LLN and CLT (Ch 4.2.1/4.4.2 in ER)

In the last lectures we saw that if

$$S \sim \text{Binomial}(N, q)$$

(remember this means $S = \sum_{i=1}^{N} X_i$, $X_i \sim \text{Bernoulli}(q)$)

feel to remind us this is r.v.

then $P(S=k)$ has a Bell curve shape which becomes narrower as N becomes large

This implies for large N

$$\bar{X} = S/N \approx \mathbb{E}[X_i] = q \quad \text{for large } N$$

In other words, the sample avg. tends towards the expected value.

This is true for (almost) any r.v. and is made precise by the LLN:

Law of Large Numbers (Thm 4.2.1)

Let $X_1, X_2, \ldots X_N$ be iid and $\mathbb{E}[X_i] < \infty$ then for any $\varepsilon > 0$

$$P(|\bar{X} - \mathbb{E}[X_i]| > \varepsilon) \to 0 \quad \text{as } N \to \infty$$

Note in the textbook they call $\bar{X}$, M_N

We also saw that the distribution of the <u>standardized variable</u> (or z-score)

$$Z = \frac{S - \mathbb{E}[S]}{\sqrt{Var(S)}} = \frac{S - Nq}{\sqrt{Nq(1-q)}}$$

looks the same for any value of N.

The limiting behavior of Z is given by an example of a continuous distribution, so we need to take a detour into <u>Ch 2.4</u>

<u>Detour</u> (Continuous distributions)

<u>Def 2.4.3</u> a continuous (or technically absolutely continuous) r.v. is one that can take on real # values and hence can not be described by a probability function $P(X=x)$. Instead we define a curve $f_X(x)$ called a <u>density</u> and defined the probabilities by ← see def 2.3.2

$$P(a < X < b) = \int_a^b f_X(x)\,dx$$

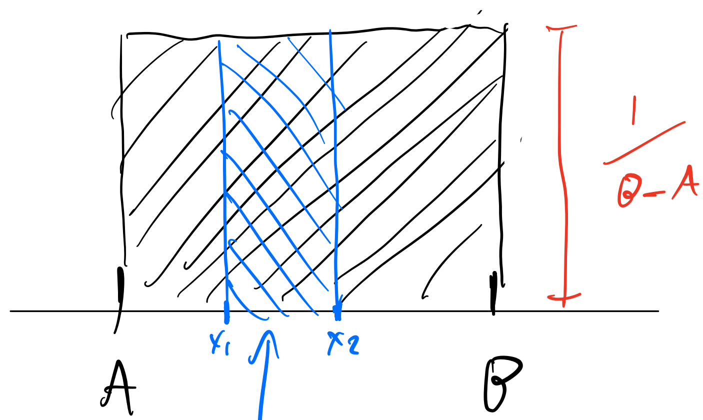
Note that $P(-\infty < X < \infty) = 1$

Example

The r.v. $U \sim \text{Uniform}(A, B)$ has density

$$f_U(x) = \begin{cases} \frac{1}{B-A} & \text{if } x \in [A,B] \\ 0 & \text{if } x \notin [A,B] \end{cases}$$

This means any value of $x \in [A,B]$ is equally likely (also $P(U=x)=0$ for all x, which may seem odd)



$\text{Area} = (x_2 - x_1) \times \frac{1}{B-A}$

$= P(x_1 < U < x_2)$

Note that the density is the mathematical idealization of a histogram

Back to the CLT

The CLT says that the standardized variable Z approaches a special random variable called a standard Normal random variable which has density (Example 2.4.7)

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Central Limit Theorem (thm 4.4.3)

Let $X_1, \ldots, X_N$ be iid w/ $\mu = \mathbb{E}[X_i] < \infty$ and $\sigma^2 = \text{var}(X) < \infty$, let $S = \sum_{i=1}^{N} X_i$

and $Z = \frac{S - N\mu}{\sigma\sqrt{N}}$, then

$$P(a < Z < b) \longrightarrow \int_a^b \phi(x)\,dx$$

Basically the histogram becomes normal

Examples to focus on: 4.4.7, 4.4.9, 4.4.10

I will also cover general Normal random variables and properties of Normal r.v. if time permits