Lecture 2: Conditional Probability

Ziyu Shao

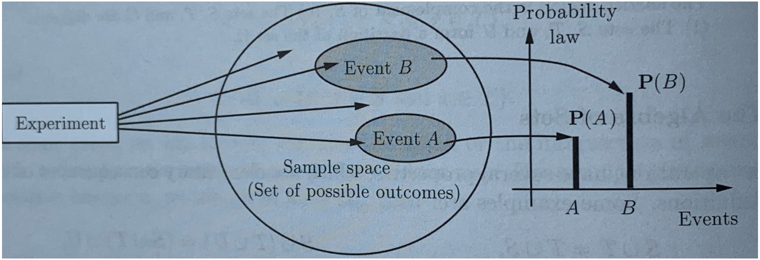
School of Information Science and Technology
ShanghaiTech University

September 29, 2024

Previous Lecture

Ω $\mathcal{F}$ $(\Omega, \mathcal{F}, P)$

S ; σ-field P



Outline

Outline

Thinking Conditionally

- New data & information may affect our belief on uncertainties
- Conditional probability: how to update our belief?
- All probabilities are conditional! (explicit/implicit background info or assumption)

Definition of Conditional Probability

Definition

If A and B are events with $P(B) > 0$, then the *conditional probability* of A given B, denoted by $P(A|B)$, is defined as

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

- $P(A)$: *prior probability* of A.
- $P(A|B)$: *posterior probability* of A.

data/info.

Example: Rolling the Fair Die

Sample space $S = \{1,2,3,4,5,6\}$

① $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$

$P(B) = \frac{|B|}{|S|} = \frac{3}{6} = \frac{1}{2}$.

$P(A \cap B) = \frac{|A \cap B|}{|S|} = \frac{2}{6} = \frac{1}{3}$

$A \cap B = \{1,3\}$

Rolling a fair die. Let A be the event that the outcome is an odd number, i.e, $A = \{1,3,5\}$. Also let B be the event that the outcome is less than or equal to 3, i.e, $B = \{1,2,3\}$. Find $P(A|B)$ and $P(B|A)$.

② $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} > \frac{1}{2} = P(B)$

$P(A) = \frac{|A|}{|S|} = \frac{3}{6} = \frac{1}{2}$

$P(B \cap A) = P(A \cap B) = \frac{1}{3}$

Conditional Probability is Probability

$1^\circ.\ 0 \le \hat{P}(\cdot) \le 1$

$\hat{P}(S) = 1$

$\hat{P}(\phi) = 0.$

$\hat{P}(\cdot) = P(\cdot|E)$

$2^\circ.\ \hat{P}(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \hat{P}(A_j)$

- Condition on an event E, we update our beliefs to be consistent with this knowledge.

$3^\circ.\ \hat{P}(A^c) = 1 - \hat{P}(A)$

- $P(\cdot|E)$ is also a probability function with sample space S:
  - $0 \le P(\cdot|E) \le 1$ with $P(S|E) = 1$ and $P(\emptyset|E) = 0$.
  - If events $A_1, A_2, \ldots$ are disjoint, then $P(\cup_{j=1}^{\infty} A_j|E) = \sum_{j=1}^{\infty} P(A_j|E)$.
  - $P(A^c|E) = 1 - P(A|E)$.
  - Inclusion-exclusion: $P(A \cup B|E) = P(A|E) + P(B|E) - P(A \cap B|E)$.

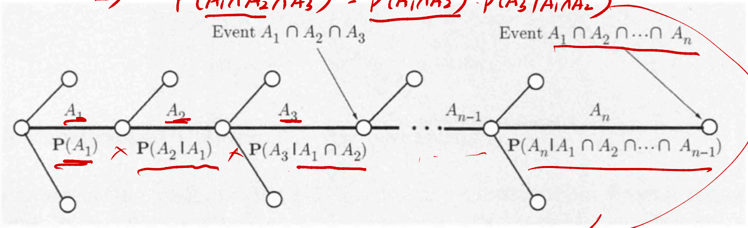
$4^\circ.\ \hat{P}(A \cup B) = \hat{P}(A) + \hat{P}(B) - \hat{P}(A \cap B)$

Conditional Probability: Chain Rule

$$P(A|B) \triangleq \frac{P(A\cap B)}{P(B)} \Rightarrow P(A|B)\cdot P(B) = P(A\cap B) = P(A)P(B|A)$$

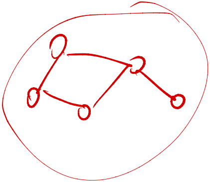
$$\Rightarrow P(A_1\cap A_2) = P(A_1)\cdot P(A_2|A_1)$$

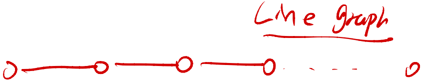
$$\Rightarrow P(A_1\cap A_2\cap A_3) = P(A_1\cap A_2)\cdot P(A_3|A_1\cap A_2)$$



$$= P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1\cap A_2)$$

Conditional Probability: Chain Rule





Theorem

For any events $A_1, \ldots, A_n$ with positive probabilities,

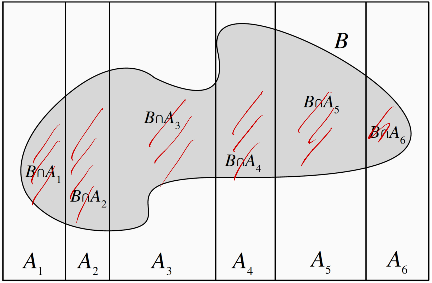
$$P(A_1, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1}).$$

$$\log P(A_1, \ldots A_n) = \log P(A_1) + \log P(A_2|A_1) + \cdots + \log P(A_n|A_1, \ldots A_{n-1})$$

Thinking Conditionally

- Conditioning :a powerful problem-solving strategy
- Reducing a complicated probability problem to a bunch of simpler conditional probability problems
- First-step analysis: obtain recursive solution to multi-stage problems
- **Conditioning is the soul of statistics!**

The Law of Total Probability (LOTP)



$B = \bigcup_{j=1}^{6} (B \cap A_j)$

$P(B) = \sum_{j=1}^{6} P(B \cap A_j)$

$= \sum_{j=1}^{6} P(A_j) P(B|A_j)$

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i) > 0$ for all i. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Outline

$P(A|B)$ ✓

$P(A|B\cap C)$ ✓

$P(A|B,C)$ ✓

$P(A|B|C)$ ✗

$P(A|B|C,D)$ ✗

Independence of Two Events

Definition

Events A and B are *independent* if

$$P(A \cap B) = P(A)P(B).$$

If $P(A) > 0$ and $P(B) > 0$, then this is equivalent to

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Independence vs. Disjointness

A, B

$P(A|B) = P(A)$

A, B disjoint

$P(A \cap B) = 0$

Sample space $S = \{1, 2, 3, .., 10\}$

1^0. $A \cap B = \{2, 4, 6\}$

Rolling a fair 10-sided die. Let A be the event that the outcome is less than 7, i.e, $A = \{1, 2, 3, 4, 5, 6\}$. Also let B be the event that the outcome is an even number, i.e, $B = \{2, 4, 6, 8, 10\}$. Are A and B independent?

