Conditional Probability & Independence

Conditional Probabilities

• **Question**: How should we modify $P(E)$ if we learn that event $F$ has occurred?

• **Definition**: the conditional probability of $E$ given $F$ is

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}, \quad \text{for} \quad P(F) > 0$$

Condition probabilities are useful because:

• Often want to calculate probabilities when some partial information about the result of the probabilistic experiment is available.

• Conditional probabilities are useful for computing ”regular” probabilities.
Example 1. 2 random cards are selected from a deck of cards.

(a) What is the probability that both cards are aces given that one of the cards is the ace of spaces?

(b) What is the probability that both cards are aces given that at least one of the cards is an ace?

\[
\mathcal{N} = \{ \text{all unordered pairs of cards} \} \quad \text{uniform prob dist'n}
\]

\[
\begin{align*}
(a) \quad \Pr(\text{AA} \mid \text{AQ}) &= \frac{\Pr(\text{AQ and another ace})}{\Pr(\text{AQ})} \\
&= \frac{3}{51} \approx 0.059
\end{align*}
\]

\[
\begin{align*}
(b) \quad \Pr(\text{AA} \mid \geq 1 \text{ A}) &= \frac{\Pr(\text{AA})}{\Pr(\geq 1 \text{ A})} \\
&= \frac{\binom{4}{2}}{\binom{52}{2} - \binom{48}{2}} = \frac{4 \cdot 3}{52 \cdot 51 - 48 \cdot 47} \approx 0.03
\end{align*}
\]
Example 2. Deal a 5 card poker hand, and let
\( E = \{ \text{at least 2 aces} \} \), \( F = \{ \text{at least 1 ace} \} \),
\( G = \{ \text{hand contains ace of spades} \} \).
(a) Find \( P(E) \)

\[
P(E) = 1 - \frac{\binom{48}{5}}{\binom{52}{5}} - \frac{\binom{4}{4}}{\binom{5}{5}}
\]

(b) Find \( P(E | F) \)

\[
\frac{P(E \cap F)}{P(F)} = \frac{P(E)}{P(F)}
\]

\[
P(F) = 1 - \frac{\binom{48}{5}}{\binom{52}{5}}
\]

(c) Find \( P(E | G) \)

\[
\frac{P(E \cap G)}{P(G)} = \frac{P(A \cap G + \text{at least 1 more ace})}{P(\text{contains A at least one ace})}
\]

\[
= \frac{\binom{51}{4} - \binom{48}{4}}{\binom{51}{4}}
\]
\[ \text{Proof:} \quad \Pr(E | F) = \frac{\Pr(ENF) + \Pr(ENF^c)}{\Pr(F)} \]

\[ = \frac{\Pr(F \cap E) + \Pr(F \cap E^c)}{\Pr(F)} \]

\[ = \frac{\Pr(F \cap E)}{\Pr(F)} = 1 \]
Cond prob satisfies the usual prob axioms.

Suppose \((\mathcal{S}, P(\cdot))\) is a probability space.

Then \((\mathcal{S}, P(\cdot | F))\) is also a probability space (for \(F \subset \mathcal{S}\) with \(P(F) > 0\)).

- \(0 \leq P(\omega | F) \leq 1\)
- \(\sum_{\omega \in \mathcal{S}} P(\omega | F) = 1\)

- If \(E_1, E_2, \ldots\) are disjoint, then
  \[
  P(\bigcup_{i=1}^{\infty} E_i | F) = \sum_{i=1}^{\infty} P(E_i | F)
  \]

Thus all our previous propositions for probabilities give analogous results for conditional probabilities.

Examples
\[
P(E^c | F) = 1 - P(E | F)
\]
\[
P(A \cup B | F) = P(A | F) + P(B | F) - P(A \cap B | F)
\]

However,
\[
\Pr(E | F) + \Pr(E | F^c) \text{ not necessarily } 1!
\]
The Multiplication Rule

• Re-arranging the conditional probability formula gives

\[ P(E \cap F) = P(F)P(E | F) \]

This is often useful in computing the probability of the intersection of events.

