# Probability and Random Process

COSE 382

# The importance of thinking conditionally

Conditional probability is essential for scientific, medical, and legal reasoning, since it shows how to incorporate evidence into our understanding of the world in a logical coherent manner. In fact, a useful perspective is that all probabilities are conditional; whether or not it's written explicitly, there is always background knowledge (or assumptions) built into every probability.

*Conditioning is the soul of statistics.*

# What a misuse of probability can do

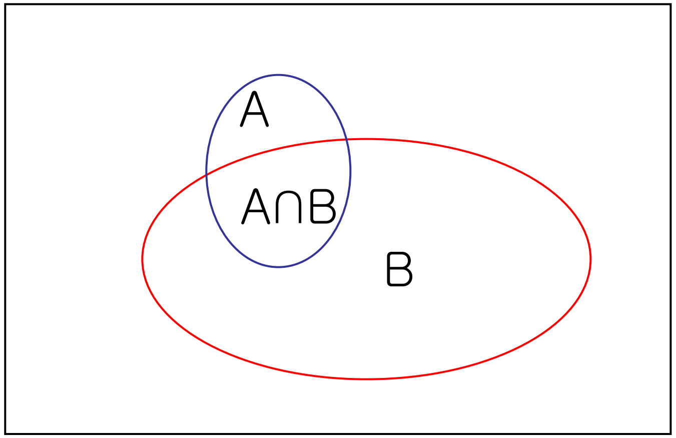
- In 1998, Sally Clark was tried for murder after two of her sons died shortly after birth.
  - During the trial, an expert witness for the prosecution testified that the probability of a newborn dying of sudden infant death syndrome (SIDS) was $1/8500$,
  - so the probability of two deaths due to SIDS in family was $(1/8500)^2$, or about one in 73 million.
  - Therefore, he continued, the probability of Clark's innocence was one in 73 million.
- Sadly, Clark was convicted of murder and sent to prison, partly based on the expert's wrongheaded testimony.
- She spent over three years in jail before her conviction was overturned.
- The outcry over the misuse of *conditional probability* in the Sally Clark case led to the review of hundreads of other cases where similar faulty logic was used by the prosecution.

# Conditional Probability

- **Definition 2.2.1** (Conditional probability)
  - If of $A$ and $B$ are events with $P(B) > 0$, then the *probability of $A$ given $B$*, denoted by $P(A|B)$, is defined as

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \text{ for } P(B) > 0$$

    Here $A$ is the event whose uncertainty we want to update, and $B$ is the evidence we observe.
  - $P(A)$ is called *prior* probability
  - $P(A|B)$ is called *posterior* proability



**Example 2.2.5**(elder is a girl vs. at least one girl)

A family has two children.

- What is the probability that both are girls, given that at least one is a girl ?
- What if it is known that the elder child is a girl ?
- What is the probability that both are girls, given that you randomly met one of the two and see that she is a girl ?

You may assume that each child is equally likely to be a boy or a girl, independently.

Sample space is $\{GG, GB, BG, BB\}$

$$P(\text{both girls}|\text{at least one girl}) = \frac{P(\text{both girls, at lesat one girl})}{P(\text{at least one girl})} = \frac{1/4}{3/4} = 1/3$$

$$P(\text{both girls}|\text{elder is a girl}) = \frac{P(\text{both girls, elder is a girl})}{P(\text{elder is a girl})} = \frac{1/4}{1/2} = 1/2$$

$$P(\text{both girls}|\text{ranom child is a girl}) = \frac{P(\text{both girls, random child is a girl})}{P(\text{random child is a girl})} = \frac{1/4}{1/2} = 1/2$$

# Bayes Rule

**Theorem 2.3.1**

For any events $A$ and $B$ with positive probabilities

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

**Theorem 2.3.2**

For any events $A_1, \cdots, A_n$ with positive probabilities.