2^0. $P(A) = \frac{|A|}{|S|} = \frac{6}{10} = 0.6$

$P(B) = \frac{|B|}{|S|} = \frac{5}{10} = 0.5$

$P(A \cap B) = \frac{|A \cap B|}{|S|} = \frac{3}{10} = 0.3 = 0.6 \times 0.5 = P(A) \cdot P(B)$

$\Rightarrow$ A and B are independent.

Independence of Complementary Set

W.L.O.G.

Without loss of Generality

$P(A)>0,\ P(B)>0.$

$\Rightarrow\ P(A|B)=P(A);$

$P(B|A)=P(B);$

$\Rightarrow\ P(B^c|A) = 1-P(B|A)$

$= 1-P(B) = P(B^c)$

Theorem

If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.

Independence of Three Events

Definition

Events A, B and C are *independent* if all of the following equations hold:

$$\left.\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A \cap C) &= P(A)P(C) \\ P(B \cap C) &= P(B)P(C) \end{aligned}\right\} \text{Pairwise independence}$$

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Pairwise Independence $\neq$ Independence

① $P(A)=P(B)=\frac{1}{2}$

② A, B, C pairwise independence

$P(A\cap B)=P(A)\cdot P(B)$ ✓

$P(A\cap C)=P(A\cap B)=\frac{1}{4}=P(A)\cdot P(C)$ ✓

$P(B\cap C)=P(A\cap B)=\frac{1}{4}=P(B)\cdot P(C)$ ✓

$P(C)=P[(A\cap B)\cup(A^c\cap B^c)]$

$=P(A\cap B)+P(A^c\cap B^c)$

$=P(A)\cdot P(B)+P(A^c)\cdot P(B^c)$

$=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$

Consider two fair, independent coin tosses.

- A: the event that the first is Heads.
- B: the event that the second is Heads.
- C: the event that both tosses have the same result.

$P(C|A,B)=1$

③ $P(A\cap B\cap C)=P(A\cap B)\cdot \underbrace{P(C|A\cap B)}_{1}$

$=P(A\cap B)$

$=\frac{1}{4} \neq \underbrace{P(A)\cdot P(B)\cdot P(C)}_{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}}=\frac{1}{8}$

Conditional Independence

$$\hat{P}(\cdot) = P(\cdot \mid E)$$

Definition

Events A and B are said to be *conditionally independent* given E if:

$$P(A \cap B|E) = P(A|E)P(B|E).$$

$$\hat{P}(A \cap B) = \hat{P}(A) \cdot \hat{P}(B)$$

Example: Conditional Independence $\not\Rightarrow$ Independence

① $P(A_1 \cap A_2 | F) = P(A_1|F) \cdot P(A_2|F)$

$P(A_1 \cap A_2 | F^c) = P(A_1|F^c) \cdot P(A_2|F^c)$

$\Rightarrow P(A_1 \cap A_2) \neq P(A_1) \cdot P(A_2)$, with $P(A_1 \cap A_2) = \frac{13}{32}$, $P(A_1) = \frac{5}{8}$, $P(A_2) = \frac{5}{8}$

② $P(A_1) \overset{LOTP}{=} P(A_1|F)P(F) + P(A_1|F^c)P(F^c)$

$= \frac{1}{2} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{2} = \frac{5}{8}$; $P(A_2) = \frac{5}{8}$ (LOTP)

- We choose either a fair coin or a biased coin (w.p. 3/4 of landing Heads).

③ $P(A_1 \cap A_2) \overset{LOTP}{=} P(A_1 \cap A_2|F)P(F) + P(A_1 \cap A_2|F^c)P(F^c)$

$= P(A_1|F) \cdot P(A_2|F) \cdot P(F) + P(A_1|F^c) P(A_2|F^c) P(F^c)$

$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{2}$

$= \frac{13}{32}$

$\neq \left(\frac{5}{8}\right)^2$

- But we do not know which one we have chosen and we flip it twice.
- Event F: "chosen the fair coin"
- Event A_1: "the first coin tosses landing Heads"
- Event A_2: "the second coin tosses landing Heads"

Example: Independence $\not\Longrightarrow$ Conditional Independence

A, B are independent

Given R, A and B

are NOT independent.

- Only Alice and Bob call me.
- Each day, they call me independently.
- Event R: "Phone rings".
- Event A: "Alice call me"
- Event B: "Bob call me"

$P(B \mid A^c, R) = 1$

Example: Conditional Independence Given E vs. E^c

$$P(W \cap A \mid E^c) = P(W \mid E^c) \cdot P(A \mid E^c)$$

$$P(W \cap A \mid E) \neq P(W \mid E) \cdot P(A \mid E)$$

- There are two classes: good & bad.
- Good: students get grade A with working hard.
- Bad: students get grades randomly regardless of their efforts.
- Event E: "Class is good".
- Event W: "Students working hard".
- Event A: "Students receive grade A".

(Conditional) Independence Simplifies Computing

- If events $A_1, A_2, \ldots, A_n$ are independent, then

$\bar{A}_1 \triangleq A_1^c$

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{j=1}^{n} P(A_j)$$

$1 - P(\overline{A_1 \cup \cdots \cup A_n}) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \cdots \bar{A}_n) = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \cdots P(\bar{A}_n)$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - \prod_{j=1}^{n} [1 - P(A_j)]$$

$\leq \sum_{i=1}^{n} P(A_i)$

- If events $A_1, A_2, \ldots, A_n$ are conditional independent given event E, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n | E) = \prod_{j=1}^{n} P(A_j | E)$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n | E) = 1 - \prod_{j=1}^{n} [1 - P(A_j | E)]$$

Outline

LOTP + independence.