Example. Draw 2 balls at random without replacement from an urn with 8 red balls and 4 white balls. Find the chance that both are red.

\[ \begin{array}{c|c}
8 & R \\
4 & W \\
\end{array} \]

draw 2 balls without replacement

\[ \Pr(\text{both } R) = \Pr(\text{first } R) \Pr(\text{2nd } R | 1\text{st } R) \]

\[ = \frac{4}{12} \cdot \frac{7}{11} \]
The General Multiplication Rule

\[ P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \times P(E_2 \mid E_1) \times P(E_3 \mid E_1 \cap E_2) \times \cdots \times P(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \]

Example 1. Alice and Bob play a game as follows. A die is thrown, and each time it is thrown it is equally likely to show any of the 6 numbers. If it shows 5, A wins. If it shows 1, 2 or 6, B wins. Otherwise, they play a second round, and so on. Find \( P(A_n) \), for \( A_n = \{ \text{Alice wins on nth round} \} \).

\[ \text{Ni: event that nobody wins on ith round} \]
\[ \Pr(A_n) = \Pr(N_1 \cap N_2 \cap \cdots \cap N_{n-1} \cap A_n) \]
\[ = \Pr(N_1) \Pr(N_2 \mid N_1) \Pr(N_3 \mid N_1 \cap N_2) \cdots \Pr(N_{n-1} \mid N_1 \cdot N_{n-2}) \Pr(A_n \mid N_1 \cdot N_{n-1}) \]
\[ = \left( \frac{2}{6} \right)^{n-1} \cdot \frac{1}{6} \]
Example 2. I have $n$ keys, one of which opens a lock. Trying keys at random without replacement, find the chance that the $k$th try opens the lock.

$E_i$: event that $i$th key opens lock

$$\Pr(E_k) = \Pr(E_1 \cap E_2 \cap \cdots \cap E_{k-1} \cap E_k)$$

$$= \Pr(E_1) \Pr(E_2 \mid E_1) \Pr(E_3 \mid E_1 E_2) \cdots \Pr(E_{k-1} \mid E_1 \cdots E_{k-2}) \Pr(E_k \mid E_1 \cdots E_{k-1})$$

$$= \frac{n-1}{n} \frac{n-2}{n-1} \cdots \frac{n-(k-1)}{n-k+2} \frac{1}{n-k+1}$$

$$= \frac{n-k+2}{n} \cdots \frac{n-2}{n} \frac{1}{n-k+1}$$

$$= \frac{1}{n}$$
The Law of Total Probability

- We know that $P(E) = P(E \cap F) + P(E \cap F^c)$.
Using the definition of conditional probability,

$$P(E) = P(E | F) P(F) + P(E | F^c) P(F^c)$$

- This is extremely useful. It may be difficult to compute $P(E)$ directly, but easy to compute it once we know whether or not $F$ has occurred.

- To generalize, say events $F_1, \ldots, F_n$ form a partition if they are disjoint and $\bigcup_{i=1}^n F_i = S$.

- Since $E \cap F_1, E \cap F_2, \ldots, E \cap F_n$ are a disjoint partition of $E$, $P(E) = \sum_{i=1}^n P(E \cap F_i)$.

- Apply conditional probability to give the law of total probability,

$$P(E) = \sum_{i=1}^n P(E | F_i) P(F_i)$$
Example 1. Eric’s girlfriend comes round on a given evening with probability 0.4. If she does not come round, the chance Eric watches *The Wire* is 0.8. If she does, this chance drops to 0.3. Find the probability that Eric gets to watch *The Wire*.

\[
\Pr(\text{watched}) = \Pr(\text{watched} | \text{GF}) \Pr(\text{GF}) + \Pr(\text{watched} | \overline{\text{GF}}) \Pr(\overline{\text{GF}})
\]

\[
= 0.3 \cdot 0.4 + 0.6 \cdot 0.6
\]

\[
= 0.48
\]
Bayes Formula

- Sometimes $P(E | F)$ may be specified and we would like to find $P(F | E)$.

**Example 2.** I call Eric and he says he is watching *The Wire*. What is the chance his girlfriend is around?

- A simple manipulation gives **Bayes’ formula**,

\[ P(F | E) = \frac{P(E | F) P(F)}{P(F)} \]

Since $P(E | F) P(F) = P(E \cap F)$

- Combining this with the law of total probability,

\[ P(F | E) = \frac{P(E | F) P(F)}{P(E | F) P(F) + P(E | F^c) P(F^c)} \]

Since denominator = $P(E)$ by law of total probability

\[ \Pr(GF | \text{Watching}) = \frac{\Pr(\text{watches} | GF) \Pr(GF)}{\Pr(\text{watches})} \]

\[ = \frac{0.3 \cdot 0.4}{0.48} = \frac{1}{4} \]
Sometimes conditional probability calculations can give quite unintuitive results.

**Example 3.** I have three cards. One is red on both sides, another is red on one side and black on the other, the third is black on both sides. I shuffle the cards and put one on the table, so you can see that the upper side is red. What is the chance that the other side is black?

- is it $1/2$, or $> 1/2$ or $< 1/2$?

