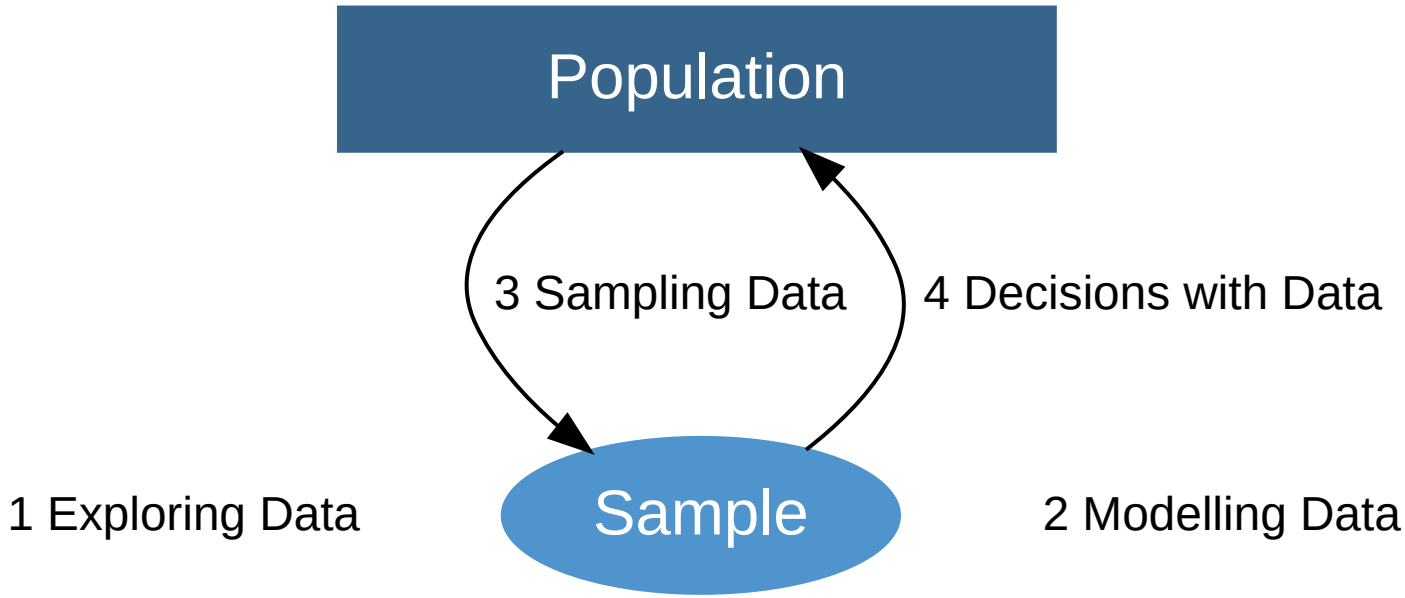
Law of Averages

Sampling Data | Chance Variability

© University of Sydney DATA1001/1901

Unit Overview



Module3 Sampling Data

Understanding Chance

What is chance?

Chance Variability

How can we model chance variability by a box model?

Sample Surveys

How can we model the chance variability in sample surveys?

Law of Averages

Data Story | Coin tossing in WWII

Chance Processes

Introducing the Box Model

Applying the Box Model for Gambling

More Examples

Summary

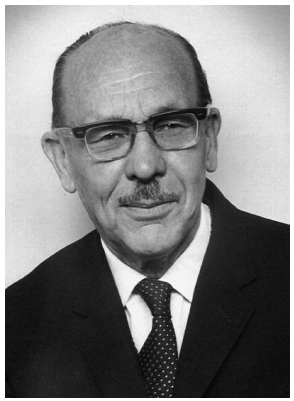
Data Story

Coin tossing in WWII

Coin tossing in WWII

- John Edmund Kerrick (1903–1985) was a mathematician noted for a series of experiments in probability which he conducted while interned in Nazi-occupied Denmark (Viborg, Midtjylland) in the 1940s.

- Kerrich had travelled from South Africa to visit his in-laws in Copenhagen, and arrived just 2 days after Denmark was invaded by Nazi Germany!



- Fortunately Kerrich was imprisoned in a camp in Jutland run by the Danish Government in a `truly admirable way'.

- With a fellow internee Eric Christensen, Kerrich set up a sequence of experiments demonstrating the empirical validity of a number of fundamental laws of probability.
  - They tossed a (fair) coin 10,000 times and counted the number of heads.
  - They made 5000 draws from a container with 4 ping pong balls (2x2 different brands), ‘at the rate of 400 an hour, with - need it be stated - periods of rest between successive hours.’
  - They investigated tosses of a “biased coin”, made from a wooden disk partly coated in lead.
- In 1946 Kerrich published his finding in a monograph An Experimental Introduction to the Theory of Probability.

