國立清華大學
NATIONAL TSING HUA UNIVERSITY

# EE 306001
# Probability

Lecture 10: discrete random variable

李祈均

# p7

- Suppose that X and Y are independent, identically distributed (iid), geometric random variables with parameter p, we want to show the following:

- $P(X = i | X + Y = n) = \frac{1}{n-1}, \text{for } i = 1,2, \ldots, n-1$

$$P(X=i|X+Y=n)=\frac{P(\{X=i\}\cap\{X+Y=n)}{P(X+Y=n)}$$

The event $\{X=i\}\cap\{X+Y=n\}$ in the numerator in equivalent to $\{X=i\}\cap\{Y=n-i\}$, taking this in combination with the fact hat X and Y are independent

$$P(X+Y=n)=\sum_{i=1}^{n-1}P(X=i)P(X+Y=n|X=i)=$$

$$\sum_{i=1}^{n-1}P(X=i)P(i+Y=n|X=i)=$$

$$\sum_{i=1}^{n}P(X=i)P(Y=n-i|X=i)$$

$$=\sum_{i=1}^{n-1}P(X=i)P(Y=n-i)$$

**Total probability theorem**

We only get non-zero probability for i=1, …, n-1 since X and Y are both geometric random variables

So now we can write it completely from the previous slides:

$$P(X = i|X + Y = n) = \frac{P(X = i)P(Y = n - i)}{\sum_{i=1}^{n-1} P(X = i)P(Y = n - i)}$$

$$= \frac{(1-p)^{i-1}p(1-p)^{n-i-1}p}{\sum_{i=1}^{n-1}(1-p)^{i-1}p(1-p)^{n-i-1}p} = \frac{(1-p)^n}{\sum_{i=1}^{n-1}(1-p)^n}$$

$$= \frac{(1-p)^n}{(1-p)^n\sum_{i=1}^{n-1} 1} = \frac{1}{n-1}$$

# Summary of conditional PMF

Conditional PMF of X given an event A with P(A) > 0, is defined by,

$$p_{X|A}(x) = P(X = x|A)$$

and satisfy:

$$\sum_x p_{X|A}(x) = 1$$

If $A_1,\ldots,A_n$ are disjoint events that form a partition of the sample space, with $P(A_i) > 0$ for all i, then

$$p_X(x) = \sum_{i=1}^{n} P(A_i) p_{X|A_i}(x)$$

- Furthermore, for any event B, with $P(A_i \cap B) > 0$ for all i, we have:

$$p_{X|B}(x) = \sum_{i=1}^{n} P(A_i|B) p_{X|A_i \cap B}(x)$$

- 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:

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

# Conditional expectation

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

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

For a function g(X), we have:

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

- Conditional expectation of X given a value of y of Y is defined as follow:

$$E[X|Y = y] = \sum\nolimits_x x p_{X|Y}(x|y)$$

- If $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space, with $P(A_i) > 0$ for all i, then:

$$E[X] = \sum_{i=1}^{n} P(A_i)\, E[X|A_i]$$

Furthermore, for any event B with $P(A_i \cap B) > 0$ for all i, we have:

$$E[X|B] = \sum_{i=1}^{n} P(A_i|B)\, E[X|A_i \cap B]$$

We have:

$$E[X] = \sum_{y} p_Y(y)\, E[X|Y = y]$$

Generalize to multiple variables should be straightforward

# Independence of random variables

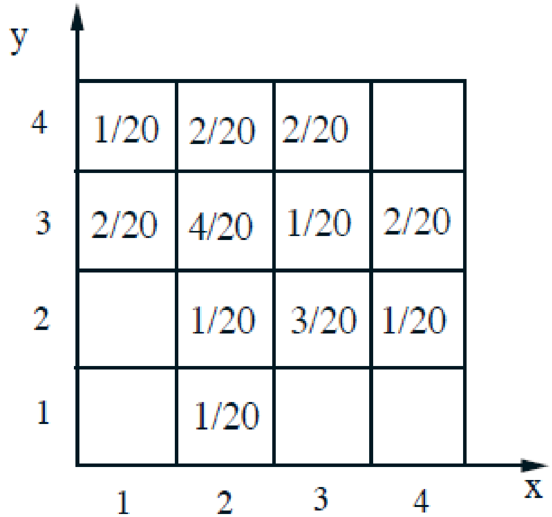
Say a three-variable multiplication rule:

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

Random variables are independent iff:

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

The intuitive content is the same as for events. Random variables are independent if knowing something about the realized values of some of these random variables does not change our beliefs about the likelihood of various values for the remaining random variables




- Are x, y independent of on another?
  - Let's try with simple inspection:
  - Say if Y=1, then X must be 2
  - Has this changed our belief? Of course!

