Lecture 02

- Equally Likely Outcomes
- Remember how to count?
- Conditional probability
- Law of total probability
- Bayes rule

Equally Likely Outcomes (等可能事件)

Some sample spaces have equally likely <u>outcomes</u> (elementary events, 基本事件).

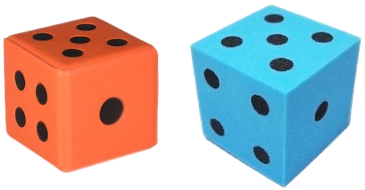
- Coin flip: $S = \{\text{Head, Tails}\}$
- Flipping two coins: $S = \{(H,H),(H,T),(T,H),(T,T)\}$
- Roll of a 6-sided die: $S = \{1,2,3,4,5,6\}$

If we have equally likely outcomes , then

$$P(\text{Each outcome}) = \frac{1}{|S|}$$

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S}$$

Roll two dice



Ex. 5a (P34).
Roll two 6-sided fair dice. What is P(sum = 7)?

$$
\begin{aligned}
S = \{ & (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\
& (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\
& (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\
& (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\
& (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\
& (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}
\end{aligned}
$$

$$E = \{(6,1), (5,2), (4,3), (3,4), (2,5), (1,6)\}$$

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S} = \frac{6}{36} = \frac{1}{6}$$

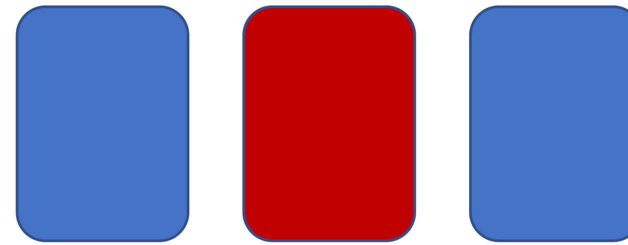
The three-card puzzle

Suppose we have three cards in a hat:

- RB - One card is painted blue on one side and red on the other.
- BB - One card is painted blue on both sides.
- RR - One card is painted red on both sides.

The setup

I pick one of the three cards at random, flip it to a random side, and place it on the table.
The color of the side of the card facing up. (B or R)



Do you want to bet?

If the other side of the card has a different color, I pay you $1;
If the other side has the same color, you pay me $1.

Why is this a fair bet?

Suppose the side facing up is R.
→ the card is either RR or RB.
→ the other side can be either R or B.
→ in this case the odds are equal.
A similar argument holds for the case where is B.

Let’s use a monte-carlo simulation (with Python!)

```
red_bck="\x1b[41m%s\x1b[0m"
blue_bck="\x1b[44m%s\x1b[0m"
red=red_bck%'R'
black=blue_bck%'B'
Cards=[(red,black),(red,red),(black,black)]
counts={'same':0,'different':0}
from random import random
for j in range(50):
    i=int(random()*3.)  # Select a random card
    side=int(random()*2.)
    C=Cards[i]
    if(side==1):        # select which side to be "up"
        C=(C[1],C[0])
    same= 'same' if C[0]==C[1] else 'different'  # counting
    counts[same]+=1
    print(''.join(C)+' %-9s'%same, end='')
    if (j+1)%5==0:
        print()
print()
print(counts)
```

The three-card puzzle

```
RR same     BB same     BR differentBB same     RR same
RR same     RB differentRB differentRB differentRR same
RB differentBR differentBB same     BB same     RB different
BB same     BR differentBB same     BB same     BR different
RB differentBB same     RB differentBB same     BB same
RR same     RR same     BR differentRR same     BB same

{'same': 36, 'different': 14}

Process finished with exit code 0
```

Lecture 02

- Equally Likely Outcomes
- Remember how to count?
- Conditional probability
- Law of total probability
- Bayes rule

Permutations (排列)

- A permutation is an ordered arrangement of sample.
- Ex. How many queues can be formed by 5 students?
- # arrangements of k elements taken from a given set of size n,

$$P(n,k) = P_k^n = A_n^k = \frac{n!}{(n-k)!}$$

where $n! = 1 \cdot 2 \cdot \ldots \cdot n$, $0! = 1! = 1$.

Details are available in P3 of textbook.

Combinations (组合)

- A combination is an unordered group of samples.

