國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 8: discrete random variable

李祈均

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

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

Variance of a constant? 0

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

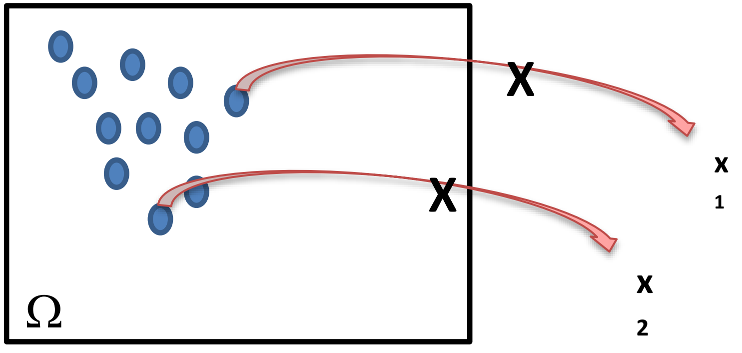
$$Y = g(X)$$

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

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

$$= \sum_y y \sum_{\{x|g(x)=y\}} p_X(x) = \sum \sum_{\{x|g(x)=y\}} y p_X(x)$$

$$= \sum_y \sum_{\{x|g(x)=y\}} g(X) p_X(x) = \sum_x g(X) p_X(x)$$



Therefore, for any "moment" of the function

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

Some PMF distribution

- 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
  - $X = \begin{cases} 1, & \text{X=0} \\ 0, & \text{X=1} \end{cases}$
  - PMF, $p_X(x) = \begin{cases} p, & \text{if k=1} \\ 1-p, & \text{if k=0} \end{cases}$
  - Cases:
    - State of telephone: free or busy
    - State of a person: healthy or sick

- Expectation of Bernoulli variable

$$E[X] = \sum_k X_k P(X = X_k) = 1 * p + 0 * (1 - p) = p$$

- Variance of Bernoulli variable

$$Var[X] = \sum_k \left(x_k \ - E(X)\right)^2 P(X = x_k)$$
$$= (1 - p)^2(p) + (0 - p)^2(1 - p) = p(1 - p)$$

