Lecture 7

§1 二维变量

1. General case

1° X can take values in a set

Y can take values in a set

2° X 与 Y 间可能存在某些关系

Examples

- X: the temperature outside tomorrow
- Y: the number of households using air conditioners

- X: the age of a random selected person
- Y: the height of this person

- X: the attendance rate of a random selected student
- Y: the final score of this student

2. joint distribution (联合分布) (离散型)

1° 用于表示 the outcome of the pair (X,Y) 的概率

2° joint pmf (联合概率质量方程):

$f(x,y)$ 表示 $X=x, Y=y$ 的概率

3° 性质

$\sum_{x,y} f(x,y) = 1$, $f(x,y) \in [0,1]$. (对于离散情况)

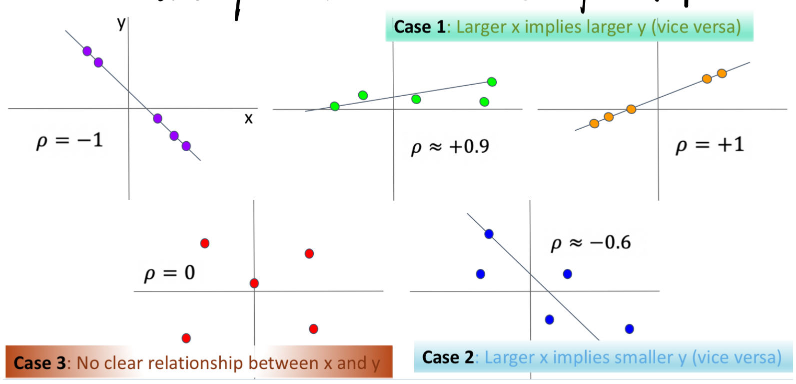
$\sum_y f(x,y)$ 表示 $X=x$ 的概率

$\sum_x f(x,y)$ 表示 $Y=y$ 的概率

§2 Correlation (相关性)

1. Graphical illustration

draw all the possible outcomes of (X,Y) in a figure



2. Mathematical definition (相关系数)

相关系数 $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$

where $Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$

分母用来 ensure $\rho \in [-1,1]$, $|\rho|$ 越大，线性相关越强

3. Cov: covariance (协方差)

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$
$$= E[XY - X\cdot E[Y] - Y\cdot E[X] + E[X]\cdot E[Y]]$$
$$= E[XY] - 2E[X]\cdot E[Y] + E[X]\cdot E[Y]$$
$$= E[XY] - E[X]\cdot E[Y]$$

Covariance (Intuition)

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

$X > E[X], Y > E[Y] \rightarrow +$ $X < E[X], Y > E[Y] \rightarrow -$

$X < E[X], Y < E[Y] \rightarrow +$ $X > E[X], Y < E[Y] \rightarrow -$

Covariance (Intuition)

- If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values (that is, the variables pair (X,Y) tend to show similar behavior), the covariance is positive.

- In the opposite case, when the greater values of one variable mainly correspond to the lesser values of the other, (that is, the variables pair (X,Y) tend to show opposite behavior), the covariance is negative.

注: 1° Correlation 不会揭示 causality (因果关系)

2° 若X与Y独立，则 $E[XY] = E[X]\cdot E[Y]$

3° 方差可理解为两变量相同的协方差

4. 多个随机变量 $(X_1, X_2, \cdots, X_n)$ 的方差计算

$$Var\left(\sum_{i=1}^{n} X_i\right) = cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} cov(X_i, X_j)$$
$$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

	X_1	X_2	X_3	X_4	X_5
X_1					
X_2					
X_3					
X_4					
X_5					

例: **Hat Check Again**

n people go to a party and leave their hat with a hat-check person. At the end of the party, she returns hats randomly since she doesn't care about her job. Let X be the number of people who get their original hat back. What is $E[X]$?

For $i = 1, \ldots, n$, let $X_i = \begin{cases} 1, & \text{if } i^{th} \text{ person got hat back} \\ 0, & \text{otherwise} \end{cases}$. Then $X = \sum_{i=1}^{n} X_i$.