- Ex. How many different groups can be selected from 5 students?

- # combinations of k elements taken from a given set of size n,

$$\binom{n}{k} = C_n^k = \frac{n!}{(n-k)!\,k!}$$

Details are available in P5 of textbook.

Some questions from textbook

Ex. 5b (P34)

If 3 balls are randomly drawn from a bowl containing 6 white and 5 black balls, what is the probability that one of the balls is white and the other two are black?

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S} = \frac{C_6^1 C_5^2}{C_{11}^3}$$

More about counting

$$P(E) = \frac{|E|}{|S|}$$

Counting tasks on n objects

- Ordered: **Sort objects (permutations)**
  - Distinct (distinguishable): $n!$
  - Some distinct: $\frac{n!}{n_1!\,n_2!\dots n_r!}$, $n_1 + n_2 + \cdots + n_r = n$
    - For each group of indistinct objects, divide by the overcounted permutations.
- Unordered: **Choose k objects (combinations)**
  - Distinct
    - 1 group: C_n^k
    - r groups: $C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n-n_1-\cdots-n_{r-1}}^{n_r} = \frac{n!}{n_1!\,n_2!\dots n_r!}$
      - Consider r group independently.
- **Put objects in r buckets**
  - Distinct: r^n
  - Indistinct: $\frac{(n+r-1)!}{n!\,(r-1)!}$
    - Separate n objects in r group by $r-1$ bars.

Unique 6-digit passcodes (Permutations)

How many unique 6-digit passcodes are possible if a phone password uses each of six distinct numbers?

Total = 6! = 720 passcodes

How many unique 6-digit passcodes are possible if a phone password uses each of five distinct numbers?

Total = $5 \times \frac{6!}{2!} = 1800$ passcodes

A trick question on combinations

How many ways are there to group 6 indistinct (indistinguishable) objects into 3 groups, where groups A, B, and C have sizes 1, 2, and 3, respectively?