- Are x, y conditionally independent
  - Say conditioning on ( $X \leq 2, Y \geq 3$ )

| 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 | | |

| | | |
|---|---|---|
| 4 | 1/9 | 2/9 |
| 3 | 2/9 | 4/9 |
| Y/X | 1 | 2 |

- Are x, y conditionally independent
  - Say conditioning on ( $X \leq 2, Y \geq 3$ )
  - New universe:
    - Ratio 1, 2, 2, 4
  - So say in this universe:
    - If we know Y=3 or Y=4, does it change our belief about the likely occurrence of X?
    - No! (33.33 vs. 66.66)

| | | |
|---|---|---|
| 4 | 1/9 | 2/9 |
| 3 | 2/9 | 4/9 |
| Y/X | 1 | 2 |

- We can try the following:

$P(X=1, Y=3 | X \leq 2, Y \geq 3) = 2/9$

$P(X=1 \mid X \leq 2, Y \geq 3) = 1/9 + 2/9 = 3/9$

$P(Y=3 \mid X \leq 2, Y \geq 3) = 2/9 + 4/9 = 6/9$

$P(X=1 \mid X \leq 2, Y \geq 3) * P(Y=3 \mid X \leq 2, Y \geq 3) = 3/9 * 6/9 = 2/9$

$P(X=1, Y=3 | X \leq 2, Y \geq 3) = P(X=1 \mid X \leq 2, Y \geq 3) * P(Y=3 \mid X \leq 2, Y \geq 3)$

Yes! Conditionally-independent

# Expectations

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

If X, Y are independent variables then:

$$E[XY] = E[X]E[Y]$$

Proof:

$$E[XY] = \sum_x \sum_y xy p_{X,Y}(x,y) = \sum_x \sum_y xy p_X(x) p_Y(y)$$

$$= \sum_x x p_X(x) \sum_y y p_Y(y) = E[X]E[Y]$$

- If X, Y are independent:

$$E[g(X)h(Y)] = E[g(X)] * E[h(Y)]$$

Proof:

Let U=g(X), V=h(Y)

$$p_{U,V}(u, v) = \sum_{\{(x,y)|g(x)=u,h(y)=v)\}} p_{X,Y}(x, y)$$

$$= \sum_{\{(x,y)|g(x)=u,h(y)=v)\}} p_X(x)p_Y(y)$$

$$= \sum_{\{x|g(x)=u)\}} p_X(x) \sum_{\{y|h(y)=v\}} p_Y(y) \quad = p_U(u)p_V(v)$$

# Variances

- If X, Y are independent,

$$Var(X+Y) = Var(X) + Var(Y)$$

Say we know from variance property:

$$Var(X+a) = Var(X)$$

Proof:

Define: $\tilde{X} = X - E[X], \tilde{Y} = Y - E[Y]$

$$Var(X+Y) = Var\left(\tilde{X}+\tilde{Y}\right) = E\left[\left(\tilde{X}+\tilde{Y}\right)^2\right]$$

$$= E\left[\tilde{X}^2\right] + 2E\left[\tilde{X}\tilde{Y}\right] + E[\tilde{Y}^2]$$

$$E\left[\tilde{X}\tilde{Y}\right] = E\left[\tilde{X}\right]E\left[\tilde{Y}\right] = 0$$

Means are zero

$$Var(X+Y) = Var(\tilde{X}+\tilde{Y}) = E\left[(\tilde{X}+\tilde{Y})^2\right]$$

$$= E[\tilde{X}^2] + 2E[\tilde{X}\tilde{Y}] + E[\tilde{Y}^2] = E[\tilde{X}^2] + E[\tilde{Y}^2]$$

$$= var(\tilde{X}) + var(\tilde{Y}) = var(X) + var(Y)$$

- Examples:

Think about another variance property that we talked about:

$$var(aX) = a^2 var(X)$$

If X = Y, Var(X+Y)=?

$$Var(X+Y) = Var(X+X) = Var(2X) = 4Var(X)$$

If X = -Y, Var(X+Y)=?

$$Var(X+Y) = Var\big(X+(-X)\big) = Var(0) = 0$$

If X, Y are independent, and Z = X – 3Y

$$Var(Z) = Var(X-3Y) = Var\big(X+(-3Y)\big)$$
$$= Var(X) + Var(-3Y) = Var(X) + (-3)^2 Var(Y)$$
$$= Var(X) + 9Var(Y)$$

# 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,

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

- We want to compute E[X] not E[$X_i$]
- Are the $X_i$ independent ?

- Let's think about the case:
  - If we know that out of 10 people, 9 of them got their own hat back, does that tell you something about $10^{th}$ person?
  - Of course!
  - That means Xi are not really independent of one another!

