4. Basic Probability Theory

Transfer exploration seminar: Statistics and Data Science

Dr. Uma Ravat

PSTAT 194TR

We completed a module

Introduction to R

We covered a lot but really just scratched the surface.

You keep learning programing tips and tricks each day!

I still do! And that too, after more than 3 decades coding and working in industy as a software and database engineer too!

Study skills:

Have you gone over lecture, code and made a glossary of functions covered?

Next we will see. . .

Today: New module: Introduction to Probability

- Basic Terminology in Probability
  - Random experiment
  - Sample Space
  - Events
    - Complement, Union, Intersection
    - Independent, Mutually exclusive or disjoint
- Probability approaches
  - Frequentist
  - Relative frequency (Simulation)
- Probability Rules
  - Multiplication Rule
  - Addition

Why learn probability?

- Used plots and summary statistics to explore distributions and relationships of different variables in our (observed) data/sample.
- Now, Statistics aims to generalize these findings to the entire population.

When generalizing from a sample to the population

- There's always some uncertainty about the true distributions and relationships in the population
- We should clearly specify the extent of our uncertainty. (PSTAT 120B)
- Probability is the mathematical tool used to measure and express this uncertainty. (PSTAT 120A)

Courses that build on probability fundamentals

- Measure and express uncertainty in going from sample to population (PSTAT 120B)
- Hypothesis testing (PSTAT 120B)
- Bayesian statistics (PSTAT 115)
- Linear Regression (PSTAT 126)
- Statistical Machine Learning (PSTAT 131)
- Computational statistics (PSTAT 194CS)
    - Monte Carlo methods, Social Network Analysis, AI

Dangers

- Theory not used correctly

Basic definitions and examples

Random experiment

In a **random experiment** there is more than one possible outcome and the exact outcome (value) cannot be determined with certainty before it occurs.

e.g. flip of a coin

e.g. roll of a die

e.g. the bill length of a penguin we will measure as part of our data collection

Sample Space

The set of all these possible outcomes is called the **sample space.**

Experiment	Sample space
Toss a coin	S = $\{$H, T$\}$
Roll a die	S = $\{$ 1 , 2 , 3 , 4 , 5 , 6 $\}$
Toss a single coin twice	S = $\{$ H H , T T , H T , T H $\}$
Measuring bill length	S = $[0,\infty)$ or S = (20 mm, 80 mm)

Uses set theory notation for Sample space

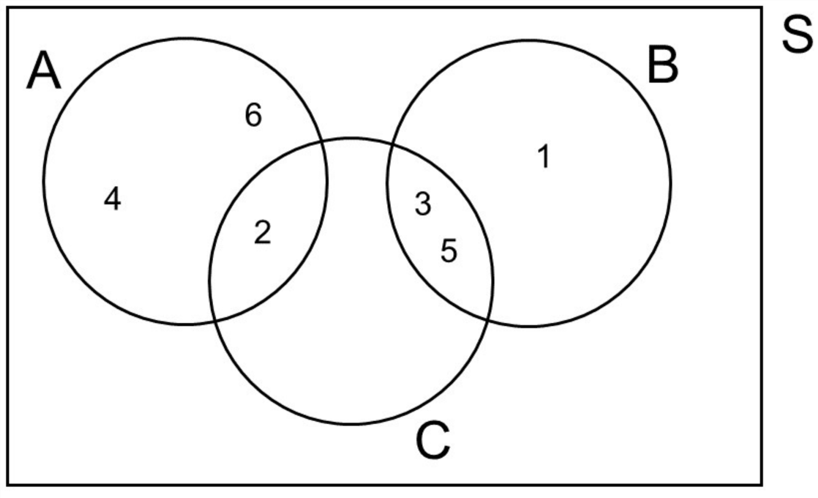
Event

An **event** is a subset of the sample space.

