國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 21:
further topics
limiting theorem
李祈均

Logistics change

- 5/3 project proposal announcement
- 5/10 homework 4 due (no class)
- 5/15 quiz 4
- 5/17 project proposal

Final Projects

- Team-based project
  - Final project: (25%)
- Find your topic
  - Present 5/17 your ideas (2)
- What does it involve
  - How do you collect data (show us you have actually collected the data)
  - How 'good' is your inference
  - Probabilistic reasoning processs
  - Q&A from every body

- Idea: probabilistic reasoning -> inference -> real data collection

Final Projects : Exemplary Topics

- 測量每一次等電梯要等幾秒才搭得到 (detla 貨梯 or 客梯)
  - 測量：每次從按電梯到真的搭到電梯的時間
  - infer: 哪一台電梯
  - Decide which distribution
    - Estimate: parameter
    - Make probabilistic reasoning (inference)
- 測量小吃部小七有幾秒收銀員結帳一個人
  - 測量： 時間
  - infer : 哪間小七

- 二十分鐘內進小七的人數
  - 測量：人的數量
  - infer: 哪一家小七

- 珍珠奶茶裏面的珍珠量
  - 測量：珍珠的數量
  - infer: 哪一家飲料店

Proposal presentation

- Present 2 – 3 different ideas that your team wish to explore (each team)
- What do you want to infer?
  - Classes? (discrete category)
  - Numbers? (continuous number)
- How do you plan to collect your data
  - How much data do you need?
- TA will be in class, find a single topic

Law of iterated Expectations

$$E[X] = E\big[E[X|Y]\big]$$

Law of total variance

$$var(X) = E[var(X|Y)] + var(E[X|Y])$$

Total variance of X = expected within-Y X variance + variance of the between-X averages

$$M_Y(s) = E[e^{sY}] = E\big[E[e^{sY}|N]\big] = E[(M_X(s))^N] = \sum_{n=0}^{\infty}\big(M_X(s)\big)^n p_N(n)$$

$$M_Y(s) = \sum_{n=0}^{\infty}\big(M_X(s)\big)^n p_N(n) = \sum_{n=0}^{\infty} e^{n\log M_X(s)} p_N(n)$$

Now let's look at:

$$M_N(s) = E[e^{sN}] = \sum_{n=0}^{\infty} e^{sn} p_N(n)$$

Comparing these two:

$$M_Y(s) = M_N(\log M_X(s))$$

Example

Romeo and Juliet have a date at a given time, and each, independently will be late by amounts of time X and Y, respectively, that are exponentially distributed with parameter λ

Find the PDF of $Z = X - Y$ using two step approach

- Let's break into 2 regions ($z \geq 0$, $z < 0$)

$$F_Z(z) = P(X - Y \leq z) = P(X \leq Y + z)$$

$$= \int_0^\infty \int_0^{y+z} f_{X,Y}(x, y) dx dy$$

$$= \int_0^\infty \lambda e^{-\lambda y} \int_0^{y+z} \lambda e^{-\lambda x} dx\, dy$$

$$= 1 - \frac{1}{2} e^{-\lambda z}, z \geq 0$$

- $z < 0$
- By symmetry
  - We know the distribution $Z = X - Y$ is the same as $-Z = Y - X$

$$F_Z(z) = P(Z \le z) = P(-Z \ge -z) = P(Z \ge -z)$$
$$= 1 - F_Z(-z)$$

Put it together:

$$F_Z(z) = \begin{cases} 1 - \frac{1}{2} e^{-\lambda z}, \text{if } z \ge 0 \\ \frac{1}{2} e^{\lambda z}, \text{if } z < 0 \end{cases}$$

- Simply differentiate it to get the PDF:

$$f_Z(z) = \begin{cases} \frac{\lambda}{2} e^{-\lambda z}, \text{if } z \geq 0 \\ \frac{\lambda}{2} e^{\lambda z}, \text{if } z < 0 \end{cases} = \left(\frac{\lambda}{2}\right) e^{-\lambda |z|}$$

This is called two-sided exponential PDF, also note as Laplace PDF

Example

How do we find the moment transform of Binomial?

Knowing that binomial distribution essentially is a summation of Bernoulli,

First, let's find the moment transform of Bernoulli

$$M(s) = \sum_{x} e^{sX} p_X(x)$$

$$M_{X_i}(s) = (1-p)e^{0s} + pe^{1s} = 1 - p + pe^{s}$$

Summation of random variable?

