

Probability — the language of randomness

The field of statistics is concerned with decision making under uncertainty. We observe a random process (generically, an *experiment*), where multiple results (*outcomes* or *events*) are possible.

- (a) Flip a coin — heads or tails
- (b) The weather on a particular day — rainy, sunny, etc.
- (c) People arriving for class — the number of people, their genders, heights, weights, etc.

The set of all possible outcomes is called the *sample space*. The sample space depends on the context within which the process is observed.

- (a) Flip a coin?
- (b) Roll a die?
- (c) Roll two dice?

What do we mean by probability? The probability that an event will occur is the chance that it will happen. But what does that really mean? There are different kinds of “chance”:

- (1) What is the chance of drawing an ace from a well-shuffled deck of cards?
- (2) What is the chance that it will rain tomorrow?
- (3) What is the chance that cancer will be cured in the next fifty years?

Each of these is a valid probability question, but they clearly refer to different notions of probability. No one definition of probability can handle all of these possibilities, so different definitions have been developed as appropriate. All of these definitions have some things in common:

- (a) $0 \leq p \leq 1$
- (b) $p = 0$ means that the event never occurs
- (c) $p = 1$ means that the event always occurs

Equally likely rule

Consider a situation where you can enumerate all of the elements of the sample space. If there are n equally likely possibilities, and s are “favorable” to the event, then the probability of the event is

$$\frac{s}{n}.$$

- (1) The probability of drawing an ace from a well-shuffled deck: $4/52$
- (2) The probability of throwing a specific sum when tossing two dice:

		<i>Red die</i>					
		1	2	3	4	5	6
<i>White die</i>	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

What is the probability of throwing a 6? A “hard” 6 versus an “easy” one?

		<i>Red die</i>					
		1	2	3	4	5	6
<i>White die</i>	1					6	
	2				6		
	3			6			
	4		6				
	5	6					
	6						

But, what if events are not equally likely? What if they should be, but aren't?

Frequency theory

In situations where it is reasonable to think of a random process as being repeatable, a reasonable definition of the probability of an event is as the proportion of the time (the relative frequency) that similar events occur in the long run.

- (1) Probabilities from a loaded (unfair) die. Can a coin be unfair?
- (2) The weather (the probability of rain at any given point in the forecast area over the specified time period).

But, what if no “long run” can be defined? How do you define “similar events”?

Personal (subjective) probability

These definitions of probability cannot address the kinds of informal assessments of chance that we all do all the time. (Why don't you cross against a red light on a rainy

day? Because you think the chance that you'll slip, fall, and get run over is too high.) If you would bet \$a on the event occurring to get \$b, your personal probability that the event will occur is

$$\frac{a}{a+b}.$$

- (1) An “even money” bet (say, heads on the flip of a coin) says that you would bet \$1 on “heads” to get back your dollar plus one more dollar; that is, the probability is $1/(1+1) = 1/2$.
- (2) Parimutuel betting (horse racing, greyhound racing, jai alai).

But, different people will get different probabilities for the same events; this is inevitable, since each person's personal probability is strongly affected by that person's individual life experience. And, some people are risk averse, and don't like betting, while others are risk takers, and like it (maybe too much), which affects the bets that they are willing to take and make. We all make personal probability assessments, but sometimes they are not close to objective measures that are available.

All of these probability rules must satisfy certain conditions. In all cases, $0 \leq p \leq 1$. In addition, any probability rule must satisfy the *addition rule*: for any two events A and B ,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

(in any probability statement, “ A or B ” means “either A , or B , or both”). Say you roll one fair die. What is the probability of getting an odd number? an even number? at least a 4? both an odd and an even number (*mutually exclusive* events)? either an odd or an even number? either an odd number or at least a 4?

Conditional probability

In many situations, we are interested in several events, and how they relate to each other. For example, does training improve performance on the job? Does the weather affect retail sales? Does seeing a commercial change the chance that someone will purchase the advertised product? We are interested in *conditional probability* — the probability of one event **given** the occurrence of another event.

- (1) marginal probability: $P(\text{receive A in class})$
conditional probability: $P(\text{receive A in class} \mid \text{at least 90 on first exam})$
- (2) marginal probability: $P(\text{purchase product})$

conditional probability: $P(\text{purchase product} \mid \text{see commercial})$

It is natural to think in terms of time ordering, in that a conditional probability of $P(B|A)$ means that A occurred first, but that's not a requirement. For example, in discrimination lawsuits, an interesting probability is $P(\text{over age 50} \mid \text{were fired})$.

