Maths for Economics

10. Introduction to probability

Vera L. te Velde

31. August 2015

1 Introduction

Flipping coins, rolling dice, and spinning blackjack tables. These are classic random processes designed such that we can fully understand their behavior, yet the same tools for understanding these simple processes also apply to long-tail macroeconomic risks, movements in stock prices, pricing of insurance, and most other economic phenomena involving risk. Probability and statistics is an enormous field but just a few concepts are enough to allow you to analyze a huge number of situations that people instinctively get wrong.

2 Notation

As usual, there are new ways of writing concepts very tersely. Here are some to keep an eye out for:

1. x: a single event whose probability we are interested in. E.g., a coin flip coming up heads.
2. X: a set of events whose probabilities, either separate or joint, we might be interested in. E.g., a dice roll coming up even.
3. $P(x)$: The probability that event x will happen.
4. $P(x|y)$: The probability that event x will happen, given that event y happens.
5. $P(X)$: The probability that some event x in the set of possible events X will happen.
6. X^c: The **negation** of the set X. $P(X^c)$ is the probability that X does *not* happen.

And if we think of events as sets of possibilities, we can combine set notation with probability notation to write things like $P(X \cap Y)$, which is the probability that *both* X and Y happen (that is, an event happens that is a member of both sets X and Y, such as rolling a 2 if X is "rolling an even number" and Y is "rolling a number less than 3."

3 Probabilities

The probability of an event is the fraction of times that you can expect the event to happen, if you repeat the same situation many times. The probability of flipping a coin and getting heads is 50%. The probability of rolling a die and getting a 1 or 2 is 1/3.

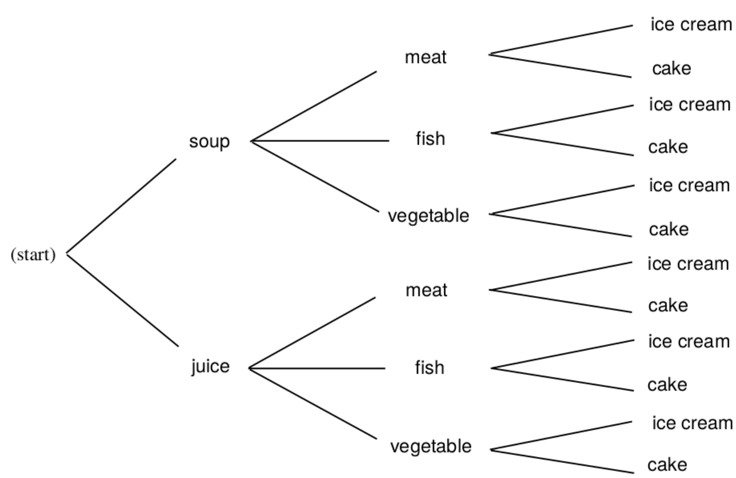
Probabilities necessarily reflect long-run averages. It's entirely possible for five coin flips to all come up heads; in fact, this outcome is just as likely as any other outcome and happens 1 in 32 times. But, if we flip a coin millions of times, the fraction of times we get heads should be very close to 1/2.

This is the **law of large numbers**: If you repeat an experiment more and more times, the measured probability of an event will get closer and closer to the true probability.

3.1 Counting

If the events we're interested in are built up out of individual outcomes that are equally probably and mutually exclusive, such as numbers on a dice, we can use counting methods to calculate probabilities.

The basic idea of counting is very simple. If there are m_1 possibilities in the first stage of an event, and each of those leads to m_2 possibilities in the second stage, and each of those leads to m_3 possibilities in the third stage, and so on... then there are $m_1 \cdot m_2 \cdot m_3 \cdots$ total possibilities. For example, if there are two appetizers, 3 main courses, and two deserts on the menu, there are a total of $2 \cdot 3 \cdot 2$ possible meals.



The birthday paradox is a good example of this process. If there are 25 people in a room, what is the probability that two of them share a birthday? To solve this problem we have to count how many possible ways two people can share a birthday, and divide by the total number of possible birthdays. The latter is simple: each person has one of 365 birthdays. Two people have one of $365 \cdot 365 = 365^2$ pairs of birthdays, and 25 people have one of 365^{25} sets of birthdays.