Multiplication in transform,

Binomial is a sum of ‘n’ independent Bernoulli (parameter: n, p), expectation of joint factors, hence, moment function multiplies

$$M_Z(s) = (1 - p + pe^{s})^n$$

Example

We toss n times a biased coin whose probability of heads, denoted by q, is the value of a random variable Q, with mean μ and positive variance, σ^2

Let X_i be a Bernoulli rv. That models the outcome of the i^{th} toss (i.e., $X_i = 1$ if i^{th} toss is a head).

We assume that $X_1, \ldots, X_n$ are conditionally independent, given $Q = q$. Let X be the number of heads obtained in the n tosses

a) Find $E[X_i]$ and $E[X]$

From law of iterated expectations:

First:

$$E[X_i|Q] = Q$$

So,

$$E[X_i] = E\big[E[X_i|Q]\big] = E[Q] = \mu$$

$$E[X] = E[X_1] + \cdots + E[X_n] = n\mu$$

b) Find $cov(X_i, X_j)$, are X_i's independent of one another?

First for $i \neq j$

With conditional independent:

$$E\left[X_i X_j | Q\right] = E[X_i | Q] E\left[X_j \middle| Q\right] = Q^2$$

Using law of iterated expectation:

$$E[X_i X_j] = E\left[E\left[X_i X_j | Q\right]\right] = E[Q^2]$$

Now we can compute covariance:

$$cov\left(X_i, X_j\right) = E\left[X_i X_j\right] - E[X_i] E\left[X_j\right] = E[Q^2] - \mu^2 = \sigma^2$$

We are told that σ^2 is strictly positive,

So $cov\left(X_i, X_j\right) \neq 0$, hence, they are not independent

For the case of $i = j$

First noting that $X_i^2 = X_i$

$$\begin{aligned} var(X_i) &= E\left[X_i^2\right] - (E[X_i])^2 \\ &= E[X_i] - (E[X_i])^2 \\ &= \mu - \mu^2 \end{aligned}$$

New topic: limit theorem

Imagine a case:

- Assume there is a **population** of penguins down at the south pole

You are asked to perform a task:

- Find the expected value of the penguin's height

What is that?

- Pick a penguin at random and measure their height
- Expected value (with equally-likely) essentially is just the average

That is an extremely tedious time-consuming task (impossible actually!):

Alternatively,

- Pick a number n (that is less than the total), the take the average of the height
- We call this the '*estimate*' of the expected value

This is also called *sample mean*

- Mean value within the samples that you measure (not all)
- Not the same as the *true* expected value (that's over the entire population)

- Expected value is a *number*
- Sample mean is a random variable
  - Why? Because the sample that you compute is random!

Okay, we can think that is **sample mean** is a reasonable way of **estimating** the **expectation**

So we are thinking,

In the limit, n goes to infinity,

Does it (sample mean – a function) get close to the true expected value (a number)?

What does this mean to get close?

- Get close in what manner (in what sense?)
- Is it true that we actually get close ? (converged?)

This will the topic centered around limit theorem

$$X_1 + \cdots + X_n \text{ iids}, M_n = \frac{X_1 + \cdots + X_n}{n}$$

What happens as $n \to \infty$

Why bother

- First:
  - Of course if you are in the sampling business (e.g., marketing, polls, etc), you want to know this whether this way of estimating the true expected value gets you close to the true answer

- Second:
  - In terms of probabilistic reasoning: it's easy to work out probability problems if you have say 1 or 2 random variables that you can write down their mass functions, joint pdf, etc

- If you have lots of variables, this calculation quickly becomes intractable!

- Now, if you assume you are working with millions of variables, then you are in the realm of ‘taking the limits’

- By taking the limits, many formulas start to simplify, and you can get useful answers easily!

We are going to talk about different ways of ‘getting close’

– First, we will introduce a few *tools*

Outline

- Tools: Markov and Chebyshev inequality
  - Easy to obtain mean and variance
  - PDF is not
- Define the meaning of convergence for random variables
- Convergence "in probability"
- Then we will see about that sample mean's convergence
  - Weak law of large numbers
- Central limit theorem

Markov Inequality

If a random variable X can only take non-negative values, then

$$P(X \geq a) \leq \frac{E[X]}{a}$$

- It relates **expected values** to **probabilities**
- Loosely speaking, it asserts that if a *nonnegative* random variable has a small mean, then the probability that it takes a large value must also be small

Let’s justify a little

- Let fix a positive number a and consider the following variable

$$Y_a = \begin{cases} 0, \text{if } X < a \\ a, \text{if } X \geq a \end{cases}$$

