Introduction to Probability

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 308

1. Conditional Probability and Independence
   - Conditional Probabilities
   - Bayes' Formula
   - Independent Events
   - $P(\bullet|F)$ is a Probability

Subsection 1

Conditional Probabilities

Conditional Probability: An Example

- Suppose that we toss 2 dice, and suppose that each of the 36 possible outcomes is equally likely to occur and hence has probability $\frac{1}{36}$.
- Suppose further that we observe that the first die is a 3.
- Then, given this information, what is the probability that the sum of the 2 dice equals 8?
- Given that the initial die is a 3, there can be at most 6 possible outcomes of our experiment, namely, $(3,1)$, $(3,2)$, $(3,3)$, $(3,4)$, $(3,5)$ and $(3,6)$.
- Since each of these outcomes originally had the same probability of occurring, the outcomes should still have equal probabilities.
- That is, given that the first die is a 3, the (conditional) probability of:
  - each of $(3,1)$, $(3,2)$, $(3,3)$, $(3,4)$, $(3,5)$ and $(3,6)$ is $\frac{1}{6}$;
  - each of the other 30 points in the sample space is 0.
- Hence, the desired probability will be $\frac{1}{6}$.

Conditional Probability

- Let:
  - E be the event that the sum of the dice is 8;
  - F be the event that the first die is a 3.

  Then the probability just obtained is called the **conditional probability that E occurs given that F has occurred** and is denoted by $P(E|F)$.
- A general formula for $P(E|F)$ that is valid for all events E and F is derived in the same manner:
  - If the event F occurs, then, for E to occur, it is necessary that the actual occurrence be a point both in E and in F, i.e., it must be in EF.
  - Moreover, since F has occurred, F becomes the new sample space.
  - Hence, the probability that the event EF occurs will equal the probability of EF relative to the probability of F.

Definition

If $P(F) > 0$, then $P(E|F) = \dfrac{P(EF)}{P(F)}$.

Example

- A student is taking a one-hour-time-limit makeup examination. Suppose the probability that the student will finish the exam in less than x hours is $\frac{x}{2}$, for all $0 \leq x \leq 1$.
  Then, given that the student is still working after 0.75 hour, what is the conditional probability that the full hour is used?
  Let L_x denote the event that the student finishes the exam in less than x hours, $0 \leq x \leq 1$.
  Let F be the event that the student uses the full hour.
  Because F is the event that the student is not finished in less than 1 hour, $P(F) = P(L_1^c) = 1 - P(L_1) = 1 - \frac{1}{2} = 0.5$.
  Now, the event that the student is still working at time 0.75 is the complement of the event $L_{0.75}$.
  So the desired probability is obtained from
  $$P(F|L_{0.75}^c) = \frac{P(FL_{0.75}^c)}{P(L_{0.75}^c)} = \frac{P(F)}{1 - P(L_{0.75})} = \frac{0.5}{0.625} = 0.8.$$

The Case of Equally Likely Outcomes

- If each outcome of a finite sample space S is equally likely, then, conditional on the event that the outcome lies in a subset $F \subseteq S$, all outcomes in F become equally likely.
- In such cases, it is often convenient to compute conditional probabilities of the form $P(E|F)$ by using F as the sample space.
- Indeed, working with this reduced sample space often results in an easier and better understood solution.

Example

- A coin is flipped twice.

  Assume that all four points in the sample space $S = \{(h,h),(h,t),(t,h),(t,t)\}$ are equally likely.

  What is the conditional probability that both flips land on heads, given that:

  (a) the first flip lands on heads?

  (b) at least one flip lands on heads?

  Let $B = \{(h,h)\}$ be the event that both flips land on heads.

  Let $F = \{(h,h),(h,t)\}$ be the event that the first flip lands on heads.

  Let $A = \{(h,h),(h,t),(t,h)\}$ be the event that at least one flip lands on heads.

  Then we have:

$$\begin{array}{rcl} P(B|F) & = & \frac{P(BF)}{P(F)} = \frac{P(\{(h,h)\})}{P(\{(h,h),(h,t)\})} = \frac{1/4}{2/4} = 1/2; \\ P(B|A) & = & \frac{P(BA)}{P(A)} = \frac{P(\{(h,h)\})}{P(\{(h,h),(h,t),(t,h)\})} = \frac{1/4}{3/4} = 1/3. \end{array}$$

Example

- In the card game bridge, the 52 cards are dealt out equally to 4 players - called East, West, North, and South.

  If North and South have a total of 8 spades among them, what is the probability that East has 3 of the remaining 5 spades?

  We work with the reduced sample space.

  Suppose that North-South have 8 spades among their 26 cards.

  There remains a total of 26 cards, exactly 5 of them being spades, to be distributed among the East-West hands.

  Since each distribution is equally likely, it follows that the conditional probability that East will have exactly 3 spades among his/her 13 cards is:

$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} \approx 0.339.$$

Example

- A total of n balls are sequentially and randomly chosen, without replacement, from an urn containing r red and b blue balls.
  Suppose that k of the n balls are blue.
  What is the conditional probability that the 1st ball chosen is blue?
  Suppose the balls are numbered:
  - The blue balls have numbers 1 through b;
  - The red balls have numbers $b+1$ through $b+r$.

  The outcome of this experiment is a vector of distinct integers $x_1, \ldots, x_n$, where each x_i is between 1 and $r+b$.
  Moreover, each such vector is equally likely to be the outcome.
  We are given that the vector contains k values between 1 and b.
  It follows that each of these outcomes is equally likely.
  The first ball chosen is, therefore, equally likely to be any of the n chosen balls, of which k are blue.
  Hence, the desired probability is $\frac{k}{n}$.

Example (Cont'd)

- Suppose we did not choose to work with the reduced sample space. Let:
  - B be the event that the first ball chosen is blue;
  - B_k be the event that a total of k blue balls are chosen.

  Then
  $$P(B|B_k) = \frac{P(BB_k)}{P(B_k)} = \frac{P(B_k|B)P(B)}{P(B_k)}.$$
  Now, $P(B_k|B)$ is the probability that a random choice of $n-1$ balls from an urn containing r red and $b-1$ blue balls results in a total of $k-1$ blue balls being chosen. Consequently, $P(B_k|B) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}}$.

  Moreover, $P(B) = \frac{b}{r+b}$ and $P(B_k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{r+b}{n}}$.

  Hence $P(B|B_k) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}} \cdot \frac{b}{r+b} \cdot \frac{\binom{r+b}{n}}{\binom{b}{k}\binom{r}{n-k}} = \frac{k}{n}$.

Probability of the Intersection

- Multiplying both sides of $P(E|F) = \frac{P(EF)}{P(F)}$ by $P(F)$, we obtain
$$P(EF) = P(F)P(E|F).$$

Example: Celine is undecided as to whether to take a French course or a chemistry course.
She estimates that her probability of receiving an A grade would be $\frac{1}{2}$ in a French course and $\frac{2}{3}$ in a chemistry course.
If Celine decides to base her decision on the flip of a fair coin, what is the probability that she gets an A in chemistry?
Let C be the event that Celine takes chemistry.
Let A be the event that she receives an A in whatever course she takes.
Then the desired probability is $P(CA)$, which is calculated as follows:
$$P(CA) = P(C)P(A|C) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Example

- Suppose that an urn contains 8 red balls and 4 white balls.
  We draw 2 balls from the urn without replacement.
  (a) If we assume that at each draw each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red?
  (b) Now suppose that the balls have different weights, with each red ball having weight r and each white ball having weight w.
  Suppose that the probability that a given ball in the urn is the next one selected is its weight divided by the sum of the weights of all balls currently in the urn.
  Now what is the probability that both balls are red?

  Let R_1 and R_2 denote, respectively, the events that the first and second balls drawn are red.

  Now, given that the first ball selected is red, there are 7 remaining red balls and 4 white balls. So $P(R_2|R_1) = \frac{7}{11}$.
  Moreover, $P(R_1) = \frac{8}{12}$.
  Hence, $P(R_1R_2) = P(R_1)P(R_2|R_1) = \frac{2}{3} \cdot \frac{7}{11} = \frac{14}{33}$.

