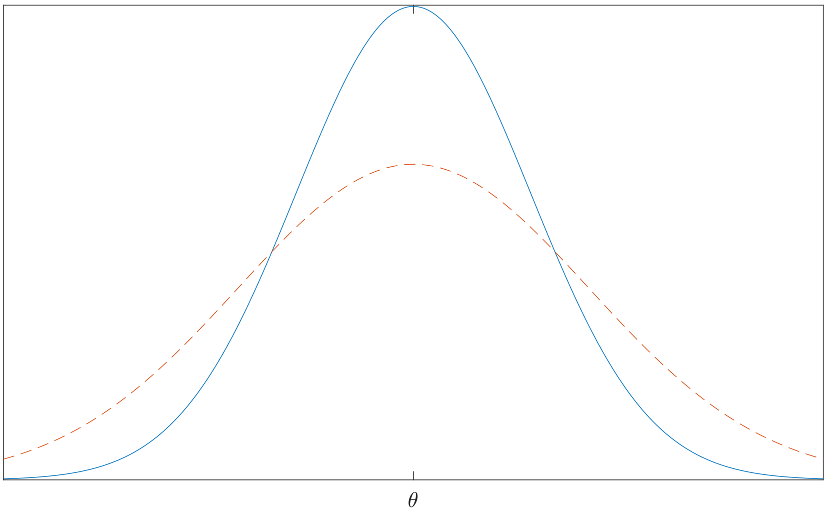
$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is described by the sample variable

$$\mu^* = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

For this variable we have that

$$E[\mu^*] = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E\left[X_i\right]}_{=\mu} = \mu.$$



Figur 2: Exemples of density functions of estimates of θ. The estimate with the denisty function in drawn with a solid line is to be preferred, since its distribution is more concentrated around θ.

Since $E[\mu^*] = \mu$ the estimate $\mu^*_{obs} = \bar{x}$ is an unbiased estimate of μ. Furthermore, we have that

$$\begin{aligned} V(\mu^*) &= V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \{\text{independence}\} \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{V\left(X_i\right)}_{=\sigma^2} = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n} \to 0 \quad \text{as } n \to \infty \end{aligned}$$

Recall Chebychev's inequality

$$P(|X-\mu| > k\sigma) \leq \frac{1}{k^2} \qquad \text{for all } k > 0.$$

Recall Chebychev's inequality

$$P(|X-\mu| > k\sigma) \leq \frac{1}{k^2} \qquad \text{for all } k > 0.$$

If we let $k = \epsilon/\sigma$ then Chebychev's inequality reads

$$P(|X-\mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

so with $X = \mu^*$ we obtain

$$P(|\mu^*-\mu| > \epsilon) \leq \frac{V(\mu^*)}{\epsilon^2} \to 0 \quad \text{as } n \to \infty.$$

This implies that $\mu^*_{obs} = \bar{x}$ is a consistent estimate of μ.

It can be shown that $(\sigma^2)^*_{obs} = s^2$ is an unbiased estimate of σ^2, i.e. that $E[S^2] = \sigma^2$ (this is the reason for dividing by $n-1$ rather than by n). See Theorem 2, Chapter 12 in [2]. $\square$

Example 4 Continuing with Example 2 we see that the estimate

$$p^*_{obs} = p^*(x) = \frac{x}{n} = \frac{1053}{3002} = 0.3508$$

is described by the sample variable

$$p^* = p^*(X) = \frac{X}{n} \qquad \text{where } X \in Bin(n, p)$$

(since the sampling is performed without replacement X is really hypergeometrically distributed, but since the electorate is big in comparison to the sample size a binomial approximation shoud work well). We have that

$$E[p^*] = E\left[\frac{1}{n} \cdot X\right] = \frac{1}{n}E[X] = \frac{1}{n} \cdot np = p$$

i.e. $p^*_{obs} = x/n$ is an unbiased estimate of p. Furthermore we have that

$$V(p^*) = V\left(\frac{1}{n} \cdot X\right) = \frac{1}{n^2}V(X) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n} \to 0 \quad \text{as } n \to \infty$$

This means that you can show that $p^*_{obs} = x/n$ is a consistent estimate of p using Chebychev's inequality.