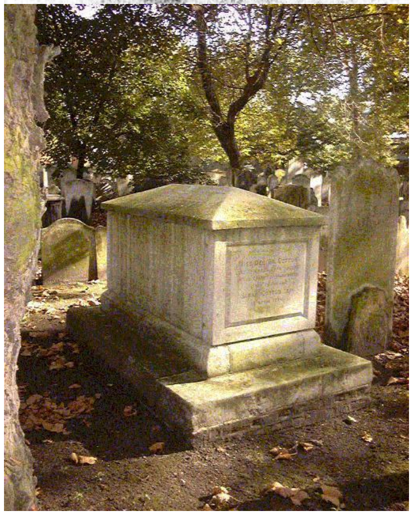
$$P(A, A_2, \cdots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \cdots, A_{n-1})$$

The commas denote intersections $((A, B) = A \cap B)$.

**Theorem 2.3.3** (Bayes' rule)

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

- Thomas Bayes (1702-1761)



*English theologian and mathematician who was the first to use probability inductively and who established a mathematical basis for probability inference (a means of calculating, from the frequency with which an event has occurred in prior trials, the probability that it will occur in future.*

Bunhill Fields Cemetery
(the cemetery of Puritan England),
City of London

# Law of Total Probability (LOTP)

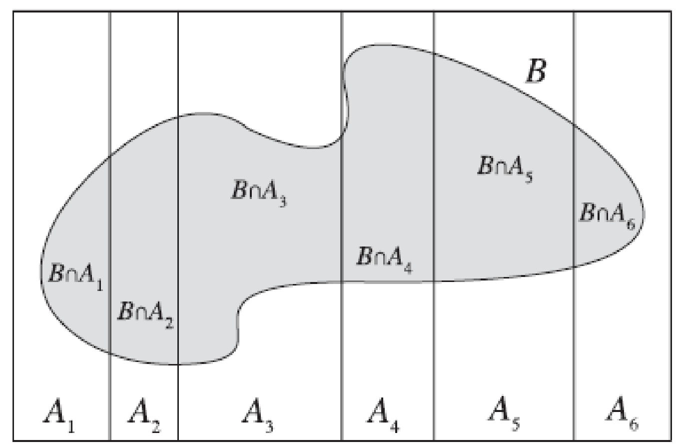
**Theorem 2.3.3** (Law of total probability (LOTP))

Let $A_1, \cdots, A_n$ be ta partition of the sample space $S$, i.e. $S = \dot{\bigcup}_{i=1}^{n} A_i$, then

$$P(B) = \sum P(B|A_i)P(A_i)$$

Proof: From

$$\begin{aligned} B &= B \cap S = B \cap \left(\dot{\bigcup}_{k=1}^{n} A_k\right) = \dot{\bigcup}_{k=1}^{n} (B \cap A_k) \\ P(B) &= \sum_{k=1}^{n} P(B \cap A_k) = \sum_{k=1}^{n} P(B|A_k)P(A_k) \end{aligned}$$



# Random Coin Example

**Fair and biased Coin.**

- There are two coins:
  - one is fair coin $P(H) = 1/2$ and
  - the other one is biased $P(H) = 1/3$.
- We randomly selct one coin and toss it once.
- What is the probability that we see a Head ?

*Solution:*

Let $F = \{\text{Fair coin is picked}\}$ and
$B = \{\text{Biased coin is picked}\}$

$$P(H) = P(H|F)P(F) + P(H|B)P(B)$$
$$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{12}$$

**Fair and biased Coin (Bayes).**

- There are two coins:
  - one is fair coin $P(H) = 1/2$ and
  - the other one is biased $P(H) = 1/3$.
- We randomly selct one coin and toss it once.
- We have observed $H$
- What is the probability that we picked a fair coin ?