What is parameter of this PMF? *p*

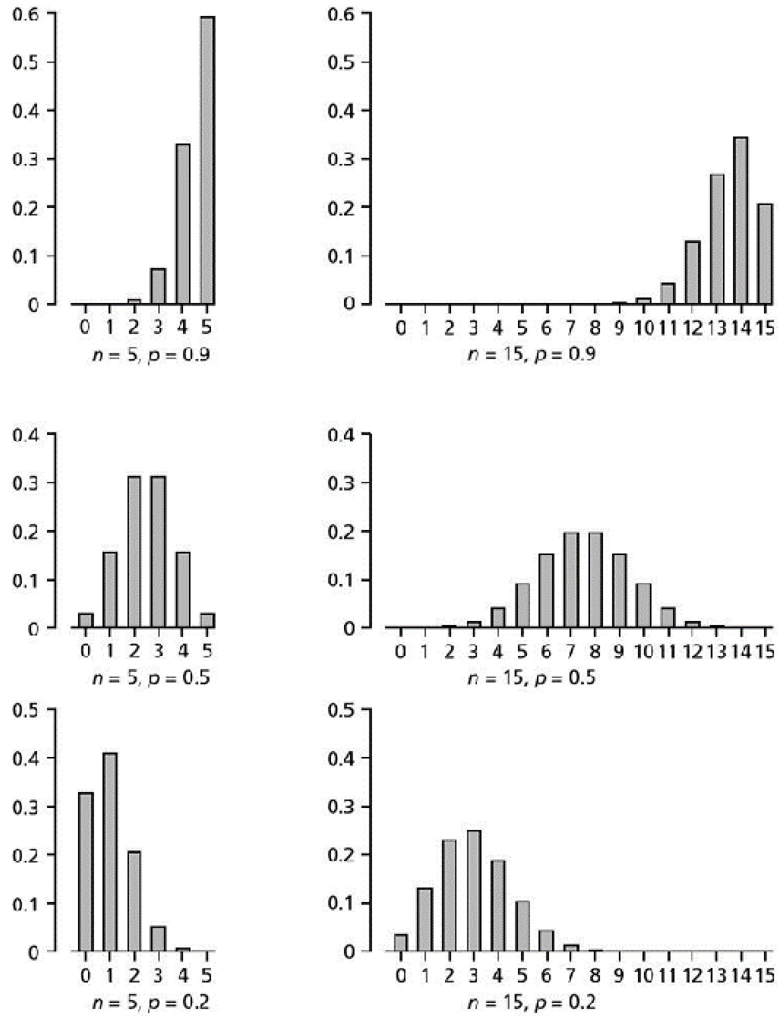
Binomial PMF

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

What the E[X]?

$$E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} p = np$$

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



- Var(Y) = ?
  - $\sum_i p(1-p)$
  - n p q , where q= (1-p)

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.

Poisson distribution

- $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 1,2,3 \ldots$
- λ is the parameter of this distribution (number of events occurring within an interval)

Approximates Binomial:

Take n=100, p=0.01 (=> k is 5, λ = np =1)

Using Binomial: $\frac{100!}{95!5!}0.01^5(1-0.01)^{95}= 0.00290$

Using Poisson: $e^{-1}\frac{1}{5!} = 0.00306$

p1

Let X and Y be independent random variables. Random variable X has mean μ_X and variance σ_X^2, and random variable has mean μ_Y and variance σ_Y^2.

Let $Z = 2X - 3Y$, find the mean and variance of Z in terms of means and variances of X and Y

Solution

Straightforward application for linearity of expectation
$E[\alpha X + \beta Y + c] = \alpha E[X] + \beta E[Y] + c$, therefore,

$$E[Z] = 2E[X] - 3E[Y]$$

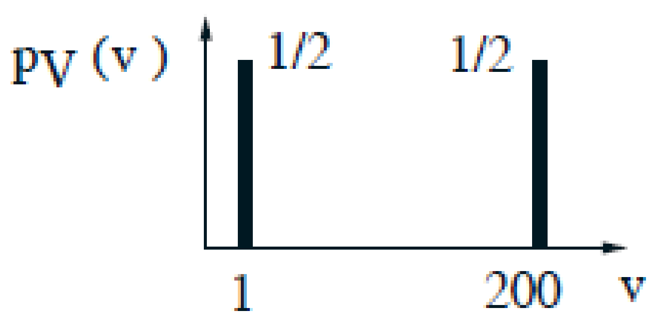
Independent random variables:

$$var(Z) = \alpha^2 var(X) + \beta^2 var(Y)$$

α = 2, β = 3

Random speed

- Let's try an example:
  - Say you are going to go from here to New York, 200 miles
  - You have two alternatives:
    - 1: get your private plan go at a speech 200 miles/hr, constantly
    - 2: decide walk slowly 1 mile/hr
    - You pick the speed at random by doing this experiment
    - Let V be an random variable



- If you are interested in how much time it's going to take you to get there
  - Time equals to distance divided by speed
  - d=200, T=t(V) = 200/V

Let's do some trivial examples:

- E[V]= ?
  - $0.5 * 200 + 0.5 * 1 = 100.5$
- Var[V]= ?
  - $0.5 * (200 - 100.5)^2 + 0.5 * (1 - 100.5)^2 = 9900.25$
  - A better intuitive understanding is to take the square root to get standard deviation σ_V = 99.5

Average speed vs. average time

- T=t(V)=200/V
- $E[V] = 100.5$
- $E[T] = E[t(V)] = \sum_v t(v) p_V(v) = 0.5 * 200 + 0.5 * 1 = 100.5$ (on average, how much time it takes)
- $E[TV] =$? T*V = 200 no matter what so :
  - $E[TV] = 200 \neq E[T]E[V]$ = 100000
  - $E\left[\frac{200}{V}\right] = E[T] \neq 200/E[V]$

