Reminders

- Axiomatic theory of probability
- The set of all possible outcomes of an experiment is the sample space \( \Omega \)
- Events are subsets of \( \Omega \)
- An event is said to have occurred if the outcome of the experiment is a member of the event (that is, subset of \( \Omega \))
- \( A \) and \( A^c \) are a partition of \( \Omega \)
- On every trial, one of \( A \) and \( A^c \) must occur

20-20 hindsight is a wonderful thing

- Once the experiment has been performed and the outcome is known, we have perfect and complete knowledge
- For each pair of events \( A \) and \( A^c \), we can tell which one occurred and which one didn’t — just check which set \( A \) or \( A^c \) the observed outcome belongs to!
- There is no probabilistic consideration any more, and we do not need to think about the chances of \( A \) (or \( A^c \)) occurring

A little learning is a dangerous thing...

- Now suppose that the experiment has been performed, but we do not know the outcome exactly
- All we know is that the outcome is some member of the event \( A \), but we do not know which member of \( A \) it is
- Put another way, we are told that the event \( A \) has occurred, but nothing else
- To avoid trivial cases, assume that \( A \) is not a singleton or elementary event

... B or \( \sim B \)? That is the question...

- The experiment has been performed and we know that the event \( A \) has occurred, that is, the outcome is some member of \( A \)
- Did the event \( B \) occur? or did \( B^c \) occur?
- Unlike the case of perfect knowledge, we cannot tell whether \( B \) or \( B^c \) occurred
- \( AB \) and \( AB^c \) are a partition of \( A \)
- If the outcome \( \in AB \), then \( B \) occurred; if the outcome \( \in AB^c \), then \( B^c \) occurred

The exorcism did not work...

- The experiment has been performed and we know that the event \( A \) occurred, that is, the outcome is some member of \( A \)
- We cannot tell for sure whether \( B \) occurred
- We have not exercised the probability from the problem as yet — A probability question still continues to plague us
- Question: What are the chances that \( B \) occurred? in view of the new knowledge that event \( A \) is known to have occurred?

A mind like a steel trap...

- Question: What are the chances that \( B \) occurred? in view of the new knowledge that event \( A \) is known to have occurred?
- One (stupid?) answer to this question is that the chances that \( B \) occurred are still what they always were, viz. \( P(B) \)
- “Don’t bother me with facts; my mind is made up!”
- This ostrich-like approach is wrong, at least in some cases
Refusing to face the facts…

- Question: What are the chances that B occurred? In view of the new knowledge that event A is known to have occurred?
  - Left diagram: AB = ∅. Obviously, if A occurred, B cannot have occurred.
  - Right diagram: AB = A. Obviously, if A occurred, B must also have occurred.

Swaying every which way...

- More generally, even when the special cases AB = ∅ or AB = A do not hold, it is still reasonable to change the value of P(B) to reflect the information obtained from knowing that A occurred.
- Probabilities are beliefs, and we can and should revise them as we grow wiser.
  - Growing older is mandatory; growing wiser is optional.

What’s in a name?

- The (original) value of P(B) is called the unconditional or a priori probability of B.
- Here a priori means “before the fact” or prior to the experiment being performed.
- Given that A occurred, the revised chances of B occurring are called the conditional probability of B given A and denoted by P(B|A).
- Read this as “probability of B given A”.

Definition of conditional probability

- The conditional probability of B given A is denoted by P(B|A).
- Read this as “the probability of B given A” or “the probability of B conditioned on A”.
- A is called the conditioning event.
- Definition: If P(A) > 0, P(B|A) is defined as
  \[ P(B|A) = \frac{P(AB)}{P(A)} \]

Special cases

- Left diagram: AB = ∅. Obviously, if A occurred, B cannot have occurred.
  - P(B|A) = P(AB)/P(A) = 0
- Right diagram: AB = A. Obviously, if A occurred, B must also have occurred.
  - P(B|A) = P(AB)/P(A) = P(A)/P(A) = 1

An example

- Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 6 given that the dice are showing different faces?
- Critique of statement: Problems on conditional probability are often stated in the careless manner above. A conditional probability is being asked for, but the word conditional has been left out!
Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 6 given that the dice are showing different faces?

Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 6 given that the dice are showing different faces?

Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 7 given that the dice are showing different faces?

Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 7 given that the dice are showing different faces?

Example (continued):

Example (changed a bit):

Example (changed a bit, continued):

- Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 6 given that the dice are showing different faces?
  - \( P(A) = \frac{30}{36} = \frac{5}{6} \)
  - \( P(B) = \frac{5}{36} \)
  - \( AB = \{(1,5), (2,4), (4,2), (5,1)\} \)
  - \( P(AB) = \frac{4}{36} \)
  - \( P(B|A) = \frac{P(AB)}{P(A)} = \frac{4/36}{30/36} = \frac{4}{30} = \frac{2}{15} < P(B) \)

- Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 6 given that the dice are showing different faces?
  - \( A = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\} \)
  - \( B = \{(1,5), (2,4), (3,3), (4,2), (5,1)\} \)
  - \( AB = \{(1,5), (2,4), (4,2), (5,1)\} \)
  - \( P(A) = \frac{30}{36} = \frac{5}{6} \)
  - \( P(B) = \frac{5}{36} \)
  - \( P(AB) = \frac{4}{36} \)
  - \( P(B|A) = \frac{P(AB)}{P(A)} = \frac{4}{30} = \frac{2}{15} < P(B) \)

- Example: Two fair dice are rolled. What is the probability that the sum of the two faces is 7 given that the dice are showing different faces?
  - \( A = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\} \)
  - \( B = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \)
  - \( C = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \)
  - \( P(C) = \frac{6}{36} = \frac{1}{6} < P(B) \)
  - \( AC = C \)
  - \( P(C|A) = \frac{P(AC)}{P(A)} = \frac{6/36}{30/36} = \frac{1}{5} < P(C) \)
Example: Two fair dice are rolled. What is the probability that the first die is showing a 6 given that the dice are showing different faces?

- \( A = \{ \text{dice are showing different faces} \} \)
- \( P(A) = \frac{30}{36} = \frac{5}{6} \)
- \( D = \{ \text{first die shows a 6} \} = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\} \)
- \( P(D) = \frac{6}{36} \)
- \( AD = \{(6,1), (6,2), (6,3), (6,4), (6,5)\} \)

\[
P(D | A) = \frac{P(AD)}{P(A)} = \frac{\frac{5}{36}}{\frac{30}{36}} = \frac{1}{6} = P(D)
\]

So, did we learn anything?

- The three different problems comprising the example showed that a conditional probability can be smaller, larger, or the same as, the unconditional probability
- They did?
  - \( P(B | A) = \frac{2}{15} < P(B) = \frac{5}{36} \)
  - \( P(C | A) = \frac{1}{5} > P(C) = \frac{1}{6} \)
  - \( P(D | A) = \frac{1}{6} = P(D) = \frac{1}{6} \)

Classical approach

- The definition of conditional probability is motivated by considerations arising from the classical probability viewpoint
- If \( |\Omega| = n \) and \( |A| = k \), then the reduced sample space consists of just the set \( A \) viewed as a sample space
- Outcomes in \( A \) have prob. \( \frac{1}{k} \), not \( \frac{1}{n} \)
- What is the probability of \( B \) in this reduced sample space?

Reduced Sample Spaces

- Each outcome in \( A \) has probability \( \frac{1}{k} \) (instead of \( \frac{1}{n} \))
- What is the probability of \( B \) in this reduced sample space?
  - The only elements in \( B \) that are in \( A \) are members of \( AB \) whose size is \( |AB| \)
  - “New” \( P(B) = \frac{|AB|}{k} = \frac{|AB|}{|A|} = \frac{P(AB)}{P(A)} = P(B|A) \)
  - I hate reduced sample spaces!

Relative frequencies?

- The definition of conditional probability is motivated by considerations arising from the relative frequency viewpoint
- Suppose that \( N \) independent trials of the experiment have been performed
- Let \( N_A \) denote the number of trials on which event \( A \) occurred
- Let \( N_B \) denote the number of trials on which event \( B \) occurred
Focus on trials where $A$ occurred...

- Events $A$ and $B$ respectively occurred on $N_A$ and $N_B$ trials out of $N$
- Event $AB$ occurred on $N_{AB}$ trials
- $N_A \approx P(A)\cdot N$
- $N_B \approx P(B)\cdot N$, etc
- Consider only those $N_A$ trials on which $A$ occurred and ignore the rest
- On how many of these did $B$ also occur?
- If $B$ occurred, then $AB$ must have occurred (since $A$ occurred on all these $N_A$ trials)

You forgot a few, didn’t you?

- Consider only those $N_A$ trials on which $A$ occurred and ignore the rest
- On how many of these did $B$ also occur?
- If $B$ occurred, then $AB$ must have occurred (since $A$ occurred on all these $N_A$ trials)
- $B$ occurred on $N_{AB}$ trials out of the $N_A$ trials on which $A$ occurred
- The set of $N_{AB}$ trials on which $AB$ occurred must be a subset of the $N_A$ trials on which $A$ occurred

No, I didn’t!

- Consider only those $N_A$ trials on which $A$ occurred and ignore the rest
- $B$ occurred on $N_{AB}$ trials out of the $N_A$ trials on which $A$ occurred
- The relative frequency of $B$ on those $N_A$ trials on which $A$ occurred is $N_{AB}/N_A$
- $N_{AB}/N_A = (N_{AB}/N)/(N_A/N) \approx P(AB)/P(A) = P(B|A)$
- Thus, conditional probability $\approx$ relative frequency on a restricted set of trials

Does it work for random variables?