Example (Part (b))

- For part (b), we again let R_i be the event that the ith ball chosen is red and use $P(R_1R_2) = P(R_1)P(R_2|R_1)$.
  Now, number the red balls, and let B_i, $i = 1, \ldots, 8$ be the event that the first ball drawn is red ball number i.
  $$P(R_1) = P(\bigcup_{i=1}^{8} B_i) = \sum_{i=1}^{8} P(B_i) = 8\frac{r}{8r+4w}.$$
  Moreover, given that the first ball is red, the urn then contains 7 red and 4 white balls.
  $$P(R_2|R_1) = \frac{7r}{7r+4w}.$$
  Hence, the probability that both balls are red is
  $$P(R_1R_2) = \frac{8r}{8r+4w} \cdot \frac{7r}{7r+4w}.$$

The Multiplication Principle

The Multiplication Principle

Let E_1, E_2, …, E_n be events in a sample space S. Then

$$P(E_1 E_2 E_3 \cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \cdots P(E_n|E_1 \cdots E_{n-1}).$$

- To prove the multiplication rule, just apply the definition of conditional probability to its right-hand side:

$$\begin{aligned} P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \cdots P(E_n|E_1 \cdots E_{n-1}) \\ = P(E_1)\frac{P(E_1 E_2)}{P(E_1)}\frac{P(E_1 E_2 E_3)}{P(E_1 E_2)} \cdots \frac{P(E_1 E_2 \cdots E_n)}{P(E_1 E_2 \cdots E_{n-1})} \\ = P(E_1 E_2 \cdots E_n). \end{aligned}$$

Example

- In the match problem, it was shown that P_N, the probability that there are no matches when N people randomly select from among their own N hats, is given by $P_N = \sum_{i=0}^{N} \frac{(-1)^i}{i!}$.
  What is the probability that exactly k of the N people have matches?
  We fix a particular set of k people and determine the probability that these k individuals have matches and no one else does.
  - Let E be the event that everyone in this set has a match.
  - Let G be the event that none of the other $N-k$ people have a match.

  We have $P(EG) = P(E)P(G|E)$.
  Let F_i, $i = 1, \ldots, k$, be the event that the ith member has a match. Then

$$\begin{aligned} P(E) &= P(F_1 F_2 \cdots F_k) \\ &= P(F_1)P(F_2|F_1)P(F_3|F_1F_2)\cdots P(F_k|F_1\cdots F_{k-1}) \\ &= \frac{1}{N}\frac{1}{N-1}\frac{1}{N-2}\cdots\frac{1}{N-k+1} = \frac{(N-k)!}{N!}. \end{aligned}$$

Example (Cont'd)

- Given that everyone in the set of k has a match, the other $N-k$ people will be randomly choosing among their own $N-k$ hats. So the probability that none of them has a match is equal to the probability of no matches in a problem having $N-k$ people choosing among their own $N-k$ hats. Therefore,

$$P(G|E) = P_{N-k} = \sum_{i=0}^{N-k} \frac{(-1)^i}{i!}.$$

Thus, the probability that a specified set of k people have matches and no one else does is $P(EG) = \frac{(N-k)!}{N!} P_{N-k}$.

There will be exactly k matches if the preceding is true for any of the $\binom{N}{k}$ sets of k individuals. Hence, the desired probability is

$$P(\text{exactly } k \text{ matches}) = \binom{N}{k} \frac{(N-k)!}{N!} P_{N-k} = \frac{P_{N-k}}{k!}.$$

Example

- An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each.

  Compute the probability that each pile has exactly 1 ace.

  Define events E_i, $i = 1, 2, 3, 4$, as follows:

$$\begin{aligned} E_1 &= \{\text{the ace of spades is in any one of the piles}\}; \\ E_2 &= \{\text{the ace of spades and the ace of hearts are in different piles}\}; \\ E_3 &= \{\text{the aces of spades, hearts and diamonds are all in different piles}\}; \\ E_4 &= \{\text{all 4 aces are in different piles}\}. \end{aligned}$$

  The desired probability is $P(E_1E_2E_3E_4)$. By the multiplication rule,

$$P(E_1E_2E_3E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3).$$

Example (Cont'd)

- We have:
  - $E_1 = S$.
    Hence, $P(E_1) = 1$.
  - The pile containing the ace of spades will receive 12 of the remaining 51 cards.
    Hence, $P(E_2|E_1) = \frac{39}{51}$.
  - The piles containing the aces of spades and hearts will receive 24 of the remaining 50 cards.
    Hence, $P(E_3|E_1E_2) = \frac{26}{50}$.
  - $P(E_4|E_1E_2E_3) = \frac{13}{49}$.

  Therefore, the probability that each pile has exactly 1 ace is

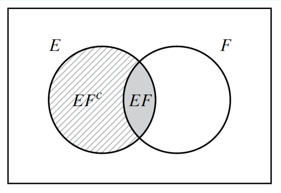
$$P(E_1E_2E_3E_4) = \frac{39}{51} \cdot \frac{26}{50} \cdot \frac{13}{49} \approx 0.105.$$

Subsection 2

Bayes' Formula

Probability of an Event as Weighted Average

- Let E and F be events. We may express E as $E = EF \cup EF^c$ for, in order for an outcome to be in E, it must either be in both E and F or be in E but not in F.



As EF and EF^c are clearly mutually exclusive, we have, by Axiom 3,

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)]. \end{aligned}$$

This equation states that the probability of the event E is a weighted average of the conditional probability of:

- E given that F has occurred;
- E given that F has not occurred;

each conditional probability being given as much weight as the event on which it is conditioned has of occurring.

Example

- Insurance Inc. believes that people can be divided into two classes:
  - Those who are accident prone;
  - Those who are not.

  The company's statistics show that:
  - An accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4;
  - This probability decreases to 0.2 for a person who is not accident prone.

  If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

  Let A_1 denote the event that the policyholder will have an accident within a year of purchasing the policy.

  Let A denote the event that the policyholder is accident prone.

  The desired probability is given by

$$\begin{aligned} P(A_1) &= P(A_1|A)P(A) + P(A_1|A^c)P(A^c) \\ &= (0.4)(0.3) + (0.2)(0.7) = 0.26. \end{aligned}$$

Example (Cont'd)

- Suppose that a new policyholder has an accident within a year of purchasing a policy.

  What is the probability that he or she is accident prone?

  The desired probability is

$$\begin{aligned} P(A|A_1) &= \frac{P(AA_1)}{P(A_1)} \\ &= \frac{P(A)P(A_1|A)}{P(A_1)} \\ &= \frac{(0.3)(0.4)}{0.26} \\ &= \frac{6}{13}. \end{aligned}$$

Example

- Consider the following game played with an ordinary deck of 52 playing cards:
  - The cards are shuffled and then turned over one at a time.
  - At any time, the player can guess that the next card to be turned over will be the ace of spades.
    - If it is, then the player wins.
    - The player also wins if the ace of spades has not yet appeared when only one card remains and no guess has yet been made.

  What is a good strategy? What is a bad strategy?

  Every strategy has probability $1/52$ of winning!

  We use induction to prove the stronger result that, for an n card deck, one of whose cards is the ace of spades, the probability of winning is $\frac{1}{n}$, no matter what strategy is employed.

  This is clearly true for $n = 1$.

  Assume it to be true for an $n - 1$ card deck.

Example (Cont'd)

- Consider an n card deck.

  Fix any strategy, and let p denote the probability that the strategy guesses that the first card is the ace of spades.

  Given that it does, the player's probability of winning is $\frac{1}{n}$.

  If the strategy does not guess that the first card is the ace of spades, then the probability that the player wins is the product of:
  - the probability that the first card is not the ace of spades: $\frac{n-1}{n}$;
  - the conditional probability of winning given that the first card is not the ace of spades, i.e., the probability of winning when using an $n-1$ card deck containing a single ace of spades: $\frac{1}{n-1}$ (by the induction hypothesis).

  Hence, given that the strategy does not guess the first card, the probability of winning is $\frac{n-1}{n} \cdot \frac{1}{n-1} = \frac{1}{n}$.

