Chance and Chance Simulation

Sampling Data | Chance and Chance Simulation

STAT5002

The University of Sydney

Mar 2025

CRICOS 00026A TEQSA PRV12057

Sampling Data

Topic 5: Understanding chance and chance simulation

Topic 6: Chance variability

Topic 7: Central limit theorem

Outline

Definitions

Basic properties

Some Exercises

Simulations

Why study chance?

- Why is understanding "chance" important?
  - ➟ An important modelling tool
  - ➟ Probability (or the p-value) is crucial for decision making in Hypothesis Testing
- What are some examples of probability in everyday life?
  - ➟ Card games, gambling, stock market, ...
- How can we define chance?
  - ➟ Basics of probability
- How can we investigate chance using R?
  - ➟ Simulations

Definitions

Probability

The frequentist definition of **probability** (or chance) is the percentage of time a certain event is expected to happen, if the same process is repeated long-term (infinitely often).

This differs from the Bayesian definition of probability which relates to the degree of belief that an event will occur (later study).

Basic properties of Probability

1. Probabilities are between 0 (impossible) and 1 (certain)

$$\mathrm{P}(\text{Impossible event}) = 0$$

$$\mathrm{P}(\text{Certain event}) = 1$$

Or, in terms of percentages, 0% and 100%

2. Complement

The probability of something equals 1 minus the probability of its opposite

$$\mathrm{P}(\text{Event}) = 1 - \mathrm{P}(\text{Complement event})$$

For example, we randomly toss a coin (two possible events: Head or Tail), P(Head) - 1 - P(Tail).

3. Conditional probability

The chance that a certain event (1) occurs, *given* another event (2) has occurred.

$$\mathrm{P}(\mathrm{Event\ 1}|\mathrm{Event\ 2})$$

For example, what is the chance of having good weather tomorrow given the weather today is good?

Multiplication rule and addition rule

Multiplication rule

The probability that two events occur is the chance of the 1st event **multiplied** by the chance of the 2nd event, given the 1st has occurred.

$$\mathrm{P}(\text{Event1 and Event2}) = \mathrm{P}(\text{Event1}) \times \mathrm{P}(\text{Event2} \mid \text{Event1})$$

Addition Rule

The probability at least one of two events occurs is the chance of the 1st event **plus** the chance of the 2nd event **minus** the probability that both events occur.

$$\mathrm{P}(\text{Event1 or Event2}) = \mathrm{P}(\text{Event1}) + \mathrm{P}(\text{Event2}) - \mathrm{P}(\text{Event1 and Event2})$$

Mutually exclusive

Two events are **mutually exclusive** when the occurrence of one event prevents the occurence of the other.

- For example, when tossing a coin, the event of getting Head and the event of getting Tail are mutually exclusive.

Independence

Two events are **independent** if the chance of the 1st given the 2nd is the same as the 1st, ie.

$$\mathrm{P}(\text{Event } 1 \mid \text{Event } 2) = \mathrm{P}(\text{Event } 1)$$

What's the difference between mutually exclusive and independence?

Term	Definition
Mutually exclusive	the occurrence of Event 2 prevents Event 1 occuring
Independence	the occurrence of Event 2 does not change the chance of Event 1

Summary of addition and multiplication rule

Addition rule: P(At least 1 of 2 events occurs)

Formula	Condition
P(Event1) + P(Event2) - P(Both events occur)	always
P(Event1) + P(Event2)	if mutually exclusive

Multiplication rule: P(Both events occur)

Formula	Condition
P(Event1) x $P(\mathrm{Event2} \vert \mathrm{Event1})$	always
P(Event1) x P(Event2)	if independent

Exercises

Example 1

A coin is tossed twice. If the coin lands on head on the 2nd toss, you win \$1.

- Are the tosses independent?
- If the 1st coin is a head, what is the chance of winning \$1?
- If the 1st coin is a tail, what is the chance of winning \$1?
- What is the chance of winning \$1?