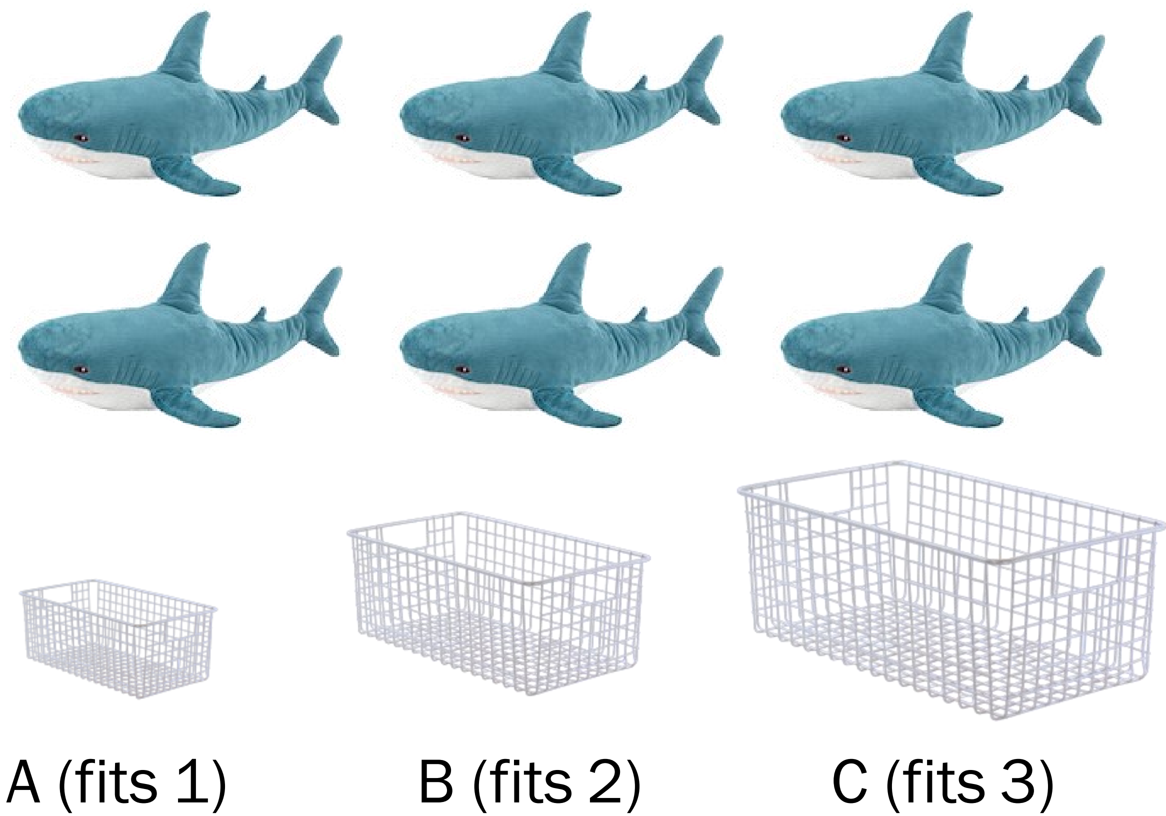
A. $\frac{6!}{3!2!1!}$

B. $C_6^3 C_3^2 C_1^1$

C. 0

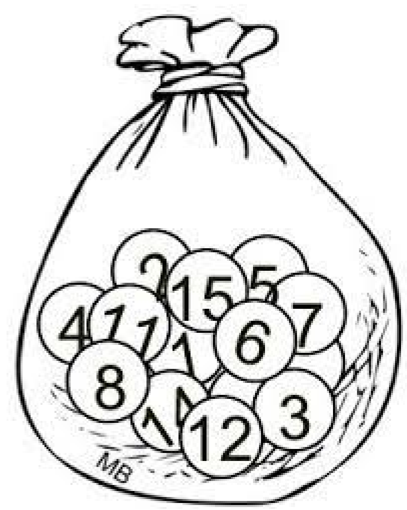
D. 1

E. Something else



Lecture 02

- Equally Likely Outcomes
- Remember how to count?
- Conditional probability
- Law of total probability
- Bayes rule



Conditional Probability (条件概率)

The conditional probability of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F.

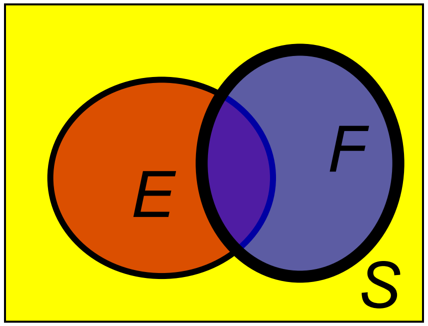
Written as: $P(E|F)$

Meaning: $P(E,\ \text{given } F \text{ already observed})$

Sample space: all outcomes in F

Event: all outcomes in $E \cap F$

Calculation: $P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{|EF|}{|F|}$



Spam email problem

Ex. 24 emails are sent, 6 each to 4 users.

- 10 of the 24 emails are spam.
- All possible outcomes are equally likely.

Let E = user1 receives 3 spam emails.
What is $P(E)$?

Let F = user2 receives 6 spam emails.
What is $P(E|F)$?

Let G = user3 receives 5 spam emails.
What is $P(G|F)$?

Try by yourselves.

Spam email problem

Ex. 24 emails are sent, 6 each to 4 users.

- 10 of the 24 emails are spam.
- All possible outcomes are equally likely.

Let E = user1 receives 3 spam emails.
What is $P(E)$?

$$P(E) = \frac{C_{10}^{3} C_{14}^{3}}{C_{24}^{6}}$$

Let F = user2 receives 6 spam emails.
What is $P(E|F)$?

$$P(E|F) = \frac{P(EF)}{P(F)}$$

$$= \frac{C_{10}^{6} C_{4}^{3} C_{14}^{3} / C_{24}^{6} C_{18}^{6}}{C_{10}^{6} / C_{24}^{6}}$$

Let G = user3 receives 5 spam emails.
What is $P(G|F)$?

0

No way to choose 5 spam from 4 remaining spam emails!

Some questions from textbook

Ex. 2b (P59, textbook)

A coin is flipped twice. Assuming that all four points in the sample space S={(h,h), (h,t), (t,h),(t,t)} are equally likely, what is the conditional probability that both flips land on heads, given that

(a) The first flip lands on heads?

(b) At least one flip lands on heads?