Example (Cont'd)

- Thus, letting G be the event that the first card is guessed, we obtain:

$$\begin{aligned} P\{\text{win}\} &= P\{\text{win}|G\}P(G) + P\{\text{win}|G^c\}(1 - P(G)) \\ &= \frac{1}{n}p + \frac{1}{n}(1 - p) \\ &= \frac{1}{n}. \end{aligned}$$

Example

- In answering a question on a multiple choice test, a student either knows the answer or guesses.

  Let p be the probability that the student knows the answer and $1-p$ be the probability that the student guesses.

  Assume that a student who guesses at the answer will be correct with probability $\frac{1}{m}$, where m is the number of multiple choice alternatives. What is the conditional probability that a student knew the answer to a question given that he or she answered it correctly?
  - Let C be the event that the student answers the question correctly;
  - Let K be the event that he or she actually knows the answer.

$$\begin{aligned} P(K|C) &= \frac{P(KC)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K)+P(C|K^c)P(K^c)} \\ &= \frac{p}{p+(1/m)(1-p)} = \frac{mp}{1+(m-1)p}. \end{aligned}$$

Example

- A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present.

  However, the test also yields a "false positive" result for 1 percent of the healthy persons tested, i.e., if a healthy person is tested, then, with probability 0.01, the test result will imply that he or she has the disease.

  Suppose 0.5 percent of the population actually has the disease.

  What is the probability that a person has the disease given that the test result is positive?

  - Let D be the event that the person tested has the disease;
  - Let E be the event that the test result is positive.

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E|D)P(D)+P(E|D^c)P(D^c)} \\ &= \frac{(0.95)(0.005)}{(0.95)(0.005)+(0.01)(0.995)} = \frac{95}{294} \approx 0.323. \end{aligned}$$

Example

- At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect.
  Suppose a new piece of evidence which shows that the criminal has a certain characteristic (e.g., left-handedness) is uncovered.
  Assume 20 percent of the population possesses this characteristic.
  How certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?
  - Let G denote the event that the suspect is guilty;
  - Let C the event that he possesses the characteristic of the criminal.

  Assume that the probability of the suspect having the characteristic if he is innocent is the proportion of the population possessing the characteristic: $P(C|G^c) = 0.2$.

$$\begin{aligned} P(G|C) &= \frac{P(GC)}{P(C)} = \frac{P(C|G)P(G)}{P(C|G)P(G)+P(C|G^c)P(G^c)} \\ &= \frac{1(0.6)}{1(0.6)+(0.2)(0.4)} \approx 0.882. \end{aligned}$$

Example

- Urn 1 initially has n red molecules and urn 2 has n blue molecules. Molecules are randomly removed from urn 1 in the following manner:

  After each removal from urn 1, a molecule is taken from urn 2 (if urn 2 has any molecules) and placed in urn 1.

  The process continues until all the molecules have been removed.

  (Thus, there are $2n$ removals in all.)

  Find $P(R)$, where R is the event that the final molecule removed from urn 1 is red.

Example (Cont'd)

- Fix a particular red molecule.

  Let F be the event that this molecule is the final one selected.

  In order for F to occur, the molecule in question must still be in the urn after the first n molecules have been removed (at which time urn 2 is empty).

  Let N_i be the event that the fixed molecule is not the ith molecule to be removed.

  The conditional probability that the molecule under consideration is the final molecule to be removed, given that it is still in urn 1 when only n molecules remain, is, by symmetry, $\frac{1}{n}$.

$$\begin{aligned} P(F) &= P(N_1 \cdots N_n F) \\ &= P(N_1)P(N_2|N_1)\cdots P(N_n|N_1\cdots N_{n-1})P(F|N_1\cdots N_n) \\ &= (1-\tfrac{1}{n})\cdots(1-\tfrac{1}{n})\tfrac{1}{n}. \end{aligned}$$

Example (Cont'd)

- Now number the n red molecules.

  Let R_j be the event that red molecule number j is the final molecule removed.

  Then it follows from the preceding formula that

$$P(R_j) = \left(1 - \frac{1}{n}\right)^n \frac{1}{n}.$$

  Because the events R_j are mutually exclusive, we obtain

$$P(R) = P(\bigcup_{j=1}^{n} R_j) = \sum_{j=1}^{n} P(R_j) = \left(1 - \frac{1}{n}\right)^n \approx e^{-1}.$$

Odds

Definition

The **odds** of an event A are defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}.$$

That is, the odds of an event A tell how much more likely it is that the event A occurs than it is that it does not occur.

Example: Suppose $P(A) = \frac{2}{3}$. What are the odds of A?
We have

$$\frac{P(A)}{1 - P(A)} = \frac{2/3}{1/3} = 2.$$

So the odds are 2.

- If the odds are equal to α, then it is common to say:

The odds are "α **to** 1" in favor of the hypothesis.

Odds Under Conditioning

- Consider a hypothesis H that is true with probability $P(H)$.
- Suppose that new evidence E is introduced.
- Then the conditional probabilities, given the evidence E, that H is true and that H is not true are, respectively, given by:

$$P(H|E)=\frac{P(E|H)P(H)}{P(E)}, \quad P(H^c|E)=\frac{P(E|H^c)P(H^c)}{P(E)}.$$

- Therefore, the new odds after the evidence E has been introduced are:

$$\frac{P(H|E)}{P(H^c|E)}=\frac{P(H)}{P(H^c)}\frac{P(E|H)}{P(E|H^c)}.$$

- That is, the new value of the odds of H is the product of:
  - the old value of the odds of H;
  - the ratio of the conditional probability of the new evidence given that H is true to the conditional probability given that H is not true.

Example

- An urn contains two type A coins and one type B coin.
  When a type A coin is flipped, it comes up heads with probability $\frac{1}{4}$.
  When a type B coin is flipped, it comes up heads with probability $\frac{3}{4}$.
  A coin is randomly chosen from the urn and flipped.
  Given that the flip landed on heads, what is the probability that it was a type A coin?
  - Let A be the event that a type A coin was flipped;
  - Let $B = A^c$ be the event that a type B coin was flipped.

  We want $P(A|H)$, where H is the event that heads occurred.

$$\begin{aligned} \frac{P(A|H)}{P(A^c|H)} &= \frac{P(A)}{P(B)} \cdot \frac{P(H|A)}{P(H|B)} \\ &= \frac{2/3}{1/3}\frac{1/4}{3/4} = \frac{2}{3}. \end{aligned}$$

  Hence, the odds are $2/3 : 1$.
  Equivalently, the probability that a type A coin was flipped is $\frac{2}{5}$.

Generalization of the Sum Formula

- Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\bigcup_{i=1}^{n} F_i = S$.
- In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ must occur.
- Write $E = \bigcup_{i=1}^{n} EF_i$ and use the fact that the events EF_i, $i = 1, \ldots, n$, are mutually exclusive:

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i).$$

- The equation shows how, for given events $F_1, F_2, \ldots, F_n$, of which one and only one must occur, we can compute $P(E)$ by first conditioning on which one of the F_i occurs.

Example

- Consider the probability that, for a randomly shuffled deck, the card following the first ace is some specified card.

  Let E be the event that the card following the first ace is some specified card, say, card x.

  To compute $P(E)$, we ignore card x and condition on the relative ordering of the other 51 cards in the deck.

  Let $\boldsymbol{O}$ be the ordering:

$$P(E) = \sum_{\boldsymbol{O}} P(E|\boldsymbol{O})P(\boldsymbol{O}).$$

Example (Cont'd)

- Now, given $\boldsymbol{O}$, there are 52 possible orderings of the cards, corresponding to having card x being the ith card in the deck, $i = 1, \ldots, 52$.

  But because all 52! possible orderings were initially equally likely, it follows that, conditional on $\boldsymbol{O}$, each of the 52 remaining possible orderings is equally likely.

  Because card x will follow the first ace for only one of these orderings, we have $P(E|\boldsymbol{O}) = \frac{1}{52}$.