From this, we know

$$Y_a \leq X$$

Always holds, and therefore,

$$E[Y_a] \leq E[X]$$

$$E[Y_a] = aP(Y_a = a) = aP(X \geq a)$$
$$aP(X \geq a) \leq E[X]$$
$$P(X \geq a) \leq \frac{E[X]}{a}$$

Quick example:

Let X be uniformly distributed in the interval $[0,4]$ and note that $E[X] = 2$. Then, the Markov inequality assert that:

$$P(X \geq 2) \leq \frac{2}{2} = 1, P(X \geq 3) \leq \frac{2}{3} = 0.67, P(X \geq 4) \leq \frac{2}{4} = 0.5$$

Compare this with exact probability

$$P(X \geq 2) \leq 0.5, P(X \geq 3) \leq 0.25, P(X \geq 4) = 0$$

Markov inequality is a very conservative bound

Chebyshev's inequality

- Definition

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

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

- Loosely speaking, if a random variable has a small variance, the probability that it takes a value far from its mean is also small
- Chebyshev does not require the random variable to be nonnegative

Justify a little

- Consider this random variable:

$$(X-\mu)^2$$

Now apply the Markov inequality with $a = c^2$

$$P\big((X-\mu)^2 \geq c^2\big) \leq \frac{E[(X-\mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

Now observing that the event $(X-\mu)^2 \geq c^2$ is identical to the event $|X-\mu| \geq c$, so that:

$$P(|X-\mu| \geq c) = P\big((X-\mu)^2 \geq c^2\big) \leq \frac{E[(X-\mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

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

$$P(|X-\mu| \geq k\sigma) \leq \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2},$$

Now, a quick example:

Let X be exponentially distributed with parameter $\lambda = 1$, so that $E[X] = var(X) = 1$

$$P(X \geq c) = P(X-1 \geq c-1) \leq P(|X-1| \geq c-1) \leq \frac{1}{(c-1)^2}$$

This is still conservative compared to the exact answer

$$P(X \geq c) = e^{-c}$$

Now let's talk about convergence…

- Deterministic limits

- Given:
  - There is a sequence of values, a_n
  - There is a value, a

- We say a_n converges to a

$$\lim_{n \to \infty} a_n = a$$

"a_n eventually gets and stays (arbitrarily) close to a"

Officially define this form of convergence:

For every $\epsilon > 0$,

There exists a n_0

Such that for every $n > n_0$

We have $|a_n - a| \leq \epsilon$

Convergence in probability

- Say we have a sequence of random variables Y_n

- Converges in probability to a number a:
  - "(almost all) of the PMF/PDF of Y_n eventually gets concentrated (arbitrarily) close to a number a"

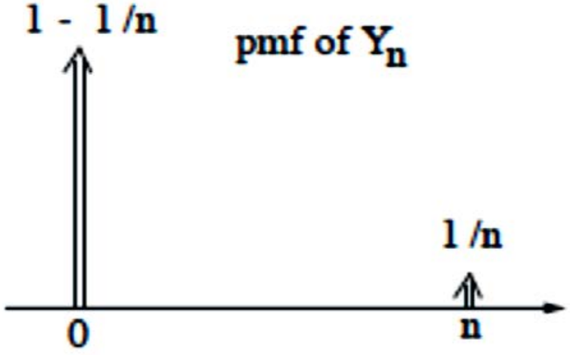
- That means: for every $\epsilon > 0$

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

In words:

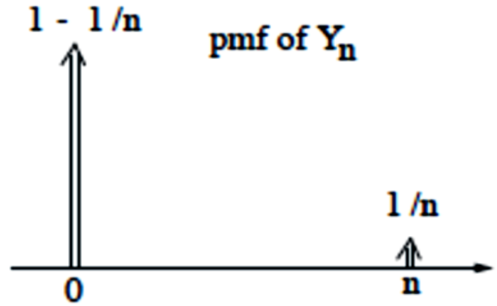
For every $\epsilon > 0$, there exists a n_0, such that for every $n > n_0$, we have$P(|Y_n - a|) \leq \epsilon$

Simple example



Does Y_n converge?

- Is there a convergence in probability?
  - We can see that as n gets larger and larger, more and more probability concentrates around 0
  - If we have an ε-band around 0, the probability of falling outside of this band as n goes to infinity, is 0



Does Y_n converge?