Experiment	Sample space	Subset of sample space	Event in words
Toss a coin	S = {H, T}	E = {H}	Getting a head while tossing a coin
Roll a die	S = { 1 , 2 , 3 , 4 , 5 , 6 }	E = {4}	Rolling a 4 while rolling a die
		A = {2, 4, 6}	Getting an even number
		B = {1, 3, 5}	Rolling an odd number
Measuring bill length	S = (20 mm, 80 mm)	E = (20 mm , 25 mm)	Penguins with small bills

Visualizing events: Venn Diagrams

Events A, B, C for rolling a die example



Your Turn: Example

We roll a six sided die twice and look for a double.

What's the experiment, sample space and event as a subset of the sample space?

Example

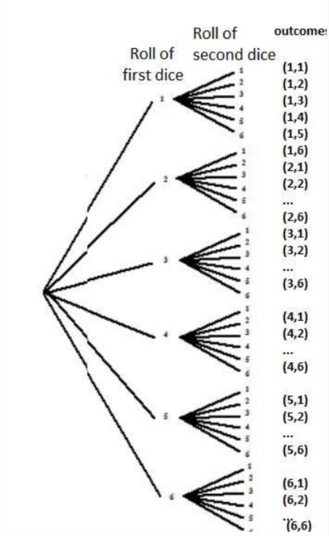
We roll a six sided die twice and look for a double.

Experiment: Roll a six sided die twice OR roll two dice.

Sample space: $S = \{(1, 1), (1, 2), ... , (1, 6),$
$(2, 1), . . . ,. . ., . .(2,6),$
$(6,1). . . ,. . . ,. . . .. (6, 6)\}$ (36 outcomes)

Event: $E = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$

Visualizing two stage experiments: Tree diagram



Classical Approach to Probability

If the number of outcomes is finite and they are **all equally likely to occur**, then

$$\mathbb{P}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of possible outcomes}}$$

Example

- **Experiment:** Roll a fair six sided die once.
- **Sample space:** $\{1, 2, 3, 4, 5, 6\}$
- Find the probability of scoring a 4.
  - Let Event E = scoring a 4 =$\{4\}$
  - $\mathbb{P}(E) = 1/6$.

Example

What's the probability of getting a double when rolling two dice?

- **Experiment:** Roll two dice
- **Sample space:** S = $\{(1, 1), (1, 2), \ldots, (1, 6),$
  $(2, 1), \ldots, \ldots, \ldots (2,6),$
  $(6,1) \ldots, \ldots, \ldots (6, 6)\}$ (36 outcomes)
- **Event:** E = $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$
- Find the probability of getting a double.
  - Let Event E = getting a double
  - $\mathbb{P}(E) = 6/36 = 1/6.$

Simulating probabilities in R

Example: Toss 10 coins

```r
#10 coin tosses
(x<-sample( c("H", "T"), 10, replace = TRUE))
```

```
##  [1] "H" "H" "H" "H" "T" "T" "H" "H" "T" "T"
```

```r
table(x) # How many H vs T
```

```
## x
## H T
## 2 8
```

```r
table(x)/10 #Estimated probabilities
```

```
## x
##   H   T
## 0.2 0.8
```

What happens with 100 coin tosses, 1000 coin tosses?

```
x<-sample( c("H", "T"), 100, replace = TRUE)
table(x)/100
```

```
## x
##    H    T
## 0.55 0.45
```

```
x<-sample( c("H", "T"), 1000, replace = TRUE)
table(x)/1000
```

```
## x
##     H     T
## 0.508 0.492
```

Relative frequency approach to probability

1. repeat the relevant experiment over and over again, say n times
2. count how many times, say e times, the event E occurs in these n reps,
3. The "relative frequency" of the event E is the proportion e/n
4. The *probability* of E is this proportion e/n when we take the number of reps , $n \to \infty$ i.e

$$P(E) = \lim_{n \to \infty} \frac{e}{n}$$

We can simulate the relative frequency approach in `R` using the `sample` function

Simulation based approach to probability

What is the probability of getting "H" when you toss a fair coin?