Solution 1

1. Yes.
2. Event HH out of {HH,HT}; P(2nd H| 1st H) = $\frac{1}{2}$.
3. Event TH out of {TH,TT}; P(2nd H| 1st T) = $\frac{1}{2}$.
4. $\frac{1}{2}$ regardless of the first event.

Example 2

A die is rolled twice. What is the chance that:

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}}$$

- Both the 1st and 2nd rolls are 1s?
- Either the 1st roll is a 1 or the 2nd roll is a 1?

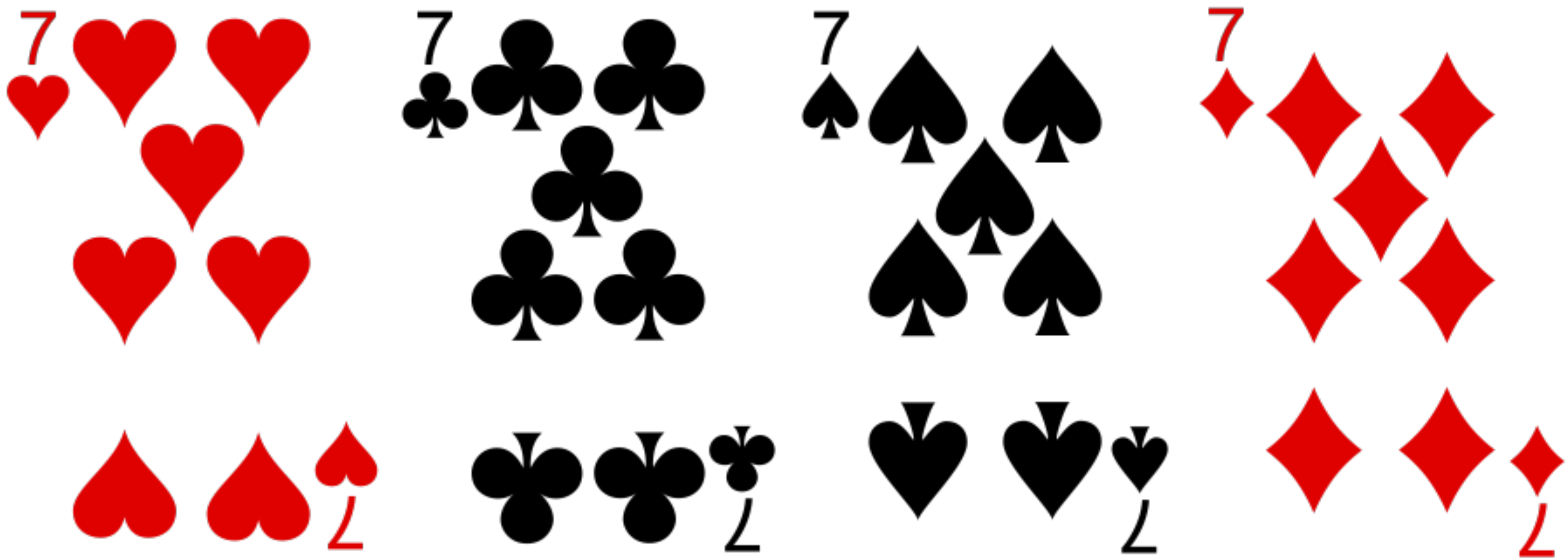
Solution 2

Two rolls are independent.

1. P(1st roll is 1) × P(2nd roll is 1 | 1st roll is 1) = $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$
2. P(1st roll is 1) + P(2nd roll is 1) - P(1st roll is 1 and 2nd roll is 1) = $\frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$

Example 3

A standard deck of 52 playing cards, consisting of 13 cards of each suit (four suits). A deck of cards is shuffled.



What is chance that:

- Both the top card and the bottom card are the ace of spades?
- The top card is the ace of spades or the bottom card is the ace of spades?