Note: the change of sample space.

(a) Let $B = \{(h,h)\}$ be the event that both flips land on heads;

let $F = \{(h,h),(h,t)\}$ be the event that the first flip lands on heads;

let $A = \{(h,h),(h,t),(t,h)\}$ be the event that at least one flip lands on heads.

$$P(B|F) = \frac{P(BF)}{P(F)}$$

$$= \frac{P(\{h,h\})}{P(\{(h,h),(h,t)\})} = \frac{1/4}{1/2} = 1/2$$

(b)

$$P(B|A) = \frac{P(BA)}{P(A)}$$

$$= \frac{P(\{h,h\})}{P(\{(h,h),(h,t),(t,h)\})} = \frac{1/4}{3/4} = 1/3$$

A tricky ball drawing exercise

Ex. There are 10 balls in a bag, including 3 black balls and 7 white balls. In each of two trails, we draw one ball from the bag without replacement.

1) Given that a black ball is drawn in the first time, find the probability that a black ball is drawn in the second time;
2) Given that a black ball is drawn in the second time, find the probability that the black ball is drawn in the first time.

Try by yourselves.

Ex. There are 10 balls in a bag, including 3 black balls and 7 white balls. In each of two trails, we draw one ball from the bag without replacement.

1) Given that a black ball is drawn in the first time, find the probability that a black ball is drawn in the second time;

Sol. Denote the event "a black ball is drawn in the first time" as A_1, and "a black ball is drawn in the second time" as A_2.

$$P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)} = \frac{\frac{3}{10}\times\frac{2}{9}}{\frac{3}{10}} = \frac{2}{9}$$

Ex. There are 10 balls in a bag, including 3 black balls and 7 white balls. In each of the two trails, we draw one ball from the bag without replacement.

2) Given that a black ball is drawn in the second trial, find the probability that the black ball is drawn in the first trial.

$$P(A_2) = P(A_1A_2 \cup \overline{A_1}A_2) = P(A_1A_2) + P(\overline{A_1}A_2)$$

$$= \frac{3\cdot 2}{10\cdot 9} + \frac{7\cdot 3}{10\cdot 9} = \frac{3}{10}$$

$$P(A_1|A_2) = \frac{P(A_1A_2)}{P(A_2)} = \frac{\frac{3}{10}\times\frac{2}{9}}{\frac{3}{10}} = \frac{2}{9}$$

The probabilities of (a) and (b) are equal!

Conditional probability in general

A general definition of conditional probability:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

The multiplication rule, (a.k.a. Chain rule, Product rule (概率的乘法定理)):

$$P(EF) = P(E|F)P(F)$$

In general,

$$P(E_1 E_2 E_3 \ldots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \ldots$$
$$P(E_n|E_1 \ldots E_{n-1})$$

Monty Hall Problem
a.k.a. Let's Make a Deal

https://en.wikipedia.org/wiki/Monty_Hall_problem

Behind one door is a prize (equally likely to be any door).
Behind the other two doors are two goats.