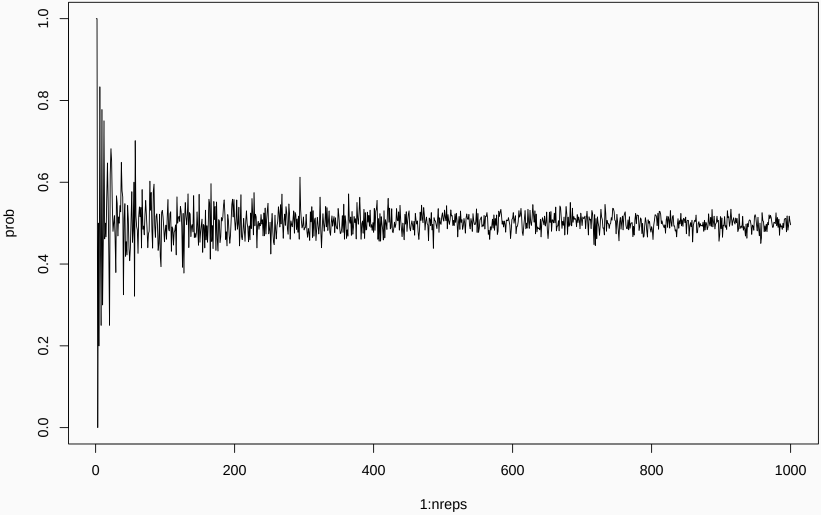
```
nreps <- 1000
prob <- rep(0,nreps)
for(i in 1:nreps){
  coin <- sample(c(1, 0), i, replace = TRUE)
  prob[i] <- sum(coin)/i
}
round(prob[1:10], 2)
```

```
##  [1] 1.00 1.00 0.00 0.50 0.20 0.83 0.43 0.25 0.78 0.30
```

```
round(prob[990:1000], 2)
```

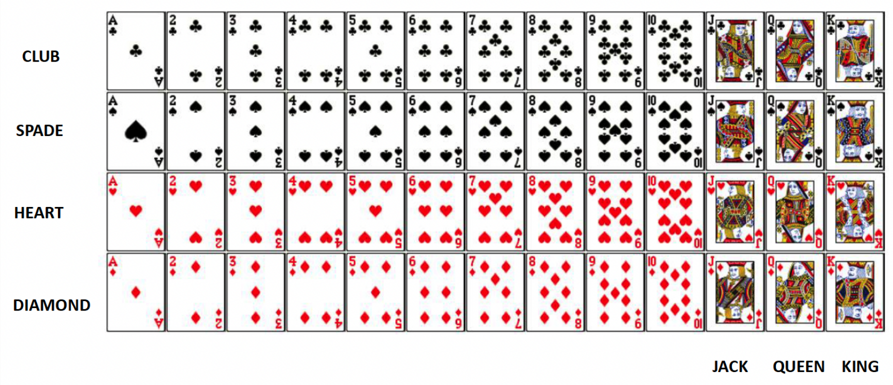
```
##  [1] 0.49 0.51 0.51 0.51 0.48 0.52 0.48 0.49 0.52 0.50 0.50
```

```
plot(1:nreps, prob, type='l')
```



Defining more events and probability rules

Standard Deck of Cards



Probability, Certain & Impossible Events

The probability of any event A, $P(A)$, **is a number between 0 and 1**; $0 \leq P(A) \leq 1$

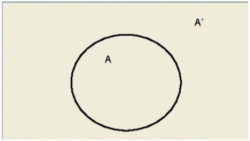
The probability of an event that is **CERTAIN** to occur is 1.

The probability of an **IMPOSSIBLE** event is 0.