Solution 3

1. 0 as they are mutually exclusive.
2. P(top card is the ace of spades) + P(bottom card is the ace of spades) - P(top card and the bottom card are the ace of spades) = $\frac{1}{52} + \frac{1}{52} - 0 = \frac{2}{52} = \frac{1}{26}$

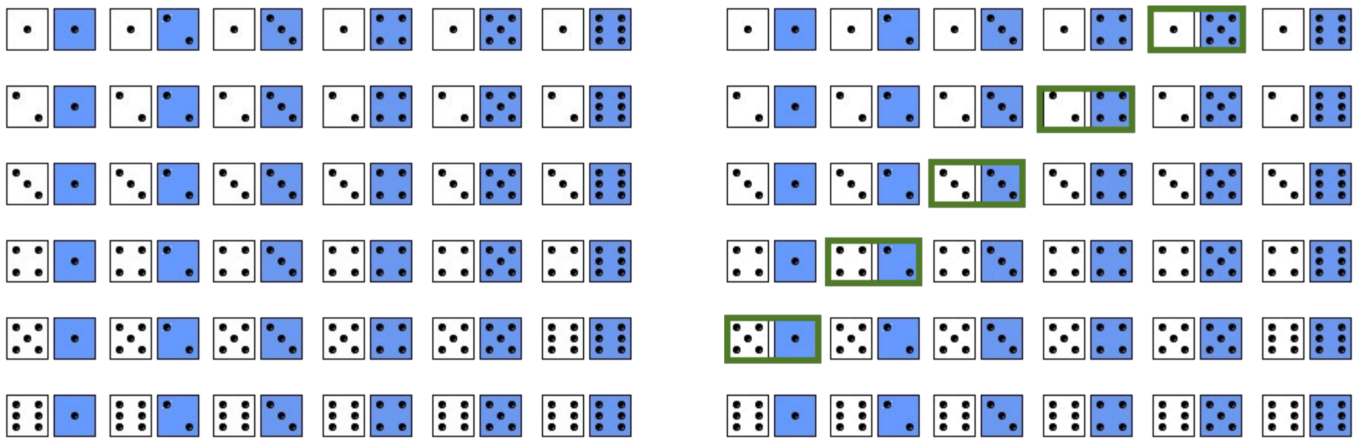
Chance simulation

Enumeration

- Very often, we want to determine the probability of a certain event of interest. For example, what are the chances of winning gambling games?
- For simple chance problems, a good way to start is to enumerate all the possible outcomes.

Example:

Two dice are thrown. What is the chance that their sum is 6?



So the chance is 5/36 (approx 0.14). But this may not work for complicated problems (even with advanced counting techniques such as permutation and combination).

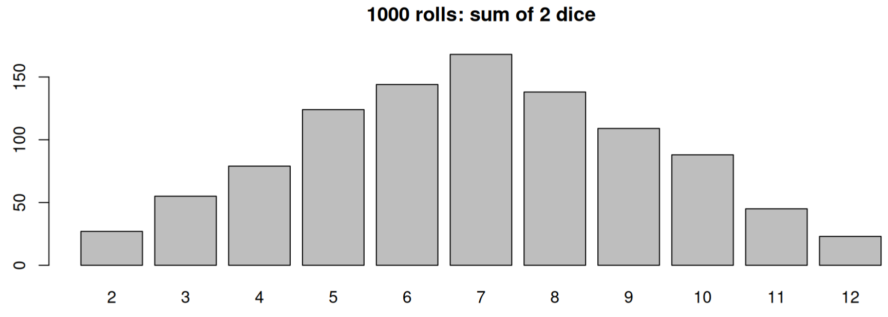
Simulate (in R)

- We can **use R** to perform many computer experiments to estimate the chance.
- Simulate throwing 2 dice m times and record the findings.

```r
set.seed(23)  # set the random seed
die1 = sample(1:6, 1000, rep = T)
die2 = sample(1:6, 1000, rep = T)
totals = die1 + die2
table(totals)
```

```
totals
  2   3   4   5   6   7   8   9  10  11  12 
 27  55  79 124 144 168 138 109  88  45  23 
```

```r
barplot(table(totals), main = "1000 rolls: sum of 2 dice")
```



So the (simulated) chance of getting a total of 6 is 144/1000 = 0.144, very close to the exact answer of 5/36 = 0.139.

