Unit Schedule: Modules

Module	Week	Content	Ross
1.	1	Introduction to modelling for data science and to R	1,2
2.	2	**Probabilities**	3
	3	Expectations	4
	4	Distributions	5
3.	5	Statistical inference	6&7
	6	Hypothesis testing	7&8
4.	7	Dependence and linear regression	9
	8	classification and clustering	
5.	9	Comparing means	10
	10	Random number generation and simulation	
6.	11	Validation and complexity	15
	12	Modelling	

FIT5197 Statistical Data Modelling

Module 2

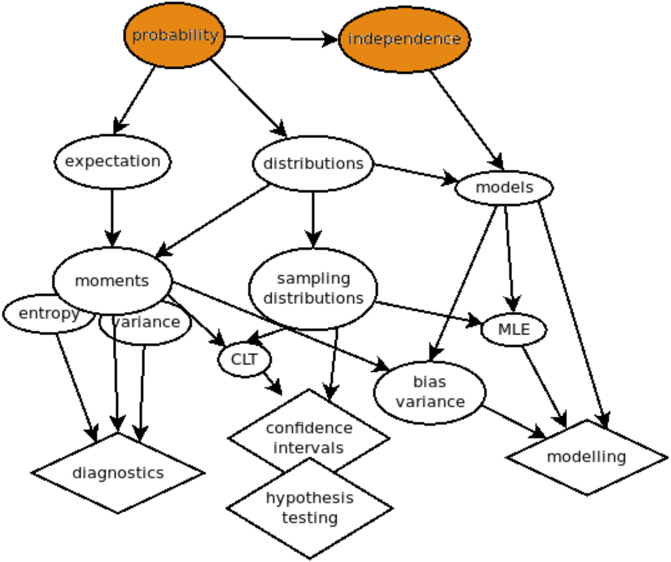
Probability Distribution Theory

2020 Lecture 2

Monash University

polls at *https://flux.qa/43FMK4*

Concept Map for This Unit



Probability and Statistics Notation

Using Probabilities I

First, do the poll at *https://flux.qa/43FMK4*

Consider the statement:

What is the probability you have colon cancer?

- what do we mean exactly?
  - probability a person in this room has it?
  - probability for a random person in Caulfield has it?
    - what is a “random person” in Caulfield?
  - probability for *you* in particular
    - a probability specialised to you
- what do you mean by “have colon cancer”?
  - have a cancerous tumor at least 1mm across?
  - a biopsy confirms the presence of cancerous cells?
  - are expected to be sick from cancer within 6 months?

Using Probabilities II

You walk into the oncologists office and she says,

> *"after reviewing your test results, I would say you have moderate chance of having colon cancer"*

- Why does she say "moderate" and not, say 15%?
- She says the advice is contingent on having reviewed your test results. Presumably, if she hadn't read your test results, the probability would be a lot lower.
- Again, what does she mean by "have colon cancer"?

Making Decisions

- As a data scientist your job is to build systems to support decision making:
  - your built system may make the decisions itself
  - or it may summarise appropriate evidence so that others can make decisions
- So you need to understand what is the intellectual apparatus we use to support decision making:
  - alternative outcomes,
  - probabilities,
  - costs and benefits
- We introduce this with the dry but precise notation of probabilities, sets, events and so forth.

Some Notation

Basic set notation

- We use $\{a, b, c\}$ to denote a set with elements a, b and c
- We use $x \in \mathcal{X}$ to denote that x is an element of the set $\mathcal{X}$
  - Example: $3 \in \{1, 2, 3, 4, 5\}$
- We use $A \subseteq \mathcal{X}$ to denote that A is a subset of the set $\mathcal{X}$
  - Example: $\{2, 3, 4\} \subseteq \{1, 2, 3, 4, 5\}$

Some important sets:

- $\mathbb{Z}$ is the set of all integers;
- $\mathbb{Z}_+$ is the set of non-negative integers;
- $\mathbb{R}$ is the set of all real numbers;
- $\mathbb{R}_+$ is the set of non-negative numbers;
- $[0, 1]$ is the subset of $\mathbb{R}$ between 0 and 1, including 0 and 1.

