國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 22: limiting theorem

李祈均

Quick review

Chebyshev inequality

If X is a random variable with mean μ, and variance σ^2 then,

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}, \text{for all } c > 0$$

Convergence

Convergence of a deterministic sequence:

Let $a_1, a_2, \ldots$ be a sequence of real numbers, and let a be another real number. We say that the sequence a_n converges to a, or $\lim_{n \to \infty} a_n = a$, if for every $\epsilon > 0$ there exists some n_0 such that:

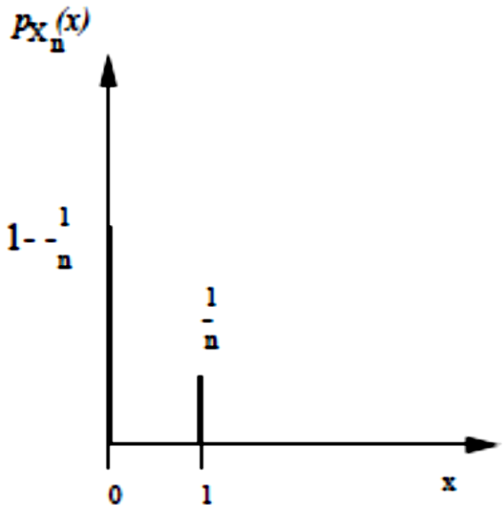
$$|a_n - a| \le \epsilon, \text{for all } n \ge n_0$$

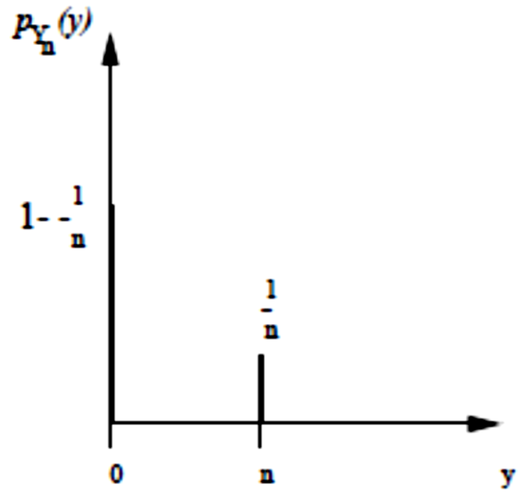
Convergence in probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence, Y_n, **converges** to a **in probability**, if for every $\epsilon > 0$, we have:

$$\lim_{n\to\infty} P(|Y_n - a| \geq \varepsilon) = 0$$

Example





- Find the expected value and variance of X_n, Y_n

$$E[X_n] = 0 * \left(1 - \frac{1}{n}\right) + 1 * \frac{1}{n} = \frac{1}{n}$$

$$var(X_n) = \left(0 - \frac{1}{n}\right)^2 * \left(1 - \frac{1}{n}\right) + \left(1 - \frac{1}{n}\right)^2 * \left(\frac{1}{n}\right) = \frac{n-1}{n^2}$$

- Find the expected value and variance of X_n, Y_n

$$E[Y_n] = 0 * \left(1 - \frac{1}{n}\right) + n * \frac{1}{n} = 1$$

$$var(Y_n) = (0-1)^2 * \left(1 - \frac{1}{n}\right) + (n-1)^2 * \left(\frac{1}{n}\right) = n - 1$$

What does the Chebyshev inequality tells us about the convergence:

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

$$\lim_{n \to \infty} P\left(\left|X_n - \frac{1}{n}\right| \geq \epsilon\right) \leq \lim_{n \to \infty} \frac{n-1}{n^2 \epsilon^2} = 0$$