*Solution:*

$$P(F|H) = \frac{P(H|F)P(F)}{P(H)}$$
$$= \frac{P(H|F)P(F)}{P(H|F)P(F) + P(H|B)P(B)}$$
$$= \frac{1}{2} \cdot \frac{1}{2} / \frac{5}{12} = \frac{3}{5}$$

**Example 2.3.9**(Testing for a rare disease)

Fred is tested for disease that afflicts 1% of population.

The test result is positive.

The test is 95% accurate, i.e. $P(T|D) = P(T^c|D^c) = 0.95$, where $D$ is the event that Fred has the disease and $T$ is the event that the test result is positive

- What is the probability that Fred has the disease given that the test is positive?

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\ &= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \\ &\approx 0.16 \end{aligned}$$

# Conditional Probabilities are Probabilities

For a given event $E$ with $P(E) > 0$,

- $P(\cdot|E)$ is a probability function, i.e.

$$\begin{aligned} P(S|E) &= 1 \\ P\left(\dot{\bigcup}_{i=1}^{n} A_i|E\right) &= \sum_{i=1}^{n} P(A_i|E) \end{aligned}$$

Hence, we have

$$\begin{aligned} P(A^c|E) &= 1 - P(A|E) \\ P(\phi|E) &= 0 \\ P(A \cup B|E) &= P(A|E) + P(B|E) - P(A \cap B|E) \end{aligned}$$

- $P(E|\cdot)$ is not a probability function, but called a *likelihood function*
  - $P(E|\phi)$ cannot be defined and, furthermore,
  - $P(E|S) = P(E) \neq 1$

Likelihood function is a useful tool in interference

# Extra Conditioning

**Theorem 2.4.2**(Bayes' rule with extra conditioning)

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

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

**Proof:** Three ways to interpret extra conditioning

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

where $(A,B)$ denotes $(A \cap B)$

**Theorem 2.4.4**(LOTP with extra conditioning)

For a partition of $S$, $A_1, \cdots A_n$ with $P(A_i \cap E) > 0$,

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

**Example 2.4.4**(Random Coin, continued)

You have one fair coin ($P(H) = 1/2$), and one biased coin with $P(H) = 3/4$

You pick one randomly and flip it three times: 3 Heads

Given this information, what is the probability that the coin will land Head one more if we toss it a fourth time ?

Let $A$ be the event that the chosen coin lands Head three times.

Let $F$ be the event that we picked the fair coin.

Let $H$ be the event that the coin lands Head on the fourth toss.

$$\begin{aligned} P(H|A) &= P(H|F,A)P(F|A) + P(H|F^c,A)P(F^c|A) = \frac{1}{2} \cdot 0.23 + \frac{3}{4}(1 - 0.23) \\ &\approx 0.69 \end{aligned}$$

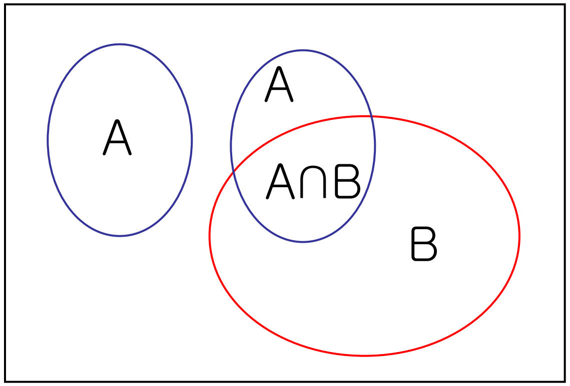
# Independence

**Definition 2.5.1** (Independence of two events) Events $A$ and $B$ are *independent* if

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

- For independent events,

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




- Mutually exclusive events ($A \cap B = \phi$) are not independent
  - Mutually exclusive: $P(A|B) = 0$
  - Equal: $P(A|B) = 1$
  - Independent: $P(A|B) = P(A)$

**Proposition 2.5.3** If $A$ and $B$ are independent, so do $A$ and $B^c$, $A^c$, and $B$, and $A^c$ and $B^c$.

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

**Definition 2.5.6** (Independence of many events) For $n$ events $A_1, A_2, \cdots, A_n$ to be *independent* , for any sub-index set $\{i_1, \cdots i_k\}$ of $\{1, \cdots, n\}$

