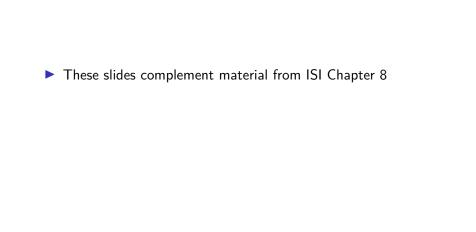
Data part 2 STAT-S520

Arturo Valdivia

02-21-23



The sum and the average of a random sample

Let $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} \mathbb{P}$ be a random sample with $EX_i = \mu$ and $VarX_i = \sigma^2$ for $i = 1, \ldots, n$ and let's define

$$Y = \sum_{i=1}^{n} X_i$$
 as the sum of the random sample and

$$\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

as the sample mean or the average of the random sample.

Expected value and variance of Y and \bar{X}_n

Let's use the properties of the expected value and variance to obtain:

$$EY = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} M = nM$$

►
$$VarY = Var \left(\sum_{i=1}^{\infty} X_{i} \right) = \sum_{i=1}^{\infty} Var \left(\sum_{i=1}^{\infty}$$

$$E\bar{X}_{n} = E\left(\frac{1}{N}\sum_{i=1}^{N}X_{i}\right) = \frac{1}{N}E\left(\sum_{i=1}^{N}X_{i}\right) = \frac{1}{N}NM = M$$

$$Var\bar{X}_{n} = Vaw\left(\frac{1}{N}\sum_{i=1}^{N}X_{i}\right) = \frac{1}{N}N\sigma^{2} = \frac{1$$

Simulations in R part 1

Let's work with simulations in R to show, approximately, that the expected value and variance of Y and \bar{X}_n are the ones shown above.

The Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \dots, X_n \stackrel{i.i.d}{\sim} \mathbb{P}$ with $EX_i = \mu$ and $VarX_i = \sigma^2$ for $i = 1, \dots, n$ and

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

It can be shown that \bar{X}_n converges in probability to μ , or in math notation:

$$\frac{\bar{X}_{n} \xrightarrow{p} \mu}{\text{For some small } \mathbb{Z} > 0 \text{ (arbitrarily small)}}$$

$$P(\bar{X}_{n} \in (M - \varepsilon, \mu + \varepsilon)) \longrightarrow 1$$
as $n \to \infty$

The Central Limit Theorem (CLT)

Let
$$X_1, X_2, \dots, X_n \stackrel{i.i.d}{\sim} \mathbb{P}$$
 with $EX_i = \mu$ and $VarX_i = \sigma^2$ for $i = 1, \dots, n$,

$$ar{X}_n = rac{1}{n} \sum_{i=1}^n X_i$$
 $Z_n = rac{ar{X}_n - \mu}{\sigma/\sqrt{n}}$ and $Z \sim \mathsf{Normal}(0,1)$

and we define F_n as the CDF of Z_n and Φ as the CDF of Z. The CLT states that, for any real number z,

$$F_n(z) \rightarrow \Phi(z)$$

as $n \to \infty$.

Practical use of the CLT

- ▶ The previous result tells us that when $n \to \infty$ the distribution of \bar{X}_n is normal.
- Mowever, for fairly small values of n, \bar{X}_n is already approximately normally distributed
 - The rule of thumb commonly used is $n \ge 30$

Simulations in R part 2

Let's work with simulations in R that show that the WLLN and the CLT hold, approximately.