Moreover,

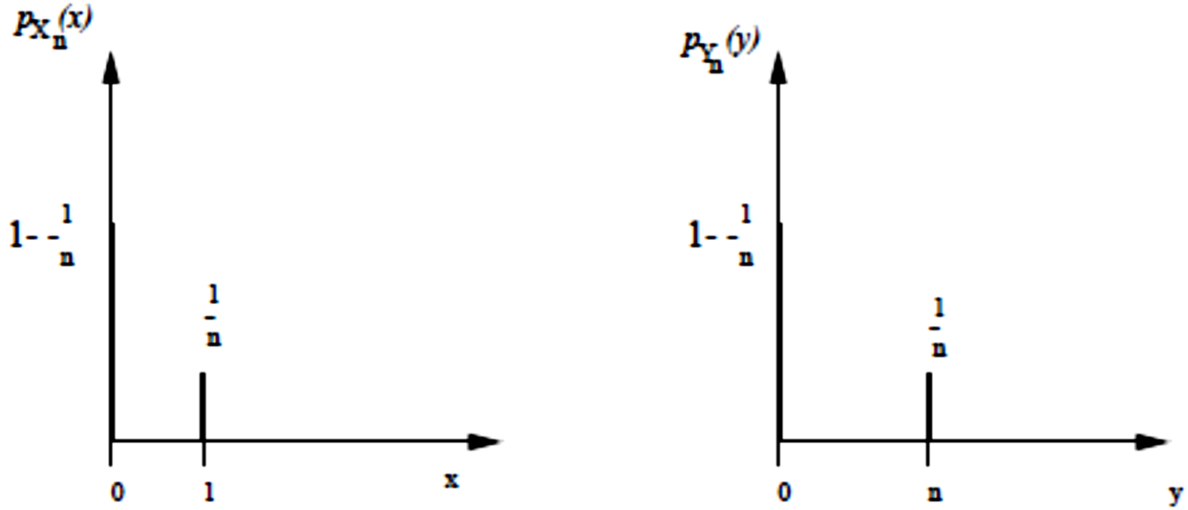
$$\lim_{n \to \infty} \frac{1}{n} = 0$$

So, X_n converges to 0 in probability

$$\lim_{n\to\infty} P(|Y_n - 1| \geq \epsilon) \leq \lim_{n\to\infty} \frac{n-1}{\epsilon^2} = \infty$$

We can conclude anything about the convergence using Chebyshev inequality, only Chebyshev

Lets look more closely, Is Y_n convergent in probability, if so what value?



$$\lim_{n\to\infty} P(|Y_n - 0| \geq \epsilon) \leq \lim_{n\to\infty} \frac{1}{n} = 0$$

So Y_n converges to 0 with in probability

Chebyshev only one of the tools, make sure you check (use) more

A sequence of random variable converges to a number c in the **mean-square** sense:

$$\lim_{n\to\infty} E[(X_n - c)^2] = 0$$

Show that convergence in mean square implies convergence in probability

$$P(|X_n - c| \geq \epsilon) = P\left(|X_n - c|^2 \geq \epsilon^2\right) \leq \frac{E[(X_n - c)^2]}{c^2}$$

Take limit as n goes to infinity, we can easily see that convergence in mean square implies convergence in probability

Does Y_n converge in mean square?

$$E[(Y_n - 0)^2] = 0 * \left(1 - \frac{1}{n}\right) + \left(n^2\right) * \frac{1}{n} = n$$

Take n goes to infinity,

This goes to infinity, does not converge in mean square, but converge in probability !

Converge ? What type? Makes a difference

Weak law of large numbers

Let $X_1, X_2, \ldots$ be independent identically distributed random variables with mean μ. For every $\epsilon > 0$, we have

$$P(|M_n - \mu| \geq \epsilon) = P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0,$$

as $n \to \infty$

Central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common mean μ, and variance, σ^2, and define:

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then, the CDF of Z_n converges to the standard normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

In the sense that:

$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z)$$

Proof of CLT

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common mean 0, and variance, σ^2, and associated transform $M_X(s)$ is finite when $-d < s < d$, where d is some positive number:

Let:

$$Z_n = \frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}$$

a) Show that the transform of Z_n is

$$M_{z_n}(s) = \left(M_X\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n$$