Statistical Thinking

Kerrich and Christensen tossed a coin 10,000 times.

- How many heads would you **expect** them to get?
- How many heads did they actually get (**observe**)?

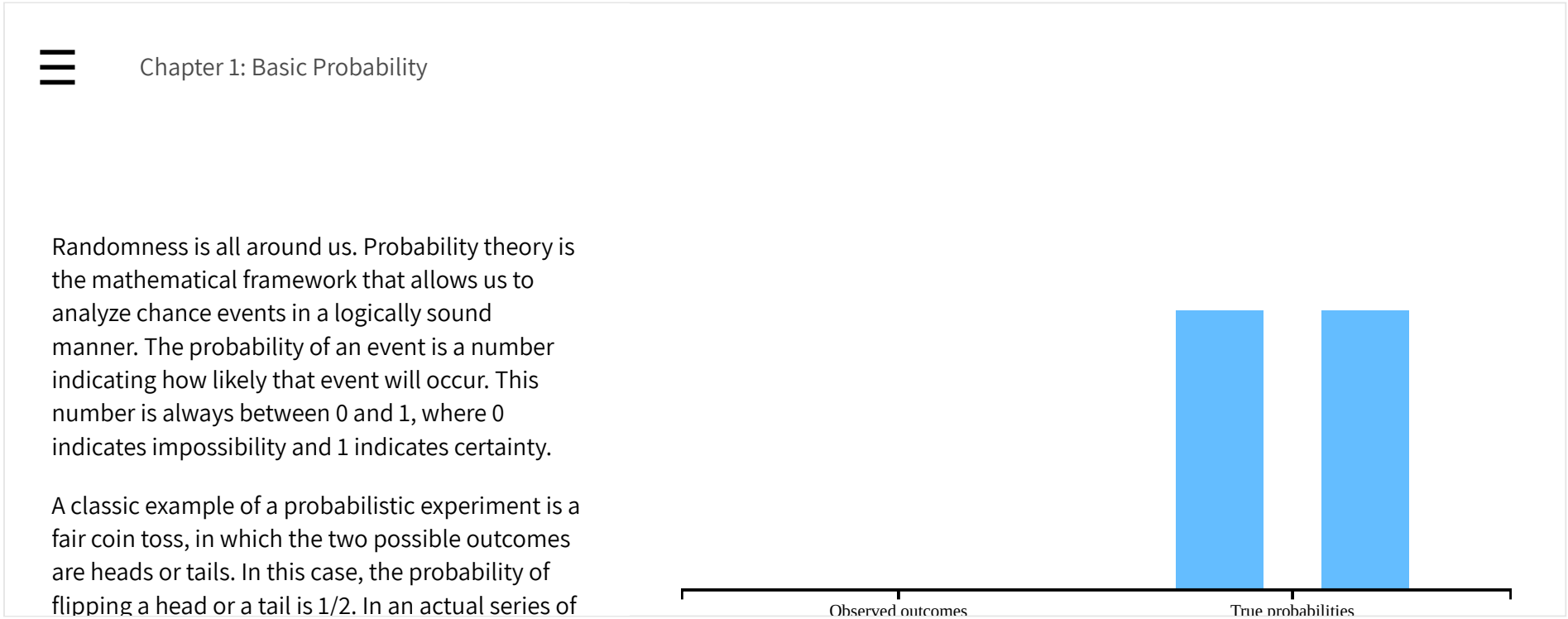
Statistical Thinking

Read each of these statements: is it true or false?

- After a sequence of 4 heads in a row, the chance of getting a tail increases.
- In the long run, eventually the number of heads and tails evens out.
- In the long run, the size of the difference between the number of heads and the expected number decreases.
- In the long run, the size of the difference between the percentage of heads and 50% decreases.

Have a play

Have a play with this app, with simulations of coins and dice.



Chance Processes

Chance Processes

- Every time you toss a fair coin, there is chance variability.

Number of heads (observed value)
= half the number of tosses (expected value) + chance error

What was the chance error for Kerrich and Christensen's experiment?

The Law of Averages

Law of Large Numbers (or Law of Averages)

- The **Law of Large Numbers** states that the **proportion** of heads becomes more stable as the length of the simulation increases and approaches a fixed number called the **relative frequency**.
- The chance error in the number of heads is likely to be **large** in absolute size, but **small** relative to the number of tosses.