  This implies that

$$P(E) = \sum_{\boldsymbol{O}} P(E|\boldsymbol{O})P(\boldsymbol{O}) = \frac{1}{52}\sum_{\boldsymbol{O}} P(\boldsymbol{O}) = \frac{1}{52}.$$

Bayes' Formula

- Again, let $F_1, \ldots, F_n$ be a set of mutually exclusive and exhaustive events (meaning that exactly one of these events must occur).
- Suppose now that E has occurred and we are interested in determining which one of the F_j also occurred.

Proposition (Bayes' Formula)

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}.$$

- Think of the events F_j as being possible "hypotheses".

  Bayes' formula shows how opinions held before the experiment about the hypotheses (the $P(F_j)$) should be modified by the evidence E produced by the experiment.

Example

- A plane is missing, and it is presumed that it was equally likely to have gone down in any of 3 possible regions.

  Let $1-\beta_i$, $i=1,2,3$, denote the probability that the plane will be found upon a search of the ith region when the plane is, in fact, in that region.

  What is the conditional probability that the plane is in the ith region given that a search of region 1 is unsuccessful?

  - Let R_i, $i=1,2,3$, be the event that the plane is in region i.
  - Let E be the event that a search of region 1 is unsuccessful.

  From Bayes' formula, we obtain:

$$\begin{aligned} P(R_1|E) &= \frac{P(ER_1)}{P(E)} = \frac{P(E|R_1)P(R_1)}{\sum_{i=1}^3 P(E|R_i)P(R_i)} \\ &= \frac{(\beta_1)\frac{1}{3}}{(\beta_1)\frac{1}{3}+(1)\frac{1}{3}+(1)\frac{1}{3}} = \frac{\beta_1}{\beta_1+2}; \\ P(R_j|E) &= \frac{P(E|R_j)P(R_j)}{P(E)} = \frac{(1)\frac{1}{3}}{(\beta_1)\frac{1}{3}+\frac{1}{3}+\frac{1}{3}} = \frac{1}{\beta_1+2}, \quad j=2,3. \end{aligned}$$

Comments on the Solution

- The updated (that is, the conditional) probability that the plane is in region j, given the information that a search of region 1 did not find it, is greater than the initial probability that it was in region j when $j \neq 1$ and is less than the initial probability when $j = 1$.

  This statement is certainly intuitive, since not finding the plane in region 1 would seem to decrease its chance of being in that region and increase its chance of being elsewhere.
- The conditional probability that the plane is in region 1 given an unsuccessful search of that region is an increasing function of the probability β_1.

  This statement is also intuitive, since the larger β_1 is, the more it is reasonable to attribute the unsuccessful search to "bad luck" as opposed to the plane's not being there.
- Similarly, $P(R_j|E)$, $j \neq 1$, is a decreasing function of β_1.

Example

- Suppose that we have 3 cards that are identical in form, except that:
  - both sides of the first card are colored red;
  - both sides of the second card are colored black;
  - one side of the third card is colored red and the other side black.

  The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the up side of the chosen card is red, what is the probability that the other side is black?
  - Let RR, BB, and RB denote, respectively, the events that the chosen card is all red, all black, or the red-black card;
  - Let R be the event that the up side of the chosen card is red.

  The desired probability is:

$$\begin{aligned} P(RB|R) &= \frac{P(RB \cap R)}{P(R)} \\ &= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR)+P(R|RB)P(RB)+P(R|BB)P(BB)} \\ &= \frac{(\frac{1}{2})(\frac{1}{3})}{(1)(\frac{1}{3})+(\frac{1}{2})(\frac{1}{3})+0(\frac{1}{3})} = \frac{1}{3}. \end{aligned}$$

Example

- A new couple, known to have two children, has just moved into town.

  Suppose that the mother is seen walking with one of her children.

  If this child is a girl, what is the probability that both children are girls?

  Let us start by defining the following events:
  - G_1/B_1: the first (oldest) child is a girl/boy;
  - G_2/B_2: the second child is a girl/boy;
  - G/B: the child seen with the mother is a girl/boy.

  Now, the desired probability is:

$$P(G_1 G_2 | G) = \frac{P(G_1 G_2 G)}{P(G)} = \frac{P(G_1 G_2)}{P(G)}.$$

Example (Cont'd)

- Also,

$$\begin{aligned} P(G) &= P(G|G_1G_2)P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) \\ &\quad + P(G|B_1G_2)P(B_1G_2) + P(G|B_1B_2)P(B_1B_2) \\ &= P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) + P(G|B_1G_2)P(B_1G_2), \end{aligned}$$

  where we use $P(G|G_1G_2) = 1$ and $P(G|B_1B_2) = 0$.

  Suppose that all 4 gender possibilities are equally likely.

$$\begin{aligned} P(G_1G_2|G) &= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{P(G|G_1B_2)}{4} + \frac{P(G|B_1G_2)}{4}} \\ &= \frac{1}{1 + P(G|G_1B_2) + P(G|B_1G_2)}. \end{aligned}$$

  Thus, the answer depends on whatever assumptions we want to make about the conditional probabilities

  - the child seen with the mother is a girl given the event G_1B_2;
  - the child seen with the mother is a girl given the event G_2B_1.

Example

- A bin contains 3 different types of disposable flashlights.
  The probability that a flashlight will give over 100 hours of use is:
  - 0.7 for a type 1 flashlight;
  - 0.4 for a type 2 flashlight;
  - 0.3 for a type 3 flashlight.

  Suppose that 20 percent of the flashlights in the bin are type 1, 30 percent are type 2, and 50 percent are type 3.
  (a) What is the probability that a randomly chosen flashlight will give more than 100 hours of use?
  (b) Given that a flashlight lasted over 100 hours, what is the conditional probability that it was a type j flashlight, $j = 1, 2, 3$?

(a) We set notation:
- Let A denote the event that the flashlight chosen lasts over 100 hours;
- Let F_j be the event that a type j flashlight is chosen, $j = 1, 2, 3$.

$$\begin{aligned} P(A) &= P(A|F_1)P(F_1) + P(A|F_2)P(F_2) + P(A|F_3)P(F_3) \\ &= (0.7)(0.2) + (0.4)(0.3) + (0.3)(0.5) = 0.41. \end{aligned}$$

Example (Cont'd)

(b) The probability is obtained by using Bayes's formula:

$$P(F_j|A) = \frac{P(AF_j)}{P(A)} = \frac{P(A|F_j)P(F_j)}{0.41}.$$

Thus,

$$\begin{aligned} P(F_1|A) &= \tfrac{(0.7)(0.2)}{0.41} = \tfrac{14}{41}; \\ P(F_2|A) &= \tfrac{(0.4)(0.3)}{0.41} = \tfrac{12}{41}; \\ P(F_3|A) &= \tfrac{(0.3)(0.5)}{0.41} = \tfrac{15}{41}. \end{aligned}$$

Note that the initial probability that type 1 is chosen is only 0.2;

However, the information that the flashlight has lasted over 100 hours raises the probability of this event to $\frac{14}{41} \approx 0.341$.

Subsection 3

Independent Events

Independent Events

- In general, $P(E|F)$, the conditional probability of E given F, is not equal to $P(E)$, the unconditional probability of E.
  In other words, knowing that F has occurred generally changes the chances of E's occurrence.
- In the special cases where $P(E|F)$ does in fact equal $P(E)$, we say that E is *independent of* F.
  That is, E is independent of F if knowledge that F has occurred does not change the probability that E occurs.
- Since $P(E|F) = \frac{P(EF)}{P(F)}$, it follows that E is independent of F if $P(EF) = P(E)P(F)$.
- The fact that this is symmetric in E and F shows that whenever E is independent of F, F is also independent of E.

Definition

Two events E and F are said to be **independent** if $P(EF) = P(E)P(F)$.
Two events E and F that are not independent are said to be **dependent**.

Example

- Two coins are flipped, and all 4 outcomes are assumed to be equally likely.
  - Let E be the event that the first coin lands on heads;
  - Let F be the event that the second lands on tails.

  Then E and F are independent.