How should conditional probabilities be defined? Consider the following situation: say we observe 1000 shoppers, and notice the following pattern (after talking to them):

Saw commercial and bought product: 600

Saw commercial and didn't buy product: 50

Didn't see commercial and bought product: 250

Didn't see commercial and didn't buy product: 100

Define the events A and B as follows:

A : buying the product

B : seeing the commercial

If we consider these 1000 shoppers a "long run," what is

$$P(A)?$$

$$P(B)?$$

$$P(A \text{ and } B)?$$

What is

$$P(A|B),$$

the probability of buying the product *given* seeing the commercial?

It might be easier to remember what a conditional probability is referring to if you think of the event $A|B$ (" A given B ") as " A within B " or " A in B ," as that makes more explicit what the mechanism of conditional probability is: by conditioning on the event B , we are reducing the sample space to only B , and then asking what "part" of B also corresponds to A . This naturally implies the definition of conditional probability,

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

Generally, the definition of conditional probability results in a *multiplication rule*:

$$\begin{aligned} P(A \text{ and } B) &= P(A|B)P(B) \\ &= P(B|A)P(A) \end{aligned}$$

Note that this implies that the probability that both of two events occurs (sometimes called a *conjunction*) must be less than or equal to the probability that either occurs separately. A rearrangement of the terms in the two equalities above gives what is sometimes called *Bayes theorem*,

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

named after the Reverend Thomas Bayes, who discussed it in a posthumous paper presented to the Royal Society of London by Richard Price in 1763, and published the following year in the *Philosophical Transactions of the Royal Society of London*. Bayes theorem is a fundamental underpinning of personal (subjective) probability, as it describes how one's current views on the probability of an event B , the *prior* probability $P(B)$, can be updated based on new information (the event A occurring), to form the *posterior* probability $P(B|A)$.

Example: The book *The Job-generation Controversy: the Economic Myth of Small Business*, written by D. Hirschberg and published in 1999, argues that it is a myth that small firms play a major role in contributing to the growth of employment in the U.S. economy. Consider U.S. firms that were active in both 1989 and 1991. According to figures given in the book, 95.6% of all such firms were small firms in 1989, employing 100 or fewer people. Of firms that were large (employing more than 100 people) in 1989, 85.7% were large in 1991; of firms that were small in 1989, 0.3% were large in 1991. Consider a particular firm that was a small firm in 1991. What is the probability that the firm was large in 1989?

It is often the case that calculations like this are done most easily using a table, and imagining a large number of replications of the process (the “hypothetical 100,000”). Say there was a total of 100,000 firms active in both 1989 and 1991.

	<i>Small in 1991</i>	<i>Large in 1991</i>	
<i>Small in 1989</i>		(.003)(95600) = 287	95600
<i>Large in 1989</i>		(.857)(4400) = 3770	4400
			100000

What is $P(\text{Large in 1989} | \text{Small in 1991})$?

Independence

Sometimes, whether one event occurs doesn't depend on whether another occurs. For example, the probability of passing a statistics exam is the same, whether you wear blue socks or not; the probability of it raining is unrelated to the day of the week (although it doesn't always seem that way!). When the knowledge that one event has occurred (or will occur) gives no additional information about whether another event will occur, we say that the two events are *independent*. More formally, two events A and B are independent if and only if

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

(A and B are arbitrary here, so we could just as well write $P(B|A) = P(B)$). Equivalently (using the multiplication rule), two events A and B are independent if and only if

$$P(A \text{ and } B) = P(A)P(B).$$

- (1) Consider the Daily Numbers game, where an integer in the range 000 through 999 is chosen as the winning number. The way the number is chosen is as follows. There are three separate machines, each of which has 10 ping pong balls (numbered 0 through 9) rapidly bouncing around. The number is chosen by choosing one ball from each machine in a specified order. Does this make all 1000 possible numbers equally likely?
- (2) Pennsylvania lottery fraud; Arizona lottery mistake.
- (3) $P(\text{rain today}) = .4$, $P(\text{rain tomorrow}) = .8$. What is the probability that you'll need a raincoat sometime over the next two days? How about

$$\begin{aligned} P(\text{need raincoat}) &= P(\text{rain today or rain tomorrow}) \\ &= P(\text{rain today}) + P(\text{rain tomorrow}) \\ &\quad - P(\text{rain today and rain tomorrow}) \\ &= P(\text{rain today}) + P(\text{rain tomorrow}) \\ &\quad - P(\text{rain today})P(\text{rain tomorrow}) \\ &= .4 + .8 - (.4)(.8) \\ &= .88? \end{aligned}$$

Note that the independence characteristics of A and B are the same as those involving the complement of A (A') and the complement of B (B'); that is, for example, A and B being independent is equivalent to A' and B' being independent.

