國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 12:
Discrete random variable
Continuous random variable

李祈均

The hat problem

- n people throw their hats in a box and then pick one at random (imagining this at a cocktail party, where everybody wears identical hat, but needs to be checked in at the front gate!)
  - X: number of people who get their own hat back
  - Find E[X]?

- First we realize that X:

$$X = X_1 + X_2 + \cdots + X_n$$

$$X_i = \begin{cases} 1, \text{if i selects own hat} \\ 0, \text{otherwise} \end{cases}$$

These random variables are much easier to handle!

So what is :

$$P(X_i = 1)?$$

- That probability, which is the probability at random, the i^{th} person will get his hat back among n hat is

$$P(X_i = 1) = \frac{1}{n}$$

$$P(X_i = 0) = 1 - \frac{1}{n}$$

So,

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

Conditional vs. unconditional issue

- In absence of information, the probability is 1/n

P(2^{nd} right)= P(1^{st} right 2^{nd} right) or P(1^{st} wrong 2^{nd} right)

=

P(1st right)P(2nd right | 1st right)+P(1st wrong and got 2nd one's hat)·P(2nd right | 1st wrong and got 2nd one's hat)+P(1st wrong and didn't get 2nd one's hat)·P(2nd right | 1st wrong and didn't get 2^{nd} one's hat)

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

P2 mean and variance of sample mean

Say we wish to estimate the approval rating of a president, to be called B. To do this, we ask *n* persons drawn at random from the voter population, and we let X_i be a random variable that encodes the response of the i^{th} person:

$$X_i = \begin{cases} 1, \text{if the i}^{th} \text{ person approves B} \\ 0, \text{if the i}^{th} \text{ person disapproves B} \end{cases}$$

We can treat each of this Xi as independent Bernoulli random variables with common mean p and variance (1-p)

We can see this ‘p’ as the true approval rate of B

Let’s take an average of the response from each Xi that we get:

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

We can see this random variable Sn, as the approval rate of B within our n-person population

What is the expected value of this random variable?

$$E[S_n] = \sum_{i=1}^{n} \frac{1}{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} p = p$$

Making use of independence to compute variance:

$$var(S_n) = \sum_{i=1}^{n} \frac{1}{n^2} var(X_i) = \frac{p(1-p)}{n}$$

- Sample mean as a random variable, Sn, on average is a ‘good’ estimate of the true approval rating

- We can see the variance of this random variable, Sn, it gets smaller and smaller as we have larger and larger n

- The estimation to the true approval rating is more and more ‘pinpointing-accurate’ as we have more and more samples!

Justification of Poisson approximation property

- Consider the PMF of a binomial random variable with parameters n and p, show that asymptotically, as

n -> ∞ and p -> 0,

While np is fixed at a given value λ, this PMF approaches the PMF of a Poisson random variable with parameter λ

Set $\lambda = np$

$$p_X(k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}$$

$$= \frac{n(n-1)\ldots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)\ldots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Fix k, let $n \to \infty$

$\frac{n(n-1)\ldots(n-k+1)}{n^k}$ goes to 1

$\left(1 - \frac{\lambda}{n}\right)^{-k}$ goes to 1

$$\left(1-\frac{\lambda}{n}\right)^n \text{ goes to } e^{-\lambda}$$

Remember exponential is defined as : $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$

Put it together now we can for each k, as n -> ∞

$$p_X(k) \to \; e^{-\lambda}\frac{\lambda^k}{k!}$$

Review

- Random variable X: function from sample space to the real numbers
- PMF (for discrete random variables):
  - $p_X(x) = P(X = x)$
- Expectation:

$$E[X] = \sum_x x p_X(X)$$

$$E[g(X)] = \sum_x g(x) p_X(x)$$

$$E[\alpha X + \beta] = \alpha E[X] + \beta$$

- Variance

$$var(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 p_X(x)$$

$$= E[(x - E[X])^2 p_X(x)] = E[X^2] - (E[X])^2$$

- Standard deviation (to get the unit right, e.g., distance)
  - $\sigma_X = \sqrt{var(X)}$