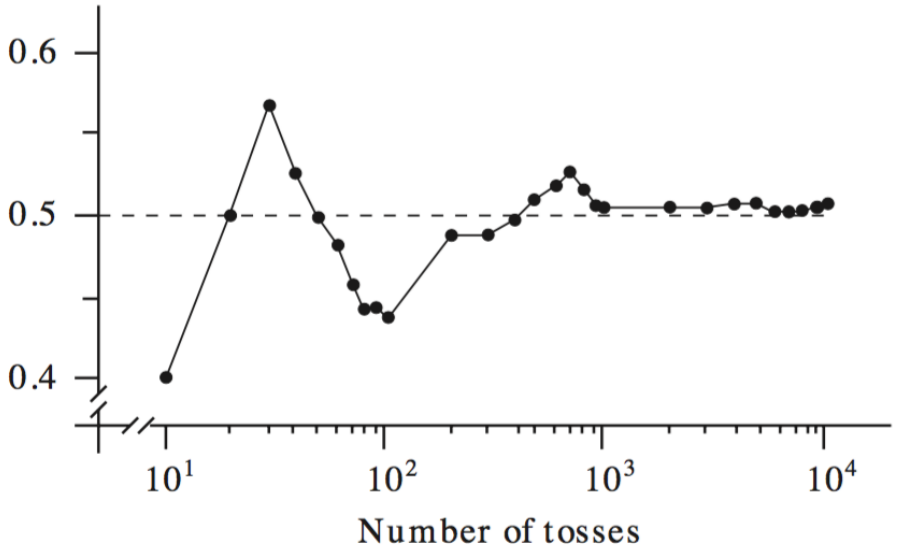
Important Facts

For a fair coin:

- Even if we observe 100 heads in a row, still P(Tail)=0.5. Misunderstanding this, leads to the Gambler's Fallacy.
- As the number of tosses **increases**
  - the absolute size of the chance error **increases**.
  - the absolute percentage size of the chance error **decreases**.
  - the proportion of the event will converge to the theoretical or expected proportion.

Kerrich's experiment over time

Cumulative proportion of heads

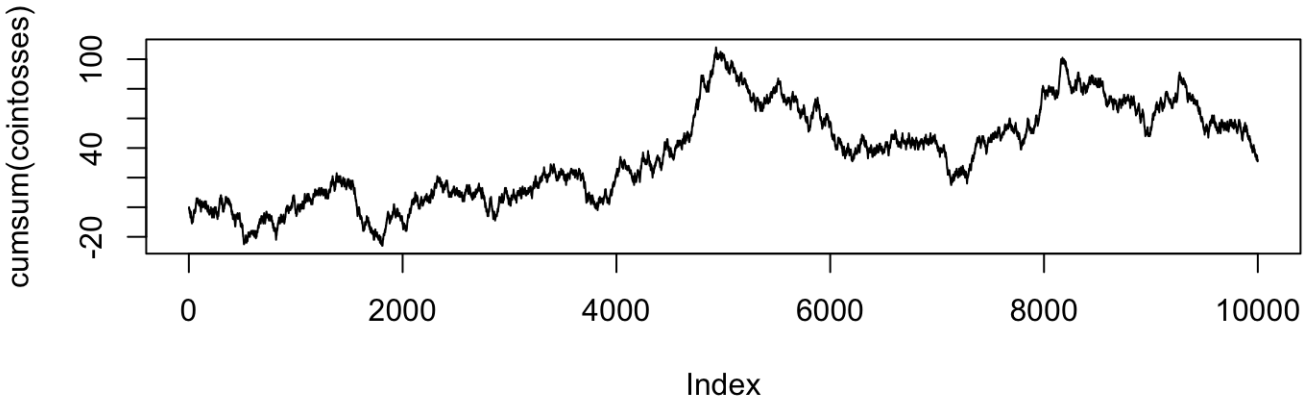


💬 When did it get close to 0.5?

Simulation of Kerrich's experiment

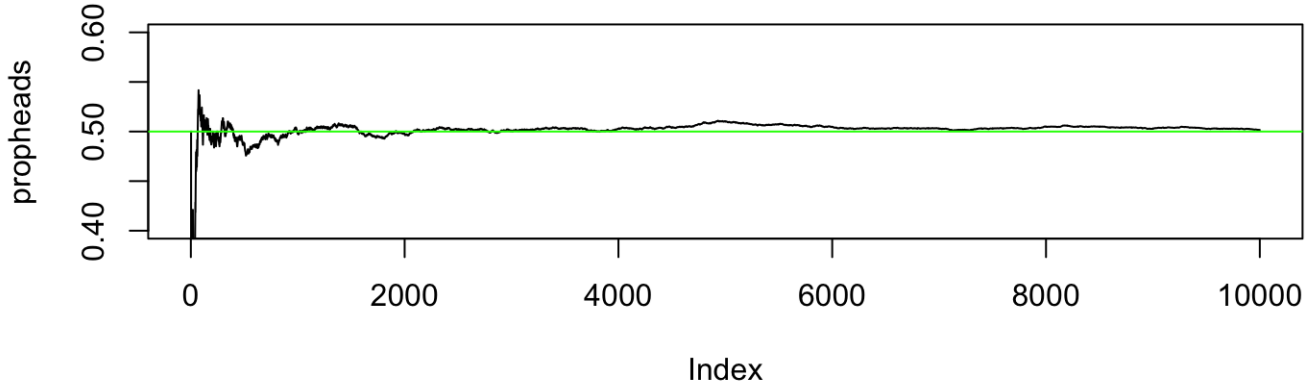
Cumulative differences

```
set.seed(1)
cointosses = sample(c(-1,1), 10000, replace = TRUE)
plot(cumsum(cointosses), type = 'l')
```