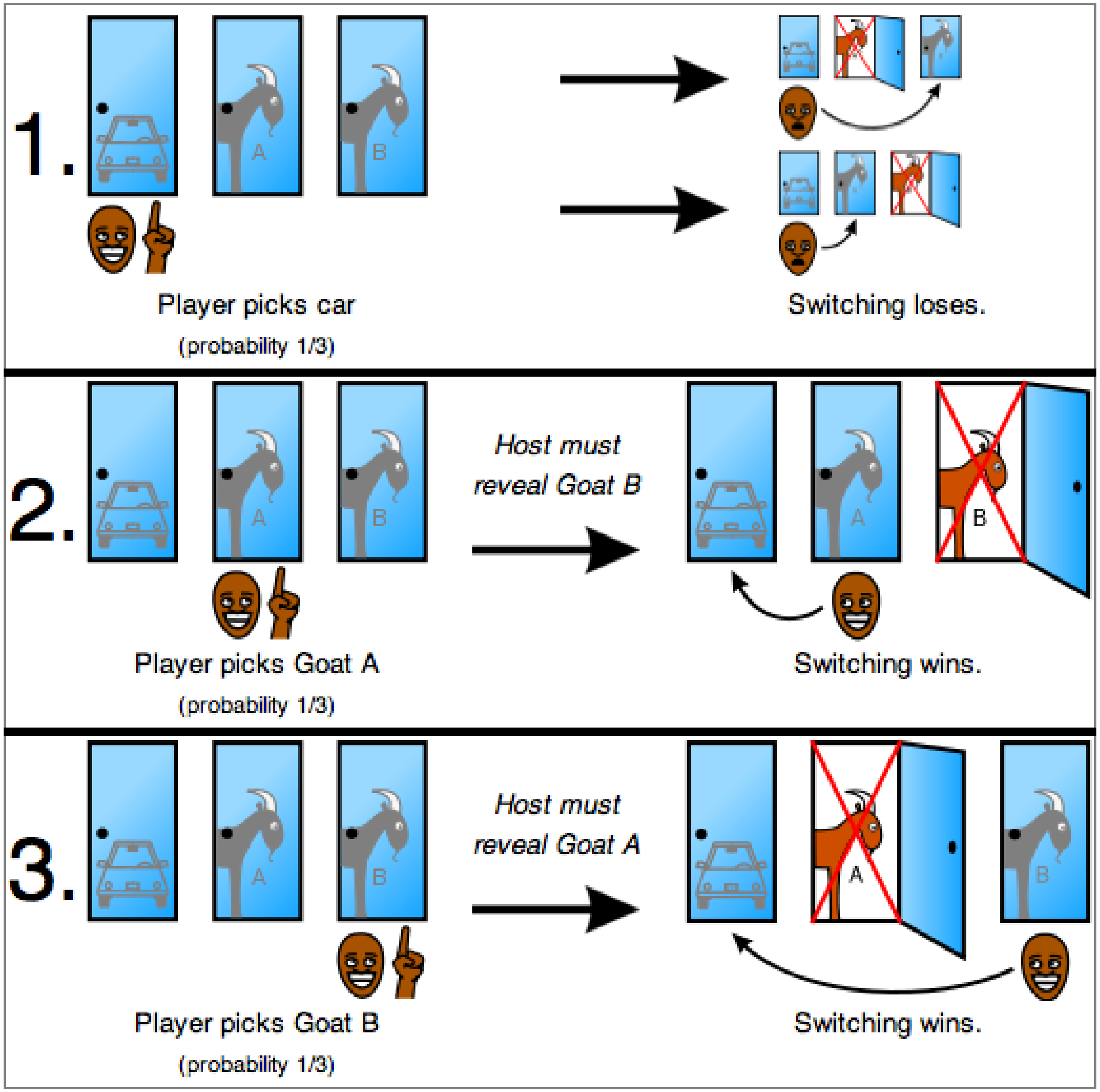
1. We choose a door.
2. **Host** opens 1 of the other 2 doors, revealing a goat.
3. We are given an option to change to the other door.

Doors A,B,C

Should we switch?

Note: If we don't switch, P(win) = 1/3 (random)

We are comparing P(win) and P(win|switch).



The player has an equal chance of initially selecting the car, Goat A, or Goat B. Switching results in a win 2/3 of the time.

Image from Wikipedia.

If we switch

Without loss of generality, say we pick A (out of Doors A,B,C).

1/3 | 1/3 | 1/3

A = prize
- Host opens B or C
- We switch
- We always lose

P(win | A prize, picked A, switched) = 0

B = prize
- Host must open C
- We switch to B
- We always win

P(win | B prize, picked A, switched) = 1

C = prize
- Host must open B
- We switch to C
- We always win

P(win | C prize, picked A, switched) = 1

P(win | picked A, switched) = 1/3×0 + 1/3×1 + 1/3×1 = 2/3

You should switch!