Some review

- Bernoulli random variable
  - Consider toss of a coin, which comes up with probability p, and a tail with probability 1-p
  - The Bernoulli random variable takes the two value
  - $X = \begin{cases} 1, \\ 0, \end{cases}$
  - PMF, $p_X(x) = \begin{cases} p, \text{ if X=1} \\ 1-p, \text{ if X=0} \end{cases}$
  - Cases:
    - State of telephone: free or busy
    - State of a person: healthy or sick

Binomial PMF

- n independent tosses, each with success probability of p
- X ~ B(n,p)
- Say Y is Bernoulli distribution
  - $X = \sum_{i=1}^{n} Y_i$

Poisson distribution

- How to imagine this?
- Imagine this as binomial random variable with very small p and very large n
  - For example, let X be the number of typos in a book with a total of n words
  - Or the number of car involved in accidents in a city on a given day
  - The number of phone calls received by a call center per hour
  - The number of taxis passing particular street corner per hour

expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event.

Multinomial distribution

A die with r faces, numbered 1, …, r, is rolled a fixed number of times, n.

The probability that the i^{th} face comes up on any one roll is denoted p_i, and the results of different rolls are assumed independent.

Let X_i be the number of times that the i^{th} face comes up,

Geometric PMF

- X: the number of independent coin tosses until first head

$$p_X(k) = (1-p)^{k-1}p, k = 1,2, \ldots$$

Negative Binomial Random Variable

- Negative binomial random variable is a generalization of geometric random variable

- Suppose that we have a sequence of independent Bernoulli trials, each with success rate of *p*

- Let X be the number of experiments **until** the r-th success occurs, then this random variable is called ‘negative binomial’

PMF:

$$p_X(x) = P(X = x)$$

Joint PMF:

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Conditional PMF:

$$p_{X|Y}(x|y) = P(X = x|Y = y)$$

Marginal PMF:

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

Joint PMF as conditional PMF:

$$p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y|x)$$

- The conditional PMF of X of given Y=y is related to the joint PMF by:

$$p_{X,Y}(x, y) = p_Y(y)p_{X|Y}(x|y)$$

- The conditional PMF of X given Y can be used to calculate the marginal PMF of X through the formula:

$$p_X(x) = \sum_y p_Y(y)p_{X|Y}(y)$$

- The conditional expectation of X given an event A with P(A) > 0 is defined by:

$$E[X|A] = \sum_x x p_{X|A}(x)$$

$$E[X] = \sum_y p_Y(y)\, E[X|Y = y]$$

- Independence

$$p_{X,Y,Z}(x, y, z) = p_X(x) p_Y(y) p_Z(z), \text{for all } x, y, z$$

Lecture outline

Readings: Section 3.1 – 3.3

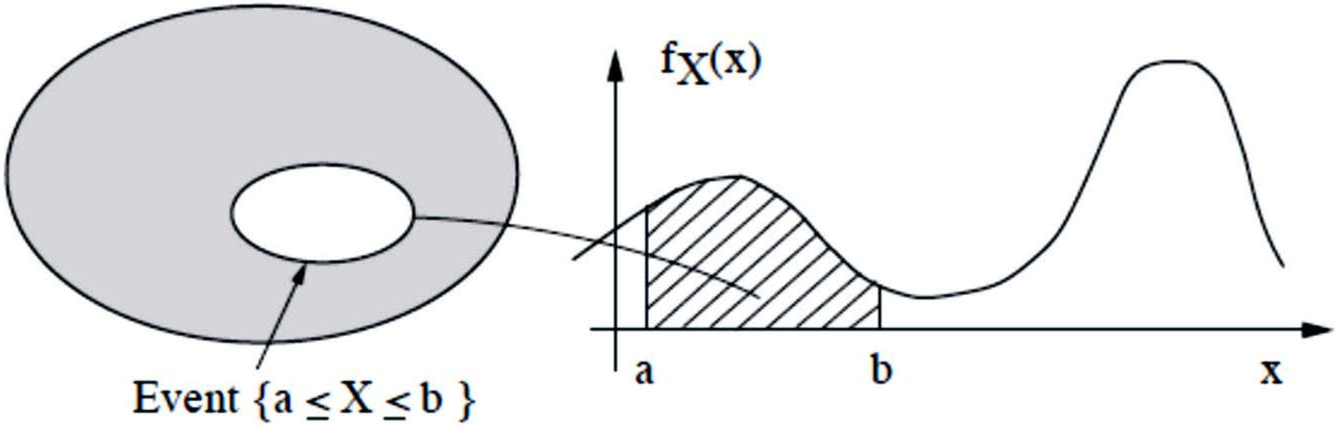
- Probability density functions
- Cumulative distribution functions
- Normal random variables