It is important to remember that the association between two random variables (or the lack of association, as when they are independent) can only be seen in the *joint* distribution of the two variables (that is, the complete set of probabilities for all pairs of outcomes from each random variable); the *marginal* distributions of each random variable provide **no** information about association.

Appendix — Background material on probability

Historical background

The first systematic examination of probability was prompted by a request from a gambler, Antoine Gombauld, the Chevalier de Méré, Sieur de Baussay, to a mathematician, Blaise Pascal, in 1654, about how to bet on dice games. Unfortunately, these origins led to the field being viewed with suspicion for many years by the Catholic Church, which attempted to have the study of probability curtailed. The rule that Pascal came up with (after discussion with mathematician Pierre de Fermat) was the equally likely rule. The formal structure of probability theory was first laid out in the 1790s by Pierre Simon, Marquis de Laplace (earlier notable work was done by Jacob Bernoulli in the early 1700s). One aspect of that theory (further developed by Jerzy Neyman and Egon Pearson in the 1920s) was the frequency theory of probability. Leonard J. (“Jimmie”) Savage constructed the formal definition of subjective probability in 1954 through the use of betting calculations.

Independence

- (1) A drug test is administered to someone who is actually not a drug user. Say each time a drug test is administered, there is probability .1 of a false positive result (a nonuser getting a positive test result). Should the test be administered two times instead, since then the probability of a false positive on both tests is $(.1)(.1) = .01$? No, because it is likely that the outcomes of the two tests are not independent. Some people are subject to *serum specificity*, wherein they consistently respond positively to a test even if they shouldn't, so repeating the test doesn't accomplish anything. This is why followup tests after a positive result should be based on a different biochemical mechanism.
- 2(a) A fair coin is flipped 100 times, with 75 heads resulting. Is the next flip more likely to be heads? tails? neither? The answer, of course, is neither.
- (b) A coin is flipped 100 times, with 75 heads resulting. Is the next flip more likely to be heads? tails? neither? Since it wasn't specified that the coin is fair, the answer here should be heads.

Note, however, that the answer is never tails, debunking any notion of a tails being "due."
- (3) We can now go back to the original question that the Chevalier de Méré asked Pascal in 1654. De Méré found that he tended to win when betting even money that a six would come up at least once in four rolls with a single die, but that he tended to

lose when betting even money that a double-six would come up at least once in 24 rolls with two dice, and was surprised by this. He reasoned that four rolls with six possibilities in each roll would be the same as 24 rolls with 36 possibilities in each roll. Pascal showed him that this was not the case:

At least one six in four rolls of a single die

$$\begin{aligned} P(\text{At least one six}) &= 1 - P(\text{no sixes on all rolls}) \\ &= 1 - P(\text{no six on any one roll})^4 \\ &= 1 - \left(\frac{5}{6}\right)^4 = 1 - .48 = .52 \end{aligned}$$

At least one double-six in 24 rolls of two dice

$$\begin{aligned} P(\text{At least one double - six}) &= 1 - P(\text{no double - sixes on all rolls}) \\ &= 1 - P(\text{no double - six on any one roll})^{24} \\ &= 1 - \left(\frac{35}{36}\right)^{24} = 1 - .51 = .49 \end{aligned}$$

Thus, de Méré was able to see (on the basis of long-run frequencies) that the first probability was greater than .5, while the second was less than .5. Imagine how many times he must have bet, and with how much concentration he must have noticed the results, to figure this out! In fact, some historians of science have claimed that this was only a theoretical question for de Méré, since he couldn't possibly have played enough, and as attentively, to detect these patterns.

Joint distributions

Here's an example of how marginal distributions don't reflect associations. The data refer to the movie reviews of Roger Ebert and the late Gene Siskel, the two well-known Chicago movie reviewers, as recorded in *Variety* from April 1995 through September 1996. Each of 160 movies were reviewed by Siskel and Ebert, and their opinions were recorded as "Con," "Mixed," or "Pro." The marginal distributions for the two reviewers were as follows:

	<i>Con</i>	<i>Mixed</i>	<i>Pro</i>
Siskel	45	32	83
Ebert	42	30	88

Obviously the two marginal distributions are very similar to each other. You might think that this implies that Siskel and Ebert's opinions were similar to each other; that is,

the review responses were associated with each other, and definitely not independent of each other. This is **not** the case, however, as it is possible to construct joint distributions consistent with these marginal distributions that exhibit a wide range of associations. Here are three examples. Note that I've given these distributions as possible observed sample counts; to convert them to probability distributions, you would divide each entry by the total number of reviews, or 160.