- `set.seed(23)`: We set the random seed so this can be reproduced
- `sample(1:6, 1000, rep=T)`
    - ➟ Sample from 1,2,3,4,5,6 (a die) with equal probability
    - ➟ Sample $m = 1000$ times
    - ➟ Sample **with replacement** using `rep=T` (independent experiments)

With/without replacement

`sample(1:6, m, rep=T)` simulates a **box model** (more in next lecture). In a box model, there are N tickets in a box, and we want to draw m tickets from the box.

- For example, two rolls of a fair die can be modeled as $m = 2$ draws from the box

  $$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}}$$

  we have to place the ticket back in the box after each draw, so the outcome of one die roll does not affect the outcome of another. In other words, the $m = 2$ draws are made **with replacement**.

  - **With replacement** means draws are **independent** to each other.
- In other situations, the draws are made **without replacement** (without using `rep=T` in `sample`).
  - For example, consider drawing four cards from a standard deck of 52 cards (without putting the drawn cards back).
  - This implies **dependent** experiments - the next outcome depends on previous ones.

Sample without replacement (using R)

Example: A company has 10,000 male employees and 11,000 female employees. A representative committee is created by randomly picking 10 employees.

- What is chance that more than 75% in the committee are female?

```r
set.seed(1)
committee = function() {
    # Write a function to simulate each committee selection
    committee = sample(c(rep(1, 11000), rep(0, 10000)), size = 10, replace = FALSE)
    condition = sum(committee)/10 > 0.75
    return(condition)
}
sim <- replicate(1e+05, committee())
sum(sim)/1e+05
```

```
[1] 0.07326
```

- `c(rep(1, 10000), rep(0, 11000))`: the box represents all employees.
- `size = 10`: size of the committee.
- `sum(committee)`: number of females in the committee.
- `condition=sum(committee)/10 > 0.75`: `TRUE` if more than 75% in the committee are female
- `replicate(10000, committee())`: repeat the experiment 100,000 times.
- `sum(sim)`: number of success (more than 75% are female)

Case study

Example: Why did the Chevalier lose money?

The Chevalier de Méré was a 17th century gambler, who played 2 games:

- Game A: Roll a die 4 times. Win = at least 1an “ace”.
- Game B: Roll a pair of dice 24 times: Win = at least a “double-ace”.
- Note: an “ace” means “1”.

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}}$$

$$\boxed{\boxed{\text{Ace}}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}}$$

He reasoned:

Game	1 roll	# rolls	Win
A	P(an Ace) = 1/6	4	P(at least an Ace) = 4x1/6 = 2/3
B	P(a Double-Ace) = 1/36	24	P(at least a Double-Ace) = 24*1/36 = 2/3

But he lost consistently in Game B. Why?

What is the actual chance of winning?

Game	1 roll	# rolls	P(no Win)	P(Win)
A	P(not Ace) = 5/6	4	P(no Aces) = (5/6)^4	1-(5/6)^4 = 0.518
B	P(not Double-Ace) = 35/36	24	P(no Double-Aces) = (35/36)^24	1-(35/36)^24=0.49

- Not having an “ace” in each roll is independent, applying the multiplication rule to work out the chance of lossing.
- Considering the **complement** event, makes each of the complements **mutually exclusive**, so the solution follows easily.
- So it’s slightly better to play Game A.

Simulate in R (using a function)

```r
gameA = function() {
    rolls = sample(1:6, size = 4, rep = TRUE)
    winning = sum(rolls == 1) > 0
    return(winning)
}
simsA = replicate(1e+05, gameA())
sum(simsA)/length(simsA)
```

```
[1] 0.51769
```

```r
gameB <- function() {
    first.die = sample(1:6, size = 24, rep = TRUE)
    second.die = sample(1:6, size = 24, rep = TRUE)
    num_double_ace = sum((first.die == second.die) & (first.die == 1))
    winning = num_double_ace > 0
    return(winning)
}
simsB = replicate(1e+05, gameB())
sum(simsB)/length(simsB)
```

```
[1] 0.49279
```

Indeed, Game A is better.