Continuous random variable and pdfs

- A continuous random variable is still a function that take values over a continuum

- A random experiment -> sample space -> sample point -> rv determines a numerical value

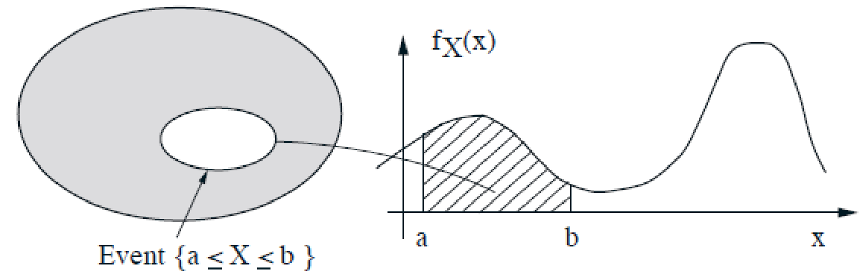
- Again, we want to say which values are more likely than others



- Say you want to find the probability of that event which have created the values you see
  - In principle, you go back to sample space, look at all outcomes
  - But we would like to essentially move away (just like chapter 2), directly work at the rv
  - Chapter 2, we introduce PMFs
    - Different masses sitting on different points
  - Here we are going to talk about PDFs (probability density function)

PDF: probability density function

- Formally define this



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

What is this?

- The probability of falling in this interval is the area under the curve (kinda similar to PMF discrete case)
- The probability of any single point?
  - Area of any single point? 0!

- But if you look at the density
  - Density is not 0, it is different from probability

- First let’s look at density function
  - Requirements:
  - Density function can not be negative, since probability is non-negative
  - $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Intuitive interpretation:

Think about density is to look at probability of small intervals (very small interval),

$$P(x \leq X \leq x + \delta) = \int_{x}^{x+\delta} f_X(s)ds \approx f_X(x)\delta$$

If δ is small, density is constant over that small interval, the probability is just base * height

Given this,

Density is probability per unit length = rates at which probabilities accumulate

Density is not probability,

does not have to be less than 1

it can be infinity even, as long as the area is 1

So, for given density function, we can compute probability of intervals, what if we want to compute probability of general set?

For ***nice*** set, it is straightforward:

$P(X \in B) = \int_B f_X(x)\,dx$ for nice set B (e.g., union of multiple interval sets)

- How to define ‘nice’ – that’s very advance
  - We can imagine B to be something like a union of intervals
  - Union of two intervals, just integrate over 1 interval, and integrate another 1 interval, then add them up!

- Density function is a complete description of any statistical information we might be interested for continuous random variable

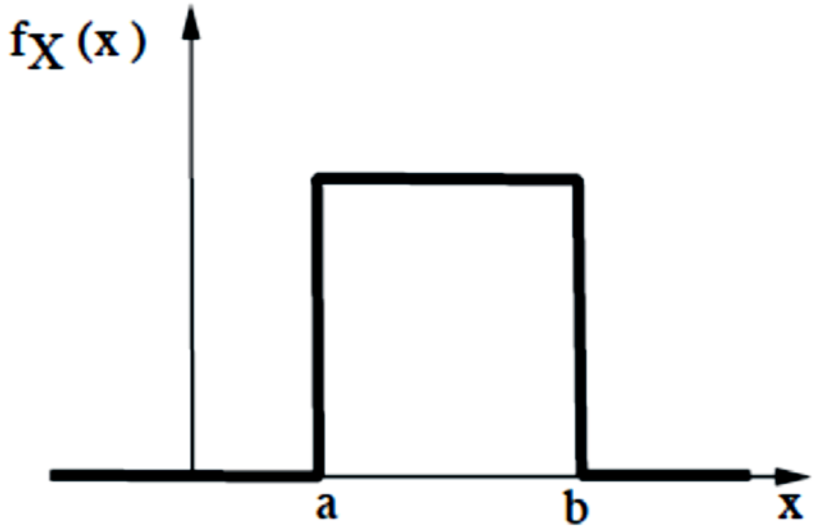
Means and variance

- Same notion as for discrete case
- $E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
  - Physics: center of gravity
  - Again: this is just like average intuitively (we will prove it rigorously later)

- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$

- $var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - E[X])^2 f_X\,(x)\, dx$

Example



- Uniform random variable
  - Note: at every point x, the probability is 0
  - If I take an interval of a given length, and if I take another interval, under the uniform distribution, two intervals are going to have the same probability

- $f_X(x) = \frac{1}{b-a}, a \le x \le b$
  - 0 otherwise

- $E[X]$?
  - You can do the integration
  - $E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} * \frac{1}{2} x^2 |_a^b = \frac{1}{b-a} \frac{b^2-a^2}{2} = \frac{a+b}{2}$
  - simply think about it as center of gravity
    - $\frac{a+b}{2}$

- $\sigma_X^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}$

- We can also do it by $E[X^2] - E[X]^2$

- $E[X^2] = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \int_a^b x^2\, dx = \frac{1}{b-a} * \frac{1}{3} x^3 \Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$

$$var(X) = E[X^2] - E[X]^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

If you look at standard deviation, the unit is the same, it's proportional to the interval – makes intuitive sense

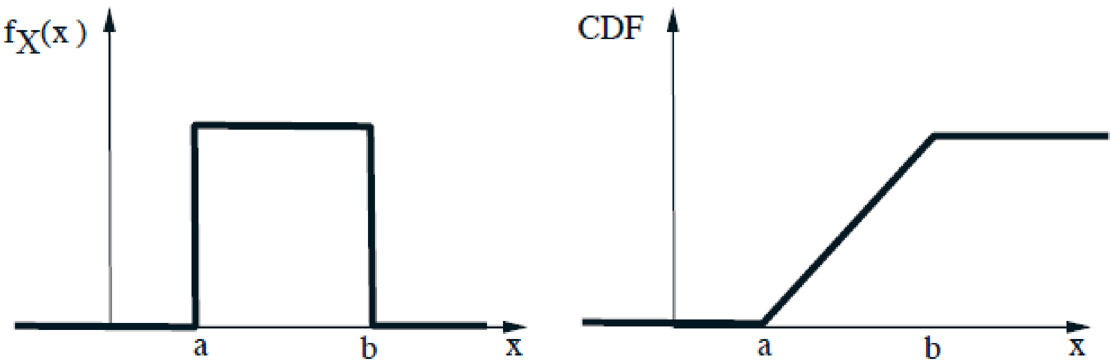
Cumulative distribution function

- A method to describe discrete PMF and continuous PDF, A unifying concept
  - Cumulative distribution function of a random variable

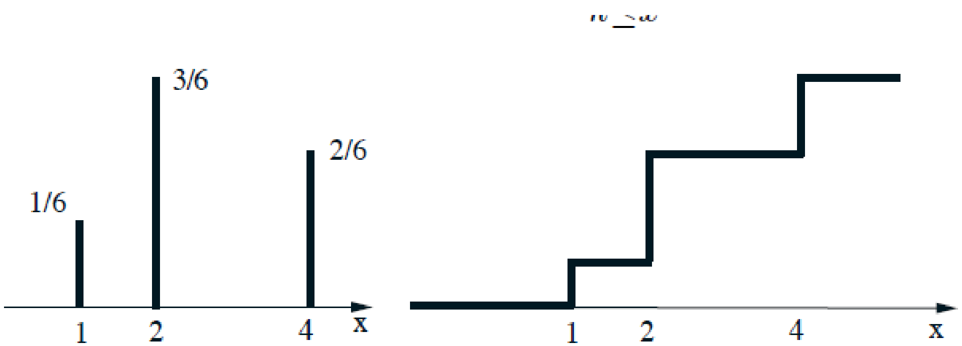
- Definition:

$$F_X(x) = P(X \leq x)$$

- Continuous case
  - $\int_{-\infty}^{x} f_X(t)dt$

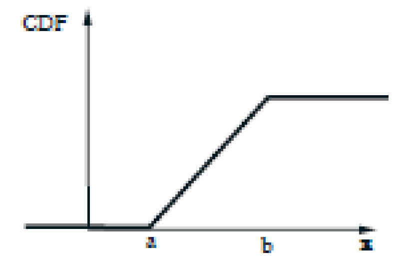


- Discrete case
  - $\sum_{k \leq x} p_X(k)$



Some CDF property

- CDF value at $x = b$ (for the right function)
  - Equals to 1, total probability equals 1



- How to find the density function if we are given CDF?
  - Essentially taking the derivative (for PDF)
  - Small caveat: derivative at the corner is undefined
    - Does it really matter? Does not matter in the integration

- CDF for PDF and PMF are both well-defined

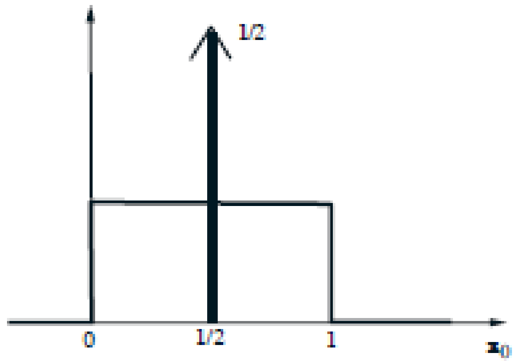
- CDF can used to specify things that are neither discrete nor continuous

Let's say you play a game, say with certain probability, you get a certain dollars in your hands.

- So with probability ½, you get reward of ½ dollar
- With probability ½, you spin a wheel in a dark room, and gives a random reward between [0,1] dollar, and any of the dollar amount is equally-likely (uniform)

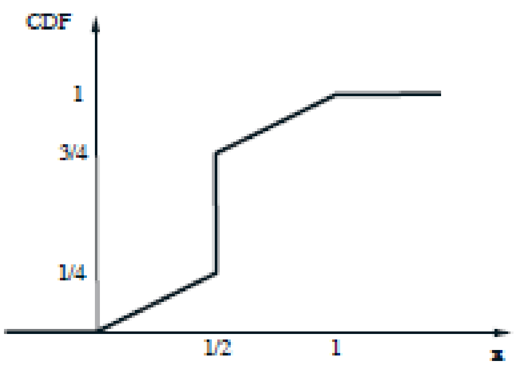
- So you flip a coin, you either get ½ dollar or some continuous value, is this random variable discrete or continuous?
  - Neither, it's a mix!

If you really want to draw this mix variable density (or mass function):



– Look at the delta function, if you have taken signal and system: that's a impulse response

- Now let's try CDF $F_X(x) = P(X \leq x)$



Simple example

Let Z be a continuous random variable with probability density function

$$f_Z = \begin{cases} \gamma\left(1 + z^2\right), \text{if } -2 < z < 1 \\ 0, \text{otherwise} \end{cases}$$

a) For what value is γ is this possible

Knowning that the PDF must integrate to 1, let's just carry out the calculation:

$$\int_{-\infty}^{\infty} f_Z(z)dz = \int_{-2}^{1} \gamma(1 + z^2) = \gamma\left(z + \frac{1}{3}z^3\right)|_{-2}^{1} = 6\gamma$$

Now, we know

$$\gamma = 1/6$$

b) Find the cumulative distribution function of Z

To find CDF, we integrate:

$$F_Z(z) = \int_{-\infty}^{z} f_Z(t)dt = \begin{cases} 0, \text{if } z < -2 \\ \frac{1}{6}\left(t + \frac{1}{3}t^3\right)\Big|_{-2}^{z}, \text{if } -2 \leq z \leq 1 \\ 1, \text{if } z > 1 \end{cases}$$

$$= \begin{cases} 0, \text{if } z < -2 \\ \frac{1}{6}\left(z + \frac{1}{3}z^3 + \frac{14}{3}\right), \text{if } -2 \leq z \leq 1 \\ 1, \text{if } z > 1 \end{cases}$$