Random Variables

A random variable (RV) is a variable that takes on a value from a set of possible values with specified probabilities

- We can let $\mathcal{X}$ denote the possible set of values
- $\mathcal{X}$ could be discrete, real, vector, structured, ...

Often use capital letters to denote a random variable

Example: let X be a random variable over $\mathcal{X} = \{1, 2, 3\}$ with:

$$X = \begin{cases} 1 & \text{with probability } 1/2 \\ 2 & \text{with probability } 1/4 \\ 3 & \text{with probability } 1/4 \end{cases},$$

Example:
Visualising random variables, Terry Tao, 13/May/2016

Random Variables, cont.

Use the language of probability distributions to describe random variables

$$P(X = x), \; x \in \mathcal{X}$$

describes the probability that the RV X takes on the value x from $\mathcal{X}$.

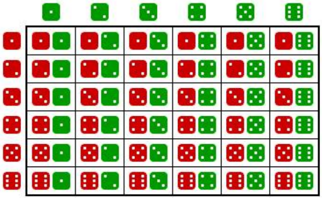
Example: use this notation to describe the example above.

$$P(X = x) = \begin{cases} 1/2 & \text{for } x = 1 \\ 1/4 & \text{for } x = 2 \\ 1/4 & \text{for } x = 3 \end{cases}$$

Sample Space

What is the space of possibilities?

- **aka** the *sample space* (when considering experimental outcomes)
- **aka** the *universal set* U (when considering set theory)
- **aka** "top" or the set of everything denoted as Ω
- **aka** the full domain $\mathcal{X}$ for values $x \in \mathcal{X}$
- for rolling two die it is given by entries in the table



- for calling the C library function `rand()` it is the set of 32 bit non-negative integers

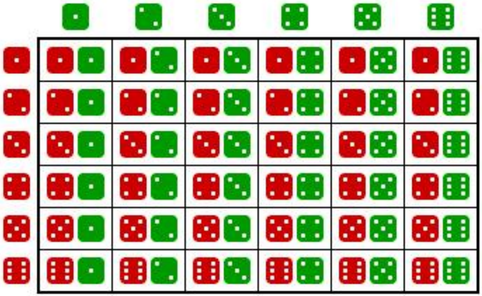
Events

- an event is any subset of the sample space
- it could be atomic (a single element), or multi-element (a subset)
- probabilities are defined on events,
  - $P(X = x)$ for $x \in \mathcal{X}$
  - $P(X \in A)$ for $A \subseteq \mathcal{X}$
- events oftentimes defined using logic

  e.g. $A = \text{height}{>}150\text{cm} \wedge \text{gender}{=}\text{Male}$
- abusing notation by mixing logic and set theory:
  - $A \cap B$ is essentially the same as $A \wedge B$
  - $A \cup B$ is essentially the same as $A \vee B$

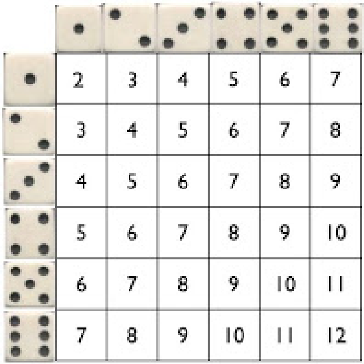
Probability and Statistics Examples

Rolling Two Dice



- each outcome for a die $\{1, 2, 3, 4, 5, 6\}$ equally likely
- each pair outcome equally likely and there are 36 outcomes
- probability of each cell is $\frac{1}{36}$

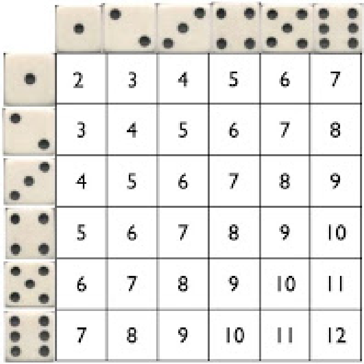
Rolling Two Dice, cont.