Bayes' Rule

$P(A \cap B) = P(A) \cdot P(B|A)$

$= P(B) \cdot P(A|B)$

Theorem

For any events A and B with positive probabilities,

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

Bayes' Rule: Learning Perspective

Coin. A_1 = " the coin is fair coin " ; $P(H) = P(T) = 0.5$

A_2 = " the coin is biased coin " ; $P(H) = 0.8$; $P(T) = 0.2$

Theorem

B = "HHHH" ; $P(B|A_1) = (0.5)^4$; $P(B|A_2) = (0.8)^4$

belief

A represents cognition, B represents data, then

$P(B|A_1) < P(B|A_2)$

Hypothesis.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \propto P(B|A) \cdot P(A)$$

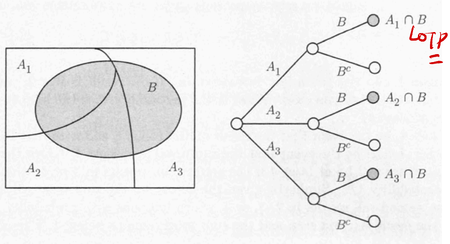
- $P(A)$: *Prior Cognition*
- $P(B|A)$: *Likelihood (info of data given prior cognition)* 似然
- $P(A|B)$: *Posterior Cognition*

B' = "HTHT" ; $P(B'|A_1) = (0.5)^4 = (0.25)^2 > P(B'|A_2)$

$P(B'|A_2) = 0.8^2 \times 0.2^2 = (0.16)^2$

Inference & Bayes' Rule

$$P(A_i|B) = \frac{P(B|A_i)\cdot P(A_i)}{P(B)} \overset{\text{LOTP}}{=} \frac{P(B|A_i)\,p(A_i)}{\sum_{j=1}^{n} P(B|A_j)\,p(A_j)}$$



Theorem

Let $A_1, \dots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i) > 0$ for all i. Then for any event B such that $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\underbrace{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)}_{P(B)}}.$$

Inference & Bayes' Rule

Example:

A_1: "coin is fair"

A_2: "coin is biased".

B: data/info tossing coins.

$P(A_1|B)$; $P(A_2|B)$;

Inference:

MLE.

MMSE;

Maximum A posterior probability. (MAP)

$P(A_1|B) > P(A_2|B)$ chose A_1

$<$ A_2

$$\log\left(\frac{P(A_1|B)}{P(A_2|B)}\right) = \frac{\frac{P(B|A_1)P(A_1)}{\cancel{P(B)}}}{\frac{P(B|A_2)P(A_2)}{\cancel{P(B)}}}$$

$$= \log\left(\frac{P(B|A_1)}{P(B|A_2)}\right) + \log\left(\frac{P(A_1)}{P(A_2)}\right).$$

$$\Rightarrow \log\frac{P(A_1|B)}{P(A_2|B)} \overset{?}{>} 0$$

✓ choose A_1

✗ A_2

$$P(A_1|B) - P(A_2|B) \overset{?}{>} 0$$

Example: Random Coin

event A: "the chosen coin lands HHH".

F: "we picked the fair coin".

① $P(F) = \frac{1}{2}$, $P(F^c) = \frac{1}{2}$; $P(F|A)$?

② $$P(F|A) = \frac{P(A|F)P(F)}{P(A|F)P(F) + P(A|F^c)P(F^c)} = \frac{(\frac{1}{2})^3 \cdot \frac{1}{2}}{(\frac{1}{2})^3 \cdot \frac{1}{2} + (\frac{3}{4})^3 \cdot \frac{1}{2}}$$

P(A) LOTP

You have one fair coin, and one biased coin which lands Heads with probability 3/4. You pick one of the coins at random and flip it three times. It lands Heads all three times. Given this information, what is the probability that the coin you picked is the fair one?

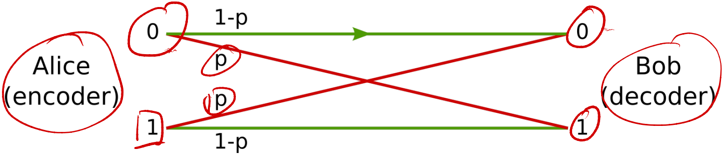
$$= \frac{8}{35} < \frac{1}{2}$$

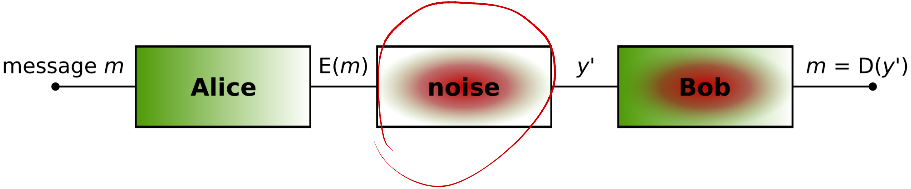
③ $$P(F^c|A) = 1 - P(F|A) = \frac{27}{35} > P(F|A)$$

MAP Inference rule: $\Rightarrow$ We picked the biased coin.

Example: Random Coin

Example: Communication Channel





Example: Communication Channel

5

event A_1 : "Alice sent a "1"".

B_1 : "Bob received a "1"".

① $P(A_1) = 0.5$; $P(A_1^c) = 0.5$;

object $P(A_1|B_1)$

② $P(B_1|A_1) = 0.95$; $P(B_1|A_1^c) = 0.05$;

Suppose that Alice sends only one bit (a 0 or 1) to Bob, with equal probabilities. If she sends a 0, there is a 5% chance of an error occurring, resulting in Bob receiving a 1; if she sends a 1, there is a 5% chance of an error occurring, resulting in Bob receiving a 0. Given that Bob receives a 1, what is the probability that Alice actually sent a 1?

③ $$P(A_1|B_1) = \frac{P(B_1|A_1)\cdot P(A_1)}{P(B_1|A_1)\cdot P(A_1) + P(B_1|A_1^c)P(A_1^c)}$$

$$= \frac{0.95\cdot 0.5}{0.95\cdot 0.5 + 0.05\cdot 0.5} = 0.95$$

Example: Bayes Spam Filter

event $\underline{G}$ = "an email is a spam"

F = "an email with "free money"".
(data)

prior

① $P(G) = 0.8$; $P(G^c) = 0.2$

object ↙ data

② $P(F|G) = 0.1$; $P(F|G^c) = 0.01$ likelihood $\underline{P(G|F)}$

A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase "free money" is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention "free money". What is the probability that it is spam?

③ $$P(G|F) = \frac{P(F|G)\cdot P(G)}{P(F|G)\cdot P(G) + P(F|G^c)P(G^c)} = \frac{0.1\times 0.8}{0.1\times 0.8 + 0.01\times 0.2} = \frac{80}{82} \approx 0.9756$$

$$\log \frac{P(F|G)P(G)}{P(F|G^c)P(G^c)} = \log\left(\frac{0.1\times 0.8}{0.01\times 0.2}\right) = \log 40 > 1$$

Example: ~~(Naive)~~ Bayes Spam Filter

(Naive)

event G = "this email is a spam",

F_1 = "email with 'free money'"

F_2 = " . . . . . 'fun' ".

$$\log\left[\frac{P(G|F_1,F_2)}{P(G^c|F_1,F_2)}\right] \overset{?}{>} 0$$

$$\Downarrow$$

$$\log\left[\frac{\dfrac{P(F_1,F_2|G)P(G)}{\cancel{P(F_1,F_2)}}}{\dfrac{P(F_1,F_2|G^c)P(G^c)}{\cancel{P(F_1,F_2)}}}\right] = \log\left[\frac{P(F_1,F_2|G)\,P(G)}{P(F_1,F_2|G^c)\cdot P(G^c)}\right]$$

Assumption: (Naive) Conditional independence

$$= \log\left[\frac{P(F_1|G)\cdot P(F_2|G)\cdot P(G)}{P(F_1|G^c)\cdot P(F_2|G^c)\cdot P(G^c)}\right] = \log\left[\frac{P(F_1|G)}{P(F_1|G^c)}\right] + \log\left[\frac{P(F_2|G)}{P(F_2|G^c)}\right]$$

log-likelihood ratio ① , . . . ②

$$+\log\left[\frac{P(G)}{P(G^c)}\right] \overset{?}{>} 0$$

log prior ratio.