How good is the estimate? The variance/standard deviation will give you an indication of this. Here we have that

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

Now a problem arises since the standard deviation depends on the parameter that we are trying to estimate! We can then try to estimate the standard deviation by for instance

$$d(p^*) = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}} = \sqrt{\frac{0.3508(1-0.3508)}{3002}} \approx 0.00871.$$

□

Definition 3 *An estimate of* $D(\theta^*)$ *is called the* **standard error** *of* θ^* *and is denoted by* $d(\theta^*)$.

Example 5 Do problem 1203 i [2]. □

1.1 Summary

Suppose that $x_1, \ldots, x_n$ are outcomes of independent random variables $X_1, \ldots, X_n$, whose distributions depend on an unknown parameter. So far we have seen the following estimates.

Parameter	Estimate	Sample variable
$\mu = E[X]$	$\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$	$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$
$\sigma^2 = V(X)$	$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
$\sigma = D(X)$	$s = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$	$S = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$
θ	$\theta^*_{obs} = \theta^*(x_1, \ldots, x_n)$	$\theta^* = \theta^*(X_1, \ldots, X_n)$

How do find suitable function of the sample values $\theta^*(x_1, \ldots, x_n)$ for your estimate? Here we will look at two standard methods.

1.2 The Maximum likelihood method (the ML method)

The Maximum likelihood estimate (ML estimate) of the parameter θ maximizes the likelihood function

$$L(\theta) = \begin{cases} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta) = p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta) \\ \{\text{independence}\} = P(X_1 = x_1; \theta) P(X_2 = x_2; \theta) \cdots P(X_n = x_n; \theta) = \prod_{i=1}^{n} p_{X_i}(x_i; \theta) \\ f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta) \\ \{\text{independence}\} = \prod_{i=1}^{n} f_{X_i}(x_i; \theta) \end{cases}$$

which in the discrete case is the probability of obatining the outcomes we actually got.

Example 6 Let $x = 972$ be an observation of X where $X \in Bin(n, p) = Bin(3002, p)$.
<u>Want:</u> The ML estimate of p.

The ML estimate maximizes the likelihood function

$$L(p) = P(X = x; p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Useful trick: the ML estimate also maximizes $\ln(L(p))$ the so called log-likelihood function. This function is often easier to work with since the logarithm of a product is sum of logarithms. Here we have that

$$\begin{aligned} \ln L(p) &= \ln\left(\binom{n}{x} p^x (1-p)^{n-x}\right) = \ln\binom{n}{x} + \ln(p^x) + \ln(1-p)^{n-x} \\ &= \ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p). \end{aligned}$$

To find the maximum we solve

$$\frac{d}{dp}\ln L(p) = 0$$

i.e.

$$x \cdot \frac{1}{p} + (n-x)\frac{-1}{1-p} = 0$$

or

$$x(1-p) - (n-x)p = 0$$

which will result in the ML estimate $p^*_{obs} = x/n$ (you should really check the second derivative to make sure that you have found a maximum, and also the boundary values). □

1.3 The method of least squares (the LS method)

Suppose that the random variables $X_1, X_2, \ldots, X_n$ all have expectations which depend on the parameter θ and that they all have the same variance, i.e. $V(X_1) = \ldots = V(X_n)$.
The least squares estimate (LS estiamte) of the parameterθ minimizes the function

$$Q(\theta) = \sum_{i=1}^{n}(x_i - E[X_i;\theta])^2.$$

Remark 2 In order to derive the LS estimate it is not necessary to know the whole distribution of the random variables X_i $i = \ldots, n$, you only need to know how their expectations depend on the parameter θ.

Example 7 Just as in Example 6 let $x = 972$ be an observation of X where $X \in Bin(n,p) = Bin(3002, p)$.

Want: the LS estimate of p.

Since $E[X] = np$ the LS estimate should minimizes the function

$$Q(p) = (x - E[X])^2 = (x - np)^2 \geq 0$$

We can make the function $Q(p)$ 0 by choosing $p = x/n$ and therefore the LS estimate of p is $p^*_{obs} = x/n$. □

Referenser

[1] Blom, G., Enger, J., Englund, G., Grandell, J., och Holst, L., (2005). Sannolikhetsteori och statistikteori med tillämpningar.

[2] Blom, Gunnar, (1989). Probability and Statistics. Theory and Applications.