2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

- now look at the numeric total of the two dice
- look at the ways of getting a 4 from 2 dice (1+3, 2+2, 3+1)
- look at the ways of getting a 6 from 2 dice (1+5, 2+4, 3+3, 4+2, 5+1)
- so what is the probability of getting a numeric total of 4 or 6 from the dice?

Rolling Two Dice, cont.



- now look at the numeric total of the two dice
- look at the ways of getting a 4 from 2 dice (1+3, 2+2, 3+1)
- look at the ways of getting a 6 from 2 dice (1+5, 2+4, 3+3, 4+2, 5+1)
- so what is the probability of getting a numeric total of 4 or 6 from the dice?

$$\frac{3+5}{36}$$

Statistics of Rolling Dice

"Statistic" definition: piece of data obtained from a study of a large quantity of other data. Lets look at *statistics of rolling dice*.

1. try just one die, and see how slowly it converges to a uniform distribution
2. try 2 dice:
   - after 1000 points the histogram resembles an inverted "V"
     - can you explain why?
   - what is the minimum possible? and what probability would it occur?
3. try 6 dice:
   - after about 1000 points the histogram resembles a Gaussian curve ("bell" curve)
     - can you explain why?
   - what is the minimum possible? and what probability?
   - what is the probability of throwing a total of 7 or 8 or 9?

Statistics of Drawing Cards

Use a full deck with no jokers, draw 5 cards without replacement.

First, do the next two polls at *https://flux.qa/43FMK4* .

Now see *Playing Card Shuffler*

Probability of Drawing Cards

What is the probability of getting no red cards?

NB. one might guess at approximately $\left(\frac{1}{2}\right)^5$

- but it is not exact ... without replacement means probability of black changes as you draw

1. probability first card is black $\frac{26}{52}$
2. include probability second card is black $\frac{26}{52}\frac{25}{51}$
3. include probability third card is black $\frac{26}{52}\frac{25}{51}\frac{24}{50}$
4. ...

$$= \frac{26}{52}\frac{25}{51}\frac{24}{50}\frac{23}{49}\frac{22}{48} = 0.025310...$$

Prob. of Drawing Cards, cont.

What is the probability of getting at least one ace?

- the 1st card is an ace
  or the 2nd card is an ace
  or the 3rd card is an ace
  or ...
- $\frac{4}{52}$
  $+ \frac{4}{52}$
  $+ \frac{4}{52}$
  $+ \ldots$

This is **wrong** because of double counting: a hand with
"1st card is an ace and 2nd card is an ace, all others are not"
is counted twice, in the first two $\frac{4}{52}$ entries

Prob. of Drawing Cards, cont.

What is the probability of getting at least one ace?

- the 1st card is an ace
  or the 1st card is not an ace but the 2nd card is
  or the 1st & 2nd cards are not an ace but the 3rd card is
  or ...
- $\frac{4}{52}$
  $+\frac{48}{52}\frac{4}{51}$
  $+\frac{48}{52}\frac{47}{51}\frac{4}{50}$
  $+\ldots$

$$= \frac{4}{52} + \frac{48}{52}\frac{4}{51} + \frac{48}{52}\frac{47}{51}\frac{4}{50} + \frac{48}{52}\frac{47}{51}\frac{46}{50}\frac{4}{49} + \frac{48}{52}\frac{47}{51}\frac{46}{50}\frac{45}{49}\frac{4}{48} = 0.341158\ldots$$

Prob. of Drawing Cards, cont.

What is the probability of getting at least one ace?

- alternatively **it is** (1- probability of getting no ace)
- second part is same logic as getting no red cards

$$= 1 - \frac{48}{52}\frac{47}{51}\frac{46}{50}\frac{45}{49}\frac{44}{48} = 0.341158...$$

Same answer as before **...** but a lot easier!

Probability theory is logically consistent, so you should get the same answer as long as the logic is correct!