We have, using the independent of X_i

$$M_{z_n}(s) = E[e^{sZ_n}]$$

$$= E\left[\exp\left\{\frac{s}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i\right\}\right]$$

$$= \prod_{i=1}^{n} E\left[e^{sX_i/\sigma\sqrt{n}}\right]$$

$$= \left(M_X\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n$$

b) Suppose that the transform $M_X(s)$ has a second order Tylor series expansion around $s = 0$, of the form (Maclaurin series):

$$M_X(s) = a + bs + cs^2 + o(s^2)$$

where $o(s^3)$ is a function that satisfies $\lim_{s \to 0} o(s^2)/s^2 = 0$, find a, b, c in terms of σ^2

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

- Using the moment generating property of the transform, we have:

$$a = M_X(0) = 1 \text{ (by definition)}$$

$$b = \frac{d}{ds} M_X(s) \bigg|_{s=0} = E[X] = 0$$

$$c = \frac{1}{2} \frac{d^2}{ds} M_X(s) \bigg|_{s=0} = \frac{E[X^2]}{2} = \frac{\sigma^2}{2}$$

$$M_X(s) = a + bs + cs^2 + o(s^2)$$

c) Combine the result from part a) and part b) to show that the transform of $M_{Z_n}(s)$ converges to the transform associated with a standard normal random variable, that is,

$$M_{Z_n}(s) = \left(M_X\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n$$

$$\lim_{n\to\infty} M_{Z_n}(s) = e^{s^2/2}$$

Combining part a) and part b)

$$M_{z_n}(s) = \left(M_X\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n$$

$$= \left(a + \frac{bs}{\sigma\sqrt{n}} + \frac{cs^2}{\sigma^2 n} + o\left(\frac{s^2}{\sigma^2 n}\right)\right)^n$$

$$= \left(1 + \frac{s^2}{2n} + o\left(\frac{s^2}{\sigma^2 n}\right)\right)^n$$

Now take the limit as $n \to \infty$

$$\lim_{n\to\infty}\left(1+\frac{c}{n}\right)^n = e^c$$

To obtain:

$$\lim_{n\to\infty} M_{z_n}(s) = e^{s^2/2}$$

So if the transform $M_{z_n}(s)$ goes to $M_Z(s)$, where Z is a standard normal, then the CDF Z_n goes to CDF of Z (this requires a separate proof… will not covered here)

Hence proved

usefulness

- Universal: only means and variances matter
  - X can be any distribution (as long as finite mean and variance)
  - You only need to know mean and variance of X

- Accurate computational shortcut

- Justification of normal models
  - Classic examples goes back in 100 years (dust in a fluid, and see the placement of that dust)
  - What is that? Brownian motion in physics
    - Movement of financial markets
    - Though recently has changed a little somewhat (Tail probability)

More notes

What exactly does it say?