- But does it matter in computing E[X]?

  – Linearity of expectation does not require independence!

$$E[X] = \sum_{i=1}^{n} E[X_i] = n\left(\frac{1}{n}\right) = 1$$

  – Out of n people in this hat problem: on average 1 person will get it back!

# Variance in the hat problem

- Since these are not independent, summation does not apply here:

$$Var(X) = E[X^2] - (E[X])^2 = E[X^2] - 1$$

What is $X^2$:

$$X^2 = (\sum_{i=1}^{n} X_i)^2 = \sum_{i} {X_i}^2 + \sum_{i,j:i\neq j} X_i X_j$$

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

$$\sum_{i=1}^{n} \frac{1}{n} = 1$$

Now look at the cross term, let's look at X1, X2 for now:

To compute $E[X_1 X_2]$, only thing that matters is when $X_1 = 1$, $X_2 = 1$

$$P(X_1X_2 = 1) = P(X_1 = 1)P(X_2 = 1|X_1 = 1)$$
$$= \frac{1}{\mathrm{n}} * \left(\frac{1}{\mathrm{n} - 1}\right)$$

$$E[X_1X_2] = \frac{1}{n} * \left(\frac{1}{n - 1}\right)$$

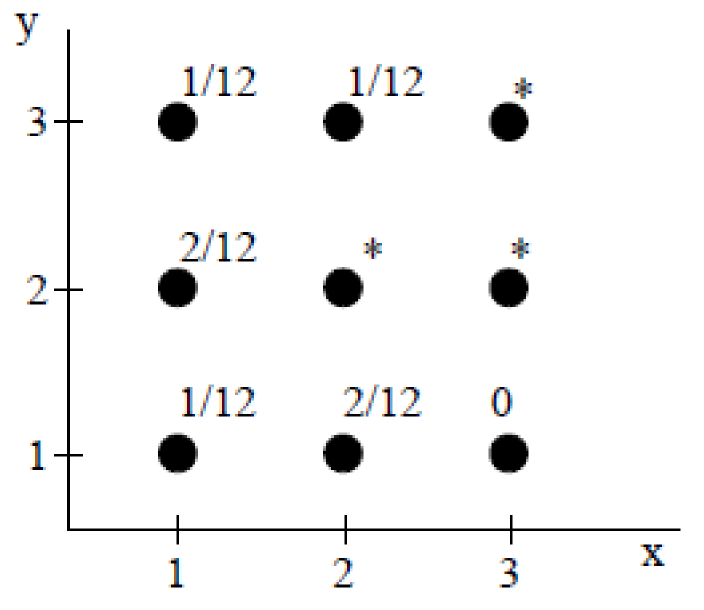
$$\sum_{i,j:i \neq j} X_iX_j = (n^2 - n)\left(\frac{1}{\mathrm{n}} * \left(\frac{1}{\mathrm{n} - 1}\right)\right)$$

$$E[X^2] = 1 + \frac{n^2 - n}{n^2 - n} = 2$$

$$Var(X) = 2 - 1 = 1$$

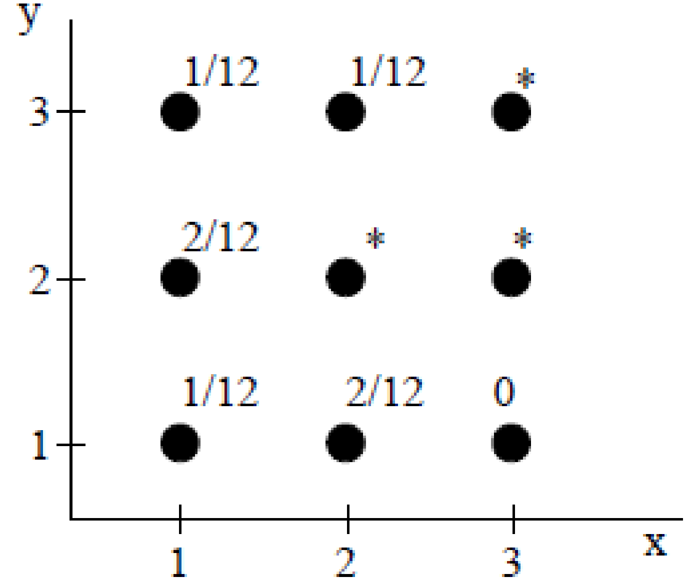
# p1

Random variable X, Y can take value in the set {1,2,3}. We are given the joint PMF on the right, the entries indicated by * are left unspecified