  We have
$$\begin{aligned} P(EF) &= P(\{(H,T)\}) = \tfrac{1}{4}; \\ P(E) &= P(\{(H,H),(H,T)\}) = \tfrac{1}{2}; \\ P(F) &= P(\{(H,T),(T,T)\}) = \tfrac{1}{2}. \end{aligned}$$
  So $P(EF) = P(E)P(F)$.

Example

- Suppose that we toss 2 fair dice.
  - Let E_1 be the event that the sum of the dice is 6;
  - Let F be the event that the first die equals 4.

  Then
  $$\begin{array}{rcl} P(E_1F) & = & P(\{(4,2)\}) = \frac{1}{36}; \\ P(E_1)P(F) & = & (\frac{5}{36})(\frac{1}{6}) = \frac{5}{216}. \end{array}$$
  Hence, E_1 and F are not independent.

  Intuitively, the reason for this is clear because our chance of getting a total of 6 depends on the outcome of the first die.
- Let E_2 be the event that the sum of the dice equals 7.

  Is E_2 independent of F?

  The answer is yes, since
  $$\begin{array}{rcl} P(E_2F) & = & P(\{(4,3)\}) = \frac{1}{36}; \\ P(E_2)P(F) & = & (\frac{1}{6})(\frac{1}{6}) = \frac{1}{36}. \end{array}$$

Independence and Complements

Proposition

If E and F are independent, then so are E and F^c.

- Assume that E and F are independent.
  We have
  - $E = EF \cup EF^c$;
  - EF and EF^c are mutually exclusive.

  Hence, we obtain

$$P(E) = P(EF) + P(EF^c) = P(E)P(F) + P(EF^c).$$

  Equivalently,

$$P(EF^c) = P(E) - P(E)P(F) = P(E)[1 - P(F)] = P(E)P(F^c).$$

  This proves the result.

Independence and Intersections

- Suppose now that E is independent of F and is also independent of G.

  Then E is not necessarily independent of FG.
- Two fair dice are thrown.
  - Let E denote the event that the sum of the dice is 7;
  - Let F denote the event that the first die equals 4;
  - Let G denote the event that the second die equals 3.

  From a previous example, we know that E is independent of F:

$$P(EF) = \frac{1}{36} = \frac{6}{36} \cdot \frac{6}{36} = P(E)P(F).$$

  The same reasoning shows that E is also independent of G.

  Clearly, E is not independent of FG, since $P(E|FG) = 1 \neq P(E)$.

Independence of Three Events

Definition

Three events E, F and G are said to be **independent** if

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F), \quad P(EG) = P(E)P(G), \quad P(FG) = P(F)P(G).$$

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

  For instance, E is independent of $F \cup G$, since:

$$\begin{aligned} P[E(F \cup G)] &= P(EF \cup EG) \\ &= P(EF) + P(EG) - P(EFG) \\ &= P(E)P(F) + P(E)P(G) - P(E)P(FG) \\ &= P(E)[P(F) + P(G) - P(FG)] \\ &= P(E)P(F \cup G). \end{aligned}$$

Independence of Many Events

- We extend the definition of independence to more than three events.
- The events $E_1, E_2, \ldots, E_n$ are said to be **independent** if, for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leq n$, of these events,

$$P(E_{1'} E_{2'} \cdots E_{r'}) = P(E_{1'}) P(E_{2'}) \cdots P(E_{r'}).$$

- Finally, we define an infinite set of events to be independent if every finite subset of those events is independent.

Independent Subexperiments and Trials

- Sometimes, a probability experiment under consideration consists of performing a sequence of **subexperiments**.
- For instance, if the experiment consists of continually tossing a coin, we may think of each toss as being a subexperiment.
- In many cases, it is reasonable to assume that the outcomes of any group of the subexperiments have no effect on the probabilities of the outcomes of the other subexperiments.
- If such is the case, we say that the subexperiments are *independent*.
- More formally, we say that the subexperiments are **independent** if $E_1, E_2, \ldots, E_n, \ldots$ is necessarily an independent sequence of events whenever E_i is an event whose occurrence is completely determined by the outcome of the ith subexperiment.
- If each subexperiment has the same set of possible outcomes, then the subexperiments are often called **trials**.

Example

- An infinite sequence of independent trials is to be performed.
  Each trial results in:
  - a success with probability p;
  - a failure with probability $1 - p$.

  What is the probability that:
  (a) at least 1 success occurs in the first n trials;
  (b) exactly k successes occur in the first n trials;
  (c) all trials result in successes?

(a) To determine the probability of at least 1 success in the first n trials, we compute the probability of the complementary event, i.e., of no successes in the first n trials.

Let E_i be the event of a failure on the ith trial.

Then the probability of no successes is, by independence,

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2)\cdots P(E_n) = (1-p)^n.$$

Hence, the answer to part (a) is $1-(1-p)^n$.

Example (Cont'd)

(b) To compute the answer to part (b), consider any particular sequence of the first n outcomes containing k successes and $n-k$ failures.
Each one of these sequences will, by the assumed independence of trials, occur with probability $p^k(1-p)^{n-k}$.
The number of sequences with k successes is $\binom{n}{k}$.
Thus, the desired probability in part (b) is

$$P\{\text{exactly } k \text{ successes}\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

(c) Note that, by part (a), the probability of the first n trials all resulting in success is given by $P(E_1^c E_2^c \cdots E_n^c) = p^n$.
Thus, the desired probability is given by

$$\begin{aligned} P(\bigcap_{i=1}^{\infty} E_i^c) &= P(\lim_{n\to\infty} \bigcap_{i=1}^{n} E_i^c) = \lim_{n\to\infty} P(\bigcap_{i=1}^{n} E_i^c) \\ &= \lim_{n\to\infty} (p^n) = \begin{cases} 0, & \text{if } p < 1 \\ 1, & \text{if } p = 1 \end{cases} \end{aligned}$$

Example

- A system composed of n separate components is said to be a parallel system if it functions when at least one of the components functions.

  Suppose that, for such a system, component i, which is independent of the other components, functions with probability p_i, $i = 1, \ldots, n$.

  What is the probability that the system functions?

  Let A_i denote the event that component i functions.

  Then

$$\begin{aligned} P\{\text{system functions}\} &= 1 - P\{\text{system does not function}\} \\ &= 1 - P\{\text{all components do not function}\} \\ &= 1 - P(\textstyle\bigcap_i A_i^c) \\ &= 1 - \textstyle\prod_{i=1}^n (1 - p_i) \\ &\qquad \text{(by independence)} \end{aligned}$$

Example

- Independent trials consisting of rolling a pair of fair dice are performed.

  Suppose the outcome of a roll is the sum of the dice.

  What is the probability that an outcome of 5 appears before an outcome of 7?

  - Let E_n denote the event that no 5 or 7 appears on the first $n-1$ trials and a 5 appears on the nth trial.

  Then the desired probability is $P(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P(E_n)$.

  But $P\{5 \text{ on any trial}\} = \frac{4}{36}$ and $P\{7 \text{ on any trial}\} = \frac{6}{36}$.

  We now obtain, by independence, $P(E_n) = \left(1 - \frac{10}{36}\right)^{n-1} \frac{4}{36}$.

  Thus,

$$P(\bigcup_{n=1}^{\infty} E_n) = \frac{1}{9} \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} = \frac{1}{9} \frac{1}{1 - \frac{13}{18}} = \frac{2}{5}.$$

Example (Using Conditional Probabilities)

- If we let E be the event that a 5 occurs before a 7, then we can obtain the desired probability, $P(E)$, by conditioning on the outcome of the first trial, as follows:
  - Let F be the event that the first trial results in a 5;
  - Let G be the event that it results in a 7;
  - Let H be the event that the first trial results in neither a 5 nor a 7.

