

MATH 1051 - Lecture 07

Probability

Kat Clark & Wesley Burr

Trent University

Today

- lots of definitions
- hopefully some intuition!

Defining Probability

Random processes

MP3 Players > Stories > iTunes: Just how random is random?

iTunes: Just how random is random?

By David Braue on 08 March 2007

- Introduction
- Say You, Say What?
- A role for labels?
- The new random

Think that song has appeared in your playlists just a few too many times? David Braue puts the randomness of Apple's song shuffling to the test -- and finds some surprising results.

Quick -- think of a number between one and 20. Now think of another one, and another, and another.

Starting to repeat yourself? No surprise: in practice, many series of random numbers are far less random than you would think.

Computers have the same problem. Although all systems are able to pick random numbers, the method they use is often tied to specific other numbers -- for example, the time -- that means you could get a very similar series of 'random' numbers in different situations.

This tendency manifests itself in many ways. For anyone who uses their iPod heavily, you've probably noticed that your supposedly random 'shuffling' iPod seems to be particularly fond of the Bee Gees, Melissa Etheridge or Pavarotti. Look at a random playlist that iTunes generates for you, and you're likely to notice several songs from one or two artists, while other artists go completely unrepresented.



- A **random process** is a situation in which we know the possible outcomes, but we don't know which particular outcome will happen.
- **Examples:** coin tosses, die rolls, iTunes shuffle¹ whether the stock market goes up or down tomorrow, etc.
- It can be helpful to model a process as random even if it is not truly random.

Probability

- There are several possible interpretations of probability but they (almost) completely agree on the mathematical rules probability must follow.

~~✗~~ ■ $P(A)$ = Probability of event A ← something that could happen.
 ■ $0 \leq P(A) \leq 1$

→ Probability has to be between 0 and 1
 → 200%. probability not possible
 → -60%. probability not possible.

50%. → 0.5

Probability: Frequentist

- **Frequentist interpretation:**

- The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.
- relies on a “multiverse” concept for most problems since we would want to observe the random process with the exact same conditions (hard to confirm that *everything* is the same between two different times).

(there is another philosophical way of thinking about probability called the Bayesian interpretation; we will not use this in this course)

Practice: Intuition

Which of the following events would you be most surprised by?

1. exactly 3 heads in 10 coin flips

→ we expect SH, ST

2. exactly 3 heads in 100 coin flips

→ we expect 500H
500T

3. exactly 3 heads in 1000 coin flips

→ 3H
997T

Practice: Intuition

Which of the following events would you be most surprised by?

1. exactly 3 heads in 10 coin flips
2. exactly 3 heads in 100 coin flips
3. **exactly 3 heads in 1000 coin flips**

Practice: Intuition

Which of the following events would you be most surprised by?

1. exactly 3 heads in 10 coin flips
2. exactly 3 heads in 100 coin flips
3. **exactly 3 heads in 1000 coin flips**

Why did you think that?

Law of Large Numbers: Definition

The **Law of Large Numbers** states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}_n$, converges to the probability of that outcome, p .



Parameters and Estimates

$\hat{p}_n$ is the number of times the outcome was observed divided by the number of total observations (n observations).

p is the *true proportion*, a population parameter.

$\hat{p}_n$ is an *estimate* of p based on a sample of size n .

new experiment each time

As we flip the coin more and more, we expect
 $\frac{\# H}{\# \text{ coin flips}} \rightarrow \hat{p}_n$ to get closer and closer to D.S
 $p \rightarrow$

Law of Large Numbers: Definition

The **Law of Large Numbers** states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}_n$, converges to the probability of that outcome, p .



Tip

$\hat{p}_n$ is the number of times the outcome was observed divided by the number of total observations.

Therefore, if you want to estimate the probability of some outcome then:

- collect a sample and find the proportion of times that the outcome of interest was observed;
- this is an estimate of the probability;
- the bigger your sample, the “better” your estimate of the true probability;

Flipping 1000 times gives you a “better” estimate than if you flipped 10 times.

Law of Large Numbers: Example

When tossing a *fair* coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?