Cumulative proportion of heads

```
set.seed(1)
cointossesheads = sample(c(0,1), 10000, repl = T)
cumheads = cumsum(cointossesheads)
propheads = cumsum(cointossesheads)/(1:10000)
plot(propheads, ylim=c(.4, .6), type = 'l')
abline(h=0.5, col="green")
```



When does it get close to 0.5?

Law of Large Numbers - Explained and Visualized

Introducing the box model

The box model

The box model

- The **box model** is a simple way to describe many chance processes.
- We need to know:
  - the **distinct numbers** that go in the box (“tickets”).
  - the **number** of each kind of tickets in the box.
  - the number of **draws** from the box.

Why the box model?

We'll using the box model for the rest of the course as a simple way to visualise a chance process.

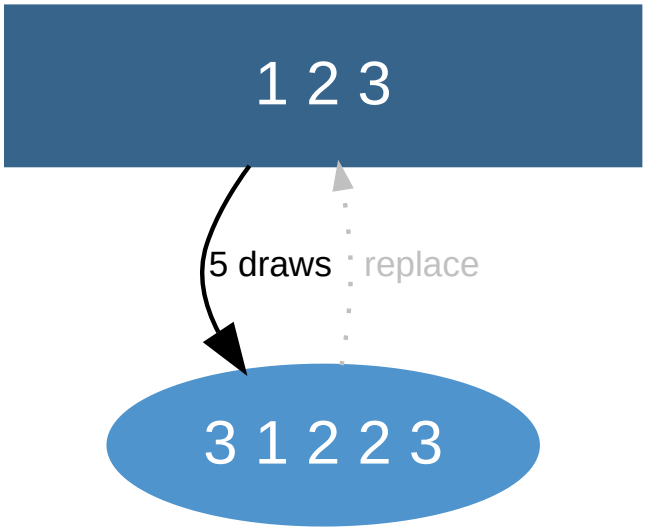
1st: Think of the **box** as a summary of the **population** (what's in there, and in what proportions).

2nd: Take draws from the box to create the **sample**.

3rd: Consider the Sum or Mean of the sample.

- what is the expected value (EV)?
- what is the observed value (OV)?
- The chance error is OV - EV, which is modelled by the standard error (SE).

Example



Consider the Sum of the sample:

- EV = 10 (on average we expect an average value of 2, for 5 draws)
- OV = 3 + 1 + 2 + 2 + 3 = 11
- Hence the chance error is 1.

Applying the Box Model to Gambling

The box model for gambling

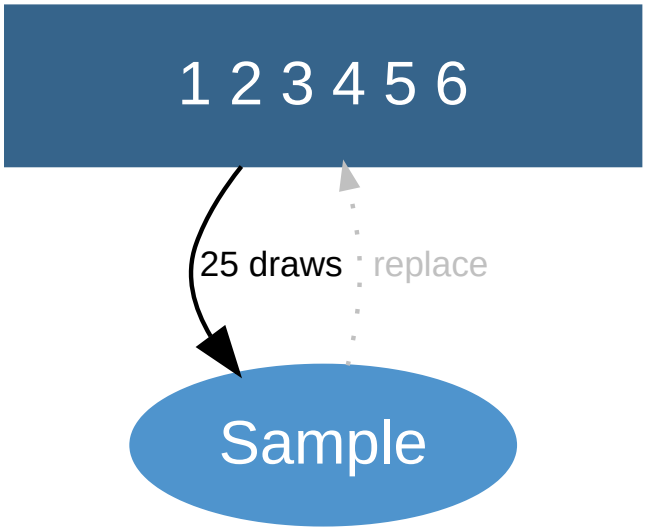
The box model for gambling

- For a box model for gambling games,
  - the tickets represent the amount won (+) and lost (-) in each play.
  - the chance of drawing a particular value, is the chance of winning that amount in 1 play.
  - the number of draws is the number of plays.
- The **net gain** is the sum of the draws from the box (sample).