Example: Bayes Spam Filter

Bayes' rule is used several times in the context of spam:

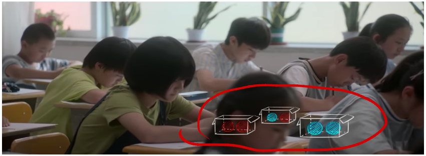
- Compute the probability that the message is spam, knowing that a given word appears in this message
- Compute the probability that the message is spam, taking into consideration all of its words (or a relevant subset of them)
- Sometimes a third time, to deal with rare words.
- Reference: `https://en.wikipedia.org/wiki/Naive_Bayes_spam_filtering`

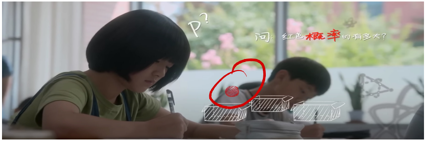
likelihood

conditional independence assumption.

Bertrand's Box Paradox

Equal prior $\not\Rightarrow$ Equal posterior.





Bertrand's Box Paradox

① event A_1: "pick box 1".
A_2: " . . . . 2"
A_3: " . . . . 3"
B: "Sampling from the box is a Gold coin".

② $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$. Prior.

$P(B|A_1) = 1$; $P(B|A_2) = 0$; $P(B|A_3) = \frac{1}{2}$

Likelihood

There are three boxes:

- a box containing two gold coins 1
- a box containing two silver coins 2
- a box containing one gold coin and a silver coin. 3

③ object

Posterior

$P(A_1|B) = \frac{2}{3}$

$P(A_2|B) = 0$

$P(A_3|B) = \frac{1}{3}$.

After choosing a box at random and withdrawing one coin at random, if that happens to be a gold coin, find the probability of the next coin drawn from the same box also being a gold coin.

$$P(A_1|B) = \frac{P(B|A_1)\cdot P(A_1)}{P(B|A_1)\cdot P(A_1) + P(B|A_2)\cdot P(A_2) + P(B|A_3)\cdot P(A_3)}$$

$$= \frac{1\cdot\frac{1}{3}}{1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{1}{3}} = \frac{2}{3}.$$

Solution

Conditional prob.

data / cognition

cognition / data

Likelihood

Posterior

$$\sum_{i=1}^{n} P(B|A_i) \neq 1$$

$\sum$ posterior $= 1$

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

Solution

Example

西汉刘向编著的《战国策》中，记录了一个“三人成虎”的故事：庞葱与太子质于邯郸，谓魏王曰:“今一人言市有虎，王信之乎?”王曰:“否.”“二人言市有虎，王信之乎?”王曰:“寡人疑之矣.”“三人言市有虎，王信之乎?”王曰:“寡人信之矣.”庞葱曰:“夫市之无虎明矣，然而三人言而成虎. 今邯郸去大梁也远于市，而议臣者过于三人矣, 愿王察之矣.”王曰:“寡人自为知.”于是辞行，而谗言先至. 后太子罢质，果不得见.

Solution

① event B = "Tiger in the street". $P(B) = 0.05$ prior

A_i = "the ith people said: "I saw a tiger"", $i = 1, 2, 3$;

$P(B^C) = 0.95$ prior

② Likelihood. $P(A_i|B) = 0.8$; $P(A_i|B^C) = 0.15$.

conditional independence assumption.

$P(A_i \cap A_j | B) = P(A_i|B) \cdot P(A_j|B)$

$P(A_1, A_2, A_3|B) = P(A_1|B) \cdot P(A_2|B) \cdot P(A_3|B)$

③ $P(B|A_1) = \dfrac{P(A_1|B) \cdot P(B)}{P(A_1|B) \cdot P(B) + P(A_1|B^C) P(B^C)}$

$= \dfrac{0.8 \times 0.05}{0.8 \times 0.05 + 0.15 \times 0.95} = 0.22$

④ $P(B|A_1, A_2) = \dfrac{P(A_1, A_2|B) \cdot P(B)}{P(A_1, A_2|B) P(B) + P(A_1, A_2|B^C) \cdot P(B^C)} = 0.60$

$= \dfrac{P(A_1|B) \cdot P(A_2|B) \cdot P(B)}{P(A_1|B) \cdot P(A_2|B) \cdot P(B) + P(A_1|B^C) \cdot P(A_2|B^C) \cdot P(B^C)} = \dfrac{0.8^2 \times 0.05}{0.8^2 \times 0.05 + 0.15^2 \times 0.95}$

Solution

(3) $$P(B|A_1,A_2,A_3) = \frac{P(A_1,A_2,A_3|B)}{P(B)}$$

$$= \frac{P(A_1|B)\cdot P(A_2|B)\cdot P(A_3|B)\cdot P(B)}{P(A_1|B)\cdot P(A_2|B)\cdot P(A_3|B)\cdot P(B) + P(A_1|B^C)\cdot P(A_2|B^C)\cdot P(A_3|B^C)\cdot P(B^C)}$$

$$= \frac{0.8^3 \times 0.05}{0.8^3 \times 0.05 + 0.15^3 \times 0.85} = \underline{0.89}$$

Solution

Example: Boy or Girl Paradox

event A = "both children are girls" $= \{(G,G)\}$

C = "the first(elder) child is girl"

$= \{(G,G),(G,B)\}$

Consider a family that has two children. We are interested in the children's genders. Our sample space is $\{(G,G),(G,B),(B,G),(B,B)\}$, where G means girl and B means boy. Also assume that all four possible outcomes are equally likely.

$A \subseteq C$

- What is the probability that both children are girls given that the first(elder) child is a girl?

$$P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{P(A)}{P(C)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$$

Example: Boy or Girl Paradox

event A = " " $\{(G,G)\}$;

event D = "At least one Girl" $= \{(G,G),(G,B),(B,G)\}$

Consider a family that has two children. We are interested in the children's genders. Our sample space is $\{(G,G),(G,B),(B,G),(B,B)\}$, where G means girl and B means boy. Also assume that all four possible outcomes are equally likely.