The Monty Hall problem in Python

```python
import random

def run_trial(switch_doors, ndoors=3):
    """
    Run a single trial of the Monty Hall problem, with or without switching
    after the gameshow host reveals a goat behind one of the unchosen doors.
    (switch_doors is True or False). The car is behind door number 1 and the
    gameshow host knows that.

    """


    # Pick a random door out of the ndoors available
    chosen_door = random.randint(1, ndoors)
    if switch_doors:
        # Reveal a goat
        revealed_door = 3 if chosen_door==2 else 2
        # Make the switch by choosing any other door than the initially-
        # selected one and the one just opened to reveal a goat.
        available_doors = [dnum for dnum in range(1,ndoors+1)
                                if dnum not in (chosen_door, revealed_door)]
        chosen_door = random.choice(available_doors)

    # You win if you picked door number 1
    return chosen_door == 1
```


https://scipython.com/book/chapter-4-the-core-python-language-ii/examples/the-monty-hall-problem/

The Monty Hall problem in Python

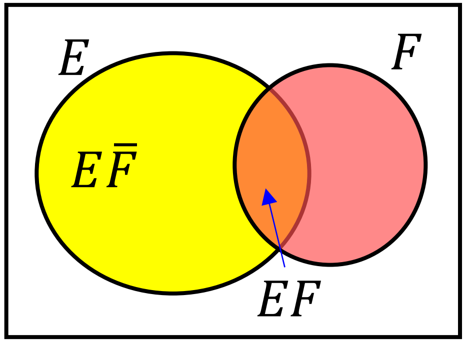
```
def run_trials(ntrials, switch_doors, ndoors=3):
    """
    Run ntrials iterations of the Monty Hall problem with ndoors doors, with
    and without switching (switch_doors = True or False). Returns the number
    of trials which resulted in winning the car by picking door number 1.

    """

    nwins = 0
    for i in range(ntrials):
        if run_trial(switch_doors, ndoors):
            nwins += 1
    return nwins

ndoors, ntrials = 3, 10000
nwins_without_switch = run_trials(ntrials, False, ndoors)
nwins_with_switch = run_trials(ntrials, True, ndoors)

print('Monty Hall Problem with {} doors'.format(ndoors))
print('Proportion of wins without switching: {:.4f}'
            .format(nwins_without_switch/ntrials))
print('Proportion of wins with switching: {:.4f}'
            .format(nwins_with_switch/ntrials))
```

Try by yourselves.

Lecture 02

- Equally Likely Outcomes
- Remember how to count?
- Conditional probability
- Law of total probability
- Bayes rule

Law of Total Probability (全概率公式)



Let $E = EF \cup E\bar{F}$, $E = (EF) \cup (E\bar{F})$

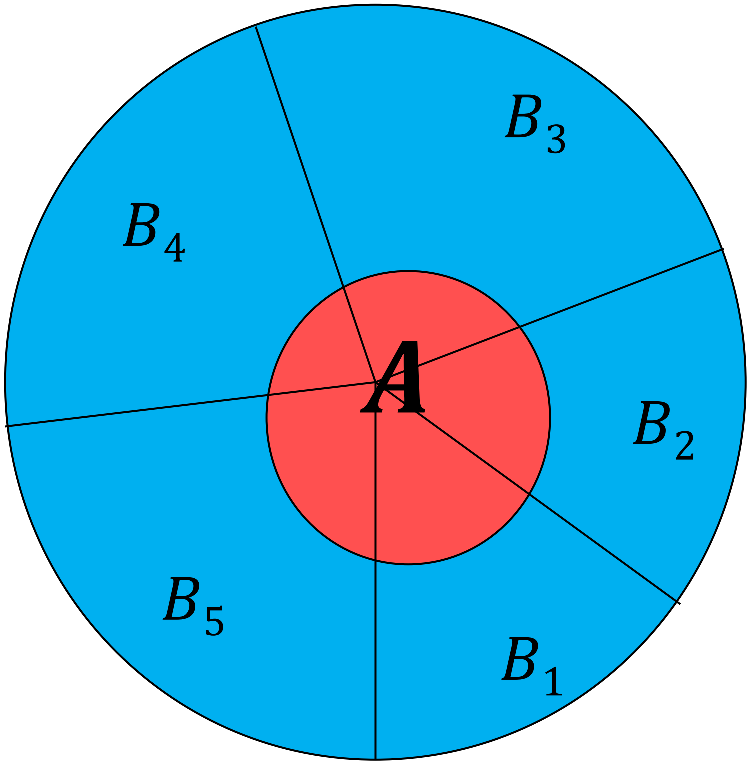
$$P(E) = P(EF) + P(E\bar{F})$$
$$= P(E|F)P(F) + P(E|\bar{F})P(\bar{F})$$

Ex. Predict the probability of construction job completion under different weathers.