Example

Example1 (Fair Dice)

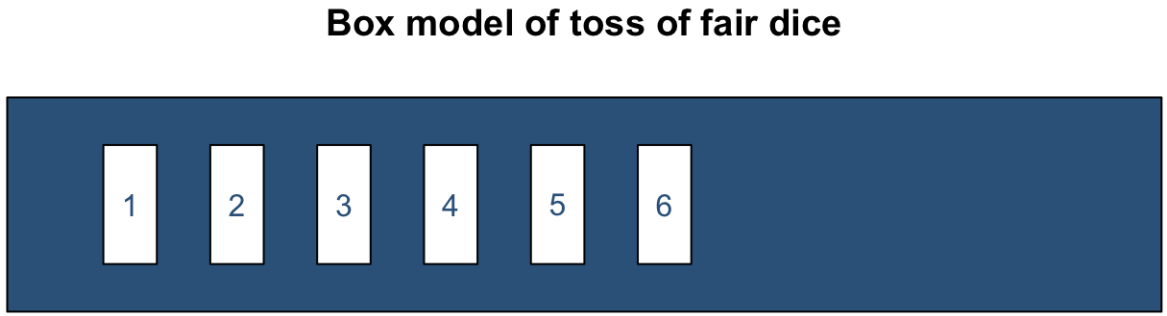
Throw a fair dice 25 times.



Box Model of 1 toss of dice

Note the composition of the box:

- The distinct tickets are 1,2,3,4,5,6, representing the 6 unique faces of the dice.
- There is 1 of each ticket, as the dice is fair.



Simulation

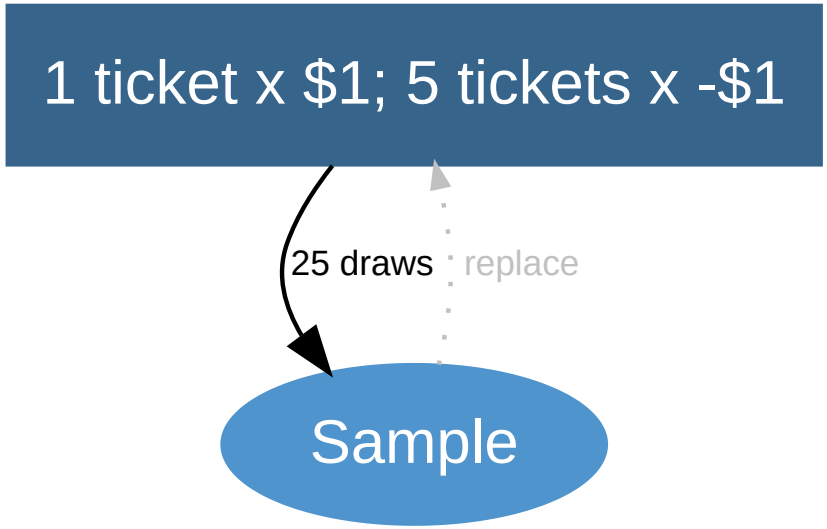
```
set.seed(1)
dietosses= sample(c(1:6), 25, repl = T)
dietosses
```

```
##  [1] 1 4 1 2 5 3 6 2 3 3 1 5 5 2 6 6 2 1 5 5 1 1 6 5 5
```

This simulation represents one possible sample formed by 25 throws of the dice.

Model a game

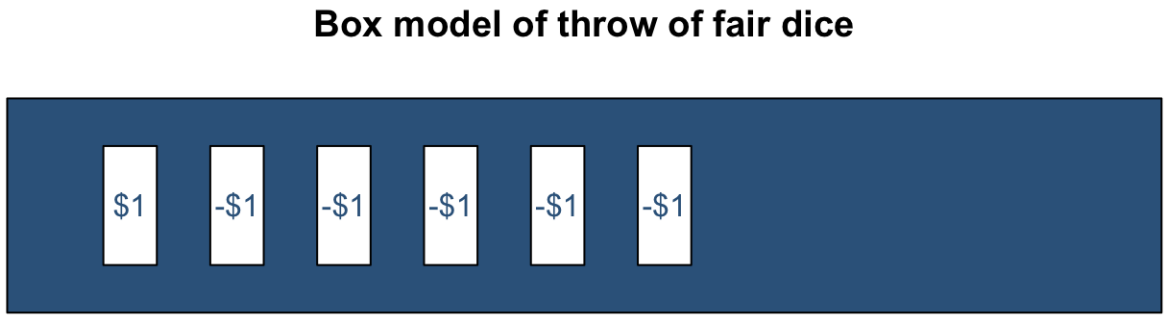
- Suppose it costs $1 to play a game.
- If you roll a “6”, you get back your $1, plus win another dollar.
- If you get any other number, you lose your $1.
- Play 25 times. What is your net gain/loss?



Box Model of 1 play of the game

Note the composition of the box:

- The distinct tickets are $1 and -$1, representing the “win” and “loss”.
- There is 1 ticket with $1 (equivalent to tossing a “6”), and 5 tickets with -$1 (equivalent to tossing a “1”,“2”,“3”,“4”,“5”).