H H H H H H H H H H ?

Law of Large Numbers: Example (cont'd)

H H H H H H H H H H ?

- The probability is still 0.5, or there is still a 50% chance that another head will come up on the next toss.

$$P(H \text{ on } 11^{th} \text{ toss}) = P(T \text{ on } 11^{th} \text{ toss}) = 0.5$$

- The coin is not “due” for a tail.
- The common **misunderstanding** of the LLN is that random processes are supposed to compensate for whatever happened in the past; this is just not true and is also called the **gambler's fallacy** (or **law of averages**)¹.

Independence

Two processes are said to be **independent** of one another if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss **does not** provide any useful information for determining what the coin will land on in the second toss. → Outcomes of two tosses of a coin are independent.

Independence

Two processes are said to be **independent** of one another if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss **does not** provide any useful information for determining what the coin will land on in the second toss. → Outcomes of two tosses of a coin are independent.
- Knowing that the first card drawn from a deck is an ace **does** provide useful information for determining the probability of drawing an ace in the second draw. → Outcomes of two draws from a deck of cards (**without replacement**) are dependent.

Independence

Two processes are said to be **independent** of one another if knowing the outcome of one provides no useful information about the outcome of the other.

Important

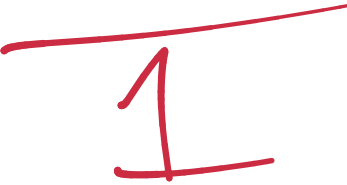
Independence is an idea we will come back to *constantly* over the rest of MATH 1051 and throughout MATH 1052.

Probability Distributions

A **probability distribution** lists all possible outcomes and the probabilities with which they occur.

- The probability distribution for the flip of a fair coin:

Event	Probability
Heads	0.5
Tails	0.5



Probability Distributions

A **probability distribution** lists all possible outcomes and the probabilities with which they occur.

- Rules for probability distributions:

- The events listed must not overlap
- Each probability must be between 0 and 1
- The probabilities must total 1

ex. get an ACE, get a spade

$$0 \leq P(A) \leq 1$$

consider all possible events,
something is guaranteed
(probability 100%) to happen

Coin flips: Why You Should Care¹

The fairness of a coin might be hugely consequential for high stakes games, but it isn't often in life that we flip coins and care about the outcome. So why bother studying the statistics of coin flips? Because coin flips are a surrogate for myriad other real-life events that we do care about.

Coin flips: Why You Should Care¹

For a given type of heart surgery, we may classify the patient outcome as survived more than a year or not, and we may want to know what is the probability that patients survive more than one year. For a given type of drug, we may classify the outcome as having a headache or not, and we may want to know the probability of headache. For a survey question, the outcome might be agree or disagree, and we want to know the probability of each response. In a two-candidate election, the two outcomes are candidate A and candidate B, and before the election itself we want to estimate, from a poll, the probability that candidate A will win. Or perhaps you are studying arithmetic ability by measuring accuracy on a multi-item exam, for which the item outcomes are correct or wrong.

Coin Flips: Why You Should Care¹

Whenever we are discussing coin flips, which might not be inherently fascinating to you, keep in mind that we could be talking about some domain in which you are actually interested! The coins are merely a generic representative of a universe of analogous applications.

Probability Distributions: Example

A **probability distribution** lists all possible outcomes and the probabilities with which they occur.

- The probability distribution for the outcome of two coin flips:

Event	HH	TT	HT	TH
Probability	0.25	0.25	0.25	0.25

Note

The order is important!

Observing a **H** on the first flip and a **T** on the second *is not the same outcome* as observing a **T** on the first flip and a **H** on the second.

Practice

In a survey, 52% of respondents said that they voted in the last election. What is the probability that a randomly selected respondent *from this sample* **did not vote**?

1. 0
2. 0.48
3. more than 0.48
4. less than 0.48
5. cannot calculate using only the information given

Practice

In a survey, 52% of respondents said that they voted in the last election. What is the probability that a randomly selected respondent *from this sample* **did not vote**?