The former is slightly trickier. The list of 25 birthdays in which at least two of them are the same is very long and hard to write down, but counting the number of ways that all birthdays are different is easier. The first person has one of 365 birthdays. In order to avoid a collision, the 2nd person must have one of the 364 other birthdays. And in order to avoid a collision with the first two, the 3rd person must have one of the 363 other birthdays. And so on. Altogether there are $365 \cdot 364 \cdot 363 \cdots 341 = \prod_0^{24}(365 - i)$ sets of birthdays with no collisions, and the remaining $365^{25} - \prod_0^{24}(365 - i)$ must have at least one collision. This, divided by the total 365^{25}, yields a probability of 57% of a shared birthday.

3.1.1 Permutations

Permutations are useful for counting arrangements of possibilities. We actually used permutations in the birthday problem above: we counting all of the arrangements of any 25 out of 365 possible birthdays. This is sometimes written as ${}_{365}P_{25}$. In general:

$$ {}_nP_k = n \cdot (n-1) \cdot (n-2) \cdot s(n-k+1) $$

This can be more compactly written with **factorial** notation:

$$ n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 $$

And, 0! is defined to be 1. With this we can rewrite ${}_nP_k = \frac{n!}{(n-k)!}$

k is the number of objects we are putting into an arrangement, and n is the number of options we have to choose from. With $k = n$, we calculate how many ways there are to arrange *all* n choices. There are, for example, $3! = 6$ ways to arrange the numbers 1, 2, and 3:

$$ 1,2,3 \quad 1,3,2 \quad 2,1,3 \quad 2,3,1 \quad 3,1,2 \quad 3,2,1 $$

Notice that permutations count how many ways there are to slot items into position in a specific order. ${}_3P_2 = 3!/1! = 6$ counts the ways to choose two out of three elements and put them in a specific order:

$$ 1,2 \quad 1,3 \quad 2,1 \quad 2,3 \quad 3,1 \quad 3,2 $$

3.1.2 Combinations

Combinations, on the other hand, are useful when we don't care about the order. They let us count how many different ways we can pick out a certain number of items from a set, regardless of what order we choose them in. There are, for example, exactly 3 ways we can choose 2 items from the set $\{1, 2, 3\}$:

$$ 1,2 \quad 2,3 \quad 1,3 $$

The general formula is:

$$ {}_nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!} $$

This makes sense because it is ${}_nP_k$ divided by the number of ways the k chosen elements might be rearranged. Notice the parenthetical notation for combinations, which is more common than the capital C.

4 Laws of probability

If we want to move away from cases that can be analyze by counting possibilities, we need to incorporate probabilities into our diagrams of possibilities. Just between a situation has two possible outcomes doesn't mean that they are equally probably.

There are a few intuitive laws of probability that apply in *any* situation:

4.1 Probabilities add up to 1

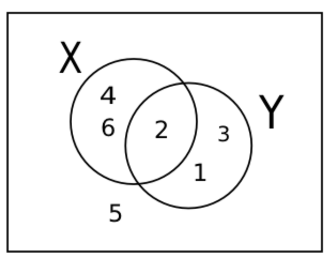
There is always a 100% probability that *something* will happen (either the coin will come up heads or it will come up tails; there's no way it can fail to come up anything), so the total probability of X and X^c has to be 1, no matter what X is.

$$P(X^c) = 1 - P(X)$$

Notice also that the probability of anything always must be between 0 and 1.