Simulation

```
set.seed(1)
dietosses= sample(c(1,-1,-1,-1,-1,-1), 25, repl = T)
# Or: dietosses= sample(c(1,-1), 25, repl = T,prob=c(1/6,5/6))
sum(dietosses)
```

```
## [1] -13
```

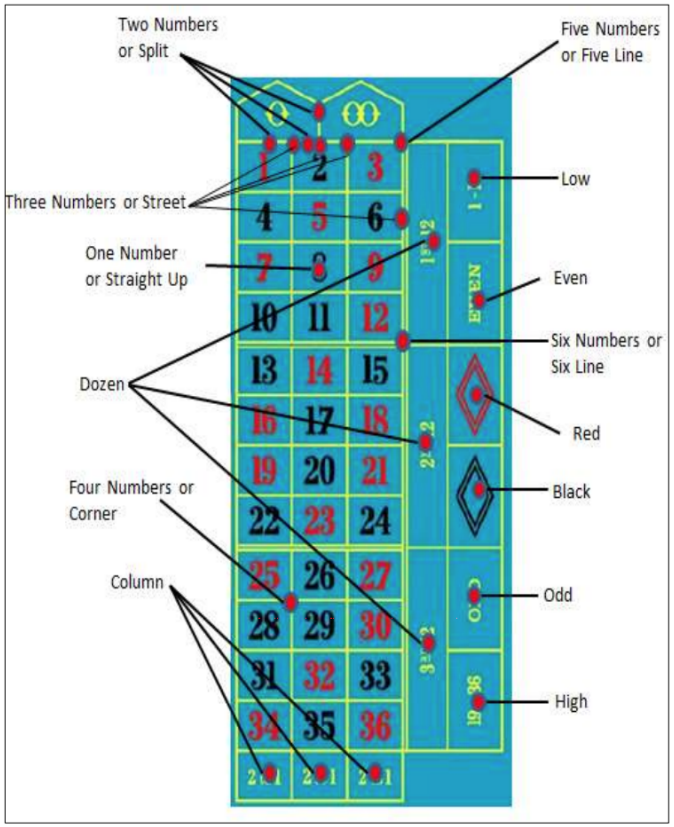
This simulation represents one possible net “winnings”, after playing the games 25 times.

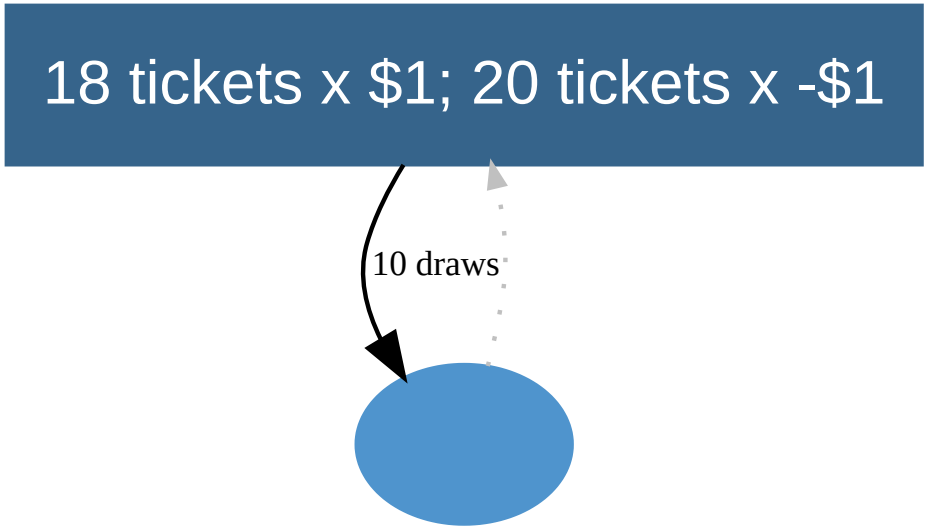
More Examples

Example2 (Star Casino 00 Roulette on red)

- The Star Casino 00 roulette wheel has 38 pockets,numbered 0 (green), 00 (green) and 1-36 (alternate red and black).
- Suppose you place a bet on red. This costs $1.
- The croupier spins the wheel until a ball lands in a pocket.
  - If it lands on a red, you get your $1 back plus an extra $1.
  - If it doesn't land on red, you lose your $1.
- You play 10 times. What is your expected net gain/loss?