  Then, conditioning on which one of these events occurs gives

  $$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H).$$

  However, $P(E|F) = 1,\ P(E|G) = 0$ and $P(E|H) = P(E),$ the latter because if the first outcome results in neither a 5 nor a 7, then at that point the situation is exactly as it was when the problem first started (taking account of independence).
  Also, $P(F) = \frac{4}{36},\ P(G) = \frac{6}{36}$ and $P(H) = \frac{26}{36}$. Now we get:

  $$P(E) = \frac{1}{9} + P(E)\frac{13}{18} \Rightarrow \frac{5}{18}P(E) = \frac{1}{9} \Rightarrow P(E) = \frac{2}{5}.$$

Example

- There are n types of coupons, and each new one collected is independently of type i with probability p_i, $\sum_{i=1}^{n} p_i = 1$.

  Suppose k coupons are to be collected.

  If A_i is the event that there is at least one type i coupon among those collected, then, for $i \neq j$, find:

  (a) $P(A_i)$;
  (b) $P(A_i \cup A_j)$;
  (c) $P(A_i | A_j)$.

(a) Each coupon is, independently, not of type i with probability $1 - p_i$. $P(A_i) = 1 - P(A_i^c) = 1 - P\{\text{no coupon is type } i\} = 1 - (1 - p_i)^k$.

(b) Each coupon is, independently, neither of type i nor j with probability $1 - p_i - p_j$. Hence, $P(A_i \cup A_j) = 1 - P((A_i \cup A_j)^c) =$ $1 - P\{\text{no coupon is either type } i \text{ or type } j\} = 1 - (1 - p_i - p_j)^k$.

Example (Cont'd)

(c) To determine $P(A_i|A_j)$, we will use the identity

$$P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i A_j),$$

which yields

$$P(A_i A_j) = P(A_i) + P(A_j) - P(A_i \cup A_j).$$

Hence, taking into account Parts (a) and (b),

$$\begin{aligned} P(A_i A_j) &= 1 - (1 - p_i)^k + 1 - (1 - p_j)^k - [1 - (1 - p_i - p_j)^k] \\ &= 1 - (1 - p_i)^k - (1 - p_j)^k + (1 - p_i - p_j)^k. \end{aligned}$$

Consequently,

$$P(A_i|A_j) = \frac{P(A_i A_j)}{P(A_j)} = \frac{1 - (1 - p_i)^k - (1 - p_j)^k + (1 - p_i - p_j)^k}{1 - (1 - p_j)^k}.$$

The Problem of the Points

- Independent trials resulting in a success with probability p and a failure with probability $1-p$ are performed.
  What is the probability that n successes occur before m failures?
  Think of Players A and B playing a game.
  - A gains 1 point when a success occurs;
  - B gains 1 point when a failure occurs.

  The desired probability is the probability that A would win if the game were to be continued in a position where A needed n and B needed m more points to win.
  Pascal's Solution: Let $P_{n,m}$ be the probability that n successes occur before m failures.
  By conditioning on the outcome of the first trial, we obtain
  $$P_{n,m} = pP_{n-1,m} + (1-p)P_{n,m-1}, \quad n \geq 1, m \geq 1.$$
  Using the obvious boundary conditions $P_{n,0} = 0$, $P_{0,m} = 1$, we can solve these equations for $P_{n,m}$.

The Problem of the Points (Cont'd)

Fermat's Solution: Fermat argued that, in order for n successes to occur before m failures, it is necessary and sufficient that there be at least n successes in the first $m+n-1$ trials.

(Even if the game were to end before a total of $m+n-1$ trials were completed, we could still imagine that the necessary additional trials were performed.)

This is true, for if there are at least n successes in the first $m+n-1$ trials, there could be at most $m-1$ failures in those $m+n-1$ trials.

Thus, n successes would occur before m failures.

If, however, there were fewer than n successes in the first $m+n-1$ trials, there would have to be at least m failures in that same number of trials.

Thus, n successes would not occur before m failures.

The Problem of the Points (Cont'd)

- As shown in a previous example, the probability of exactly k successes in $m+n-1$ trials is

$$\binom{m+n-1}{k} p^k (1-p)^{m+n-1-k}.$$

It follows that the desired probability of n successes before m failures is

$$P_{n,m} = \sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k (1-p)^{m+n-1-k}.$$

The Probabilistic Method

- Suppose that we are given a set of elements.

  We want to determine whether at least one member of the set has a certain property.
- We can attack this question probabilistically:
  - Choose randomly an element of the set in such a way that each element has a positive probability of being selected;
  - Consider the probability that the randomly selected element does not have the property of interest.
    - If this probability is equal to 1, then none of the elements of the set have the property;
    - If it is less than 1, then at least one element of the set has the property.
- This method of attack is called the **probabilistic method**.

Example

- The **complete graph** having n vertices is defined to be a set of n points (called **vertices**) in the plane and the $\binom{n}{2}$ lines (called **edges**) connecting each pair of vertices.

  Suppose that each edge in a complete graph having n vertices is to be colored either red or blue.

  For a fixed integer k, a question of interest is whether there exists a way of coloring the edges so that no set of k vertices has all of its $\binom{k}{2}$ connecting edges the same color.

  It can be shown by a probabilistic argument that if n is not too large, then the answer is yes.

  Suppose that each edge is, independently, equally likely to be colored either red or blue.

  That is, each edge is red with probability $\frac{1}{2}$.

Example (Cont'd)

- Number the $\binom{n}{k}$ sets of k vertices by $i = 1, \ldots, \binom{n}{k}$. Define the events E_i, $i = 1, \ldots, \binom{n}{k}$, as follows:

$$E_i = \{\text{all of the connecting edges of the } i\text{th set of } k \text{ vertices are the same color}\}.$$

Each of the $\binom{k}{2}$ connecting edges of a set of k vertices is equally likely to be either red or blue.

It follows that the probability that they are all the same color is $P(E_i) = 2(\frac{1}{2})^{\frac{k(k-1)}{2}}$.

The probability that there is a set of k vertices all of whose connecting edges are similarly colored is $P(\bigcup_i E_i) \leq \sum_i P(E_i)$.

We now get

$$P(\bigcup_i E_i) \leq \binom{n}{k} \left(\frac{1}{2}\right)^{\frac{k(k-1)}{2} - 1}.$$

Example (Cont'd)

- We conclude that, if $\binom{n}{k}(\frac{1}{2})^{\frac{k(k-1)}{2}-1} < 1$, or, equivalently, if

$$\binom{n}{k} < 2^{\frac{k(k-1)}{2}-1},$$

then the probability that at least one of the $\binom{n}{k}$ sets of k vertices has all of its connecting edges the same color is less than 1.

Consequently, under the preceding condition on n and k, it follows that there is a positive probability that no set of k vertices has all of its connecting edges the same color.

This conclusion implies that there is at least one way of coloring the edges for which no set of k vertices has all of its connecting edges the same color.

Subsection 4

$P(\bullet|F)$ is a Probability

Conditional Probability and Axioms

Proposition

Let E, F be events in a sample space S.

(a) $0 \leq P(E|F) \leq 1$.

(b) $P(S|F) = 1$.

(c) If E_i, $i = 1, 2, \ldots$, are mutually exclusive events, then $P(\bigcup_1^\infty E_i|F) = \sum_1^\infty P(E_i|F)$.

(a) We must show that $0 \leq \frac{P(EF)}{P(F)} \leq 1$. The left-side inequality is obvious. For the right side, note $EF \subseteq F$. Hence, $P(EF) \leq P(F)$.

(b) We have $P(S|F) = \frac{P(SF)}{P(F)} = \frac{P(F)}{P(F)} = 1$.

(c) We have

$$\begin{aligned} P(\textstyle\bigcup_{i=1}^\infty E_i|F) &= \frac{P((\bigcup_{i=1}^\infty E_i)F)}{P(F)} = \frac{P(\bigcup_1^\infty E_iF)}{P(F)} \\ &= \frac{\sum_1^\infty P(E_iF)}{P(F)} = \textstyle\sum_1^\infty P(E_i|F), \end{aligned}$$

where the next-to-last equality follows because $E_iFE_jF = \emptyset$.