$A \subset D$

- We ask the father: "Do you have at least one daughter?" He responds "Yes!" Given this extra information, what is the probability that both children are girls? In other words, what is the probability that both children are girls given that we know at least one of them is a girl?

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A)}{P(D)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Solution

Example: Boy or Girl Paradox

Likelihood

$P(R|BB) = 0$;

$P(R|BG) = \frac{1}{2}$;

$P(R|GB) = \frac{1}{2}$;

$P(R|GG) = 1$;

event R (data) $\neq$ event "at least one girl"

Consider a family that has two children. We are interested in the children's genders. Our sample space is $\{(G,G),(G,B),(B,G),(B,B)\}$, where G means girl and B means boy. Also assume that all four possible outcomes are equally likely.

- If we randomly ran into one of the two in the shopping mall, and see that she is a girl. Given this extra information, what is the probability that both children are girls?

$P(GG|R)$ posterior

$$= \frac{P(R|GG) \times P(GG)}{P(R)}$$

LOTP

Solution

$$P(R) \overset{\text{LOTP}}{=} P(R|GG)\cdot P(GG) + P(R|GB)\cdot P(GB)$$

$$+ P(R|BG)\cdot P(BG) + P(R|BB)\cdot P(BB)$$

$$= 1\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{4} + 0\cdot\frac{1}{4}$$

$$= \frac{1}{2}$$

$$\Rightarrow \quad P(GG|R) = \frac{P(R|GG)\cdot P(GG)}{P(R)} = \frac{1\cdot\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

Example: Boy or Girl Paradox

event L (data)

likelihood. $P(L|BB) = 0$; $P(L|BG) = 1 \cdot \alpha = \alpha$

$P(L|GB) = \alpha$; $P(L|GG) = 1-(1-\alpha)^2 = 2\alpha-\alpha^2$

Consider a family that has two children. We are interested in the children's genders. Our sample space is $\{(G,G),(G,B),(B,G),(B,B)\}$, where G means girl and B means boy. Also assume that all four possible outcomes are equally likely.

- We ask the father, "Do you have at least one daughter named Catherine?" He replies, "Yes!" What is the probability that both children are girls? In other words, we want to find the probability that both children are girls, given that the family has at least one daughter named Catherine. Here we assume that if a child is a girl, her name will be Catherine with probability α independently from other children's names. If the child is a boy, his name will not be Catherine.

Solution

$$\Rightarrow P(L) \overset{LOTP}{=} P(L|BB)\cdot P(BB) + P(L|BG)\cdot P(BG)$$

$$+ P(L|GB)\cdot P(GB) + P(L|GG)\cdot P(GG)$$

$$= 0\cdot\frac{1}{4} + \alpha\cdot\frac{1}{4} + \alpha\cdot\frac{1}{4} + (2\alpha-\alpha^2)\cdot\frac{1}{4}$$

$$= \frac{4\alpha-\alpha^2}{4}$$

$$\Rightarrow P(GG|L) = \frac{P(L|GG)\cdot P(GG)}{P(L)} = \frac{(2\alpha-\alpha^2)\cdot\frac{1}{4}}{(4\alpha-\alpha^2)\cdot\frac{1}{4}}$$

posterior

$$= \frac{2-\alpha}{4-\alpha}, \quad 0 \le \alpha \le 1$$

$$\in \left[\frac{1}{3}, \frac{1}{2}\right]$$

Bayes' Rule with Extra Conditioning

$\hat{P}(\cdot) = P(\cdot|E)$

$\hat{P}(A|B) \overset{\checkmark}{=} P(A|B,E) \neq P(A|B|E)$

Theorem

Provided that $P(A \cap E) > 0$ *and* $P(B \cap E) > 0$, *we have*

$$P(A|B,E) = \frac{P(B|A,E)P(A|E)}{P(B|E)}.$$

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

LOTP with Extra Conditioning

$\tilde{P}(\cdot) = P(\cdot | E)$

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i \cap E) > 0$ for all i. Then

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E).$$

$$\tilde{P}(B) = \sum_{i=1}^{n} \tilde{P}(B|A_i) \cdot \tilde{P}(A_i)$$

Approaches for Finding $P(A|B, C)$

- Think of B, C as the single event $B \cap C$.
- Use Bayes' rule with extra conditioning on C.
- Use Bayes' rule with extra conditioning on B.

Example: Random Coin

history data

① event A = " the chosen coin lands Heads three times"

new data H = " land Head on the fourth time"

Cognition F = "we picked the fair coin"

You have one fair coin, and one biased coin which lands Heads with probability 3/4. You pick one of the coins at random and flip it three times. It lands Heads all three times. If we toss the coin a fourth time, what is the probability that it will land Heads once more?

② $P(H|A) \overset{\text{LOTP with extra conditioning}}{=} P(H|F,A)P(F|A) + P(H|F^c,A)\cdot P(F^c|A)$

$= \underbrace{P(H|F)}_{\frac{1}{2}} \cdot \frac{8}{35} + \underbrace{P(H|F^c)}_{\frac{3}{4}} \cdot \frac{27}{35}$

$= \frac{97}{140} > \frac{1}{2}$

Example: Random Coin

Outline

Example: Laplace's Rule of Succession

① event C_i = "the i^{th} coin is selected initially", $i = 0, 1, \ldots, k$
(cognition)

(History data) F_n = "the first n flips all land on Heads"

(new data) H = "the $(n+1)^{th}$ flip lands on Head".

There are $k+1$ coins in a box. When flipped, the i^{th} coin will turn up heads with probability i/k, $i = 0, 1, \ldots, k$. A coin is randomly selected from the box and is then repeatedly flipped. If the first n flips all result in heads, what is the conditional probability that the $(n+1)^{st}$ flip will do likewise?