4.2 Probabilities of unions

Similarly, if we want to know the probability that a die roll will come up *either* even ($X = \{2, 4, 6\}$) or less than 4 ($Y = \{1, 2, 3\}$), we can illustrate the possibilities with a Venn diagram:

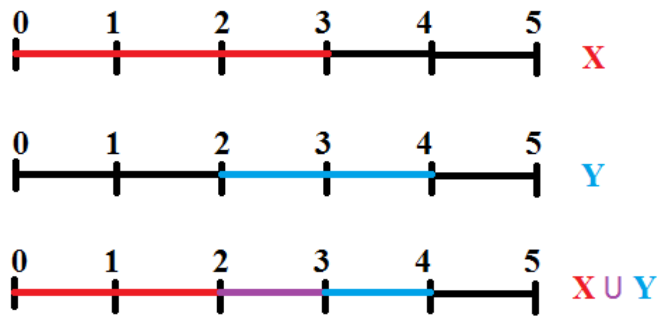


Probability, unlike simple sets, is a measure of how big those sets are. If we add together the side of the sets X and Y, we are double counting their overlap. So we have:

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

In our example, this is $.5 + .5 - .1\overline{6} = 0.\overline{6}$, or 2/3.

Another way of looking at this is to think about events that we can draw with a number line directly. For example, if we pick any random real number between 0 and 5, what's the chance that it will be between 0 and 3 ($P(\{x \in X = [0, 3]\})$? 60%. And what's the chance it will be between 2 and 4 ($P(\{x \in Y = [2,4]\})$? 40% And what is $P(X \cup Y) = P(\{x \in [0,4]\})$? It's 80%, which happens to be $P(X) + P(Y) - P(X \cap Y)$:



For more than two events, this rule can be generalized:

$$P(\cup_{i=1}^{n} X_i) = 1 - P(\cap_{i=1}^{n} X_i^c)$$

This formula first adapts the notation for set unions to combine many sets at once, sort of like the summation symbol $\sum$ adds many things together at once. Specifically,

$$\cup_i^n X_i = X_1 \cup X_2 \cup X_3 \cup \cdots \cup X_n$$

and

$$\cap_i^n X_i = X_1 \cap X_2 \cap X_3 \cap \cdots \cap X_n.$$

Interpreting the probability formula, this thus says that the probability of *any* of many events is 1 minus the probability that all of them *don't* happen. This makes sense because the only way we might *not* end up in the situation where one of the events happened, is for none of them to happen.

4.3 Nested events

An important consequence of this simple law is that the probability of two things happening simultaneously is always less than (or equal to) the probability of just one thing happening. But this is a common intuitive mistake in everyday situations!

Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. Which of the following is more probable?

1. Linda is a bank teller.

2. Linda is a bank teller and is active in the feminist movement.

While many people intuitively feel that the latter feels more like something Linda would be, simple probability means that the former must be more likely, simply because it contains the second possibility entirely.

5 Conditional probability

Another law of probability tells you how to update your probabilities of other events once one event has already happened:

$$P(X|Y) = P(X \cap Y)/P(Y)$$

What this says is that once Y has happened, we have to calculate the probability of X within the subset of possibilities in which Y also happens. The universe of possibilities gets smaller, from 1 to $P(Y)$, and then everything proceeds as usual.

For example, what is the probability that you roll a 1 given that you have rolled an odd number? Our universe of possibilities now excludes $\{2, 4, 6\}$. Of the three possibilities remaining there is a 1/3 chance of rolling a 1. More formally, $P(1|\text{odd}) = P(1 \text{ and odd})/P(\text{odd}) = (1/6)/(1/2) = 1/3$.

Another way of writing this rule is also very useful:

$$P(X|Y)P(Y) = P(X \cap Y)$$

simply because oftentimes we know the conditional probability and want to figure out the joint probability. For example, if I know that the probability of going bankrupt if the stock market crashes is .2, and I know that the probability of a crash is .05, then I can forecast the probability that I will be bankrupt due to market conditions as $.2 * .05 = .01$

5.1 Total probability

The concept of conditional probability also lets us analyze the world of possibilities by breaking them into different categories. If we can partition the possibilities in **mutually exclusive** events (that can never happen at the same time, such as rolling a 2 and rolling a 3 on a die, or rolling a 2 and *not* rolling a 2) labeled $\{A_i\}$, then $P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3) + \cdots = \sum_{i=1}^{n} P(B|A_n) \cdot P(A_n)$.