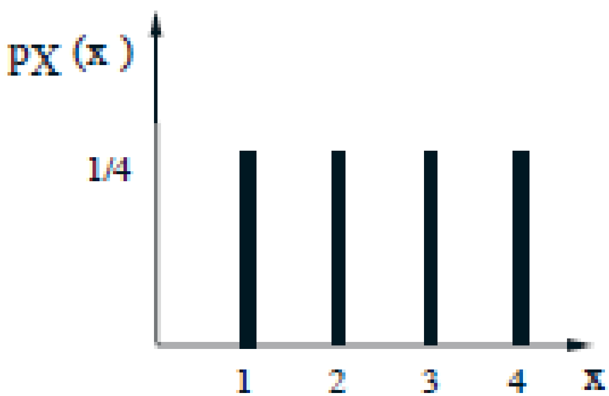
Lecture outline

Readings: Section 2.5 -

- Conditional PMF and expectation
- Geometric PMF
- Total expectation theorem
- Joint PMF of two random variables

Conditional PMF and expectation

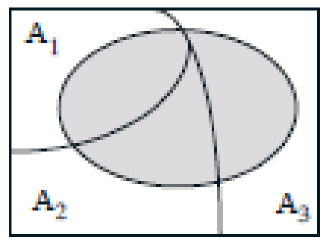
- $p_{X|A}(x) = P(X = x|A)$
- $E[X|A] = \sum_x x p_{X|A}(x)$



- Let A={X ≥ 2}
- $p_{X|A(x)} = \begin{cases} \frac{1}{3}, x = 2 \\ \frac{1}{3}, x = 3 \\ \frac{1}{3}, x = 4 \end{cases}$
- $E[X|A] = \frac{2}{3} + 1 + \frac{4}{3} = 3$
- Everything you know about expectation holds when you goes to conditional prob.

Total expectation theorem

- Partition sample space into n points



- $P(B) = P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)$
- $p_X(x) = P(A_1)p_{X|A_1}(x) + \cdots + P(A_n)p_{X|A_n}(x)$
- $E[X] = P(A_1)E[X|A_1] + \cdots + P(A_n)E[X|A_n]$

Geometric PMF

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

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

$$E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

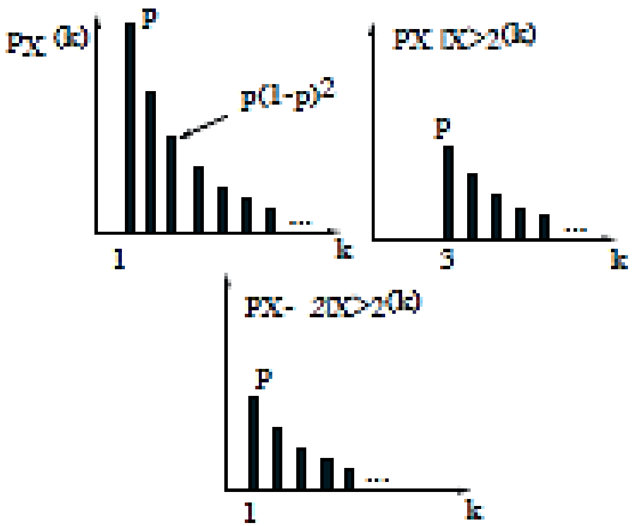
Back to geometric case

- A1: {X=1}, A2:{X>1}

- $E[X] = P(X = 1)E[X|X = 1] + P(X > 1)E[X|X > 1]$
$$= p * 1 + (1 - p)E[\mathrm{X}|\mathrm{X} - 1 > 0]$$
$$= p + (1 - p) * E[X + 1|X > 0]$$
$$= p + (1 - p)(E[X] + 1)$$
- E[X] – (1-p)E[X] = 1
- E[X] = 1/p

- Two people tossing, X & Y
  - X tosses twice first before Y, and X get both tails

- Memoryless property: Given that $X > 2$, the r.v. $X - 2$ has same geometric PMF



After two toss, if we imagine is as a complete new experiment

So what is it formally?

$$p_{X-2|\{X>2\}}(k) = p_X(k)$$

Memoryless:

Geometric variable are memoryless;
Whatever happens in the future is
Independent from whatever happens
In the past…

Joint PMF

- Joint PMF, association between random variables (height and weight, for example)
- $p_{X,Y}(x, y) = P(X = x \text{ and } Y = y)$
  - The probabilities that certain x goes with certain y
  - This is just a notation on the 'joint event' that we described in chapt 1