② $P(H \mid F_n) = \sum_{i=0}^{k} P(H \mid C_i, F_n) \cdot P(C_i \mid F_n)$

$\cancel{P(H)}$

$= \sum_{i=0}^{k} P(H \mid C_i) \cdot P(C_i \mid F_n)$

$\left(\frac{i}{k}\right)$

Solution of Laplace's Rule of Succession

$$P(C_i|F_n) = \frac{P(F_n|C_i)\cdot P(C_i)}{P(F_n)} = \frac{P(F_n|C_i)\cdot P(C_i)}{\sum_{j=0}^{k} P(F_n|C_j)\cdot P(C_j)}$$

$$= \frac{\left(\frac{i}{k}\right)^n \cdot \frac{1}{k+1}}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n \cdot \frac{1}{k+1}} = \frac{\left(\frac{i}{k}\right)^n}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n}$$

③ $$P(H|F_n) = \sum_{i=0}^{k} \frac{i}{k} \cdot \frac{\left(\frac{i}{k}\right)^n}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n} = \frac{\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n}$$

$k \gg 1$

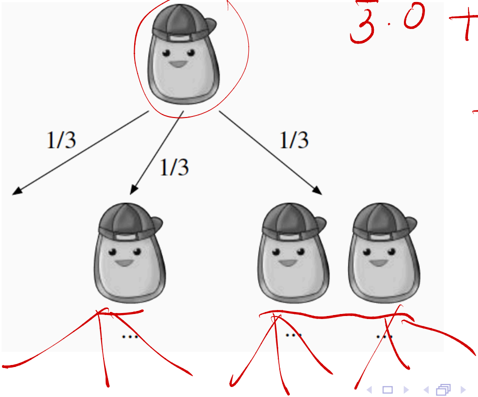
$$\frac{1}{k}\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1} \approx \int_0^1 x^{n+1}dx = \frac{1}{n+2}$$

$$\frac{1}{k}\sum_{j=0}^{k}\left(\frac{j}{k}\right)^{n} \approx \int_0^1 x^{n}dx = \frac{1}{n+1}$$

$$\Rightarrow P(H|F_n) \overset{k\gg 1}{\approx} \frac{n+1}{n+2}$$

Example: Branching Process

A single amoeba, Bobo, lives in a pond. After one minute Bobo will either die, split into two amoebas, or stay the same, with equal probability, and in subsequent minutes all living amoebas will behave the same way, independently. What is the probability that the amoeba population will eventually die out?



$$\frac{1}{3}\cdot 0+\frac{1}{3}\cdot 1+\frac{1}{3}\cdot 2 = 1$$

First-Step Analysis: Branching Process

① event D = " the population eventualy dies out"

B_i = " Bobo turns into i amobobas after the first minute ", $i=0,1,2$

$P(B_i)=\frac{1}{3}$

② $P(D|B_0)=1$; $P(D|B_1)=P(D)$; $P(D|B_2)$ $=P(D)\cdot P(D)$ $=P^2(D)$

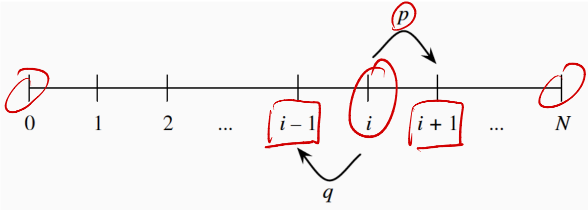
③ By LOTP $\Rightarrow$ $P(D)=P(D|B_0)\cdot P(B_0)+P(D|B_1)\cdot P(B_1)+P(D|B_2)\cdot P(B_2)$

$=1\times\frac{1}{3}+P(D)\cdot\frac{1}{3}+P^2(D)\cdot\frac{1}{3}$

$\Rightarrow (P(D)-1)^2=0$ $\Rightarrow P(D)=1$

Example: Gambler's Ruin

Two gamblers, A and B, make a sequence of dollar 1 bets. In each bet, gambler A has probability p of winning, and gambler B has probability $q = 1 - p$ of winning. Gambler A starts with i dollars and gambler B starts with $N - i$ dollars; the total wealth between the two remains constant since every time A loses a dollar, the dollar goes to B, and vice versa. The game ends when either A or B is ruined (run out of money). What is the probability that A wins the game (walking away with all the money)?



First-Step Analysis: Gambler's Ruin

① event A_i = "A starts with i dollars".

prob. $P_i = P(\text{A wins the game} \mid A_i)$; $P_0 = P(\text{A wins} \mid A_0) = 0$;

$P_N = P(\text{A wins} \mid A_N) = 1$;

② $\underline{1 \le i \le N-1}$; event W_1 = "A wins the first bet";

$P(W_1) = p$.

By LoTP with extra conditioning

$$P_i = P(\text{A wins} \mid A_i) = P(\text{A wins} \mid W_1, A_i) \cdot P(W_1 \mid A_i) + P(\text{A wins} \mid W_1^c, A_i) \cdot P(W_1^c \mid A_i)$$

$= P(\text{A wins} \mid A_{i+1}) = P_{i+1}$

p

$(1-p=q)$

$P(\text{A wins} \mid A_{i-1}) = P_{i-1}$

$$\Rightarrow \begin{cases} P_i = P_{i+1} \cdot p + P_{i-1}(1-p), & 1 \le i \le N-1 \\ P_0 = 0; \quad P_N = 1; \end{cases}$$

First-Step Analysis: Gambler's Ruin

$p_i = p_{i+1} \cdot p + p_{i-1} \cdot q$ $\qquad$ $\underline{p+q=1}$

Method 1: $p_i(p+q) = p_{i+1} \cdot p + p_{i-1} \cdot q$

$\Rightarrow (p_{i+1} - p_i) \cdot p = (p_i - p_{i-1}) \cdot q$ $\qquad$ $\underline{(p \overset{?}{=} q)}$

$\alpha_i = \underline{p_i - p_{i-1}} \Rightarrow \alpha_{i+1} \cdot p = \alpha_i \cdot q$ $\qquad$ $\Rightarrow \underline{\alpha_{i+1} = \frac{q}{p} \cdot \alpha_i}$

$\Downarrow$

Solve α_i

$\Rightarrow p_i - p_{i-1} = \alpha_i$

$\Rightarrow p_i$

$$\Rightarrow p_i = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}, & p \neq q \\[2ex] \dfrac{i}{N}, & p = q = \frac{1}{2} \end{cases}$$

First-Step Analysis: Gambler's Ruin

$a_{n+2} = a_{n+1} + a_n$

Method 2 : $f_{i+1} = bf_i + af_{i-1}$ $i \geq 1$

Characteristic equation $x^2 = bx + a$; two roots r_1, r_2

$$\begin{cases} \text{if } r_1 \neq r_2 ; & f_i = c \cdot r_1^i + d \cdot r_2^i \\ \text{if } r_1 = r_2 ; & f_i = c \cdot r^i + d \cdot i \cdot r^i \end{cases}$$

$0 < x < 1$

$$F(x) = \sum_{i=0}^{\infty} f_i \cdot x^i$$

$$\frac{Ax+B}{Cx+d}$$

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$$

Outline

Simpson's Paradox

- A phenomenon in probability and statistics in which a trend appears in multiple groups of data but disappears or reverses when the groups are combined
- Mentioned earlier by Karl Pearson in 1899 & Udny Yule in 1903
- Officially proposed by Edward H. Simpson (1922-2019) in 1951
- Delicate connection between probabilistic reasoning and causal inference
- Reference: `https://plato.stanford.edu/entries/paradox-simpson/`

Example of Simpson's Paradox

if $\frac{a_1}{b_1} < \frac{c_1}{d_1}$; $\frac{a_2}{b_2} < \frac{c_2}{d_2}$

possible $\rightarrow$ $\frac{a_1+a_2}{b_1+b_2} > \frac{c_1+c_2}{d_1+d_2}$

	Heart	Band-Aid
	90	10
Success	70	10
Failure	20	0

Dr. Hibbert

	Heart	Band-Aid
Success	2	81
Failure	8	9

Dr. Nick

Heart : $\frac{70}{90} = \frac{7}{9} > \frac{2}{10}$

Band-Aid : $\frac{10}{10} = 1 > \frac{81}{90} = 0.9$

$\frac{80}{100} = 0.8 < \frac{83}{100} = 0.83$

Math Behind Simpson's Paradox

event A = " a Successful surgey".