For example, division 1 in a company contains 20% of employees, and 80% of them are satisfied with their jobs. Division 2 is 40% of the company and 50% satisfied, and division 3 is 20% of the company and 75

5.2 Independent events

Each time you flip a coin, the result is completely independent of any other coin flip that has previously happened. Even if the coin has come up heads the last 10 times in a row, the chance of getting another head is still 50%. Events like this, that don't influence each other at all, are called **independent** events. If we represent the possibilities as h and t for heads and tails, we can write $P(h) = P(t) = 50\%$.

If we have two independent sets of events X and Y, the occurrence of one has no effect on the probability of the other. That is $P(X|Y) = P(X)$ and $P(Y|X) = P(X)$. If we look back to our definition of conditional probability above, this gives us the most famous and useful attribute of independent events:

$$P(X \cap Y) = P(X)P(Y)$$

.

The probability of flipping two heads simultaneously is 25%. The probability of flipping heads, heads, heads, heads, heads is $\left(\frac{1}{2}\right)^5 = \frac{1}{32}$, and the probability of flipping heads, tails, heads, heads, tails is $\left(\frac{1}{2}\right)^5 = \frac{1}{32}$. This is a common misconception: independent events *cannot* influence each other.

Independent events *don't* add together though. In an extreme example this is obvious: if $P(x) = .75$ and $P(y) = .5$ and x and y are independent, it's not true that $P(x \cup y) = 1.25$ because probabilities must be between 0 and 1. But we do know that $P(x \cap y) = .5 * .75 = .375$, and the rule for unions of events tells us that $P(x \cup y) = P(x) + P(y) - P(x \cap y) = .5 + .75 - .375 = .875$.

In finance, this pattern is very useful, especially if we use the general version of the law of probabilities of unions of events. Say there is an (independent) 5% probability of a business going bankrupt each year. Then after 5 years, what is the probability it has gone bankrupt? The only way to avoid this outcome is to avoid going bankrupt every single year:

$$\begin{aligned}
P(\text{bankrupt in 5 years}) &= 1 - P(\text{avoid bankruptcy every year}) \\
&= 1 - P((\text{avoid in year 1}) \cap (\text{avoid in year 2}) \cdots) \\
&= 1 - P(\text{avoid in year 1}) \cdot P(\text{avoid in year 2}) \cdots \\
&= 1 - .95 \cdot .95 \cdot .95 \cdot .95 \cdot .95 \\
&= 1 - .95^5 = 22.6\%
\end{aligned}$$

5.3 Dependent events

Most events in the world are intertwined, however, and then we must use the standard definition of conditional probabilities. A company may normally have only a 5% chance of going bankrupt, and the stock market may crash with only a 2% probability, but if the stock market has already crashed, the likelihood of bankruptcy is now much higher. In the most extreme case, the probability of rolling an even number on a die given that you roll an odd number is 0, or the probability of rolling an even number given that you roll a 2 is 1.

