Chapter 2: Probability

Yingbo Li

Southern Methodist University

STAT 2331

Random processes

- A *random process* is a situation in which we know what outcomes could happen, but we don't know which particular outcome will happen.
- Examples: coin tosses, die rolls, iTunes shuffle, 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.


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.

Probability

- Mathematical rules of probability
  - $P(A)$ = Probability of event A
  - $0 \leq P(A) \leq 1$

- Interpretation of probability:
  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.

Question

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

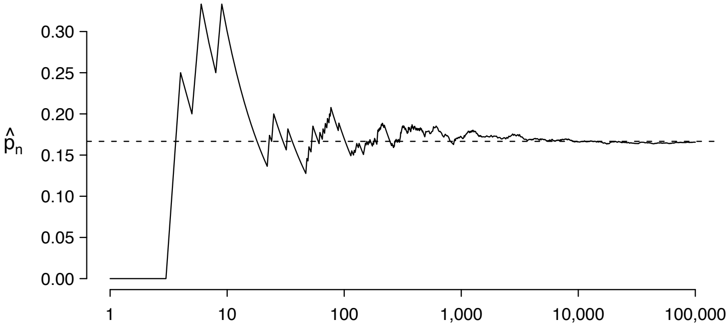
(a) 3 heads in 10 coin flips

(b) 3 heads in 100 coin flips

(c) *3 heads in 1000 coin flips*

Law of large numbers

- *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.
- The plot below shows the cumulative proportion of 1s in n rolls of a die. The probability settles around 0.167 ($\frac{1}{6}$), which is in fact the probability of rolling a 1.



Law of large numbers (cont.)

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?

$$\underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{?}$$

- 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 (mis)understanding 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 *gambler's fallacy* (or *law of averages*).

Disjoint (mutually exclusive)

- Two outcomes are called *disjoint* or *mutually exclusive* if they cannot both happen.
  - The outcome of a single coin toss cannot be a head and a tail.
  - A student cannot fail and pass a class.
  - A card drawn from a deck cannot be an ace and a queen.
- When events are disjoint, it's easy to calculate the probability of one event *or* the other happening.
  - The probability of rolling a 1 or a 2:

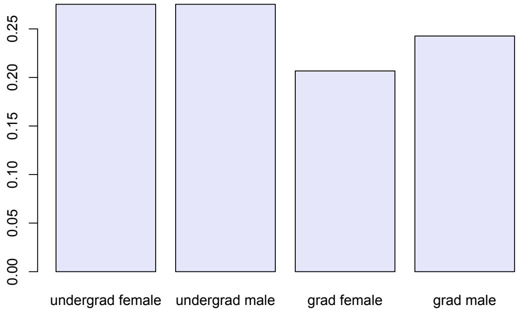
$$P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \approx 0.33$$

  - The probability of tossing a head or a tail:

$$P(H) + P(T) = 0.5 + 0.5 = 1$$

Disjoint events

What is the probability that a randomly selected student in SMU is a graduate student?

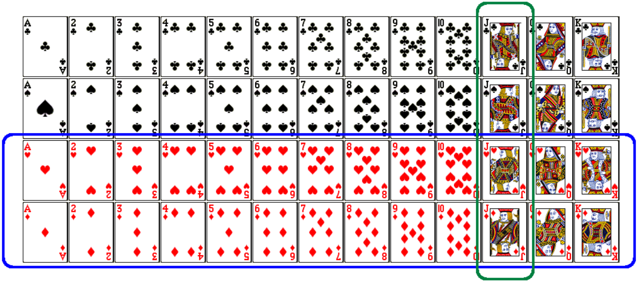


$$P(\textit{grad female or grad male}) \approx 0.21 + 0.24 = 0.45$$

Data from *https://www.smu.edu/AboutSMU/Facts* .

Non-disjoint events

What is the probability of drawing a jack or a red card from a well shuffled full deck?



$$P(jack\ or\ red) = P(jack) + P(red) - P(jack\ and\ red)$$
$$= \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

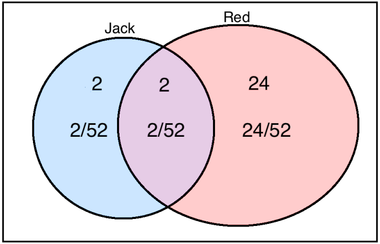
Figure from *http://www.milefoot.com/math/discrete/counting/cardfreq.htm* .

Non-disjoint events

General addition rule

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

Venn diagram



Note: When events are disjoint $P(A \text{ and } B) = 0$, hence the above formula simplifies to $P(A \text{ or } B) = P(A) + P(B)$.

Venn diagram: does it look like ...

Question

2. What is the probability that a randomly sampled student is a sophomore or is taking a statistics course this semester?

		stat course		
		no	yes	Total
	first-year	17	62	79
class	sophomore	37	40	77
	junior	22	26	48
	senior	3	4	7
	Total	79	132	211

(a) $\frac{77+132}{211}$

(b) $\frac{77+132-40}{211}$

(c) $\frac{77+132+40}{211}$

(d) $\frac{40}{132}$

Question

3. At a large apartment complex, 58% of the units have a washer and dryer, 32% have double parking, and 20% have both washer and dryer and double parking. What percent of apartments have neither double parking nor washer and dryer?