		Ebert review			
		<i>Con</i>	<i>Mixed</i>	<i>Pro</i>	
Siskel review	<i>Con</i>	12	8	25	45
	<i>Mixed</i>	8	6	18	32
	<i>Pro</i>	22	16	45	83
		42	30	88	160

This table is consistent with independence between Siskel and Ebert's reviews. The way to see this is to note that the relative distribution of Siskel's reviews is the same, no matter what Ebert's review was; in every case, about 28% of Siskel's reviews are "Con", 20% are "Mixed," and 52% are "Pro," which are the overall proportions.

Now consider this table:

		Ebert review			
		<i>Con</i>	<i>Mixed</i>	<i>Pro</i>	
Siskel review	<i>Con</i>	42	0	3	45
	<i>Mixed</i>	0	30	2	32
	<i>Pro</i>	0	0	83	83
		42	30	88	160

This table reflects almost perfect agreement between Siskel and Ebert. In contrast, the following table reflects almost perfect disagreement between them:

		Ebert review			
		<i>Con</i>	<i>Mixed</i>	<i>Pro</i>	
Siskel review	<i>Con</i>	0	0	45	45
	<i>Mixed</i>	0	0	32	32
	<i>Pro</i>	42	30	11	83
		42	30	88	160

Thus, a wide range of associations, from strong agreement to independence to strong disagreement, are all consistent with the same marginal distributions. The only way to assess if two random variables are related to each other is to look at the joint distribution of the two random variables.

The following table gives the actual joint distribution of movie reviews. As you can see, there is reasonably strong agreement between the two reviewers, although in more than one-third of the movies they did disagree.

		Ebert review			
		<i>Con</i>	<i>Mixed</i>	<i>Pro</i>	
Siskel review	<i>Con</i>	24	8	13	45
	<i>Mixed</i>	8	13	11	32
	<i>Pro</i>	10	9	64	83
		42	30	88	160

Here’s an example of how independence conditions can imply a perhaps unintuitive result. Consider a sporting event that is divided into two halves, with the final score for a team being the total of the scores for each half. Examples of such events include soccer, football and basketball. Assume that the two teams are precisely evenly matched. Also, assume that the result of play in the first half does not influence play in the second half; that is, teams don’t play harder because they’re behind, or give up because they’re behind. Given that a team is ahead at halftime, what is the probability that they will win the game?

We will ignore here the possibility of ties (so this simplified argument is least applicable to soccer, and most applicable to basketball). The intuitive answer would seem to be .5, with the argument being that if team A happens to be ahead by 5 points at the end of the first half (i.e., they “win” the first half), since the two teams are equally matched, it’s just as likely that team B will be ahead by more than 5 points for the second half (i.e., they “win” the second half), and thus wins the game. Is this the correct way to look at the problem?

Let’s write out the possible outcomes of a game. In the table below, “A” means that team A won the half, while “B” means that team B won the half. The possible outcomes of the game can then be summarized as follows:

Winner?	
First half	Second half
A	A
A	B
B	A
B	B

Since the teams are evenly matched, the probability of A or B being winning either

half is .5. Since the results of the first and second halves are independent, the probability of each of the four possibilities above is .25.

- (1) When either team A wins both halves, or team B wins both halves, clearly they have won the game, so the results AA and BB are consistent with the team that is ahead at halftime winning the game.
- (2) Say situation AB occurs. Since teams A and B are evenly matched, it is equally likely that the difference favoring A in the first half will be larger than the difference favoring B in the second half, as that it will be smaller than the difference favoring B in the second half. That is, given that AB occurs, half of the time A will win the game, and half of the time B will win the game.
- (3) The same argument as in (2) applies if situation BA occurs.

Thus,

$$\begin{aligned} P(\text{Team ahead at halftime winning game}) &= P(\text{AA}) + P(\text{BB}) + \left(\frac{1}{2}\right) P(\text{AB}) \\ &\quad + \left(\frac{1}{2}\right) P(\text{BA}) \\ &= .25 + .25 + \left(\frac{1}{2}\right) (.25) + \left(\frac{1}{2}\right) (.25) \\ &= .75 \end{aligned}$$

That is, when two evenly matched teams play, the team that is ahead at halftime can be expected to ultimately win the game 75% of the time. Note that this result has nothing to do with teams giving up at halftime, or becoming overconfident, or any other possibility of this type, since we have assumed that the first and second half results are independent. If such factors were important, we might expect the actual proportion of teams that win when ahead at halftime to be different from .75 (for example, perhaps it would be larger, if you think being ahead gives a psychological advantage to a team). Another reason that the actual proportion might be larger than .75 is if there was great disparity in quality of teams, since then the probability of one team winning both halves is greater (can you prove that the probability that one team wins both halves is minimized when the two teams are evenly matched?). Professor Hal Stern of the University of California at Irvine examined 493 National Basketball Association games from January to April 1992, and found that in 74.8% of the games, the team that was ahead at halftime ultimately won the game.