P(done) = P(done | rain) P(rain) + P(done | ¬rain) P(¬rain)

Compute the total probability of E from sum of different conditional probabilities.

Law of Total Probability (全概率公式)



More generally, given

- $B_1 \cup B_2 \cup \cdots \cup B_n = S$ (完备划分)
- $B_i B_j = \phi, i \neq j,\ j = 1,2,\cdots,n$

we have

$$P(A) = \sum_{j=1}^{n} P(A|B_j) P(B_j)$$

The king and a smart prisoner

- A prisoner was sentenced to death in ancient times. Before execution, the king gave him a final chance to avoid death.
- The king ordered the prisoner to randomly put 50 white balls and 50 black balls into two jars with the same appearance, and then asked the guards to change the two jars at will, until the prisoner could not distinguish the jars. The prisoner drew a ball from one of the jars.
- If the white ball is picked, he will be released immediately; otherwise, he will be executed immediately. As a result, the smart prisoner quickly put 100 balls into the jars and maximized his chances of escape.

The king and a smart prisoner

- After a careful thinking, the prisoner put only 1 white ball in the first jar, and then put the remaining 49 white balls and 50 black balls in the second jar.

- The probability of getting one white ball is

$$P(\text{escape}) = P(\text{jar1})\, P(\text{escape} \mid \text{jar1}) + P(\text{jar2})\, P(\text{escape} \mid \text{jar2})$$

$$= 1/2 \times 1 + 1/2 \times 49/99 = 74/99 \approx 0.747$$

Bayes' Formula

$P(E|F)$ → $P(F|E)$

For any events E and F where $P(E) > 0$ and $P(F) > 0$,

$$P(F|E) = \frac{P(EF)}{P(E)} = \frac{P(E|F)P(F)}{P(E)}$$

Or, by the total probability law

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

Bayes' Formula

$P(E|F)$ ⇨ $P(F|E)$

E: having diabetes, F: causes of diabetes
e.g., F_1: diet, F_2: genetic, F_3: others.

$$P(F_1|E) = \frac{P(EF_1)}{P(E)} = \frac{P(E|F_1)P(F_1)}{P(E)}$$

$P(E) =$ P(diabetes)
$P(F_1) =$ P(bad diet) $\Longleftarrow$ Prior (先验), general knowledge
$P(E|F_1) =$ P(diabetes, given bad diet)
$P(F_1|E) =$ P(bad diet, given diabetes)
$\Uparrow$ Posterior (后验)

Bayes' Formula: compute $P(F_1|E)$ from $P(E|F_1), P(E), P(F_1)$.

Bayes' Formula

posterior

prior

$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

total probability

Mathematical view: $P(E|F) \Rightarrow P(F|E)$

P(diabetes, given bad diet) ⇒ P(bad diet, given diabetes)

Practical view: Given new evidence E, update belief of F

Prior belief → Posterior belief

$P(F) \rightarrow P(F|E)$, find the **cause** of an event!

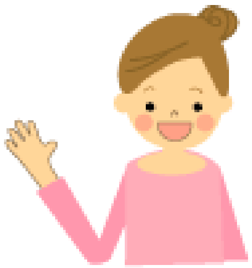
Zika, an autoimmune disease

A disease spread through mosquito bites.
Usually no symptoms; worst case paralysis.
During pregnancy: may cause birth defects.

Ziika Forest, Uganda

If a test returns positive, what is the likelihood that you have the disease?

Confusion matrix



Take test

Fact, F Has disease
or $\bar{F}$ No disease

Evidence, E Test positive
or $\bar{E}$ Test negative

		Fact	
		F, disease +	$\bar{F}$, disease -
Evidence	E, test +	True Positive $P(E\|F)$	False Positive $P(E\|\bar{F})$
	$\bar{E}$, test -	False Negative $P(\bar{E}\|F)$	True Negative $P(\bar{E}\|\bar{F})$

Confusion matrix

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Fact</td></tr>
<tr><td>F, disease +</td><td>$\bar{F}$, disease -</td></tr>
<tr><td rowspan="2">evidence</td><td>E, test +</td><td>True Positive
$P(E|F)$</td><td>False Positive
$P(E|\bar{F})$</td></tr>
<tr><td>$\bar{E}$, test -</td><td>False Negative
$P(\bar{E}|F)$</td><td>True Negative
$P(\bar{E}|\bar{F})$</td></tr>
</table>

If a test returns positive, what is the likelihood you have the disease?

