Introduction

- The conditional probability of an a event B given that event A occurred is our revised estimate of the chances that B occurred in light of partial knowledge of the outcome of the experiment, viz. knowing that A occurred
- To avoid trivialities, we assume that A, sometimes called the conditioning event, has nonzero probability

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"
- Definition: If $P(A) > 0$, $P(B|A)$ is defined as

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

- $P(B|A)$ can be larger than, smaller than, or the same as $P(B)$

Consistent with various models

- The definition of conditional probability is consistent with
  - classical approach to probability
  - relative frequency approach
- Conditional probabilities can also be discussed for events defined in terms of random variables
- $P\{\mathbf{X} = k \mid \mathbf{X} > n\}$? or $P\{\mathbf{X} \le k \mid a < \mathbf{X} < b\}$?

Geometric RVs are memoryless

- Let **X** denote a geometric random variable with parameter p
- For $k > 0$, $P\{\mathbf{X} = k+r \mid \mathbf{X} > r\} = P\{\mathbf{X} = k\}$
- Given that the event $\{\mathbf{X} > r\}$ 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 $P\{\mathbf{X} = k\}$
- It's as if the first r trials are forgotten!

Binomial random variables

- Let **X** denote a binomial random variable with parameters (n, p)
- GIven the event $\{\mathbf{X} = k\}$ has occurred, the conditional probability that the j-th trial resulted in a success is k/n, independent of the value of p
- The conditional probability of successes on the i-th and j-th trials is $k(k–1)/[n(n–1)]$
- and so on

Axioms are satisfied

- Conditional probabilities are a probability measure, that is, they satisfy the axioms of probability theory
- All the consequences of the axioms (rules of probability) also apply to conditional probabilities
- Caveat: Everything must be conditioned on the same event. No mixing and matching allowed

Rules? What rules?

- P(Ω|A) = 1 • P(|A) = 0
- P(B^c|A) = 1 – P(B|A)
- If B C, then P(B|A) ≤ P(C|A)
- If BC = , then
  P((B C)|A) = P(B|A) + P(C|A)
- More generally,
  P((B C)|A) = P(B|A) + P(C|A) – P(BC|A)

Left side versus right side

- An expression such as P((B C)|(A D)) is commonly written as P(B C|A D)
- 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
- Even if A, B, C, and D are disjoint,
  P(B C|A D) ≠ P(B) + P(C|A) +P(D)

Is that all there is to it?

- OK, so you can update your probabilities to conditional probabilities if you know that event A occurred
  - Is that all there is to it?
  - Is the notion of conditional probability just a one-trick pony?
  - Surely life holds more than that?
- Actually, conditional probabilities are fundamental tools in probabilistic analyses

The chain rule or product rule

- P(B|A) = P(AB)/P(A)
- P(AB) = P(B|A)P(A)
- Note that P(AB) can also be expressed as P(A|B)P(B)
- The conditional probability P(B|A) can be used to compute the joint probability P(AB)
- Conditional probability P(B|A) times P(A), the probability of the conditioning event

Generalization of the chain rule

- More generally,
  P(ABCD…)=P(A)P(B|A)P(C|AB)P(D|ABC)…
- Product of first two terms is P(AB)
- P(C|AB)P(AB) = P(ABC), so that the product of the first three terms is P(ABC), and so on …
- For ABCD… to occur, A must occur, and if A has occurred, so must B (with probability P(B|A)); if both A and B, then C must …

Applications of the chain rule

- Example: A random sample of size k is drawn without replacement from the set {1, 2, … , n}. What is the probability that the sample is exactly {1, 2, 3, … , k–1, n}?
- Simple answer: There are $\binom{n}{k}$ equally likely subsets that could have been drawn, and so the desired probability is just $\binom{n}{k}^{-1}$

Example (continued)

- Example: A random sample of size k is drawn without replacement from the set {1, 2, … , n}. What is the probability that the sample is exactly {1, 2, 3, … , k–1, n}?
- Let A_i be the event that the i-th item drawn is from the wish list {1, 2, 3, … , k–1, n}
- If $A_1, A_2, A_3, \ldots, A_k$ all occur, then every item drawn is from the wish list, and the sample drawn must be exactly the list
- $A_1A_2A_3\ldots A_k \Leftrightarrow$ sample drawn is the list

Enter the chain rule …

- Example: A random sample of size k is drawn without replacement from the set {1, 2, … , n}. What is the probability that the sample is exactly {1, 2, 3, … , k–1, n}?
- $P(A_1A_2\ldots A_k)$
  $= P(A_1)P(A_2|A_1)\ldots P(A_k|A_1A_2\ldots A_{k-1})$
- $P(A_1)$ = P{first item drawn is on list} = k/n because any of k numbers out of n is OK
- $P(A_2|A_1) = (k-1)/(n-1)$ because there are still k–1 desirable numbers (out of n–1)