**Solution**

**prob model 1:**

pick random card
put R side up if it has a red side

$$\Pr(RB|\text{see } R) = \frac{\Pr(RB \cap \text{see } R)}{\Pr(\text{see } R)} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

**prob model 2:**

pick random card
pick random side to show

$$\Pr(RB|\text{see } R) = \frac{\Pr(RB \cap \text{see } R)}{\Pr(\text{see } R)} = \frac{\Pr(\text{see } R | RB) \Pr(RB)}{\Pr(\text{see } R | RB) \Pr(RB) + \Pr(\text{see } R | RR) \Pr(RR) + \Pr(\text{see } R | BB) \Pr(BB)}$$

$$= \frac{\frac{1}{3} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{1}{3}$$
Example: Spam Filtering

- 60% of email is spam.
- 10% of spam has the word "Viagra".
- 1% of non-spam has the word "Viagra".
- Let $V$ be the event that a message contains the word "Viagra".
- Let $J$ be the event that the message is spam.

What is the probability of $J$ given $V$?

**Solution.**

$$Pr(J|V) = \frac{Pr(V|spam)Pr(spam)}{Pr(V|spam)Pr(spam) + Pr(V|not\,spam)Pr(not\,spam)}$$

$$= \frac{0.1 \cdot 0.6}{0.1 \cdot 0.6 + 0.01 \cdot 0.4}$$
Discussion problem. Suppose 99% of people with HIV test positive, 95% of people without HIV test negative, and 0.1% of people have HIV. What is the chance that someone testing positive has HIV?

$$Pr(HIV^+ | Test^+) = \frac{Pr(Test^+ | HIV^+) Pr(HIV^+)}{Pr(Test^+ | HIV^+) Pr(HIV^+) + Pr(Test^+ | HIV^-) Pr(HIV^-)}$$

$$= \frac{.99 \cdot .001}{.99 \cdot .001 + .05 \cdot .999} = 0.019 \approx 2\%$$

What if people who get tested have a 10% chance of being HIV^+? 

$$\sim 68\%$$

50% chance of being HIV^+ 

$$\sim 95\%$$
Example: Statistical inference via Bayes’ formula

Alice and Bob play a game where Alice tosses a coin, and wins $1 if it lands on H or loses $1 on T. Bob is surprised to find that he loses the first ten times they play. If Bob’s prior belief is that the chance of Alice having a two headed coin is 0.01, what is his posterior belief?

Note. Prior and posterior beliefs are assessments of probability before and after seeing an outcome. The outcome is called data or evidence.

Solution.

\[
\Pr(2\text{-headed} | \text{loses 10x}) = \frac{\Pr(2 \text{-headed} \& \text{loses 10x})}{\Pr(\text{loses 10x})}
\]

\[
= \frac{\Pr(\text{loses 10x} | 2 \text{-headed}) \Pr(2 \text{-headed})}{\Pr(\text{loses 10x} | 2 \text{-headed}) \Pr(2 \text{-headed}) + \Pr(\text{loses 10x} | \text{regular}) \Pr(\text{reg})}
\]

\[
= \frac{0.01}{0.01 + (\frac{1}{2})^{10} \cdot 0.99}
\]
Example: A plane is missing, and it is equally likely to have gone down in any of three possible regions. Let $\alpha_i$ be the probability that the plane will be found in region $i$ given that it is actually there. What is the conditional probability that the plane is in the second region, given that a search of the first region is unsuccessful?

$$
\frac{\Pr(\text{in 2$^\text{nd}$ | \text{search of first failed})}{\Pr(\text{in 2$^\text{nd}$ & search 1$^\text{st}$ failed})} = \frac{\Pr(\text{search of 1$^\text{st}$ failed} | \text{in 2$^\text{nd}$})}{\frac{1}{3} \cdot \alpha_1 + \frac{2}{3}} = \frac{\frac{1}{3}}{\frac{1}{3} \cdot \alpha_1 + \frac{2}{3}}
$$
Independence

- Intuitively, $E$ is independent of $F$ if the chance of $E$ occurring is not affected by whether $F$ occurs. Formally,
  \[ P(E \mid F) = P(E) \] (1)

- We say that $E$ and $F$ are **independent** if
  \[ P(E \cap F) = P(E)P(F) \] (2)

**Note.** (2) and (1) are equivalent.

**Note 1.** It is clear from (2) that independence is a symmetric relationship. Also, (2) is properly defined when $P(F) = 0$.

**Note 2.** (1) gives a useful way to think about independence; (2) is usually better to do the math.
Proposition. If $E$ and $F$ are independent, then so are $E$ and $F^c$.

Proof.

\[
\begin{align*}
\Pr(F | E) &= \Pr(F) \\
\Pr(F^c | E) &= 1 - \Pr(F | E) \\
&= 1 - \Pr(F) \\
&= \Pr(F^c)
\end{align*}
\]
Example 1: Independence can be obvious

Draw a card from a shuffled deck of 52 cards. Let 
$E =$ card is a spade and $F =$ card is an ace. Are 
$E$ and $F$ independent?