B = " Nick is the surgeon"

C = " the surgery is a heart surgery".

If

$$P(A|B, C) < P(A|B^c, C),$$
$$P(A|B, C^c) < P(A|B^c, C^c),$$

then is it possible that

$$P(A|B) = P(A|B,C)P(C|B) + P(A|B,C^c)P(C^c|B)$$
$$P(A|B^c) = P(A|B^c,C)P(C|B^c) + P(A|B^c,C^c)P(C^c|B^c)$$

$$P(A|B) > P(A|B^c)?$$

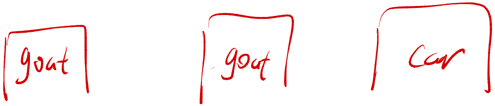
$$P(C^c|B) >> P(C^c|B^c)$$

Nick: $\frac{90}{100} = 0.9$ >> Hibert $\frac{10}{100} = 0.1$

Another Example of Simpson's Paradox

Gender discrimination in college admissions: In the 1970s, men were significantly more likely than women to be admitted for graduate study at the University of California, Berkeley, leading to charges of gender discrimination. Yet within most individual departments, women were admitted at a higher rate than men. It was found that women tended to apply to the departments with more competitive admissions, while men tended to apply to less competitive departments.

Monty Hall Problem



On the game show Let's Make a Deal, hosted by Monty Hall, a contestant chooses one of three closed doors, two of which have a goat behind them and one of which has a car. Monty, who knows where the car is, then opens one of the two remaining doors. The door he opens always has a goat behind it (he never reveals the car!). If he has a choice, then he picks a door at random with equal probabilities. Monty then offers the contestant the option of switching to the other unopened door. If the contestant's goal is to get the car, should she switch doors?

Remarks

Erdos Number

- Even great mathematician Paul Erdos & Persi Diaconis made mistakes.
- Originally proposed by Steve Selvin, American Statistician 1975.
- Famous when proposed in Marilyn vos Savant's "Ask Marilyn" column, Parade magazine 1990.
- Approximately 10000 readers, including nearly 1000 with PhDs, said no need to switch.

Solution of Monty Hall Problem

(1) Label doors 1, 2, 3. W.L.O.G. player picks door 1.

(2) Cognitive event C_i = "the car is behind the ith car", $i = 1, 2, 3$.

$$P(C_1) = P(C_2) = P(C_3) = \frac{1}{3}.$$

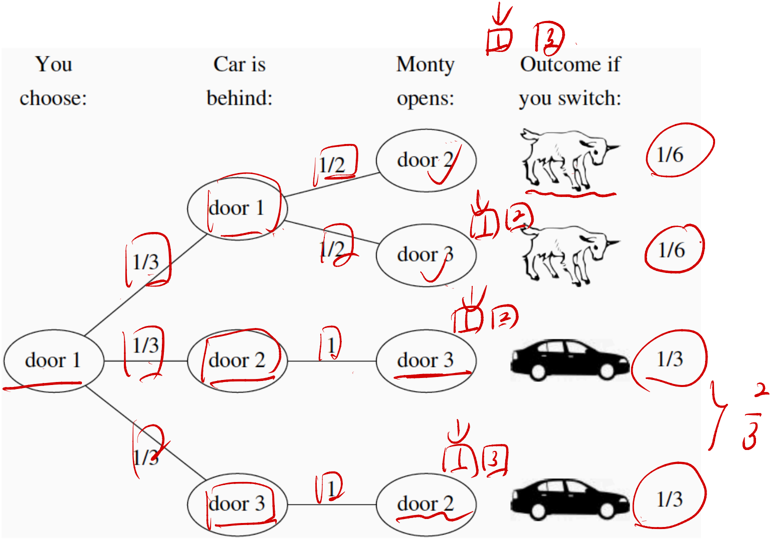
(3). $P(\text{"get a car"}) = P(\text{win}) \overset{\text{LOTP}}{=} P(\text{win} \mid C_1) P(C_1) + P(\text{win} \mid C_2) P(C_2) + P(\text{win} \mid C_3) P(C_3)$

1°. stay. $P^{\text{stay}}(\text{win}) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{3}$

2°. switch ✓ $P^{\text{switch}}(\text{win}) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$

Tree Diagram of Monty Hall Problem

door 1, 2, 3



Information Conveyed by Monty's Action

(data) event M_j = " Monty opened the jth door " $j = 1, 2, 3$

(cognitive) event C_i = " Car is behind the ith door ", $i = 1, 2, 3$.

$$P(M_2) \overset{\text{LOTP}}{=} P(M_2|C_1)P(C_1) + P(M_2|C_2)P(C_2) + P(M_2|C_3)P(C_3)$$

$$= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2}$$

Before Monty open door 2 : $P(C_1) = P(C_2) = P(C_3) = \frac{1}{3}$

After M_2 : $P(C_1|M_2) = \dfrac{P(M_2|C_1)P(C_1)}{P(M_2)} = \dfrac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$

$$P(C_2|M_2) = \frac{P(M_2|C_2)P(C_2)}{P(M_2)} = 0$$

$$P(C_3|M_2) = \frac{P(M_2|C_3)P(C_3)}{P(M_2)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Outline

Setting of Classical Monty Hall Problem

- Strategies of the contestant: switching & no switching
- Monty knows the location of the car
- Homogeneous doors
- Three doors

Variant 1: New Strategy

In this scenario, when Monty has choice on picking the door, he picks door 2 with probability p, and door 3 with probability $1-p$. Now besides the strategies of switching and no switching, you (the contestant) have the third option on your strategy:

- You first pick the door 1.
- If the door 2 is opened, you do not switch.
- If the door 3 is opened, you switch.

What is the winning probability with this strategy?

Variant 2: Unknown Car Location for Monty

As before, Monty shows you three identical doors. One contains a car, the other two contain goats. You choose one of the doors but do not open it. This time, however, Monty does not know the location of the car. He randomly chooses one of the two doors different from your selection and opens it. The door turns out to conceal a goat. He now gives you the options either of sticking with your original door or switching to the other one. What should you do?