(a) 0.10
(b) 0.1856
(c) 0.20
(d) *0.30*

	w&d	*no w&d*	*total*
dbl prk	*0.20*	*0.12*	*0.32*
no dbl prk	*0.38*	*0.30*	*0.68*
total	*0.58*	*0.42*	*1*

$P(w\&d \ or \ dbl \ prk)$
$= 0.58 + 0.32 - 0.20$
$= 0.70$
$P(neither \ w\&d \ nor \ dbl \ prk)$
$= 1 - 0.70 = 0.30$

Sample space

Sample space is the collection of all possible outcomes of a trial.

- A couple has one kid, what is the sample space for the gender of this kid? $S = \{B, G\}$
- A couple has two kids, what is the sample space for the gender of these kids? $S = \{BB, GG, GB, BG\}$

Probability distributions

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

- The probability distribution for the gender of one kid:

Event	B	G
Probability	0.5	0.5

- Rules for probability distributions:
  1. The events listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1
- The probability distribution for the genders of two kids:

Event	*BB*	*GG*	*BG*	*GB*
Probability	*0.25*	*0.25*	*0.25*	*0.25*

Complementary events

- A couple has one kid. If we know that the kid is not a boy, what is gender of this kid?
  $\{$ ~~B~~, *G* $\}$ $\rightarrow$ Boy and girl are *complementary* outcomes.
- A couple has two kids, if we know that they are not both girls, what are the possible gender combinations for these kids?
  $\{$ *BB*, ~~*GG*~~, *GB*, *BG* $\}$

Question

4. In a survey, 52% of respondents said they are Democrats. What is the probability that a randomly selected respondent from this sample is a Republican?

(a) 0.48

(b) more than 0.48

(c) less than 0.48

(d) *cannot calculate using only the information given*

While (a) and (c) are also possible, (b) is definitely not possible

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?

Yes, that's the definition of complementary, e.g. heads and tails.

Independence

- Two processes are *independent* 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 since coin tosses are independent.
  $\rightarrow$ Outcomes of two tosses of a coin are independent.
- On the other hand, 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.
  $\rightarrow$ Outcomes of two draws from a deck of cards (without replacement) are dependent.

Independent events

- You toss a coin twice, what is the sample space for the outcomes of these tosses?

$$S = \{HH, TT, HT, TH\}$$

- You toss a coin twice, what is the probability of getting two tails in a row? Since one out of four possible outcomes match this definition, the probability is $\frac{1}{4}$.
- While it is possible to calculate this probability by first obtaining the sample space, this would be an inefficient way if the number of trials was much higher.

Product rule for independent events

Product rule for independent events

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

Or more generally, $P(A_1 \text{ } and \text{ } \cdots \text{ } and \text{ } A_k) = P(A_1) \times \cdots \times P(A_k)$

You toss a coin twice, what is the probability of getting two tails in a row?

$$P(\text{T on the first toss}) \times P(\text{T on the second toss}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Question

5. SMU has 14.2% international students. What is the probability that two randomly selected students are both international?

(a) 14.2^2

(b) 0.142^2

(c) 0.142×2

(d) $(1 - 0.142)^2$

Data from *https://www.smu.edu/AboutSMU/Facts* .

Question

6. SMU has 14.2% international students. What is the probability that among two randomly selected students, at least one of them is international?

(a) 0.142

(b) $0.142 \times (1 - 0.142) + 0.142^2$

(c) $(1 - 0.142)^2$

(d) $1 - (1 - 0.142)^2$

Data from *https://www.smu.edu/AboutSMU/Facts* .

P(at least 1 success)

If we were to randomly select 5 students, what is the probability that at least one is international?

- If we were to randomly select 5 students, the sample space for the number of applicants who are uninsured would be:

$$S = \{0, 1, 2, 3, 4, 5\}$$

- We are interested in instances where at least one student is international:

$$S = \{0, {\color{red}1, 2, 3, 4, 5}\}$$

- So we can divide up the sample space intro two categories:

$$S = \{0, \textit{\color{red}at least one}\}$$

P(at least 1 success) (cont.)

Since the probability of the sample space must add up to 1:

$$\begin{aligned} P(\text{at least 1 international}) &= 1 - P(\text{none international}) \\ &= 1 - [(1 - 0.142)^5] \\ &= 1 - 0.858^5 \\ &= 1 - 0.465 \\ &= 0.535 \end{aligned}$$

At least 1

$$P(at\ least\ one) = 1 - P(none)$$

Addition rule vs product rule

1. General addition rule:
   - Formula:

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

   - When to use: A and B are outcomes,
     e.g. (1) head, tail of coin toss, or (2) Jack, red card.
2. Product rule:
   - Formula:

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

   - When to use: A and B are independent events,
     e.g. (1) 1st coin toss, 2nd coin toss, or (2) drawing 1st card, drawing 2nd card.