- **Experiment:** Roll a fair six sided die once.
- **Sample space:** $\{1, 2, 3, 4, 5, 6\}$
- Let Event A = scoring a 1, 2, 3, 4, 5, OR 6 = $\{1, 2, 3, 4, 5, 6\}$
  - $\mathbb{P}(A) = 6/6 = 1$.
- Let Event B be the event of 'rolling a 7' when a fair six sided die is rolled; then $P(B) = 0/6 = 0$.

Complement of an event

The **complement** of event A, denoted by (A^c or by $\bar{A}$ or A'), is the set of outcomes in the sample space S, that are not included in the outcomes of event A.



Example When drawing a card from a deck, if A consists of an ace, then A' consists of all those cards that are not aces.

Probability of NOT getting an ace is
$P(A') = 1 - P(A) = 1 - 4/52 = 48/52$

Independent Events

DEFINITION: Two events are *independent* if the occurrence of one of the events does not affect the chance(probability) that the other event occurs.

Two events are dependent if the chance of the second event changes depending on whether or not the first event happened.

Example Drawing Cards from a Deck with and without replacement.

- with replacement : independent
- without replacement: dependent.

Are these independent or dependent

- Rolling a fair die twice
- Guessing on each answer for a 10-question True/False test
- Flipping a coin 4 times
- Drawing students to take a survey randomly (without replacement)

Multi-event probabilities: Multiplication Rule

The multiplication rule is used to calculate the probability of two things happening at the same time.

Multiplication Rule (Independent Events)

If we have two *independent events* then the multiplication rule is:

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

Example: Rolling Two Dice

The probability of rolling two dice and getting the first to be "1" and the second to be "2"

$$P(1 \text{ and then a } 2) = P(1) * P(2) = 1/6 * 1/6 = 1/36$$

Multiplication Rule (Dependent Events)

For dependent events, the multiplication rule is

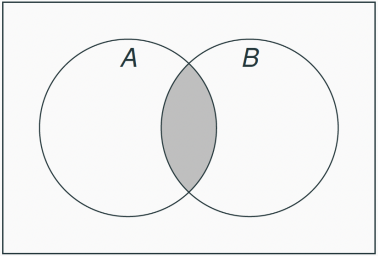
$P(\text{A and B}) = P(A)P(B|A)$, where $P(B|A)$ is the probability of event B given that event A happened.

Example: What's the probability of drawing two cards without replacement from a deck and getting a heart and then a spade?

$P(\text{heart and then spade}) = P(\text{heart})P(\text{spade}|\text{heart}) = 13/52 * 13/52$

Set theory notation: Intersection of events, AND

In general, if A and B are events in the sample space, then the intersection $A \cap B$ denotes the outcomes in "A and B"



$P(A \cap B)$ is called the **joint probability** of A and B occurring.
$P(A)$ and $P(B)$ are called **marginal probabilities**

Draw one card from a standard deck of 52 cards,

- A $=$ Spade $= \{$ AS, 2S, $\ldots$ KS $\}$,
- B $=$ Ace $= \{$AS, AC, AD, AH $\}$
- $A \cap B =$ Spade and Ace $= \{$AS $\}$
- $P(A \cap B) = P(\text{A and B}) = 1/52$

May seem complicated but just need to remember the formula and not list outcomes:

- $P(A \cap B) = P(\text{Spade})P(\text{Ace given Spade}) = 13/52 * 1/13 = 1/52$
- $P(A \cap B) = P(\text{Ace})P(\text{Spade given Ace}) = 4/52 * 1/4 = 1/52$

Multi-event probabilities: Addition Rule

The Addition Rule is used to calculate the probability that either (or both) of 2 events will happen.