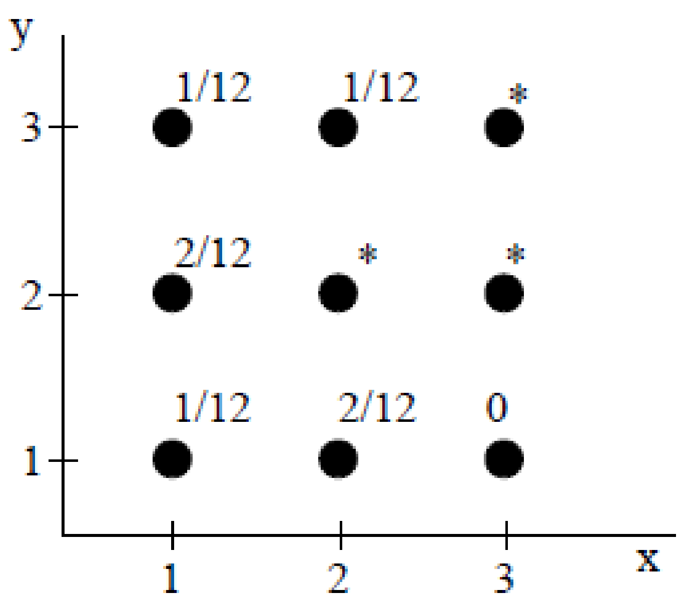
# a) What is $p_X(1)$

$$p_X(1)$$
$$= P(X=1, Y=1) + P(X=1, Y=2) + P(X=1, Y=3)$$
$$= \frac{1}{12} + \frac{2}{12} + \frac{1}{12} = 1/3$$



# b) Provide a clearly labeled sketch of the conditional PMF of Y given that X=1

$$p_{Y|X}(y|1) = \frac{p_{Y,X}(y,1)}{p_X(1)} = \begin{cases} \frac{1}{4}, y = 1 \\ \frac{1}{2}, y = 2 \\ \frac{1}{4}, y = 3 \\ 0, \text{otherwise} \end{cases}$$



# c) What is E[Y|X=1]?

$$E[Y|X = 1] = \sum_{y=1}^{3} y p_{Y|X}(y|1) = 1 * \frac{1}{4} + 2 * \frac{1}{2} + 3 * \frac{1}{4}$$
$$= 2$$

Let B be the event that $X \leq 2$ and $Y \leq 2$. We are told that conditioned on B, the random variables X and Y are independent

e) What is $P_{X,Y}(2,2)$? Or do we have enough information?

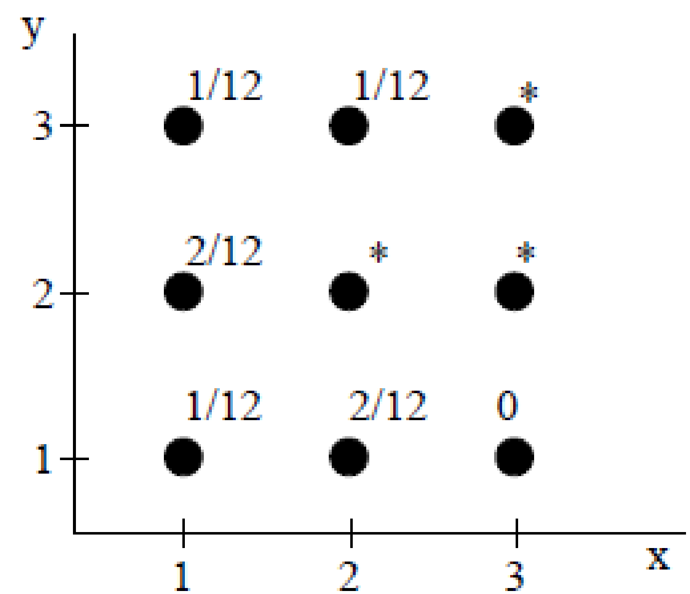
Knowing that X and Y are conditionally independent given B,

We first concentrate on 2x2 grid points

P(XY|B)=P(X|B)P(Y|B)
We can expand this or just realize:

$$\frac{p_{X,Y}(1,1)}{p_{X,Y}(1,2)} = \frac{p_{X,Y}(2,1)}{p_{X,Y}(2,2)} = \frac{1}{2}$$

$$p_{X,Y}(2,2) = \frac{4}{12}$$



# e) What is $P_{X,Y|B}(2,2|B)$

$$P(B) = \frac{9}{12} = \frac{3}{4}$$

Now just need to normalize the $P_{X,Y}(2,2)$ in the universe of B

$$p_{X,Y|B}(2,2) = \frac{p_{X,Y}(2,2)}{P(B)} = \frac{4}{9}$$

# 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:

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

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

$n \to \infty$ and $p \to 0$,

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

Set $\lambda$ = np

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

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

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

Fix k, let n->$\infty$

$\frac{n(n-1)\dots(n-k+1)}{n^k}$ 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 -> ∞

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