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

**Definition 2.5.7** (Conditional independence) Events $A$ and $B$ are said to be *conditionally independence* given $E$ if

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

– Pairwise independence doesn't imply independence (example 2.5.5)

– Conditional independence doesn't imply independence (example 2.5.9)

– Independence doesn't imply conditional independence (example 2.5.10)

**Example 2.7.1** (Monty Hall) On the game show "Let's Make a Deal", hosted by Monty Hall,

1) a contestant choose one of three doors; two have a goat and one has a car.

2) Monty, who knows where the car is, opens one of two remaining doors which has a goat.

3) Monty then offers the contestant the option of switching doors to open

- If the goal of contestant is to get the car, should he stay or switch ?

  Assume the contestant choose door 1. Let $C_i$ be the event that the car is behind door $i$. By the LOTP

$$P(\text{get car}|\text{state}) = \sum_{i=1}^{3} P(\text{get car}|C_i, \text{state}) \cdot P(C_i|\text{state})$$

  – stay:

$$P(\text{get car}|\text{stay}) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{3}$$

  – switch:

$$P(\text{get car}|\text{switch}) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$$


**Example 2.7.2** (Branching process) A single amoeba, Bobo, lives in a pond. After one minutes Bobo will either die, split into two amoebas, or stay the same, with equal probability, and in subsequent minutes all living amoeba will behave the same way, independently. What is the probability that the amoeba population will eventually die out ?

Let $D$ be the event that the population eventually dies out; we want to find $P(D)$

Let $B_i$ be the event that Bobo turns into $i$ amoebas after the first minute, for $i = 0, 1, 2$.

- $P(D|B_0) = 1$
- $P(D|B_1) = P(D)$
- $P(D|B_2) = P(D)^2$

By LOTP,

$$\begin{aligned} P(D) &= P(D|B_0) \cdot \frac{1}{3} + P(D|B_1) \cdot \frac{1}{3} + P(D|B_2) \cdot \frac{1}{3} \\ &= \frac{1}{3} + \frac{1}{3}P(D) + \frac{1}{3}P(D)^2 \end{aligned}$$

we have $P(D)^2 - 2P(D) + 1 = 0$ and, hence, $P(D) = 1$

# Pitfalls and Paradoxes

☣**2.8.1** (Prosecutor's fallacy) In 1998, Sally Clark was tried for murder after two of her sons died shortly after birth. During the trial, an expert witness for the prosecution testified that the probability of a newborn dying of sudden infant death syndrome (SIDS) was 1/8500, so the probability of two deaths due to SIDS in family was $(1/8500)^2$, or about one in 73 million. Therefore, he continued, the probability of Clark's innocence ws one in 73 million.

First, SIDS may not independent in a certain family.

Second, the expert has confused $P(\text{innocence}|\text{evidence})$ and $P(\text{evidence}|\text{innocence})$. By Bayes' rule

$$\begin{aligned}
P(\text{innocence}|\text{evidence}) &= \frac{P(\text{evidence}|\text{innocence})P(\text{innocence})}{P(\text{evidence})} \\
&= \frac{P(\text{evidence}|\text{innocence})P(\text{innocence})}{P(\text{evidence}|\text{innocence})P(\text{innocence}) + P(\text{evidence}|\text{guilt})P(\text{guilt})} \\
&= \frac{P(\text{evidence}|\text{innocence})(1-\epsilon)}{P(\text{evidence}|\text{innocence})(1-\epsilon) + P(\text{evidence}|\text{guilt})\epsilon}
\end{aligned}$$

where we set $\epsilon = P(\text{guilt})$. When $\epsilon \ll P(\text{evidence}|\text{innocence})$, $P(\text{innocence}|\text{evidence}) \to 1$.