1. 0
2. **0.48**
3. more than 0.48
4. less than 0.48
5. cannot calculate using only the information given

Event	Probability
Voted	0.52
Didn't	0.48

Definitions

The **sample space** is the set of all possible outcomes of a random experiment.

Events are subsets of the sample space.

Mutually Exclusive events are events that cannot happen at the same time.

Complementary events are two mutually exclusive events whose probabilities that add up to 1.

↳ vote / didn't vote

Sample space = $\{HH, HT, TH, TT\}$

Events = $\{2 \text{ heads}, 1 \text{ head}, \text{no heads}\}$

Definitions - Example

The **sample space** is the set of all possible outcomes of a random experiment.

Events are subsets of the sample space.

Mutually Exclusive events are events that cannot happen at the same time.

Complementary events are two mutually exclusive events whose probabilities that add up to 1.

Example - Rolling a Fair 6-sided Die

- sample space = 1, 2, 3, 4, 5, 6
- event $A = \{2, 4, 6\}$ would be the event that we roll an even number
- event $\neg A = A^c = \{1, 3, 5\}$ would be the event that we roll an odd number

Disjoint vs. Complementary

Do the sum of probabilities of two disjoint events always add up to 1?

Not necessarily, there may be more than 2 events in the sample space, e.g. party affiliation.

Disjoint vs. Complementary

Do the sum of probabilities of two disjoint events always add up to 1?

Not necessarily, there may be more than 2 events in the sample space, e.g. party affiliation.

Do the sum of probabilities of **two complementary** events always add up to 1?

A or not A

$$P(A) + P(\neg A) = 1$$

Disjoint vs. Complementary

Do the sum of probabilities of two disjoint events always add up to 1?

Not necessarily, there may be more than 2 events in the sample space, e.g., party affiliation (Liberal, Conservative, NDP, Green, BQ, ...).

Do the sum of probabilities of two complementary events always add up to 1?

Yes, that's the definition of complementary, e.g., heads and tails; or has a dog and does not have a dog.

Random Variables

Random Variables

- A **random variable** is a **numeric quantity** whose value depends on the outcome of a random event
 - We use a capital letter, like X , to denote a random variable
 - The values of a random variable are denoted with a lowercase letter, in this case x
 - We write this as $P(X = x)$

$A \nearrow$

X denotes the number of heads in 2 tosses.

Random Variables

- A **random variable** is a **numeric** quantity whose value depends on the outcome of a random event
 - We use a capital letter, like X , to denote a random variable
 - The values of a random variable are denoted with a lowercase letter, in this case x
 - We write this as $P(X = x)$
- There are two types of random variables:
 - **Discrete random variables** often take only integer values
 - **Example:** Number of credit hours, Difference in number of credit hours this term vs last
 - **Continuous random variables** take real (decimal) values
 - **Example:** Cost of books this term, Difference in cost of books this term vs last

Why Do We Care?

The **random variable** is the fundamental building block of statistics. Almost everything we do can be considered a random variable.

Gather some data and compute the mean? That *sample mean* is a random variable.

Why Do We Care?

The **random variable** is the fundamental building block of statistics. Almost everything we do can be considered a random variable.

Gather some data and compute the mean? That *sample mean* is a random variable.

In addition:

- every random variable has an associated probability distribution;
- we can't do math (calculate probabilities) when working with outcomes like “observed a blue car” or “affiliated with the NDP”;
- we use random variables to represent outcomes of a random experiment as numbers so that we **CAN** calculate “things” (like probabilities) and use statistical methods!

Outcome $\longrightarrow$ Random variable,

Summary of Random Variables

If we have random variables, then:

- they take on values
 - those values are associated with probabilities
- the arrangement of those probabilities into an organized system is called a **distribution**
- knowing the distribution is equivalent to knowing everything there is to know about a random variable
- from mathematical probability, we get certain systems that get assigned names
 - examples include the Bernoulli, binomial, normal, t, Poisson, gamma, beta, uniform ...

↳ # of H
on one
coin flip

↳ # of heads in 100 coin flips.