Probabilities: Observations

- probabilities in these cases obtained by carefully enumerating possibilities
- need to ensure you don't do double counting
- probability theory is logically consistent, so you should get the same answer no matter what your strategy was

Probabilities

Probability, Examples

Consider probabilities of:

1. *a strong virus alert will be announced for Windows in the next week*
2. *the Euro will go above AU$1.50 later in 2020*
3. *the height of a Singaporean male is between 180–185cm*
4. *a Singaporean male is tall*

- probabilities must be for well-defined events: (4) is not (what is “tall”), (1) possibly not
- those for one-off unrepeatable events, (2), cannot be sampled, so cannot be frequencies
- probabilities are always context dependent: (2) will vary during 2020, (3) changes when in a kindergarten
- continuous events need to be discretized, as done for (3)

Continuous Domains

> *"John's height is 60π cm, or 188.495559215387594307758602996770 1740... cm"*
>
> *"In Indonesia, a person with tertiary education earns an average 82% more than one with secondary qualifications"*

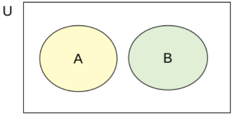
- in the real world, the only evidence we see is discrete, and the only valid statements we can make are about discrete events
- continuous domains are a useful abstraction
- integration is about taking a continuous function and getting probabilities for finite/discrete parts

Probability Basics

Probability is an additive measure:

- it behaves like a weight or an area
- it is always non-negative
- given a domain $\mathcal{X}$, we can measure probability for elements or subsets,
  - $P(x)$ for $x \in \mathcal{X}$
  - $P(A)$ for $A \subseteq \mathcal{X}$
- we can break something into separate parts and measure them independently
- if possibilities A and B cannot occur together, then $P(A \cap B) = 0$ and

$$P(A \cup B) = P(A) + P(B)$$



Probability Basics, cont.

Probability is always normalised relative to the current space of possibilities

- for universal set U,
$$P(U) = 1$$
- for domain $\mathcal{X}$ and event $A \subseteq \mathcal{X}$
$$0 \leq P(A) \leq 1$$
- for elements $x \in \mathcal{X}$
$$\sum_{x \in \mathcal{X}} P(x) = 1$$
- when throwing a 6-sided dice, $P(1, 2, 3, 4, 5 \textit{ or } 6) = 1$
- when throwing a 6-sided dice, **but** we are also told the outcome is even, $P(2, 4 \textit{ or } 6 | \textit{even}) = 1$;
moreover, $P(1, 3 \textit{ or } 5 | \textit{even}) = 0$

Conditional Probability

We change the domain of a probability by conditioning:

- given events A and B we renormalise

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

  so the domain of the conditional probability is now A
- before conditioning

$$\sum_{x \in \mathcal{X}} P(x) = 1$$

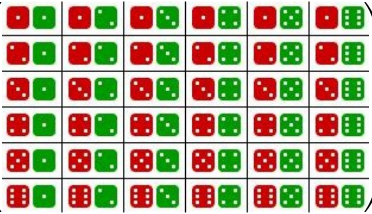
- **after** conditioning on A

$$\sum_{x \in \mathcal{A}} P(x|A) = 1$$

- the ratios are the same, just the scale changes

Conditional Probability, cont.

When rolling two die:

$P\Big($  $\Big) = 1$

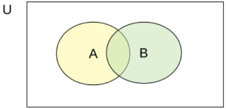
$P\Big($  $\Big) = \frac{1}{6}$

$P\big($  $\big) = \frac{1}{6}$

$P\Big($  $\Big|$  $\Big) = 1$

Probability Identities: I

the probability of the union of A and B



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

i.e. we remove double counting of $p(A \cap B)$

N.B. think if it as a result of “measure"

Probability Identities: II

the probability of the complement of A is derived from the probability of A

$$P(\overline{A}) = 1 - P(A)$$

e.g. make $B = \overline{A}$ in the union formula

N.B. think if it as a result of normalisation

Probability Axioms

The so-called *probability axioms* of Kolmogorov.