Solution

$$
\Pr(E) = \frac{1}{4}, \quad \Pr(F) = \frac{1}{13}, \quad \Pr(E \cap F) = \frac{1}{52}, \quad \checkmark
$$

Example 2: Independence can be surprising

Toss a coin 3 times. Define

$A =$ {at most one T} = \{HHH, HHT, HTH, THH\}

$B =$ {both H and T occur} = \{HHH, TTT\}.

Are $A$ and $B$ independent?

Solution

$$
\Pr(A) = \frac{4}{8} = \frac{1}{2}
$$

$$
\Pr(B) = \frac{6}{8} = \frac{3}{4}
$$

$$
\Pr(A \cap B) = \Pr(\{HHT, HTH, THH\})
$$

$$
= \frac{3}{8} = \Pr(A) \cdot \Pr(B)
$$
Independence as an Assumption

• It is often convenient to suppose independence. People sometimes assume it without noticing.

Example. A sky diver has two chutes. Let

\[ E = \{ \text{main chute opens} \}, \quad P(E) = 0.98; \]
\[ F = \{ \text{backup opens} \}, \quad P(F) = 0.90. \]

Find the chance that at least one opens, making any necessary assumption clear.

\[
1 - \Pr(\overline{E} \cap \overline{F}) = 1 - 0.02 \cdot 0.1 = 0.998
\]

Note. Assuming independence does not justify the assumption! Both chutes could fail because of the same rare event, such as freezing rain.
Independence of Several Events

- Three events $E$, $F$, $G$ are independent if
  
  \[ P(E \cap F) = P(E) \cdot P(F) \]
  \[ P(F \cap G) = P(F) \cdot P(G) \]
  \[ P(E \cap G) = P(E) \cdot P(G) \]
  \[ P(E \cap F \cap G) = P(E) \cdot P(F) \cdot P(G) \]

- If $E$, $F$, $G$ are independent, then $E$ will be independent of any event formed from $F$ and $G$.

Example. Show that $E$ is independent of $F \cup G$.

Proof.

\[
Pr(F \cup G \mid E) = Pr(F \mid E) + Pr(G \mid E) - Pr(F \cap G \mid E)
\]
\[
= Pr(F) + Pr(G) - Pr(F \cap G)
\]
\[
= Pr(F \cup G)
\]
Pairwise Independence

- $E, F$ and $G$ are **pairwise independent** if $E$ is independent of $F$, $F$ is independent of $G$, and $E$ is independent of $G$.

**Example.** Toss a coin twice. Set $E = \{HH, HT\}$, $F = \{TH, HH\}$ and $G = \{HH, TT\}$.

(a) Show that $E, F$ and $G$ are pairwise independent.

$$\Pr(E \cap F) = \frac{1}{4} = \Pr(E) \Pr(F)$$

(b) By considering $\Pr(E \cap F \cap G)$, show that $E, F$ and $G$ are NOT independent.

$$\Pr(E \cap F \cap G) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3$$

**Note.** Another way to see the dependence is that $\Pr(E | F \cap G) = 1 \neq \Pr(E)$. 

Example: Insurance policies

Insurance companies categorize people into two groups: accident prone (30%) or not. An accident prone person will have an accident within one year with probability 0.4; otherwise, 0.2. What is the conditional probability that a new policyholder will have an accident in his second year, given that the policyholder has had an accident in the first year?

\[
\Pr(\text{acc in 2\textsuperscript{nd} year} | \text{acc in first})
\]

\[
= \frac{\Pr(\text{acc in both})}{\Pr(\text{acc in first})}
\]

\[
= \frac{\Pr(\text{acc in both} | \text{AP}) \Pr(\text{AP}) + \Pr(\text{acc in both} | \overline{\text{AP}}) \Pr(\overline{\text{AP}})}{\Pr(\text{acc in first} | \text{AP}) \Pr(\text{AP}) + \Pr(\text{acc in first} | \overline{\text{AP}}) \Pr(\overline{\text{AP}})}
\]

\[
= \frac{(0.4)^2 \cdot 0.3 + (0.2)^2 \cdot 0.7}{0.4 \cdot 0.3 + 0.2 \cdot 0.7}
\]

Note that we are assuming that the event of a person having an accident this year is independent of the event of having an accident the following year.
Note: We can study a probabilistic model and determine if certain events are independent or we can define our probabilistic model via independence.

Example: Suppose a biased coin comes up heads with probability \( p \), independent of other flips.

\[
P(\text{n heads in n flips}) = p^n
\]

\[
P(\text{n tails in n flips}) = (1 - p)^n
\]

\[
P(\text{exactly k heads n flips}) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

\[
P(\text{HHTHTTT}) = p^2 (1 - p) p (1 - p)^3 = p^2 H (1 - p) T
\]