Families of Discrete Distributions

Bernoulli Trials

A Bernoulli trial is one of the simplest distributions – it has two possible outcomes:

yes or no

- success
- failure

Example 1: Flipping a coin where getting a heads is a “success” and tails is “failure”.

Example 2: Check a tire to see if it is flat (“success”), or inflated (“failure”).

We use the language of “success” and “failure”, but these are only labels on the particular outcomes that we are interested in!

Distribution of a Bernoulli R.V.

The distribution (sometimes called a *Probability Mass Function*) of a Bernoulli random variable (R.V.) is:

$$P(X = x) = p(x) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{else} \end{cases}$$

X	P(X)
x=0	1-p
x=1	p
else	0

So two cases: “success” (probability p) and “failure” (probability $1 - p$).

This is the mathematical model for our simplest probability example: coin flips! This is general enough to allow for non-fair coins - the probability of heads does not have to be the same as the probability of tails. They just have to add up to 1.

Binomial Distribution

A **very** important discrete distribution is the **binomial distribution**, which results when we have:

- a **fixed** number of trials, n
- only **two** outcomes possible in each trial
- each trial is **independent**
 - the probability of a **success** is the same on each trial

↗ success
→ failure.

Definition: the binomial distribution counts the number of successes in n independent trials of an experiment with fixed probability of success, p . In other words, a binomial is what happens when we have a number n of independent Bernoulli trials.

≥

Binomial Notation

If X is a Binomial random variable then we would use the notation:

$$X \sim \text{Binomial}(n, p)$$

- n - the fixed number of trials/experiments
- x - a specific number of successes we are interested in (must be between 0 and n , obviously)
- p - probability of success in any one trial/experiment
- $P(X = x)$ - the probability of getting exactly x successes in n trials

Things To Be Cautious Of

- make sure that x and p are referring to the same event as “success” (remember, just a label)
- make sure that all trials/experiments are independent
- when sampling **without** replacement, make sure your total number of trials/experiments is small when compared to population
 - independence!

$P(X=x)$ — the probability that variable X
 takes on value x
 X is # of flat tires
 $P(X=4)$ probability X is 4
 probability all tires are flat.

Example 1

Suppose you have no idea how to answer the questions on a True/False test with 10 questions, and you decide to just flip a coin for each one. You then count how many questions you actually got correct.

- x = number of questions correct
- $n = 10$
 - $p = 0.5$ (fair coin?)
- trials are independent, since the coin flips have nothing to do with one another

In this situation, what is the probability of getting at least 6/10?

Example 2

According to Statistics Canada, as of the last census in 2016, 86.3% of Canadians between the ages of 25 and 64 have a completed high school diploma, or equivalent. If we randomly select 15 Canadians aged 25-64, what is the probability distribution of the number of sampled people who have completed a high school diploma?

Find the probability that **at least 8 out of 15** randomly selected Canadians aged 25-64 will have such a diploma.

R Syntax for Computing Things with Binomials

Example:

```
1 sum( dbinom(x = 0:6, size = 10, prob = 0.5) )
```

```
[1] 0.828125
```

This gives the probability of getting **six or less correct** answers in ten trials, with a 50/50 probability (like Example 1).

R Syntax for Computing Things with Binomials

Example:

```
1 sum( dbinom(x = 7:10, size = 10, prob = 0.5) )
```

```
[1] 0.171875
```

This gives the probability of getting **MORE THAN SIX correct** answers in ten trials, again with a 50/50 probability (like Example 1).

Note: this is **not** “greater than or equal to”, it is strictly **greater than** 6.

General R Syntax

For the probability associated with a *specific* number of successes:

```
1 dbinom(x = [specific successes interested in],
2       size = [total number of trials],
3       prob = [probability of success in each trial])
```

For the probability associated with observing *one of* a set of successes:

```
1 sum( dbinom(x = [set of successes interested in],
2            size = [total number of trials],
3            prob = [probability of success in each trial]) )
```

We would use a vector to represent the “set of successes”.

Either `c(1,9,10)` or, if we have a sequence of successes, `3:6`.

OR, both! `c(1, 3, 8:10)`