Finite outcome

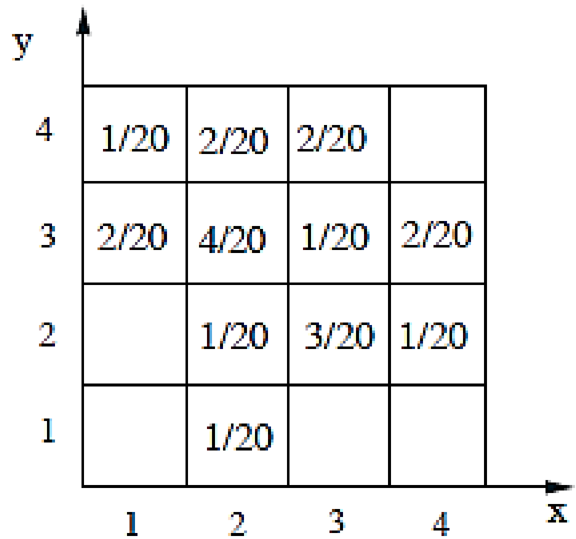
- Represent as tables

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

- First: $\sum_x \sum_y p_{X,Y}(x,y) =? 1$

If we don't care about y..

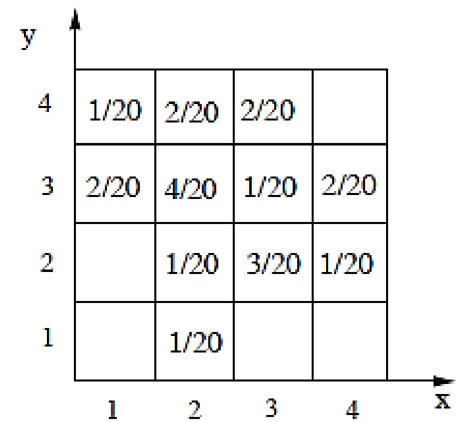
- Marginally the effect of y



- $p_X(x) = \sum_y p_{X,Y}(x, y)$
  - Take an example of finding $p_X(X=3)$

Conditional probability?

- $p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{P_y(y)}$
- Take y=2



- $p_{X|Y}(x|y = 2) = \begin{cases} \frac{1}{5}, x = 2 \\ \frac{3}{5}, x = 3 \\ \frac{1}{5}, x = 4 \end{cases}$

- $\sum_x p_{X|Y}(x|y) =?$
- 1

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,

a) Find the joint PMF

- The probability of a sequence of rolls where, for i=1…r, face i comes up k_i times is $p_1^{k_1} \dots p_r^{k_r}$.
  - k_i's sum to n
- Every such sequence determines a partition of the set of n rolls into r subsets with the i^{th} subset having cardinality k_i (this is the set of rolls for which the i^{th} face come up)
- How many total subsets are there?
  - Multinomial coefficient:
  - $\binom{n}{k_1,\dots,k_r} = \frac{n!}{k_1!\dots k_r!}$

- Thus, if $k_1 + \cdots + k_r = n$,

$$p_{X_1,\ldots,X_r}(k_1, \ldots, k_r) = \binom{n}{k_1, \ldots, k_n} p_1^{k_1} \ldots p_r^{k_r}$$

Otherwise: $p_{X_1,\ldots,X_r}(k_1, \ldots, k_r) = 0$

Joint PMF of binomial = multinomial

Binomial PMF

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

What the E[X]?

$$E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} p = np$$

b) Find the expected value and variance of X_i

- The random variable X_i is a binomial with parameters n and p_i, Therefore:

$$E[X_i] = np_i$$

$$Var(X_i) = np_i(1 - p_i)$$

c) Find E[X_i X_j] for $i \neq j$

- Suppose that $i \neq j$ and let $Y_{i,k}$ (or $Y_{j,k}$) to be Bernoulli random variable that takes the value of 1 if face i (or j, respectively) come up at the k^{th} roll, and the value of 0 otherwise
- Knowing the following:
  - $Y_{i,k}Y_{j,k} = 0$
  - If $l \neq k$, then $Y_{i,l}Y_{j,k}$ are independent -> $E[Y_{i,l}Y_{j,k}] = p_i p_j$

$$E\left[X_i X_j\right] = E\left[\left(Y_{i,1} + \cdots + Y_{i,n}\right)\left(Y_{j,1} + \cdots + Y_{j,n}\right)\right]$$
$$= (n^2 - n)E\left[Y_{i,1}Y_{j,2}\right] = n(n-1)p_i p_j$$