Professor Stern extended this model to allow investigation of this effect at points other than halftime. His model implies that the probability that the team that is ahead after

one quarter wins the game is $\frac{2}{3}$, while the probability that the team that is ahead after three quarters wins the game is $\frac{5}{6}$. These can be compared to the observed proportions of .667 and .821, respectively.

“The more numerous the observations and the less they vary among themselves, the more their results approach the truth.”

— Pierre Simon, Marquis de Laplace

“In the long run, we are all dead.”

— John Maynard Keynes

“The theory of probability is at bottom nothing but common sense reduced to calculus.”

— Pierre Simon, Marquis de Laplace

“I shall never believe that God plays dice with the universe.”

— Albert Einstein

“All Nature is but Art, unknown to thee;
All Chance, Direction, which thou canst see;
All Discord, Harmony not understood.”

— Alexander Pope

“If there is a 50–50 chance that something can go wrong, then 9 times out of 10 it will.”

— Paul Harvey

“A fool must now and then be right by chance.”

— William Cowper

“If you bet on a horse, that’s gambling. If you bet you can make three spades, that’s entertainment. If you bet cotton will go up three points, that’s business. See the difference?”

— Blackie Sherrod

“It is impossible to calculate accurately events that are determined by chance.”

— Thucydides

“But to *us*, probability is the very guide of life.”

— Bishop J. Butler

“We shall have to get used to thinking of chance not as an irritating obstacle, not as an ‘inessential adjunct to phenomena’ (as is in the philosophical dictionary), but as a source of unlimited possibilities of which even the boldest imagination can have no prescience.”

— Leonard A. Rastrigin

“A probability is a generally approved proposition; what men know to happen or not to happen, to be or not to be, for the most part thus and thus is a probability. It is the nature of probability that improbable things will happen.”

— Aristotle, *Analytica Priora*

“It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge . . . The most important questions of life are, for the most part, really only problems of probability.”

— Pierre Simon, Marquis de Laplace

“The greatest advantage from gambling comes from not playing it at all.”

— Girolamo Cardano (*Book on Games of Chance*)

“Where observation is concerned, chance favours only the prepared mind.”

— Louis Pasteur

“We know that people tend to overestimate the frequency of well-publicized, spectacular events compared with more commonplace ones; this is a well-understood phenomenon in the literature of risk assessment and leads to the truism that when statistics plays folklore, folklore always wins in a rout.”

— Donald Kennedy

“This is plain truth: everyone ought to keep a sharp eye on main chance.”

— Plautus

“We are hard-wired to react to coincidences. It goes back to primitive man. You look in the bush, it looks like stripes, you’d better get out of there before you determine the odds that you’re looking at a tiger. The cost of being flattened by the tiger is high.”

— Persi Diaconis

“Probability is too important to be left to the experts.”

— Richard W. Hamming

“Chance is a nickname for Providence.”

— Sebastien Roch Nicolas Chamfort

“ . . . Chance governs all.”

— John Milton (*Paradise Lost*)

“Nay, what should mortal fear, for whom the degree of fortune are supreme, and who hath clear foresight of nothing? 'Tis best to live at random, as one may.”

— Sophocles (*Ædipus the King*)

“Nothing is certain, and that’s for sure.”

— Marcel P.R. Van den Broecke

“Rowe’s Rule: the odds are six to five that the light at the end of the tunnel is the headlight of an oncoming train.”

— Paul Dickson

“The lot is cast into the lap; but the whole disposing thereof is of the Lord.”

— *Proverbs* 16:33

“The chances for the baby’s chromosomes being defective were 250 to 1, which sounds like reasonable odds. Except that all odds are, in reality, 50–50: it may happen and it may not.”

— *The London Times Magazine*, 2 August 1997

“Our brains are just not wired to do probability problems very well.”

— Persi Diaconis

“I long ago came to the conclusion that all life is 6 to 5 against.”

— Damon Runyon

“From where we stand the rain seems random. If we would stand somewhere else, we would see the order in it.”

— Tony Hillerman (*Coyote Waits*)

“The race is not always to the swift, nor the battle to the strong — but that’s the way to bet.”

— Damon Runyon