Probability Axioms:

1. for any event A, $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. for mutually exclusive events A_1, ... A_n
   $P(A_1 \cup A_2 \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i)$

From these, further probability rules can be derived. For the domain $\mathcal{X} \times \mathcal{Y}$ where A, B are any events:

Complement rule	$P(\overline{A}) = 1 - P(A)$
Product rule	$P(B \cap A) = P(B\|A)p(A)$
Sum rule	$P(A) = \sum_{x \in \mathcal{X}} P(x \cap A)$
Bayes Theorem	$P(x\|A) = \frac{P(A\|x)P(x)}{\sum_{x \in \mathcal{X}} P(A\|x)p(x)}$

Sum Rule

Take the bivariate discrete distribution $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{red, yellow, blue\}$.

Let $P(X, Y)$ be specified by the table

	Y=*red*	Y=*yellow*	Y=*blue*
X=0	0.05	0.15	0.1
X=1	0.25	0.15	0.3

By the sum rule,

$$\begin{aligned} P(Y{=}red) &= P(X{=}0 \cap Y{=}red) + P(X{=}1 \cap Y{=}red) \\ P(X{=}1) &= P(Y{=}red \cap X{=}1) + P(Y{=}yellow \cap X{=}1) \\ &\quad + P(Y{=}blue \cap X{=}1) \end{aligned}$$

$P(Y)$ is denoted the marginal for Y and $P(X)$ is denoted the marginal for X.

Product Rule

Take the bivariate discrete distribution $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{red, yellow, blue\}$.

Let $P(X, Y)$ be specified by the table

	$Y{=}red$	$Y{=}yellow$	$Y{=}blue$
$X{=}0$	0.05	0.15	0.1
$X{=}1$	0.25	0.15	0.3

By the product rule,

$$\begin{aligned} P(X{=}0 | Y{=}red) &= \frac{P(X{=}0 \cap Y{=}red)}{P(Y{=}red)} = 1/6 \\ P(Y{=}red | X{=}1) &= \frac{P(Y{=}red \cap X{=}1)}{P(X{=}1)} = 5/14 \end{aligned}$$

Probabilities for Discrete Random Variables

Discrete Random Variables

Random variable whose set of possible values is a sequence is said to be discrete.

- Define probability mass function $p(a)$ of X by

$$p(a) = P(X = a)$$

- The cumulative distribution function:

$$F(a) = \sum_{all\, x \leq a} p(x)$$

For any discrete probability distribution:

- for all x $0 \leq p(x) \leq 1$
- $\sum_{all\, x} p(x) = 1$

Probabilities for Continuous Domains

Continuous Random Variables

- so far we have considered only discrete random variables
- the ideas extend to the case that the values X can take on form a continuum, that is, $\mathcal{X} \subseteq \mathbb{R}$
- X now follows a probability density function (pdf) $P(X{=}x) \equiv f(x)$.
- mathematical texts usually properly distinguish between a pdf and a probability!
- a pdf $f(x)$ on domain $\mathcal{X}$ satisfies

$$f(x) \geq 0 \text{ for all } x \in \mathcal{X}$$

and

$$\int_{\mathcal{X}} f(x)dx = 1$$

Continuous RVs, cont.

- The probability that X lies in an interval (a, b) is

$$P(a < X < b) = \int_a^b f(x)dx.$$

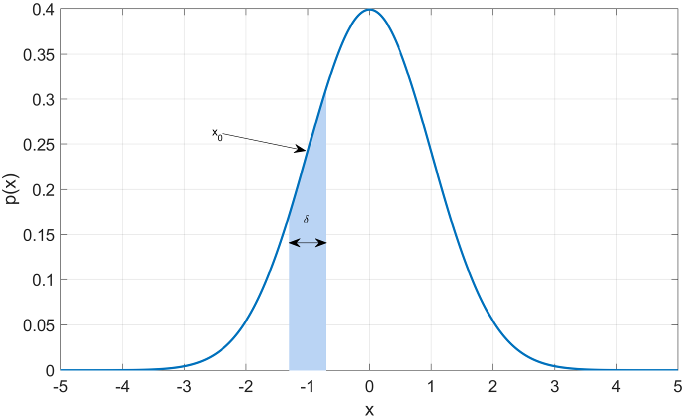
- More generally, the probability $X \in A$, where $A \subset \mathcal{X}$ is

$$P(X \in A) = \int_A f(x)dx.$$

- This implies that $P(X{=}x) = 0$
$\Longrightarrow$ One of the most confusing aspects of continous RVs

