Law of Large Numbers (LLN) Chapter 7.2

Connects the concepts of the mean of a RV and the average of n of its samples. (which is called the sample mean)

This allows us to experimentally estimate $E(X)$ by computing an average.

Suppose $X_1, X_2, \ldots, X_n$ are independent repetitions of an experiment associated with a RV X.

The X_i's are independent, identically distributed (iid or IID) RVs.

In practice, we often approximate the unknown mean μ by the sample mean.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i = M_n$$

- How accurate is that approximation?
- Does that accuracy improve as we take more samples (as n increases)?
- How fast does this accuracy improve?

Because the X_i's are RVs,
then M_n is a RV, so it has a mean and variance

① what is the mean of the sample mean?

$$E(M_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}(n E(X_i))$$
$$= E(X_i) = \mu.$$

An unbiased estimator - good!

② what is the expected error for this estimate?

$$E((M_n - \mu)^2) = E((M_n - E(M_n))^2)$$
$$= Var(M_n)$$

$$Var(M_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$
$$= \frac{1}{n^2}(n\, Var(X_i)) = \frac{Var(X_i)}{n}$$
$$= \frac{\sigma^2}{n}$$

As we take more samples the error goes down

There are 2 Laws of Large Numbers.

Both essentially state that

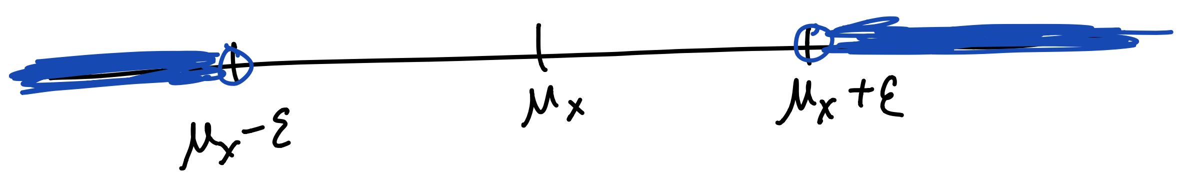
"M_n converges to $E(X)$ as $n \to \infty$"

$$\boxed{\text{The Weak Law of Large Numbers (WLLN)} \qquad \lim_{n\to\infty} P\left(\left| M_n - \mu_x \right| > \varepsilon \right) = 0}$$

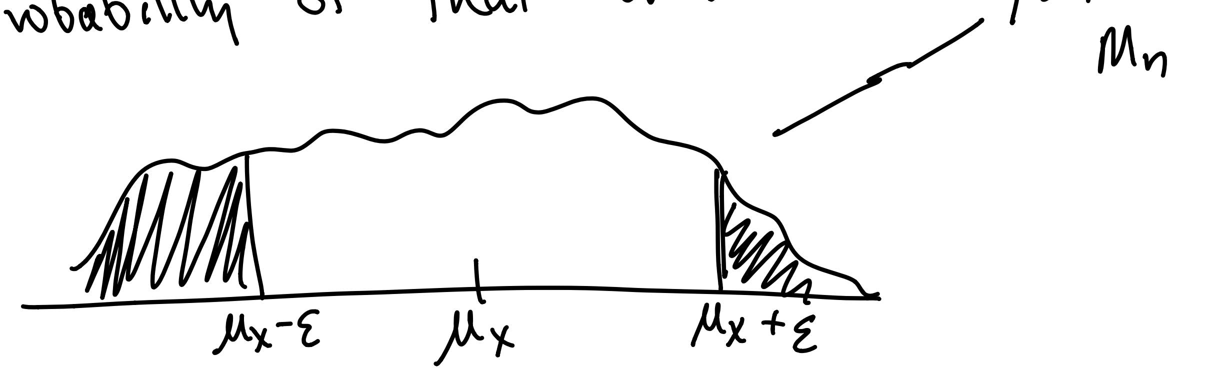
The sample mean "converges in probability" to the mean.

The probability that the sample mean, M_n, differs from the actual mean, μ_x, by more than some small value ε approaches zero as n approaches ∞

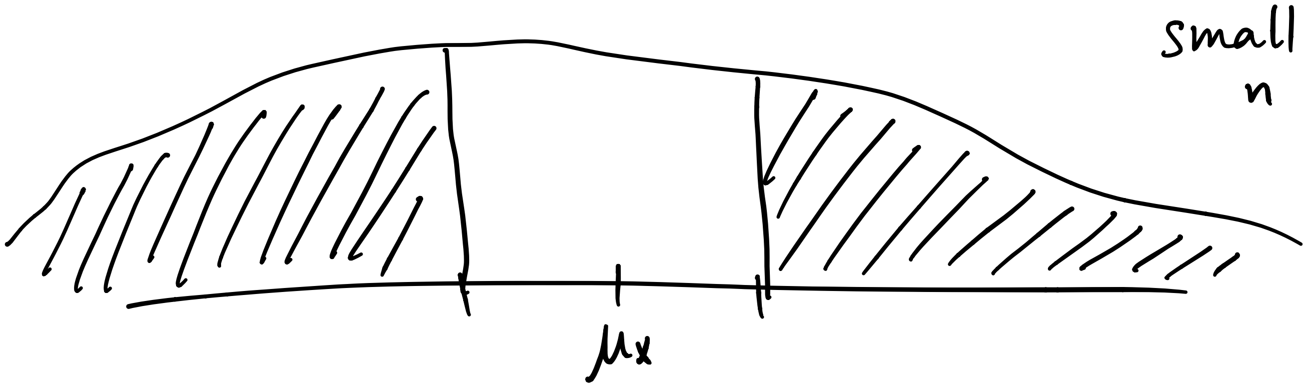
The event in question, $|M_n - \mu_x| > \varepsilon$:



The probability of that event :



The sample mean becomes more accurate as we take more samples



because the PDF of the sample mean becomes narrower as n increases

To prove the WLLN, we start by recognizing $E(M_n) = E(X_i) = \mu_x$ and $Var(M_n) = \frac{1}{n} var(X_i)$.

Then we need some more tools, which are the probability bounds in Chapter 4.6.

- Markov Bound
- Chebyshev Bound

leads to

$$P(|M_n - \mu_x| > \varepsilon) \leq \frac{\sigma_x^2/n}{\varepsilon^2}$$

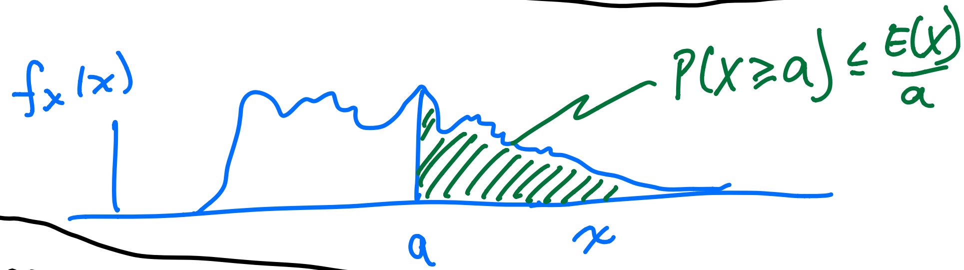
which, for a fixed ε, goes to zero as n goes to infinity.

Tukey says: An approximate answer to the right question is more important than a precise answer to the wrong question.

Markov inequality (Chapter 4.6)

$$P(X \geq a) \leq \frac{E(X)}{a} \quad \text{for any non-negative RV } X.$$

Interpretation:



Proof:

$$E(X) = \int_0^{\infty} t\, f_X(t)\, dt$$

break into 2 pieces

$$= \int_0^{a} t\, f_X(t)\, dt + \int_a^{\infty} t\, f_X(t)\, dt$$

drop one piece

$$\geq \int_a^{\infty} t\, f_X(t)\, dt$$

replace the t inside the integral with a smaller value, a

$$\geq \int_a^{\infty} a\, f_X(t)\, dt$$

a is a constant

$$= a\, P(X \geq a)$$

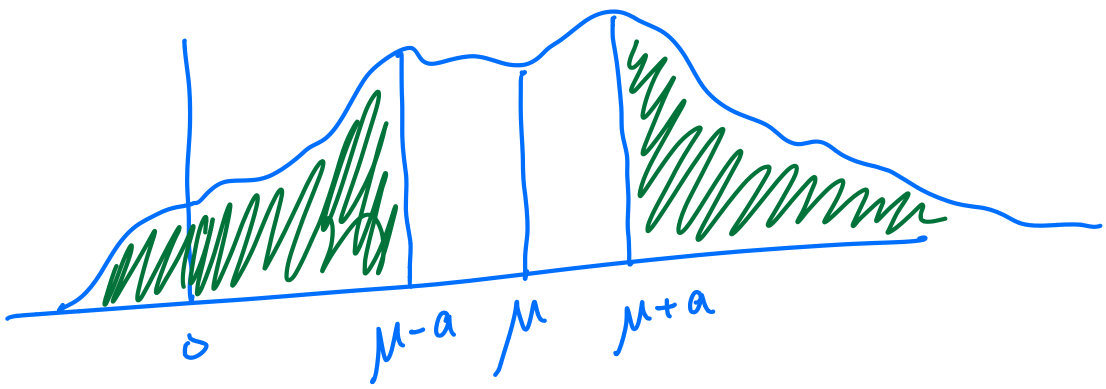
Combining,

$$P(X \leq a) = \frac{E(X)}{a}$$

Chebyshev inequality (also Ch. 4.6)

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

for a RV X with mean μ and variance σ^2



Proof:

Let $D^2 = (X-\mu)^2$ (a RV)

Apply the Markov inequality to D^2

$$P(D^2 \geq a^2) \leq \frac{E(D^2)}{a^2} = \frac{\sigma^2}{a^2}$$

Note: since event $\{D^2 \geq a^2\}$ is the same event as $\{|X-\mu| \geq a\}$

So

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

Now we can apply this to the sample mean M_n, that has mean μ_x and variance $\frac{\sigma_x^2}{n}$

Using chebyshev inequality to M_n,

$$P(|M_n - \mu_x| \geq \varepsilon) \leq \frac{\sigma_x^2/n}{\varepsilon^2}$$

As we take the limit of this as $n \to \infty$, the right-hand side goes to zero