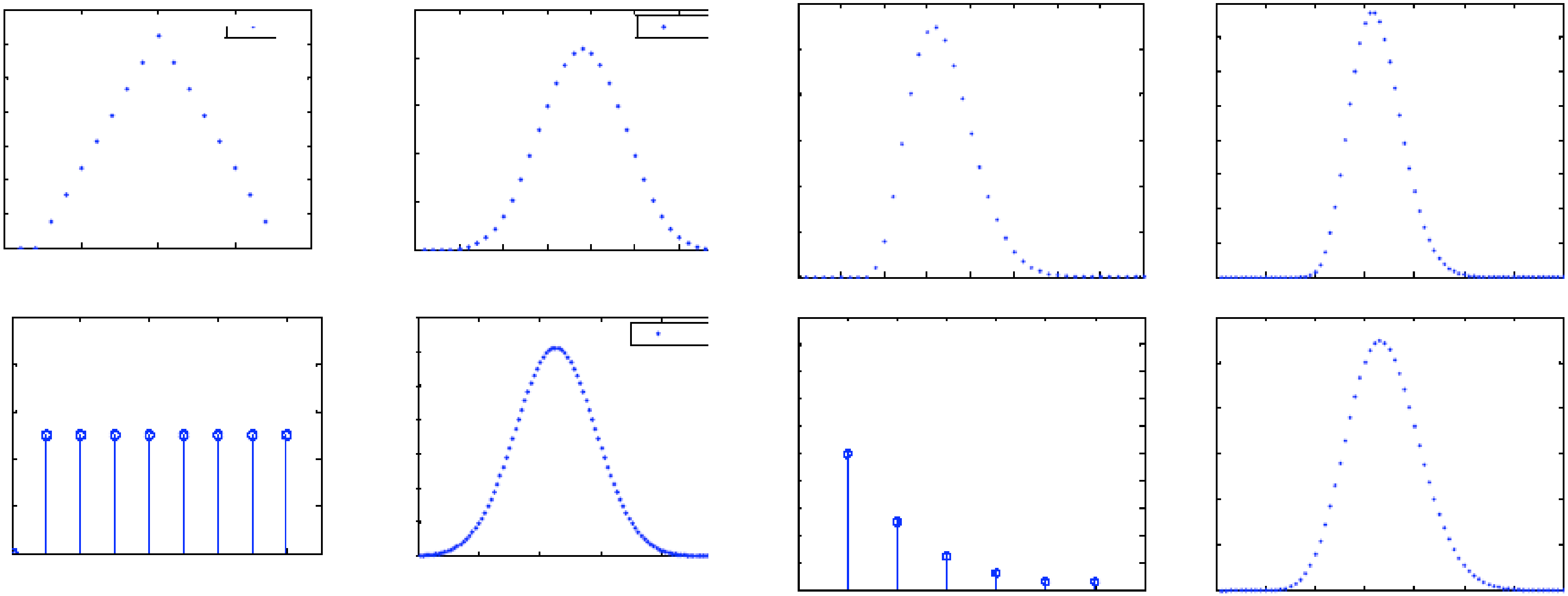
- CDF of Z_n converges to normal CDF
  - Not a statement about convergence of PDFs or PMFs

How to use:

- Treat Z_n as if it's normal variable
  - Then you can treat S_n as if it's normal
  - Linear transformation of normal rv is a normal rv (a lot of if – if – if here…)

Can we use it when n is not that large

- Yes, but there is no nice theorem to say what is a good 'n'
- Symmetry in the original distribution helps!



Back to our Coke poll's problem

- f: fraction of population that likes coke
- i^{th} (randomly selected) person polled:

$$X_i = \begin{cases} 1, \text{if yes} \\ 0, \text{if no} \end{cases}$$

$$M_n = \frac{(X_1 + \cdots + X_n)}{n}$$

- We want:

$$P(|M_n - f| \geq 0.01) \leq 0.05$$

- The event of interest:

$$|M_n - f| \geq 0.01$$

Try to massage this expression to make it look like standardize variable:

$$\left|\frac{X_1 + \cdots + X_n - nf}{n}\right| \geq 0.01$$

$$\left|\frac{X_1 + \cdots + X_n - nf}{\sqrt{n}\sigma}\right| \geq \frac{0.01\sqrt{n}}{\sigma}$$

$$P(|M_n - f| \geq 0.01) \approx P\left(|Z| \geq \frac{0.01\sqrt{n}}{\sigma}\right)$$

Since we don't know anything about σ up front

$$\sigma^2 = f * (1 - f)$$

$$\sigma \leq 0.5$$

$$P(|M_n - f| \geq 0.01) \approx P\left(|Z| \geq \frac{0.01\sqrt{n}}{\sigma}\right)$$
$$\leq P\left(|Z| \geq 0.02\sqrt{n}\right)$$

- Now, we can use normal tables to calculate probabilities of interest

$$P\left(|Z| \geq 0.02\sqrt{n}\right) = 2 * P\left(Z \geq 0.02\sqrt{n}\right)$$

How small of n can we take to make it within 5% of error

If you read it off the normal table (find z that gives you 97.5, back solve for n),