$P(\text{A or B}) = P(A) + P(B) - P(\text{A and B})$, where $P(\text{A and B})$ is subtracted to avoid double counting

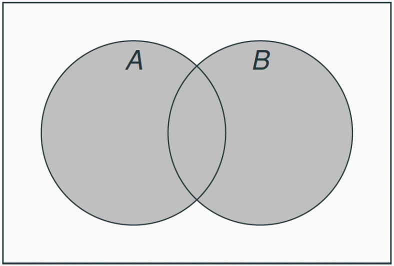
Example:

What's the probability of drawing either a queen or a heart from a standard deck of cards?

$P(\text{queen OR heart}) = P(\text{queen}) + P(\text{heart}) - P(\text{queen AND heart})$
$= 4/52 + 13/52 - 1/52 = 16/52 = 0.3077$

Set theory notation: Union of events, OR

If A and B are events in the sample space, then the union $A \cup B$ denotes the outcomes in "A or B".



Draw one card from a standard deck of 52 cards,

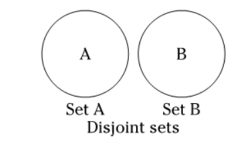
- A = Spade = { AS, 2S, ... KS },
- B = Ace = {AS, AC, AD, AH }
- $A \cup B$ = Spade or Ace = {AS, 2S, ... KS, AC, AD, AH }
- $P(A \cup B) = 16/52$

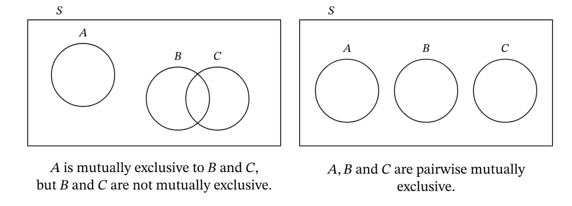
May seem complicated but just need to remember the formula and not list outcomes:

$P(A \cup B) = P(\text{Spade OR Ace})$
$= P(\text{Spade}) + P(\text{Ace}) - P(\text{Spade AND Ace})$
$= 13/52 + 4/52 - 1/52 = 16/52$

Definition: Mutually exclusive or disjoint events

Two events A and B are said to be mutually exclusive if they cannot occur together.





A and B are mutually exclusive then A and B do not share any outcomes, they are non-overlapping. i.e $A \cap B = \phi, P(A \cap B) = 0$

mutually exclusive or disjoint probabilities

If we pull a card from a deck and consider A = Spade, B = Heart

With one card selected, it is impossible to get both a heart and a spade; we may get one or the other but not both.

$A \cap B = \emptyset$ (the empty set) and the joint probability is zero: $P(A \cap B) = 0$

A and B are mutually exclusive or disjoint events.

non mutually exclusive probabilities

A $=$ Spade, B $=$ Ace then $A \cap B = \{\text{AS}\}$

$P(A \cap B) = 1/52$

So, $P(A \cap B) \neq 0$ and A and B are not mutually exclusive.

A and B are not mutually exclusive means they share some outcomes. ie they are overlapping.

Special case of Addition rule for Multi-event probabilities

Recall, The Addition Rule is used to calculate the probability that either (or both) of 2 events will happen

$P(\text{A or B}) = P(A) + P(B) - P(\text{A and B})$, where $P(\text{A and B})$ is subtracted to avoid double counting

Special case: When A and B are mutually exclusive, the addition rule simplifies to

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

Summary: Basic Probability Theory

- Experiment, Sample space, Events
- Probability - Classical(Frequentist) Definition and Simulation based approach
- Probability Properties and Rules:

1. The probability of an event A, denoted by $P(A)$, is a number between 0 and 1. $0 \leq P(A) \leq 1$
2. For the sample space $S, P(S) = 1$
3. $P(\emptyset) = 0$; $\emptyset$ is the null/empty set containing no elements.
4. Complement: $P(A^c) = 1 - P(A)$
5. Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   - **Special case:** If A and B are mutually exclusive events, that is, $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$
6. Multiplication rule: $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$, where $P(B|A)$ is the probability of event B given that event A happened.

Next we will see...

Random Variables and distribution functions