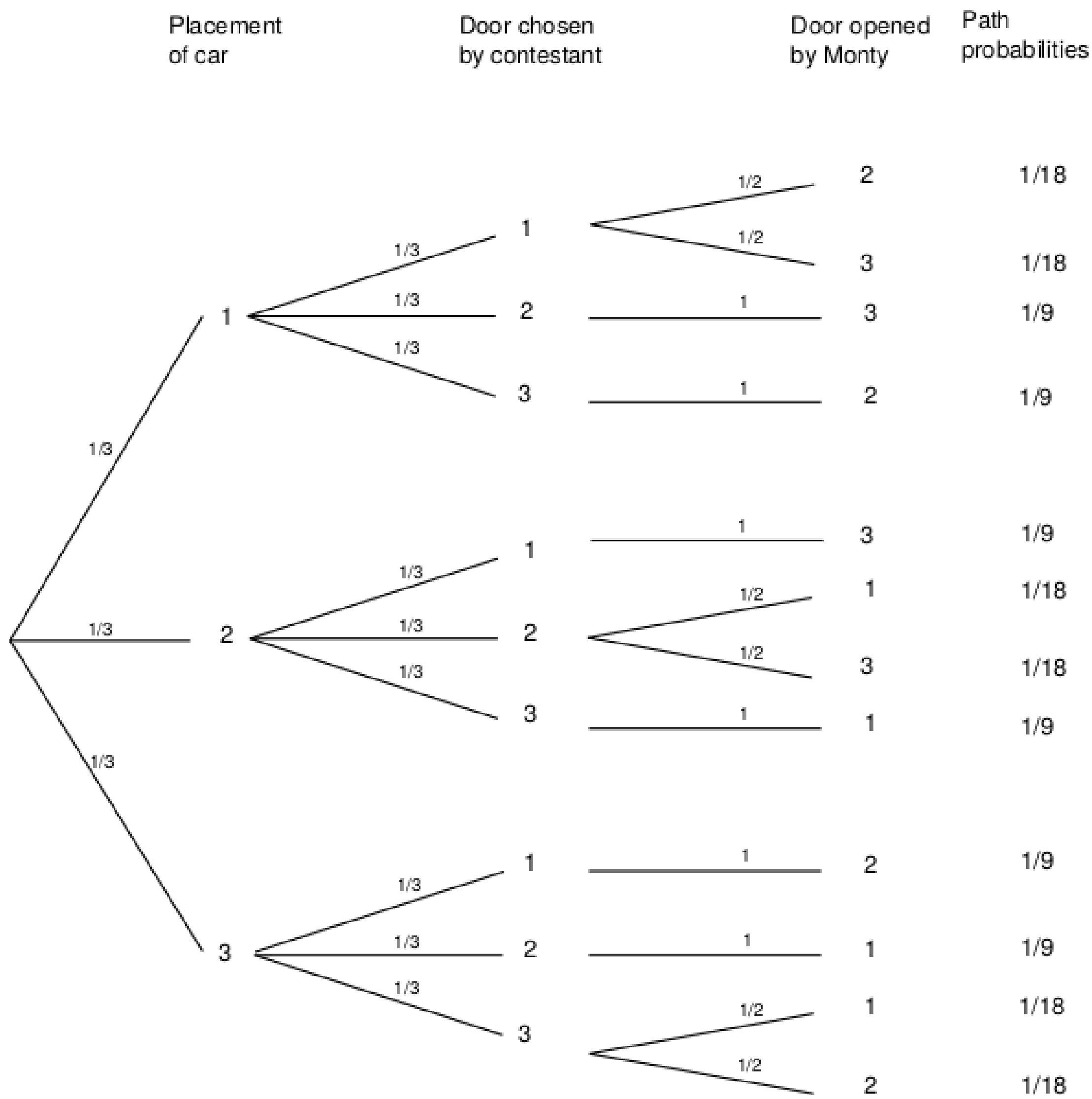
The best way to think about these problems is to carefully keep track of the order of events, their possible outcomes, and probabilities, and then to track what fraction of the time each path will be taken.

For example, if a couple has two kids, what is the probability that they are both boys if at least one of them is a boy?

A famous example is the Monty Hall problem. This is named after a gameshow in which players first guess which of three doors hides a new car. After guessing, the host opens one of the other two doors that does not contain the car. Finally, the player can choose to stick with their original guess or to change guesses before the answer is revealed.

What is your probability of winning if you switch, or if you don't switch?

The following tree traces out all of the possibilities:



5.4 Bayes's rule

One of the most important rules of probability is Bayes' rule. Bayes' rule allows us to use imperfect diagnostic tests to learn about the true state of the world. For example, we can use medical tests that sometimes give false results to learn about the true probability that someone is sick; we can run imperfect experiments and use our noisy results to learn about what's really true in the world. The formula is simple:

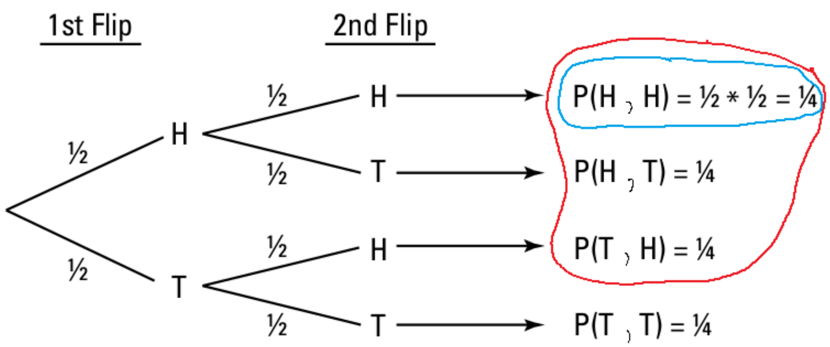
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