You can do the math and see that $n = 9604$

Much less than $50,000$ using Chebyshev inequality

CLT Apply to binomial

- Fix p, where $0 < p < 1$
- X_i: Bernoulli(p)
- $S_n = X_1 + \cdots + X_n$: Binomial(n, p)
  - Mean np, variance $np(1-p)$
- Apply CLT:

$$\text{CDF of } \frac{S_n - np}{\sqrt{np(1-p)}} \rightarrow \text{standard normal}$$

Quick example

$n = 36, p = 0.5$ find $P(S_n \leq 21)$

Exact answer using Binomial distribution

$$\sum_{k=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = 0.8785$$

Approximation

$$E[S_n] = n * p = 18$$

$$var(S_n) = n * p * (1 - p) = 9$$

$$\sigma_{S_n} = 3$$

$$P\left(\frac{S_n - 18}{3} \leq \frac{21 - 18}{3} = 1\right) = 0.841$$

By invoking central limit theorem:

Compute the mean and variance,

And pretend the distribution is normal,

Think about what we did in the previous slide, we pretend S_n to be normal

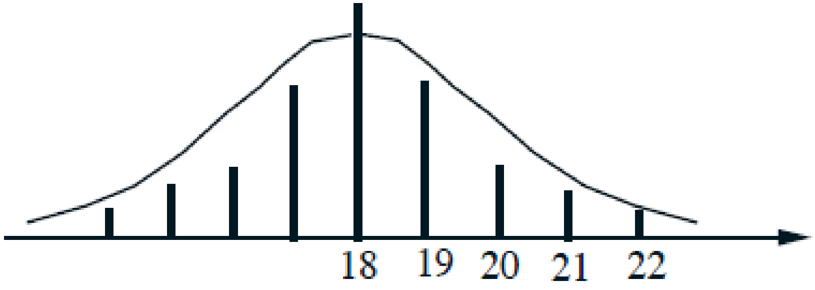
½ correction

You understand that Binomial is discrete:

$$P(S_n \leq 21) = P(S_n < 22)$$

Because S_n is integer

Let's compromise: $P(S_n \leq 21.5)$



Let’s see is this a little closer

$$P\left(\frac{S_n - 18}{3} \leq \frac{21.5 - 18}{3} = 1.167\right) = 0.8769$$

$$\sum_{k=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = 0.8785$$

It is actually a better approximation!

De Moivre-Laplace CLT for binomial

This ½ correction is used, CLT can then be used to approximate the binomial p.m.f. (not just the binomial CDF)

$$P(S_n = 19) = P(18.5 \leq S_n \leq 19.5)$$

$$18.5 \leq S_n \leq 19.5$$
$$\frac{18.5 - 18}{3} \leq (S_n - 18)/3 \leq \frac{19.5 - 18}{3}$$
$$0.17 \leq Z_n \leq 0.5$$

$$P(S_n = 19) \approx P(0.17 \leq Z \leq 0.5)$$
$$= P(Z < 0.5) - P(Z < 0.17)$$
$$= 0.6915 - 0.5675 = 0.124$$

Let's look at the exact answer:

$$\binom{36}{19}\left(\frac{1}{2}\right)^{36} = 0.1251$$

Good rule of thumb
Poisson vs. normal approximation of binomial

- Binomial(n, p)
  - p fixed, $n \to \infty$: normal
  - np fixed, $n \to \infty$, $p \to 0$: Poisson

- $p = \frac{1}{100}, n = 100$: Poisson
- $p = 1/10, n = 500$; normal

Examples

Random variable X is uniformly distributed between -1.0 – 1.0. Let $X_1, X_2, \ldots$, be independent identically distributed random variables with the same distribution as X.