Wager Placement for 00 Roulette

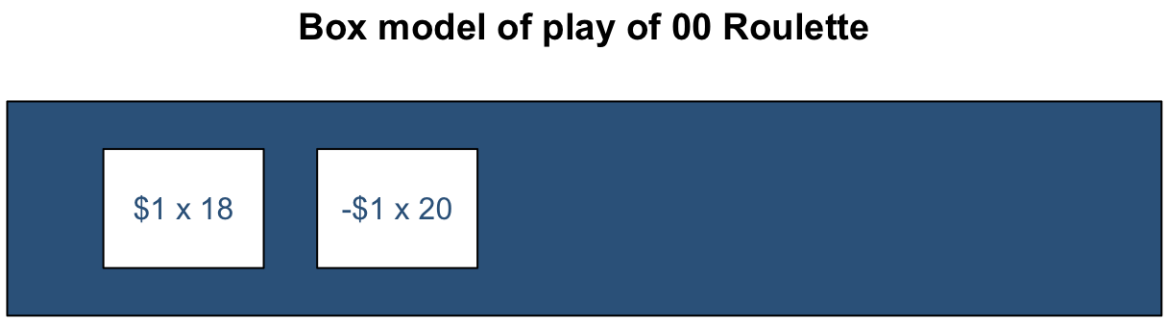



18 tickets x $1; 20 tickets x -$1
10 draws

Box Model of 1 play of the game

Note the composition of the box:

- The distinct tickets are $1 and -$1, representing the “win” and “loss”.
- There are 18 tickets with $1 (equivalent to landing on a red), and 20 tickets with -$1 (equivalent to landing on a black or green).



Simulation

```
set.seed(1)
pocket= sample(c(1,-1), 10, repl = T, prob=c(18/38,20/38))
head(pocket)
```

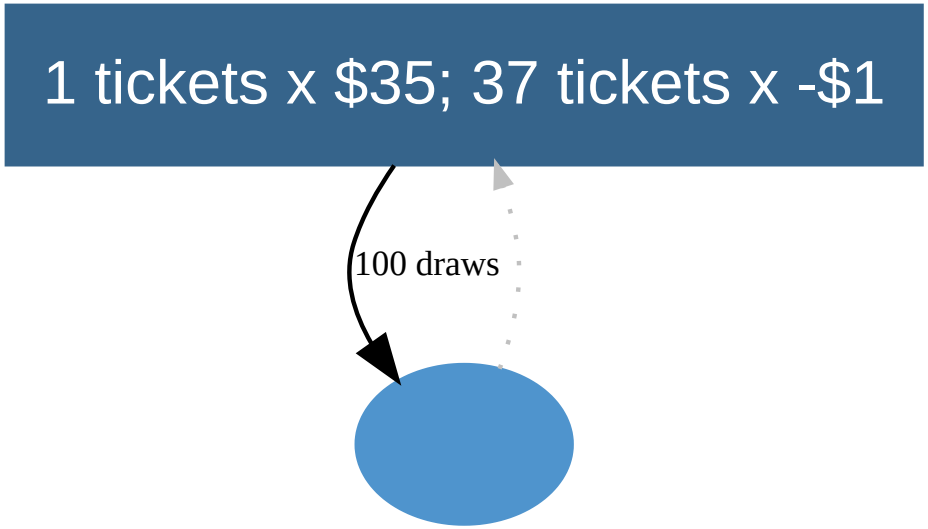
```
## [1] -1 -1  1  1 -1  1
```

```
cumsum(pocket)
```

```
##  [1] -1 -2 -1  0 -1  0  1  2  3  2
```

Example3 (Star Casino 00 Roulette on 1 number)

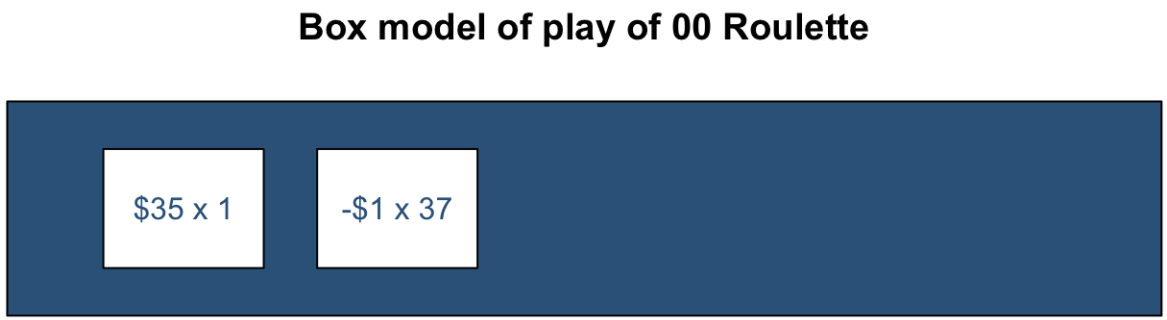
- You bet on a single number, which costs $1. This is called “Straight Up”.
  - If you win, you get $35 plus the dollar back.
  - If you lose, you lose $1.
- You play 100 times. What is your expected net gain/loss?