⇩

NOT "INDEPENDENT" RVS

We will use linearity of expectation.

$$E[X_i] = 1\cdot P(X_i = 1) + 0\cdot P(X_i = 0) = P(X_i = 1) = P(i^{th} \text{ person got hat back}) = \frac{1}{n}$$

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

Now let's compute $Var(X)$. Recall each $X_i \sim Ber\left(\frac{1}{n}\right)$. By previous proof,

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

$$X_iX_j \in \{0,1\}$$

$$E[X_iX_j] = P(X_iX_j = 1) = P(X_i = 1, X_j = 1) = P(X_i = 1)P(X_j = 1 | X_i = 1) = \frac{1}{n}\left(\frac{1}{n-1}\right)$$

$$Cov(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j] = \frac{1}{n}\left(\frac{1}{n-1}\right) - \frac{1}{n^2} = \frac{n}{n^2(n-1)} - \frac{n-1}{n^2(n-1)} = \frac{1}{n^2(n-1)}$$

$$Var(X_i) = \left(\frac{1}{n}\right)\left(1 - \frac{1}{n}\right)$$

$$Var(X) = n\left(\frac{1}{n}\right)\left(1 - \frac{1}{n}\right) + (n^2 - n)\left(\frac{1}{n^2(n-1)}\right) = 1 - \frac{1}{n} + \frac{1}{n} = 1$$

§2 Conditional probability (条件概率)

1. 定义

1° 给出X的情况，Y的分布会改变

2° 形式：$f(Y=y|X=x)$ (在 $X=x$ 情况下 $Y=y$ 的概率)

3° 性质：$f(x,y) = f(x) \cdot f(Y=y|X=x)$

$f(Y=y|X=x) = f(x,y)/f(x)$

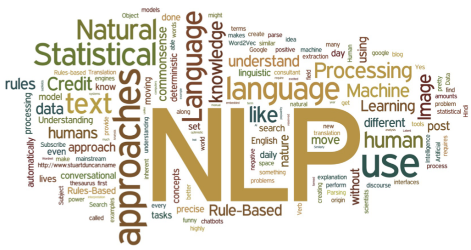
例: **Conditional Probability in NLP**

P(B|A) = P(A and B) / P(A)

- H: mention “happy” in message
- D: mention “data science course” in message

- P(H) = .1
- P(D) =.1
- P(H, D) = .08

- P(H|D) = 0.08/0.1
  = 0.8



例: **Monty Hall problem**

Suppose you're on a game show, and you're given the choice of three doors:

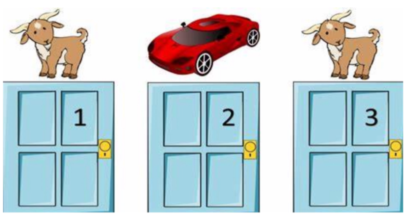
- Behind one door is a car;
- Behind the others, goats.

The car is randomly placed behind one door.

Which door you want to open?

Any will be OK

- After you choose No. 1, I randomly open a door with a goat behind, do you want to change your door?
- Now you have additional information\event!
- Given the information, you may make a better decision!



Behind door 1	Behind door 2	Behind door 3	Result if staying at door #1	Result if switching to the door offered
Goat	Goat	**Car**	Wins goat	**Wins car**
Goat	**Car**	Goat	Wins goat	**Wins car**
Car	Goat	Goat	**Wins car**	Wins goat

P(not switch)=1/3

P(Switch) = 2/3

If a car is worth 1 dollar and a goat 0 dollar, how much money will you at most pay to switch? 1/3

§3 Why Probability

A formality to make sense of the world.

- To quantify uncertainty
  - Should we believe something or not? Is it a meaningful difference?

- To be able to generalize from one situation or point in time to another.
  - Can we rely on some information? What is the chance Y happens?

- To organize data into meaningful groups or dimensions
  - Where does X belong? What words are similar to X?