Interesting point:

Y_n in probability converges to 0

Let's compute expected value:

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

As n goes to infinity, variance goes to infinity

- Convergence in probability **does not** imply anything about convergence of expected values or variances and so on … (be careful)

- Hand-wavy reason:
  - Convergence in probability tells you that tail probability is extremely small
  - It does not tell you how far does that tail go

  - In the previous example
    - That tail’s probability is extremely small, but that tail is extremely far away
    - That gives a disproportionate contribution to the expected value or variance

Now back to the sample mean we started…

- Convergence of the **sample mean**

- Let's define it

$$X_1, X_2, \ldots \text{iids}$$

Assume each has mean μ and variance σ^2

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

Find $E[M_n]$

$$E[M_n] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu$$

There is a subtle point here

- There are two kinds of averages going on here
- Sample mean is the average of the heights collected at a single round
- Expectation can be imagined as the averages over a huge number of rounds

- So what this means:
  - You do a random experiment get an average, and if you do this over zillion number of times, on average your sample mean is the overall mean

So M_n is a random variable, how random it is?

- Let's try to compute the variance

$$var(M_n) = \frac{1}{n^2} n\left(\sigma^2\right) = \frac{\sigma^2}{n}$$

- So as n gets larger, variance gets smaller smaller
- So that means if you take a large number of samples, the variance of the sample mean is very small
- Randomness of your estimate gets removed

The weak law of large number

Definition: sample mean, M_n

$$P(|M_n - \mu| \geq \epsilon) \leq \frac{var(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Now if we take the limit as n goes to infinity, that probability goes to 0

- It means sample mean, M_n converges in probability to μ

Example

Say if you are doing a sampling poll for Coke and Pepsi preference,

You want to know out of the a population of 100 million people what is the fraction, f ,of people prefer Coke?

Say the exact number is 20 million, then $f = 0.2$

However, we can't ask everybody, so we sample n people

Now let's set:

- ith (randomly selected) person polled:

$$X_i = \begin{cases} 1, \text{if prefers Coke} \\ 0, \text{otherwise} \end{cases}$$

- $M_n = \frac{(X_1 + \cdots + X_n)}{n}$
  - This sample mean essentially is the fraction of 'prefers coke' in our sample
- So it is impossible to get the true 'fraction', but you are given this task, how do you go about answering your boss' request?
- You can design a goal for this task (95% confident within 1% error)

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

How do we go about computing this:

Max variance p(1-p)/n
p=0.5

$$P(|M_n - f| \geq 0.01) \leq \frac{\sigma_{M_n}^2}{(0.01)^2} = \frac{\sigma_x^2}{n(0.01)^2} < \frac{1}{4n(0.01)^2}$$

Chebyshev

- So if $n = 50{,}000$
  - $P(|M_n - f| \geq 0.01) \leq 0.05$

- This is very conservative bound why?
  - Chebyshev inequality
  - Also 1% error (changing to 3%, it shrinks close to a factor of 10)

Different scalings of M_n

- $X_1, \dots, X_n$ i.i.d. with finite variance σ^2

- If you simply add it:

$$S_n = X_1 + \cdots + X_n$$

  - Expectation: $nE[X]$, variance: $n\sigma^2$
  - Expectation goes to infinity and variance does too

- Now, if we scale it by n

$$M_n = \frac{1}{n} S_n,$$

  - This rv **converges in probability** to $E[X]$
  - Weak law of large number

- You can think about these two as two extreme cases
  - No scaling: it flattens out, grows to infinitiy
  - n scaling: it concentrates narrowly around a number μ

- $\frac{S_n}{\sqrt{n}}$
  - This rv would have expected value: $\frac{n}{\sqrt{n}}E[X]$
  - This rv would be variance, σ^2
    - Does this one has any asymptotic shape in the limit?

The central limit theorem

- Let’s standardize this summation of random variable
  - $S_n = X_1 + \cdots + X_n$

$$Z_n = \frac{S_n - E[S_n]}{\sigma_{S_n}} = \frac{S_n - nE[X]}{\sqrt{n}\sigma}$$

  - This new r.v. has zero mean and unit variance
  - Does this look like anything?
- Let Z be a standard normal r.v.
- CLT: for every c

$$P(Z_n \leq c) \rightarrow P(Z \leq c)$$

  - $P(Z \leq c)$ is the standard normal CDF
  - Z_n CDF converges to standard normal CDF

- Z_n pdf can be extremely complicated
  - We can compute probability using this shortcut
  - Now we can pretend Z_n is normal, S_n can be pretended as normal
  - Strictly speaking, CLT talks about CDF function of Z_n NOT S_n

- We will try to prove it next time!