1 tickets x $35; 37 tickets x -$1
100 draws

Box Model of 1 play of the game

Note the composition of the box:

- The distinct tickets are $35 and -$1, representing the “win” and “loss”.
- There is 1 ticket with $35 (equivalent to landing on the chosen number), and 37 tickets with -$1 (equivalent to landing on the other numbers).



Simulation

```
set.seed(1)
pocket= sample(c(35,-1), 100, repl = T, prob=c(1/38,37/38))
head(pocket)
```

```
## [1] -1 -1 -1 -1 -1 -1
```

```
cumsum(pocket)[100]
```

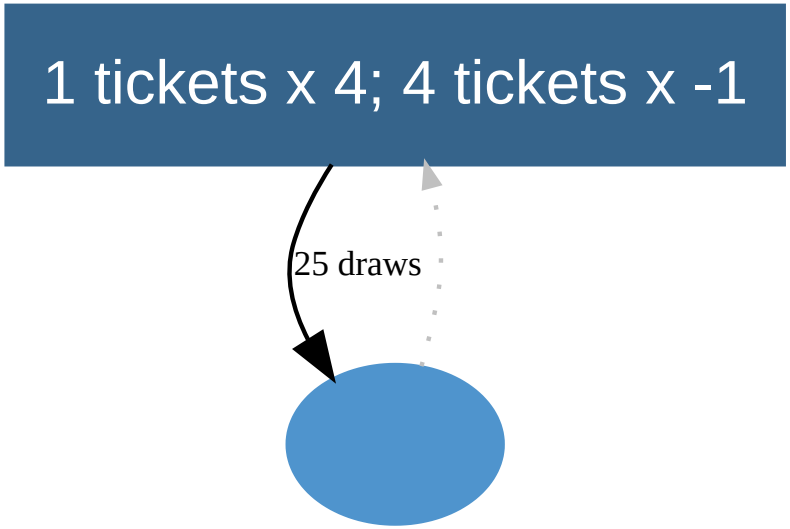
```
## [1] -64
```

```
table(pocket)
```

```
## pocket
## -1 35
## 99  1
```

Example4 (Multiple Choice Quiz)

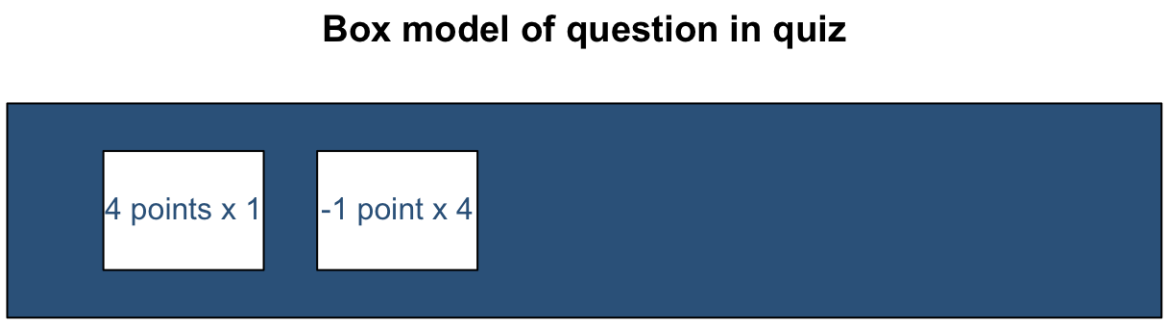
- A quiz has 25 multiple choice questions, with 5 answers each.
  - You get 4 points for a correct answer.
  - You lose 1 point for an incorrect answer.
- If you haven’t studied and guess each answer, what is your expected score?


1 tickets x 4; 4 tickets x -1
25 draws

Box Model of 1 question in the quiz

Note the composition of the box:

- The distinct tickets are 4 and -1, representing “correct answer” and “incorrect answer”.
- There is 1 tickets with 4 (equivalent to the correct answers), and 4 tickets with -1 (equivalent to the 4 incorrect answers).



Simulation

```
set.seed(1)
answer= sample(c(4,-1), 25, repl = T, prob=c(1/5,4/5))
head(answer)
```

```
## [1] -1 -1 -1  4 -1  4
```

```
cumsum(answer)[25]
```

```
## [1] 0
```

```
table(answer)
```

```
## answer
## -1  4
## 20  5
```

Summary

For independent events, it is a mistake to assume that the chance of observing a particular event changes over time, even if the event has not occurred for a long time. This is the Gambler's Fallacy, and downfall! Rather The Law of Large Numbers states that the observed **proportion** of occurences of the event in the long run approaches the expected proportion. The Box Model models a simple chance process involving drawing tickets from a fixed box (population).

Key Words

chance error, chance variability, Law of Large Numbers, box model, box model for gambling

STAT 1040--Ch 16 Box Models

Box Model Help #1 for USU STAT 1040