Determine which, if any, of the following sequences (all with $i = 1,2, \ldots$) are convergent in probability.

a) X_i

- X_i is definitely not convergent, since they are all just uniform random variable as stated in the problem definition

b) $Y_i = \frac{X_i}{i}$

- Yes it is convergent in probability to 0

$$\lim_{i\to\infty} P(|Y_i - 0| > \epsilon) = \lim_{i\to\infty} P\left(\left|\frac{X_i}{i} - 0\right| > \epsilon\right)$$

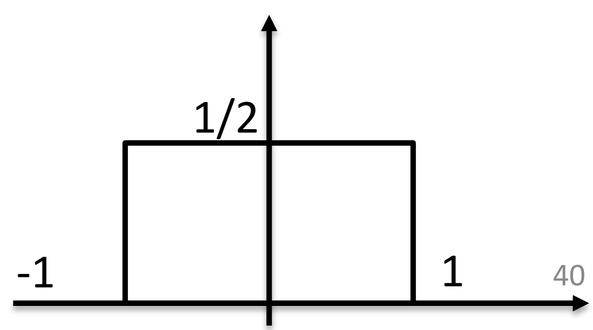
$$= \lim_{i\to\infty} [P(X_i > i\epsilon) + P(X_i < -i\epsilon)]$$

$$= 0$$

c) $Z_i = (X_i)^i$

- Yes convergent in probability to 0

$$\lim_{i\to\infty} P\left(\left|(X_i)^i - 0\right| > \epsilon\right)$$
$$= \lim_{i\to\infty}\left[P\left(X_i > \epsilon^{\frac{1}{i}}\right) + P\left(X_i < -\epsilon^{\frac{1}{i}}\right)\right]$$
$$= \lim_{i\to\infty}\left[\frac{1}{2}\left(1 - \epsilon^{\frac{1}{i}}\right) + \frac{1}{2}\left(1 - \epsilon^{\frac{1}{i}}\right)\right]$$
$$= \lim_{i\to\infty}\left(1 - \epsilon^{1/i}\right) = 0$$



Central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common mean μ, and variance, σ^2, and define:

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then, the CDF of Z_n converges to the standard normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

In the sense that:

$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z)$$

Normal approximation based on the CLT

Let $S_n = X_1 + \cdots + X_n$, where X_i are independent identically distributed random variables with mean μ, variance σ^2. If n is large, the probability $P(S_n \leq c)$ can be approximated by treating S_n as if it were normal, according to the following procedure:

1. Calculate the mean $n\mu$ and variance $n\sigma^2$ of S_n
2. Calculate the normalized value $z = (c - n\mu)/\sigma\sqrt{n}$
3. Use the approximation

$$P(S_n \leq c) \approx \Phi(z)$$

where $\Phi(z)$ is available from the standard normal CDF tables

Strong law of large numbers

- It still deals with the convergence of the sample mean to the true mean

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with mean μ. Then, the sequence of sample means $M_n = (X_1 + X_2 + \cdots + X_n)/n$ converges to μ, with <u>**probability 1**</u>

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1$$

Think about this for a little

- Recall on the sample space
- The experiment is:
  - The experiment is infinitely long
  - Each experiment generate a sequence of value (one value for each of the random variable sequence, $X_1, \dots, X_n$)
  - So sample space as a set of infinite sequences of real numbers $(x_1, x_2 \dots,)$
  - Consider the set A consisting of those sequences whose long term averages is μ

$$(x_1, x_2, \dots) \in A \leftrightarrow \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = \mu$$

- Strong law of large number
  - The collection of outcomes that do not belong to A has probability zero
  - $P\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right) = 1$

- Weak law of large number
  - $\lim_{n\to\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$

- The difference is subtle, but should be noted

Convergence with probability 1

Let Y_1, Y_2 ... be a sequence of random variables (not necessarily independent). Let c be a real value. We say that Y_n converges to c with **probability 1** (**almost surely**) if

$$P\left(\lim_{n \to \infty} Y_n = c\right) = 1$$

Sample space consisting of infinite sequences: all of the probability is concentrates on those sequences that converge to c.