- Can conditional probabilities be defined for random variables?
- Yes, we can find the conditional probability of an event defined in terms of random variables
- What is $P(X = k \mid X > n)$?
- This is just $P(\{X = k\} \cap \{X > n\})/P(X > n)$
- The event $\{X = k\} \cap \{X > n\}$ is either $\emptyset$ or $\{X = k\}$ depending on whether $k \leq n$ or not

An example with random variables

- Let $X$ denote a geometric random variable with parameter $p$
- For $k > 0$, $P(X = k+r \mid X > r)$
  - $= P(\{X = k+r\} \cap \{X > r\})/P(X > r)$
  - $= P(X = k+r)/P(X > r)$
  - $= (1-p)^{k+r-1}p/(1-p)^r$
  - $= (1-p)^{k-1}p$
  - $= P(X = k)$

Memoryless property I

- Let $X$ denote a geometric random variable with parameter $p$
- For $k > 0$, $P(X = k+r \mid X > r) = P(X = k)$
- Given that the event $\{X > r\}$ has occurred, that is, the first $r$ trials ended in a “failure”, the probability that we need to wait for an additional $k$ trials to observe the first success is the same as $P(X = k)$
- It’s as if the first $r$ trials are forgotten!
Memoryless property II

- Geometric random variables are said to be memoryless in the sense that the waiting time for a success is unchanged by previous failures.
- The chances of having the first success occur on the tenth trial from now are the same as they were 313 trials ago when we began the experiment.
- The “system” has forgotten its past failures.

Another example with RVs

- Let $X$ denote a binomial random variable with parameters $(n, p)$.
- A binomial random variable counts the number of occurrences of an event $A$ of probability $p$ on $n$ independent trials.
- Given that $X = k$, what is the conditional probability that the $j$-th trial resulted in a success? i.e., $A$ occurred on the $j$-th trial?
- Let $B = \{X = k\}$ and $C = A$ on $j$-th trial.

More on binomial RVs

- Let $B = \{X = k\}$ and $C = A$ on $j$-th trial.
- $P(C|B) = P(BC)/P(B)$.
- $BC = \{k$ total successes on $n$ trials and a success on $j$-th trial\} = $\{k–1$ total successes on $n–1$ trials and a success on $j$-th trial\}$
- $P(BC) = \binom{n-1}{k-1}p^{k-1}(1–p)^{(n–1)–(k–1)}\cdot p$

More on binomial RVs (continued)

- $P(C|B) = P(BC)/P(B)$.
- Dividing both sides by $P(A)$, we get that $P(B|A) + P(B^c|A) = 1$.
- That is, $P(B^c|A) = 1 – P(B|A)$.
- This is just like $P(B^c) = 1 – P(B)$ except that we are using conditional probabilities!

Partitions again

- $AB$ and $AB^c$ are a partition of $A$.
- $P(AB) + P(AB^c) = P(A)$.
- Dividing both sides by $P(A)$, we get that $P(B|A) + P(B^c|A) = 1$.
- That is, $P(B^c|A) = 1 – P(B|A)$.
- This is just like $P(B^c) = 1 – P(B)$ except that we are using conditional probabilities!

Just as good as regular probability

- Conditional probabilities are a probability measure, that is, they satisfy the axioms of probability theory.
- Caveat: Everything must be conditioned on the same event. No mixing and matching allowed.
- Thus, let $A$ denote the fixed conditioning event and let all the probabilities under consideration be conditional probabilities.
The axioms are satisfied …

- A is the conditioning event
- Axiom I: \(0 \leq P(B|A) \leq 1\) for all events B
- Since \(AB \subset A\), \(0 \leq P(AB) \leq P(A)\) and hence \(0 \leq P(B|A) = P(AB)/P(A) \leq 1\)
- Axiom II: \(P(\Omega|A) = 1\)
- Since \(A\Omega = A\), \(P(A\Omega) = P(A)\) and hence \(P(\Omega|A) = P(A\Omega)/P(A) = 1\)
- Similarly for Axiom III

… and all the consequences hold

- \(P(\emptyset|A) = 0\)
- \(P(B^c|A) = 1 - P(B|A)\)
- If \(B \subset C\), then \(P(B|A) \leq P(C|A)\)
- If \(BC = \emptyset\), then \(P((B \cup C)|A) = P(B|A) + P(C|A)\)
- More generally, \(P((B \cup C)|A) = P(B|A) + P(C|A) - P(BC|A)\)

You say potato, I say potato...

- An expression such as \(P((B \cup C)|A)\) is commonly written as \(P(B \cup C|A)\)
- Everything to the right of the vertical bar is the conditioning event; it is a single set
- Everything to the left of the vertical bar is the conditioned event; it is a single set
- Beginners’ mistake: If B and C are disjoint, they write \(P(B \cup C|A) = P(B) + P(C|A)\)
- NOT!

Keeping things straight

- Condition EVERYTHING on one event A and you can apply all the rules and tricks of probability that you have learned
- \(P(B \cup C|A \cup D)\) is the conditional probability of the event \(B \cup C\) conditioned on the event \(A \cup D\)
- Exercises: What is \(P(B|A \cup B)\)? What is \(P(A \cup B|A)\)? What is \(P(A^c \cup B|A)\)? What is \(P(A^c \cup B\^c|A \cup B)\)?

Summary

- We studied the notion of conditional probability as a revised estimate of the chances that an event B occurred in light of partial knowledge of the outcome of the experiment, viz. knowing that A occurred
- Some simple examples were used to illustrate the concept
- We noted that conditional probabilities are a probability measure in that they satisfy the axioms of probability