But, rather than memorizing the formula, it's easy to accomplish the same thing by using an approach like the one we took for the Monty Hall problem. Draw a tree of all of the possibilities, and keep track of what fraction end up producing A, as a fraction of the possibilities that also produce B.

For example, what is the probability that two coins both come up heads, if at least one of them comes up heads? Using the formula:

$$P(HH|\text{at least one H}) = \frac{P(\text{at least one H}|HH) \cdot P(HH)}{P(\text{at least one H})} = \frac{1 \cdot .25}{.75} = \frac{1}{3}$$

But this is easy to see from a simple tree diagram as well:



For example, if 1% of the population has AIDS and the test for AIDS produces false results 1% of the time, what is the probability that a positive result is accurate?

$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+)} = \frac{.99 \cdot .01}{.99 \cdot .01 + .01 \cdot .99} = 50\%$

That is, even though the test is extremely accurate, a positive result only means you are truly sick 50% of the time. This is why positive results are usually treated as inconclusive and a second test administered.

6 Exercises

In all exercises below, rewrite the question in mathematical notation.

1. You roll two dice.
   (a) Define the possible outcomes from the first dice roll in standard mathematical notation. Do the same for the second dice roll.
   (b) Define the following sets of possible outcomes in terms of the definitions above: first dice roll is even, both dice rolls are even, the sum of the two dice rolls is even.

(c) What is the probability that at least one dice comes up greater than 1? Illustrate with a Venn diagram or a tree.

(d) What is the probability that both dice rolls are even? Illustrate with a Venn diagram or a tree.

(e) What is the probability that the sum of the two dice rolls is even, given that the first dice roll is even? Illustrate with a tree.

(f) What is the probability that either the sum of the two dice rolls is even *or* that the first roll is equal to 1. Illustrate with a Venn diagram or a tree.

2. A lock has five numbers on it.

   (a) A valid combination is a 5-digit code in which every number has to be used once. How many possible combinations are there?

   (b) A valid combination is the same as in a), but the lock also allows combinations that start with two digits pressed simultaneously before the other three are used one at a time. How many valid combinations are there?

   (c) A valid combination is a 5-digit code, but not every digit must be used. How many possible combinations are there?

3. A bag of marbles contains 6 marbles, labeled 1, 2, 3, 4, 5, and 6. You take two of them out of the bag randomly.

   (a) What is the probability of drawing two even marbles? Illustrate with a tree.

   (b) What is the probability that the second marble drawn is even? Illustrate with a tree.

   (c) What is the probability that the second marble drawn is even, given that the first was even? Illustrate with a tree.

   (d) What is the probability that the second marble is larger than the first? Illustrate with a tree.

4. A second bag of marbles contains 6 marbles, labeled 1, 1, 1, 2, 2, 2. But, you don't know which bag of marbles you're holding.

   (a) What is the probability that you will draw a 1?

   (b) What is the probability that you will draw a 1 and then a 2?

   (c) What is the probability that you're holding the first bag of marbles, if the first one you draw is a 1?

(d) What is the probability if the second one you draw is a 2?

5. Calculate the following.

   (a) $P(A|B) = .4$, $P(B|A) = .5$ and $P(A) = .2$. What is $P(B)$?

   (b) $P(A) \geq .7$, $P(B) \geq .75$, and $P(C) \geq .8$. What possible values might $P(A \cap B \cap C)$ take on?

   (c) What is the probability of flipping HTH? What about HHH? What about a sequence of 3 flips containing exactly 1 T?

6. A coin is tossed three times. Which pairs of events are independent? A: Heads on the first toss. B: Tails on the second. C: Heads on the third toss. D: All three outcomes the same. E: Exactly one heads turns up.