The last link in the chain …

- Example: A random sample of size k is drawn without replacement from the set {1, 2, … , n}. What is the probability that the sample is exactly {1, 2, 3, … , k–1, n}?
- $P(A_k|A_1A_2\ldots) = 1/(n-k+1)$ since only one good item is left among n – (k–1)
- $P(A_1A_2\ldots A_k)$
  $= P(A_1)P(A_2|A_1)\ldots P(A_k|A_1A_2\ldots A_{k-1})$
  $$= \frac{k\,(k-1)\;\ldots\;1}{n\,(n-1)\;\ldots\;(n-k+1)} = \binom{n}{k}^{-1}$$

The birthday surprise

- Example: What is the probability that at least two persons at a fraternity party attended by N people have the same birthday?
- Throw out anyone born on February 29; they have not yet celebrated their 21st birthdays, and alcohol is being served…
- Assumption: birthdays are uniformly distributed throughout the year (this is not strictly true)

Work with the complement

- For k > 1, A_k = event that the k-th person chosen has a birthday different from the the previously chosen k–1 persons
- $A_2A_3\ldots A_N$ = event that everyone has a different birthday
- $P(A_2A_3\ldots A_N)$
  $= P(A_2)P(A_3|A_2)\ldots P(A_N|A_2A_3\ldots A_{N-1})$
  $$= \frac{364}{365}\times\frac{363}{365}\times\frac{362}{365}\times\ldots\times\frac{365-(N-1)}{365}$$

Some numerical values

- $P(A_2A_3\ldots A_{23}) < 0.5$
- $P(A_2A_3\ldots A_{40}) < 0.11$
- $P(A_2A_3\ldots A_{50}) < 0.03$
- Birthday surprise problem because many people are surprised by these probabilities
- Even though there are 365 days to choose from, at a 50-person party, it is almost certain that two or more people will have the same birthday
- Applications to computer memory failures

Further generalization of the chain rule

- $P(ABCD\ldots)$
  $= P(A)P(B|A)P(C|AB)P(D|ABC)\ldots$
- Every probability result also applies to conditional probabilities
- The chain rule applies to computation of conditional probabilities by conditioning everything on the given event H (say)
- $P(ABCD\ldots\,|H)$
  $=P(A|H)P(B|AH)P(C|ABH)P(D|ABCH)\ldots$

It's not just for breakfast any more!

- $P(AB) + P(AB^c) = P(A)$
- $P(AB)= P(A|B)P(B)$
- $P(AB^c) = P(A|B^c)P(B^c)$
- Hence, $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$
  and $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$
- These formulas are totally unlike the ones seen previously

It's not 'the same thing, only different…'

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

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

- These formulas are totally unlike the ones seen previously
- On the right side, we have probabilities conditioned on different events
- Previously, we were conditioning on the same event throughout

…it's something much much more!

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

- B and B^c cannot occur simultaneously on the same trial
- To find P(A), first imagine that B occurred
- From $P(A|B)$, we can determine P(AB)
- Next imagine that B^c occurred
- From $P(A|B^c)$, we can determine $P(AB^c)$
- The sum of these two numbers is P(A) !

Oatmeal or haute cuisine?

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

- We knew how to obtain conditional probabilities from "regular" probabilities
- $P(A|B) = P(AB)/P(B)$
- New result allows us to find unconditional probabilities from conditional probabilities
- It is a fundamentally important result
- It is also very simple (uses horse sense)

The theorem of total probability

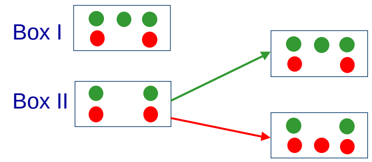
$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

- This fundamental result is called the theorem of total probability
- The probability of the event A is the weighted average of the probabilities of A conditioned on B and on B^c
- In the Ross textbook, this result is Eq.(3.1) on page 72

Applications

- Example: Box I has 3 green and 2 red balls, while Box II has 2 green and 2 red balls. A ball is drawn at random from Box I and transferred to Box II. Then, a ball is drawn at random from Box II. What is the probability that the ball drawn from Box II is green?
- Note that the color of the ball transferred from Box I to Box II is not known

Example (continued)



- The color of the ball transferred is not known, but it's either green or red for sure!