Additional Properties

- If we define $Q(E) = P(E|F)$, then, from the proposition, $Q(E)$ may be regarded as a probability function on the events of S.
- Hence, all of the propositions previously proved for probabilities apply to $Q(E)$.
- For instance, we have $Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1 E_2)$ or, equivalently,

$$P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F) - P(E_1 E_2|F).$$

Additional Properties (Cont'd)

- Suppose we define the conditional probability $Q(E_1|E_2)$ by
$$Q(E_1|E_2) = \frac{Q(E_1E_2)}{Q(E_2)}.$$
Then, we have
$$Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c).$$
- Note that
$$Q(E_1|E_2) = \frac{Q(E_1E_2)}{Q(E_2)} = \frac{P(E_1E_2|F)}{P(E_2|F)} = \frac{\frac{P(E_1E_2F)}{P(F)}}{\frac{P(E_2F)}{P(F)}} = P(E_1|E_2F).$$
Thus, the preceding equation is equivalent to
$$P(E_1|F) = P(E_1|E_2F)P(E_2|F) + P(E_1|E_2^cF)P(E_2^c|F).$$

Example

- An insurance company believes that people can be divided into two distinct classes: those who are accident prone (probability $\frac{3}{10}$) and those who are not.

  During any given year:
  - An accident-prone person will have an accident with probability 0.4;
  - The corresponding figure for a person who is not accident-prone is 0.2.

  What is the conditional probability that a new policyholder will have an accident in his or her second year of policy ownership, given that the policyholder has had an accident in the first year?
  - Let A be the event that the policyholder is accident prone.
  - Let A_i, $i = 1, 2$, be the event that he or she has had an accident in the ith year.

  The desired probability $P(A_2|A_1)$ may be obtained by conditioning on whether or not the policyholder is accident prone:

$$P(A_2|A_1) = P(A_2|AA_1)P(A|A_1) + P(A_2|A^cA_1)P(A^c|A_1).$$

Example

- Now,
$$P(A|A_1) = \frac{P(A_1 A)}{P(A_1)} = \frac{P(A_1|A)P(A)}{P(A_1)}.$$
It was shown in a previous example that $P(A_1) = 0.26$.
Hence,
$$P(A|A_1) = \frac{(0.4)(0.3)}{0.26} = \frac{6}{13}.$$
Thus, $P(A^c|A_1) = 1 - P(A|A_1) = \frac{7}{13}$.
Moreover, $P(A_2|AA_1) = 0.4$ and $P(A_2|A^c A_1) = 0.2$.
Hence,
$$P(A_2|A_1) = (0.4)\frac{6}{13} + (0.2)\frac{7}{13} \approx 0.29.$$

Example

- A female chimp has given birth.

  It is not certain, however, which of two male chimps is the father.

  Before any genetic analysis has been performed, it is felt that the probability that Male 1 is the father is p and $1-p$ for Male 2.

  DNA obtained from the mother, Male 1, and Male 2 indicate that, on one specific location of the genome:

  - The mother has the gene pair (A, A);
  - Male 1 has the gene pair (a, a);
  - Male 2 has the gene pair (A, a).

  If a DNA test shows that the baby chimp has the gene pair (A, a), what is the probability that Male 1 is the father?

Example (Cont'd)

- Let M_i be the event that Male i, $i = 1, 2$, is the father;
- Let $B_{A,a}$ be the event that the baby chimp has the gene pair (A, a).

Then

$$\begin{aligned} P(M_1|B_{A,a}) &= \frac{P(M_1 B_{A,a})}{P(B_{A,a})} \\ &= \frac{P(B_{A,a}|M_1)P(M_1)}{P(B_{A,a}|M_1)P(M_1) + P(B_{A,a}|M_2)P(M_2)} \\ &= \frac{1 \cdot p}{1 \cdot p + \frac{1}{2}(1-p)} = \frac{2p}{1+p}. \end{aligned}$$

Note that $\frac{2p}{1+p} > p$ when $p < 1$.

Thus, the information that the baby's gene pair is (A, a) increases the probability that Male 1 is the father.

Example (Theory of Runs)

- Independent trials, each resulting in a success with probability p or a failure with probability $q = 1 - p$, are performed.

  We are interested in computing the probability that a run of n consecutive successes occurs before a run of m consecutive failures.

  Let E be the event that a run of n consecutive successes occurs before a run of m consecutive failures.

  We start by conditioning on the outcome of the first trial.

  Let H denote the event that the first trial results in a success.

  $$P(E) = pP(E|H) + qP(E|H^c).$$

  Given that the first trial was successful, one way we can get a run of n successes before a run of m failures would be to have the next $n-1$ trials all result in successes.

  So, we condition on whether or not that occurs.

Example (Cont'd)

- Let F be the event that trials 2 through n all are successes.

$$P(E|H) = P(E|FH)P(F|H) + P(E|F^cH)P(F^c|H).$$

On the one hand, clearly, $P(E|FH) = 1$.
On the other hand, if the event F^cH occurs, then the first trial would result in a success, but there would be a failure some time during the next $n-1$ trials.
However, when this failure occurs, it would wipe out all of the previous successes, and the situation would be exactly as if we started out with a failure. Hence, $P(E|F^cH) = P(E|H^c)$.
Independence of trials implies that F and H are independent.
Moreover, $P(F) = p^{n-1}$.
Hence, from the first equation we get

$$P(E|H) = p^{n-1} + (1-p^{n-1})P(E|H^c).$$

Example (Cont'd)

- We now obtain an expression for $P(E|H^c)$ in a similar manner. Let G denote the event that trials 2 through m are all failures. Then

$$P(E|H^c) = P(E|GH^c)P(G|H^c) + P(E|G^cH^c)P(G^c|H^c).$$

 GH^c is the event that the first m trials all result in failures. So $P(E|GH^c) = 0$.
 Also, if G^cH^c occurs, then the first trial is a failure, but there is at least one success in the next $m-1$ trials.
 Hence, since this success wipes out all previous failures, we see that $P(E|G^cH^c) = P(E|H)$.
 But $P(G^c|H^c) = P(G^c) = 1 - q^{m-1}$.
 Therefore,

$$P(E|H^c) = (1 - q^{m-1})P(E|H).$$

Example (Cont'd)

- We solve the two derived equations:

$$P(E|H) = p^{n-1} + (1 - p^{n-1})P(E|H^c), \; P(E|H^c) = (1 - q^{m-1})P(E|H).$$

We get, by substitution,

$$\begin{aligned} P(E|H) &= p^{n-1} + (1 - p^{n-1})(1 - q^{m-1})P(E|H) \\ [1 - (1 - p^{n-1})(1 - q^{m-1})]P(E|H) &= p^{n-1} \\ [p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}]P(E|H) &= p^{n-1} \\ P(E|H) &= \frac{p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}. \end{aligned}$$

Now substitute into the second equation to get

$$P(E|H^c) = \frac{(1 - q^{m-1})p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}.$$

Example (Cont'd)

- Solving the two derived equations yields

$$P(E|H) = \frac{p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}},\ P(E|H^c) = \frac{(1-q^{m-1})p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}.$$

  Thus,

$$\begin{array}{rcl} P(E) & = & pP(E|H) + qP(E|H^c) \\ & = & \frac{p^n + qp^{n-1}(1-q^{m-1})}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} = \frac{p^{n-1}(1-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}. \end{array}$$

- Note that, by the symmetry of the problem, the probability of obtaining a run of m failures before a run of n successes is given by the same equation with p and q interchanged and n and m interchanged, i.e., by $\frac{q^{m-1}(1-p^n)}{q^{m-1}+p^{n-1}-q^{m-1}p^{n-1}}$.
  The two probabilities sum to 1. Thus, with probability 1, either a run of n successes or a run of m failures will eventually occur.

Example (Matching Revisited)

- At a party, n men take off their hats.

  The hats are then mixed up, and each man randomly selects one.

  We say that a **match** occurs if a man selects his own hat.

  What is the probability of:

  (a) no matches?

  (b) exactly k matches?

(a) Let E denote the event that no matches occur.

To make explicit the dependence on n, write $P_n = P(E)$.

Let M be the event that the first man selects his own hat.

Start by conditioning on M and M^c.

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c).$$

Clearly, $P(E|M) = 0$. So $P_n = P(E|M^c)\frac{n-1}{n}$.