Find $P(F|E)$!

Zika Testing

- A test is 98% effective at detecting Zika (“true positive”).
- However, the test has a “false positive” rate of 1%.
- 0.5% of the US population has Zika.

Say we have 1000 people

- 5 have Zika and test positive
- 10 do not have Zika and test positive
- 985 do not have Zika and test negative.

Zika Testing

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

What is the likelihood you have Zika if you test positive?

Why would you expect this number?

Let $E =$ test positive, $F =$ have the disease,

We want: P(have disease | test+) $= P(F|E)$.

Zika Testing

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

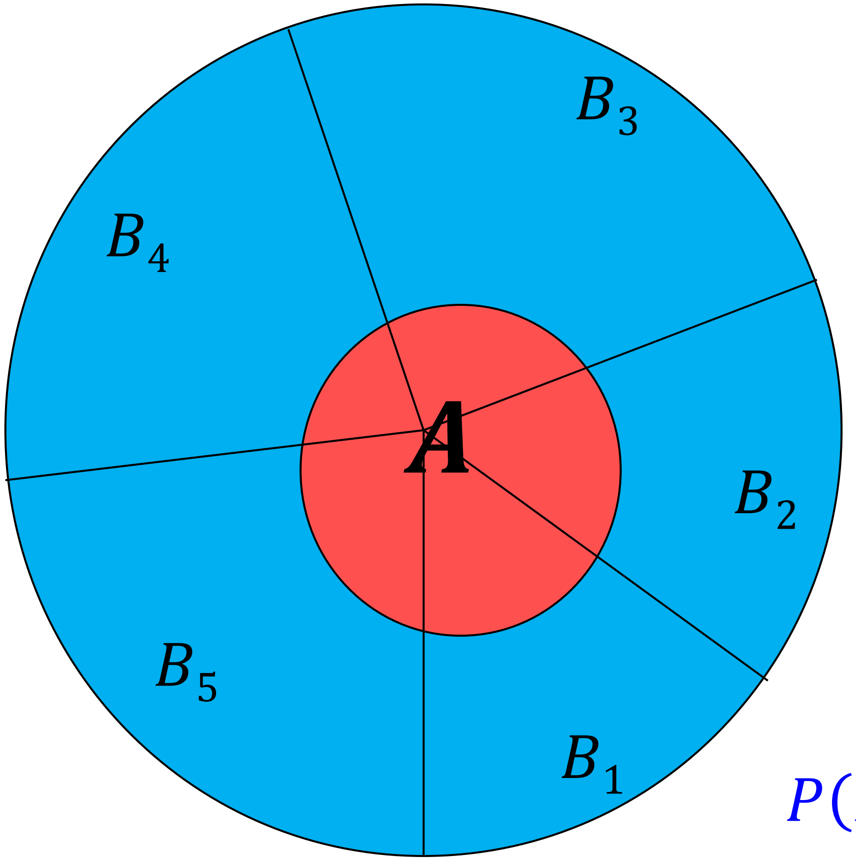
- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

Let $E =$ test positive, $F =$ have the disease,
We want: P(have disease | test+) $= P(F|E)$.

$P(E|F) = 98\%$, $P(E|\bar{F}) = 1\%$, $P(F) = 0.5\%$,

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$
$$= \frac{0.98 \cdot 0.005}{0.98 \cdot 0.005 + 0.01 \cdot 0.995} = 33.0\%$$

Bayes Rule



More generally, given

- $B_1 \cup B_2 \cup \cdots \cup B_n = S$
- $B_i B_j = \Phi, i \neq j, j = 1,2,\cdots,n;$

we have

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P\left(A\middle|B_j\right)P(B_j)} = \frac{P(AB_i)}{P(A)}$$

Quiz of Chapter 1

Date: 14-Sep-2024

Scope: Chapter 1 basics of probability theory

Open-book exam, mainly with fill-in-the-blank and multiple-choice questions.

Time: 40-50 mins