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’

- How does the PMF looks like
  - The outcome of the n-th trial is the 'r-th' success, so in the first (n-1) trial, there has been (r-1) success happens, we can then easily write the following probability

$$P(X = n) = \binom{n-1}{r-1} p^{r-1}(1-p)^{n-r}p$$
$$= \binom{n-1}{r-1} p^{r}(1-p)^{n-r}$$

  - Given this we can write the PMF of negative binomial officially:

$$p_X(x) = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}, x = r, r+1, \dots,$$

  - This is a negative binomial with parameter, (r, p)

Expectation

- As usual, we realize the following:

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

That,

X_1 = number of trials to the first success

X_2 = number of additional trial to get second success

And so on…

What are they: each is just a geometric random variables

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_r] = r * \left(\frac{1}{p}\right) = \frac{r}{p}$$

Can we do the same thing for variance?

Are they independent random variables! Of course

Independence means the variance adds!

Variance of a single geometric random variable: $\frac{1-p}{p^2}$

Variance of negative binomial? $r * \frac{1-p}{p^2}$

Example

- Sharon and Ann play a series of backgammon games until one of them wins 5 games. Suppose that the games are independent and the probability that Sharon wins a game is 0.58

1) Find the probability that the series ends in 7 games

- Let X be the number of games until Sharon wins 5 games
- Let Y be the number of games until Ann wins 5 games

X is a negative binomial rv with parameter (5, 0.58)
Y is a negative binomial rv with parameter (5, 0.42)

The probability that the series end in 7 games?

$$p_X(7) + p_Y(7)$$
$$= \binom{6}{4}(0.58)^5(0.42)^2 + \binom{6}{4}(0.58)^2(0.42)^5 \approx 0.24$$

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, \dots,$$

2) If the series ends in seven games, what is the probability that Sharon wins?

Let A be the event that Sharon wins, and B be the event ends in seven games. Then the desired probability is:

$$p(A|B) = \frac{P(AB)}{P(B)} = \frac{p_X(X=7)}{p_X(X=7) + p_Y(Y=7)} \approx \frac{0.17}{0.24}$$
$$\approx 0.71$$

Hypergeometric Random variable

- Imagine a scenario:

Suppose that, from a box containing *D* defective and *N*-*D* non-defective items, n are drawn at random and 'without replacement'.

Furthermore, suppose that the number of items drawn does not exceed the number of defectives or the number of non-defective item

$$n < \min(D, N - D)$$

Now, let X be the number of defective items drawn

- X can take on values from 0,1,2,…n
- X is called 'hypergeometric random variable'

- PMF?

$$p_X(x) = p_X(X = x) = \begin{cases} \dfrac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \text{if x} \in \{0,1,2,\dots,n\} \\ 0, \text{otherwise} \end{cases}$$

Expectation

How to compute E[X]?

Let A_i be the event that the i-th item drawn is defective. Also, for i=1, 2,…, n let:

$$X_i = \begin{cases} 1, \text{if } A_i \text{ occurs} \\ 0, \text{otherwise} \end{cases}$$

What is X?

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

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

What is $E[X_i]$=?

$$E[X_i] = 1 * P(X_i = 1) + 0 * P(X_i = 0) = P(X_i = 1)$$
$$= P(A_i) = \frac{D}{N}$$

$$E[X] = \frac{nD}{N}$$

At large N, sometimes, people just use Binomial assuming ‘with replacement’ draws

Variance: save for later class (are they independent?

Example

- In 500 independent calculations a scientist has made 25 errors. If a second scientist checks seven of these calculations randomly, what is the probability that he detects two errors? Assume that the scientist will definitely find the error of a false calculations

Solution

Let X be the number of errors found by the second scientist. Then X is hypergeometric with N=500, D=25, and n=7.

And we want to compute?

$$p_X(2) = \frac{\binom{25}{2}\binom{500-25}{7-2}}{\binom{500}{7}} \approx 0.04$$