Example (continued)

- Box I has 3g, 2r; Box II has 2g, 2r
- After the transfer, Box II has 5 balls in it
- G = event ball drawn from Box II is green
- A = event ball transferred is red
- $P(G|A) = 2/5$ • $P(G|A^c) = 3/5$
- $P(A) = 2/5$
- $P(G) = P(G|A)P(A) + P(G|A^c)P(A^c)$
  $= (2/5)(2/5) + (3/5)(3/5) = 13/25$

A built-in test for checking answers

- The probability of event A is the weighted average of $P(A|B)$ and $P(A|B^c)$
- $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$
  $= P(A|B)P(B) + P(A|B^c)[1 - P(B)]$
- The linear function $y = a\bullet x + b\bullet(1 - x)$ has value b at $x = 0$ and a at $x = 1$
- For $0 < x < 1$, y is between a and b
- P(A) is between $P(A|B)$ and $P(A|B^c)$

Example (checking our work)

- $P(G|A) = 2/5$ • $P(G|A^c) = 3/5$
- $P(G) = P(G|A)P(A) + P(G|A^c)P(A^c)$
  $= (2/5)(2/5) + (3/5)(3/5) = 13/25$

$P(G|A) = 2/5 \le P(G) = 13/25 \le P(G|A^c) = 3/5$

- If the check is satisfied, it does not imply that your work is right; there may be other mistakes, e.g. you computed $P(G) = 12/25$
- But, if the check is not satisfied, …

Generalizations of the theorem I

- $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$
- Since conditional probabilities form a probability measure, a similar result also holds for conditional probabilities
- $P(A|C) = P(A|BC)P(B|C)+P(A|B^c C)P(B^c|C)$
- All probabilities in the first equation are now conditioned on C (in addition to any previously existing conditioning)

Example

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

- A = event that a flight is late in arriving
- B = event that flight is arriving at O'Hare
- C = event that flight is an United Airlines
- $P(A|BC)$ = probability that a United Airlines flight is late arriving at O'Hare
- $P(A|B^cC)$ = probability that a United Airlines flight is late arriving elsewhere

Example (continued)

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

- A = event that a flight is late in arriving
- B = event that flight is arriving at O'Hare
- C = event that flight is on United Airlines
- $P(B|C)$ = probability that a flight arriving at O'Hare is a United Airlines flight
- $P(B^c|C)$ = probability that a flight arriving elsewhere is a United Airlines flight

Example (continued)

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

- $P(A|BC)$, $P(A|B^c C)$, $P(B|C)$, and $P(B^c|C)$ can all be estimated (for example, via relative frequencies) by United Airlines or by the FAA
- $P(A|C)$ = probability that a United Airlines flight is late can then be computed (and published in the newspapers)

Generalizations of the theorem II

Given a countable partition $A_1, A_2, \ldots A_n, \ldots$ of the sample space,

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n) + \ldots$$

The theorem as presented originally was the finite case n = 2 of this more general result

The two generalizations can also be combined: condition throughout on C!

Generalization of built-in test I

- Suppose that
  $$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n) + \ldots$$
- This is a weighted sum of the $P(B|A_i)$
- If $P(B|A_j)$ is the smallest of the $P(B|A_i)$, then replacing the $P(B|A_i)$ by $P(B|A_j)$ gives
  $$P(B) \geq P(B|A_j)\cdot[P(A_1) + P(A_2) + \ldots] = P(B|A_j)$$

Generalization of built-in test II

- Suppose that
  $$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n) + \ldots$$
- If $P(B|A_k)$ is the largest of the $P(B|A_i)$, then replacing the $P(B|A_i)$ by $P(B|A_j)$ gives
  $$P(B) \leq P(B|A_k)\cdot[P(A_1) + P(A_2) + \ldots] = P(B|A_k)$$
- Conclusion: $P(B|A_j) \leq P(B) \leq P(B|A_k)$

Another Example

- You and a friend (also taking ECE 313) are at a party with N–1 other people when suddenly a conga line forms. Assume that all (N+1)! orderings are possible
- What is the probability that your friend is ahead of you in the conga line?
- Answer: 1/2 (by symmetry)
- If there was a different (correct) answer, you would be ahead with same prob ≠ 1/2

Do it by the theorem…

- Both you and your friend are equally likely to be anywhere in the conga line
- P(you are in j-th position) = 1/(N + 1)
- P(friend ahead | you in j-th) = (j – 1)/N
- Why j–1? Why N and not N+1?
- P(friend ahead) = sum of [(j–1)/N]•[1/(N+1)] = [0 + 1 + … + N]/[N•(N + 1)] = 1/2
- 1 + 2 + … + N = N•(N + 1)/2 !!!!

Summary

- The chain rule or product rule allows us to compute a joint probability (i.e. probability of an intersection) as the product of various conditional probabilities
- The theorem of total probability allows us to find an unconditional probability from conditional probabilities
- We discussed some examples of the applications of these rules