Variant 3: Partially Known Car Location for Monty

As before, you are shown three equally likely doors. You choose door one. Monty now points to door two but does not open it. Instead he merely tells you that it conceals a goat. You know that in those cases where the car really is behind door one, Monty chooses randomly between door two and door three. You also know that when the car is behind door two or door three, it is Monty's intention to identify the car's location, but that his assertions regarding the location of the car are only correct with probability p, What should you do now?

Variant 4: Heterogeneous Doors

Suppose the car is not placed randomly behind the three doors. Instead, the car is behind door one with probability p_1, behind door two with probability p_2, and behind door three with probability p_3. Here $p_1 + p_2 + p_3 = 1$ and $p_1 \geq p_2 \geq p_3 > 0$. You are to choose one of the three doors, after which Monty will open a door he knows to conceal a goat. Monty always chooses randomly from among his options in those cases where your initial choice is correct. What strategy should you follow?

Variant 5: Progressive Monty

This time we assume there are n identical doors, where n is an integer satisfying $n \geq 3$. One door conceals a car, the other $n-1$ doors conceal goats. You choose one of the doors at random but do not open it. Monty then opens a door he knows to conceal a goat, always choosing randomly among the available doors. At this point he gives you the option either of sticking with your original door or switching to one of the remaining doors. You make your decision. Monty now eliminates another goat-concealing door (at random) and once more gives you the choice either of sticking or switching. This process continues until only two doors remain in play. What strategy should you follow to maximize your chances of winning?

Variant 6: Two Players

As usual, we are presented with three doors. This time, however, there is a second player in the game. Player one chooses a door, and then player two chooses a different door. If both have chosen goats, then Monty eliminates one of the players at random. If one has chosen the car, then the other player is eliminated. The surviving player knows the other has been eliminated, but does not know the reason for the elimination. After eliminating a player, Monty then opens that player's door and gives the surviving player the options of switching or sticking. What should the player do?

Variant 7: Two Hosts

As before, we are confronted with three identical doors, one concealing a car, the other two concealing goats. We initially choose door one, and Monty then opens door three. This time we know that there are two different hosts who preside over the show, with a coin flip deciding who hosts the show on a given night. The two hosts do not make their decisions in the same way. Coin-Toss Monty chooses his door randomly when your initial choice conceals the car. Three-Obsessed Monty always opens door three when he has the option of doing so. Under these circumstances, is there an advantage to be gained from switching to door two?

Variant 8: Many Cars

As before, this time we still have n doors, but now there are $1 \leq j \leq n-2$ cars and $n-j$ goats. After making your initial choice, Monty opens one of the other doors at random. Should you switch?

Variant 9: Many Cars

As before, this time we still have n doors, but now there are $1 \leq j \leq n-2$ cars and $n-j$ goats. This time, however, after making your initial choice, Monty tells us that he will reveal a goat with probability p, and will reveal a car with probability $1-p$. The catch is that we must make our decision to stick or switch before knowing which of these possibilities will come to pass. What should we do?

Variant 10: Many Cars, Open Many Doors

As before, this time we still have n doors, but now there are $1 \leq j \leq n-2$ cars and $n-j$ goats. This time, however, after making your initial choice, Monty opens m doors at random, revealing k cars and $m-k$ goats. What should we do?

Three Prisoners Problem

Three prisoners, A, B and C, are in separate cells and sentenced to death. The governor has selected one of them at random to be pardoned. The warden knows which one is pardoned, but is not allowed to tell. Prisoner A begs the warden to let him know the identity of one of the others who is going to be executed. “If B is to be pardoned, give me C’s name. If C is to be pardoned, give me B’s name. And if I’m to be pardoned, flip a coin to decide whether to name B or C.”

Three Prisoners Problem

The warden tells A that B is to be executed. Prisoner A is pleased because he believes that his probability of surviving has gone up from $1/3$ to $1/2$, as it is now between him and C. Prisoner A secretly tells C the news, who is also pleased, because he reasons that A still has a chance of $1/3$ to be the pardoned one, but his chance has gone up to $2/3$. What is the correct answer?

Three Prisoners Problem

- Proposed in Martin Gardner's "Mathematical Games" column, Scientific American, 1959
- Mathematically equivalent to the Monty Hall problem
  - car: freedom
  - goat: death execution
  - Monty: warden
- Very few math problems that have been immortalized in verse

The Prisoner's Paradox Revisited (Richard Bedient)

Awaiting the dawn sat three prisoners wary
    A trio of brigands named Tom, Dick and Mary
Sunrise would signal the death knoll of two
    Just one would survive, the question was who.

Young Mary sat thinking and finally spoke
    To the jailer she said, "You may think this a joke.
But it seems that my odds of surviving 'til tea,
    Are clearly enough just one out of three.

The Prisoner's Paradox Revisited

But one of my cohorts must certainly go,
    Without question, that's something I already know.
Telling the name of one who is lost,
    Can't possibly help me. What could it cost?"

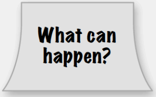
That shriveled old jailer himself was no dummy,
    He thought, "But why not?" and pointed to Tommy.
"Now it's just Dick and I," Mary chortled with glee.
    "One in two are my chances, and not one in three!"

Imagine the jailer's chagrin, that old elf.
    She'd tricked him, or had she? Decide for yourself.

Outline

Summary : Conditional Probability & Inference

Monday



events | **numbers**

A, B, C

We learn on Monday that B occurred.

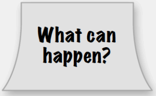
$\xrightarrow{P}$

$P(A)$
(Monday's prior for A)

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

(Monday's posterior for A)

Tuesday



events | **numbers**

A, B, C

We already know that B occurred.
We learn on Tuesday that C occurred.

$\xrightarrow{P}$

$P(A|B)$
(Tuesday's prior for A)

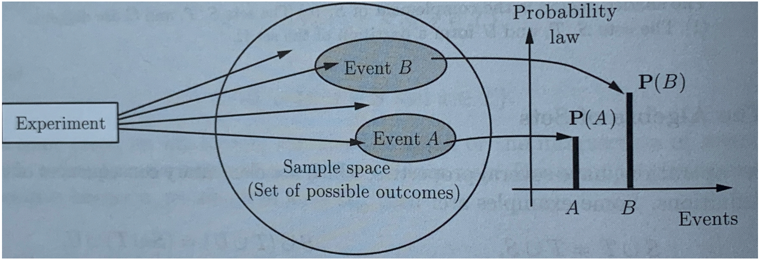
$$P(A|B,C) = \frac{P(A,B,C)}{P(B,C)} = \frac{P(C|A,B)P(A|B)}{P(C|B)}$$

(Tuesday's posterior for A)

References

- Chapter 2 of **BH**
- Chapter 1 of **BT**

Motivation for Next Lecture



Motivation for Next Lecture