Estimation of the Mean

Hans Georg Schaathun

Session 4 February 2015
Updated 20th January 2015

Exercise 1 *Discuss:*

> *What is the relationship between the* sample mean*, the* population mean*, and the* expected value*?*

1 The sample mean

The sample mean $\bar{X}$ is a stochastic variable. Considering a single observation of X, the expected value is $E(X) = \mu$ and the variance is $\mathsf{var}(X) = \sigma^2$.

Exercise 2 *Write down the formulæ for $\bar{X}$ in terms of a n observations $X_1, X_2, \ldots, X_n$.*

Exercise 3 *Give formulæ for $E(\bar{X})$ and $\mathsf{var}(\bar{X})$.*

2 Point estimation

Suppose you want to use the sample mean $\bar{X}$ as an estimator for $E(X)$.

Exercise 4 *What do we mean by an* unbiased estimator*? Is $\bar{X}$ an unbiased estimator?*

Let $X_1, X_2, \ldots, X_n$ be independent stochastic variables, the we have

$$\mathsf{var}\left(\sum X_i\right) = \sum \mathsf{var}(X_i), \tag{1}$$

$$E\left(\sum X_i\right) = \sum E(X_i). \tag{2}$$

Table 1: Two important formulæ, repeated from Session 2.

Let $X = X_1 + X_2 + \ldots + X_n$ be a sum of *identically* distributed variables X_i. regardless of the exact distribution of X_i When as $n \to \infty$, X has always the same distribution, namely the normal distribution.

Table 2: The Central Limit Theorem

Exercise 5 *Give an expression for the standard error of $\bar{X}$. (Remember to use the results from the previous section.)*

Exercise 6 *Discuss*

> *How can you estimate the standard error when you estimate $E(X)$ from a sample of size n?*

3 The Central Limit Theorem

Let X be a stochastic variable, uniformly distributed on the

Exercise 7 *Make a Matlab routine `sample(m)` which takes an integer argument m, draws m random numbers uniformly distributed on the integer range $[0, 10]$, and calculates the sample mean of these m numbers.*

Exercise 8 *Run `sample(1)` 100 times and draw a histogram of the values. Repeat this for `sample(m)` for $m = 5, 25, 125, 500$.*

Compare the histograms for different values of m. What do you see?

Exercise 9 *Consider the Central Limit Theorem (CLT) in Table 2. What evidence do you see of the CLT in the histograms in the previous exercise.*

It is the CLT which allows us to use the normal distribution to approximate the binomial distribution for large n. It also means that when the sample size is large, we can make confidence intervals for the mean based on the normal distribution, regardless of the distribution of the underlying variables.