Example (Cont'd)

- Now, $P(E|M^c)$ is the probability of no matches when $n-1$ men select from a set of $n-1$ hats that does not contain the hat of one of these men.

  This can happen in either of two mutually exclusive ways:
  - There are no matches and the extra man does not select the extra hat (this being the hat of the man who chose first);
  - There are no matches and the extra man does select the extra hat.

  The probability of:
  - the first of these events is P_{n-1}, which is seen by regarding the extra hat as "belonging" to the extra man;
  - the second of these events is $\frac{1}{n-1}P_{n-2}$.

  Hence, we have

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}.$$

Example (Cont'd)

- Substituting, we get $P_n = \frac{n-1}{n} P_{n-1} + \frac{1}{n} P_{n-2}$.
  Equivalently,
  $$P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2}).$$
  P_n is the probability of no matches when n men select among their own hats.
  Hence, $P_1 = 0$, $P_2 = \frac{1}{2}$.
  So, from the preceding equation,
  $$\begin{array}{lll} P_3 - P_2 = -\frac{P_2 - P_1}{3} = -\frac{1}{3!} & \Rightarrow & P_3 = \frac{1}{2!} - \frac{1}{3!}; \\ P_4 - P_3 = -\frac{P_3 - P_2}{4} = \frac{1}{4!} & \Rightarrow & P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}. \end{array}$$
  In general,
  $$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}.$$

Example (Cont'd)

(b) To obtain the probability of exactly k matches, we consider any fixed group of k men.

Denote by P_{n-k} the conditional probability of no matches among the remaining $n-k$ men given that they are selecting among their own hats.

The probability that the k men in the fixed group, and only they, select their own hats is

$$\frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k}=\frac{(n-k)!}{n!}P_{n-k}.$$

But there are $\binom{n}{k}$ choices of a set of k men.

Hence, the desired probability of exactly k matches is

$$\frac{P_{n-k}}{k!}=\frac{\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{n-k}}{(n-k)!}}{k!}.$$

Conditionally Independent Events

- Let E_1, E_2 and F be events in a sample space S.
- We say that the events E_1 and E_2 are **conditionally independent given** F if, given that F occurs, the conditional probability that E_1 occurs is unchanged by information as to whether or not E_2 occurs.
- More formally, E_1 and E_2 are said to be **conditionally independent given** F if
$$P(E_1|E_2F) = P(E_1|F).$$
- Equivalently, we have
$$P(E_1E_2|F) = P(E_1|F)P(E_2|F).$$

Example (Laplace's Rule of Succession)

- There are $k+1$ coins in a box.

  The ith coin turns up heads with probability $\frac{i}{k}$, $i = 0, 1, \ldots, k$.
  - A coin is randomly selected from the box;
  - The selected coin is 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?

  Denote by:
  - C_i the event that the ith coin, $i = 0, 1, \ldots, k$, is initially selected;
  - F_n the event that the first n flips all result in heads;
  - H the event that the $(n+1)$st flip is a head.

  The desired probability, $P(H|F_n)$, is obtained as

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

Example (Cont'd)

- Given that the ith coin is selected, it is reasonable to assume that the outcomes will be conditionally independent, with each one resulting in a head with probability $\frac{i}{k}$. Hence, $P(H|F_nC_i) = P(H|C_i) = \frac{i}{k}$.

  Also,

$$P(C_i|F_n) = \frac{P(C_iF_n)}{P(F_n)} = \frac{P(F_n|C_i)P(C_i)}{\sum_{j=0}^k P(F_n|C_j)P(C_j)} = \frac{(\frac{i}{k})^n \frac{1}{k+1}}{\sum_{j=0}^k (\frac{j}{k})^n \frac{1}{k+1}}.$$

  Thus, $P(H|F_n) = \frac{\sum_{i=0}^k (\frac{i}{k})^{n+1}}{\sum_{j=0}^k (\frac{j}{k})^n}$.

  If k is large, one can use the integral approximations.

$$\begin{array}{rcl} \frac{1}{k}\sum_{i=0}^k (\frac{i}{k})^{n+1} & \approx & \int_0^1 x^{n+1}dx = \frac{1}{n+2}; \\ \frac{1}{k}\sum_{j=0}^k (\frac{j}{k})^n & \approx & \int_0^1 x^n dx = \frac{1}{n+1}. \end{array}$$

  So, for k large, $P(H|F_n) \approx \frac{n+1}{n+2}$.

Updating Information Sequentially: Prior and Posterior

- Suppose there are n mutually exclusive and exhaustive possible hypotheses, with initial (sometimes referred to as **prior**) probabilities $P(H_i)$,
$$\sum_{i=1}^{n} P(H_i) = 1.$$
- If information that the event E has occurred is received, then the conditional probability that H_i is the true hypothesis (sometimes referred to as the **updated** or **posterior** probability of H_i) is
$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}.$$

Updating Information Sequentially (Cont'd)

- Suppose now that we learn first that E_1 has occurred and then that E_2 has occurred.
- Then, given only the first piece of information, the conditional probability that H_i is the true hypothesis is

$$P(H_i|E_1) = \frac{P(E_1|H_i)P(H_i)}{P(E_1)} = \frac{P(E_1|H_i)P(H_i)}{\sum_j P(E_1|H_j)P(H_j)}.$$

- Given both pieces of information, the conditional probability that H_i is the true hypothesis is $P(H_i|E_1E_2)$, which can be computed by

$$P(H_i|E_1E_2) = \frac{P(E_1E_2|H_i)P(H_i)}{\sum_j P(E_1E_2|H_j)P(H_j)}.$$

Updating Under Conditional Independence

- When is it legitimate to regard $P(H_j|E_1)$, $j \geq 1$, as the prior probabilities and then use the formula

$$P(H_i|E_1E_2) = \frac{P(E_2|H_i)P(H_i|E_1)}{\sum_j P(E_2|H_j)P(H_j|E_1)}$$

to compute the posterior probabilities?

The answer is this is legitimate, provided that, for each $j = 1, \ldots, n$, the events E_1 and E_2 are conditionally independent, given H_j.

In that case, for all $j = 1, \ldots, n$, $P(E_1E_2|H_j) = P(E_2|H_j)P(E_1|H_j)$.

Therefore, setting $Q(1,2) = \frac{P(E_1E_2)}{P(E_1)}$,

$$\begin{aligned} P(H_i|E_1E_2) &= \frac{P(E_2|H_i)P(E_1|H_i)P(H_i)}{P(E_1E_2)} = \frac{P(E_2|H_i)P(E_1H_i)}{P(E_1E_2)} \\ &= \frac{P(E_2|H_i)P(H_i|E_1)P(E_1)}{P(E_1E_2)} = \frac{P(E_2|H_i)P(H_i|E_1)}{Q(1,2)}. \end{aligned}$$

Updating Under Conditional Independence (Cont'd)

- Since the preceding equation is valid for all i, we obtain, upon summing,

$$1 = \sum_{i=1}^{n} P(H_i|E_1E_2) = \sum_{i=1}^{n} \frac{P(E_2|H_i)P(H_i|E_1)}{Q(1,2)}.$$

This shows that

$$Q(1,2) = \sum_{i=1}^{n} P(E_2|H_i)P(H_i|E_1).$$

Therefore

$$P(H_i|E_1E_2) = \frac{P(E_2|H_i)P(H_i|E_1)}{\sum_{i=1}^{n} P(E_2|H_i)P(H_i|E_1)}.$$

Significance & Advantages of Conditional Independence

- Suppose that one of two coins is chosen to be flipped.
- Suppose that when coin i is flipped, it lands on heads with probability p_i, $i = 1, 2$.
- Let H_i be the event that coin $i, i = 1, 2$, is chosen.
- Suppose we are to sequentially update the probability that coin 1 is the one being flipped, given the results of the previous flips.
- By the preceding result,

$$P(H_1|E_1E_2) = \frac{P(E_2|H_1)P(H_1|E_1)}{\sum_j P(E_2|H_j)P(H_j|E_1)},$$

  all that must be saved after each new flip is the conditional probability that coin 1 is the coin being used.
- That is, it is not necessary to keep track of all earlier results.