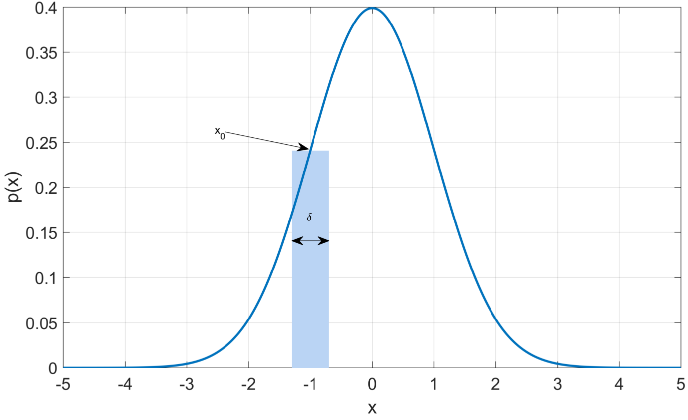
Continuous RVs, cont.

- Example: Probability of $(x_0 - \delta/2 < X < x_0 + \delta/2)$



Continuous RVs, cont.

- If δ is small enough then $\int_{x_0-\delta/2}^{x_0+\delta/2} f(x)dx \approx f(x_0)\delta$
  $\implies$ Take $\delta \to 0$ and it is clear why $P(X{=}x) = 0$.



Cumulative Distributions

The cumulative distribution function (cdf) of a continuous RV is:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x')dx'$$

that is, the probability that X is less than some value x

Then,

- $f(x) \geq 0$ for all $\infty < x < +\infty$
- $\int_{-\infty}^{+\infty} f(x) = 1$

Cumulative Distributions

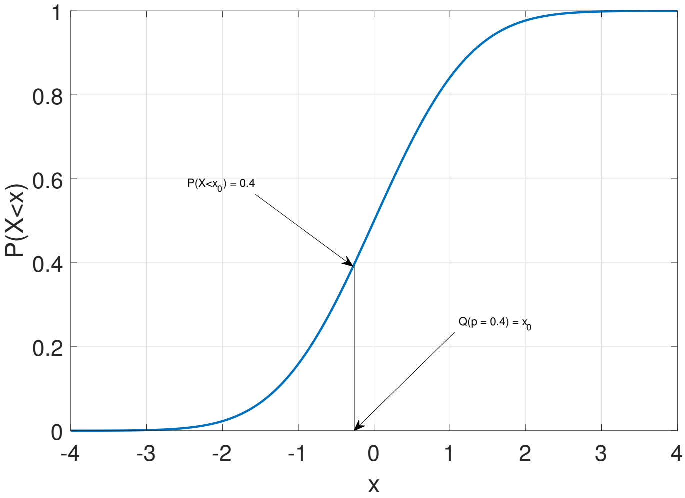
The **inverse cdf** is

$$Q(p) = \{x \in \mathcal{X} : p(X \leq x) = p\}$$

which is sometimes called the **quantile function.**

- In words, the quantile function says: find the the value x such that the probability that $X \leq x$ is p
- For example:
  - $Q(p = 1/2)$ is the median;
  - $Q(p = 1/4)$ is the first quartile; and
  - $Q(p = 3/4)$ is the third quartile.

Cumulative Distributions, cont.



Probabilities in Models

Continuous RVs, Properties

Probability density functions for X and Y: $f(x, y)$

- The probability of some set $A \subseteq \mathcal{X} \times \mathcal{Y}$

$$F(A) = \int_A f(x, y)\, dx\, dy$$

- The sum rule is then given by

$$f(x) = \int_{\mathcal{Y}} f(x, y) dy$$

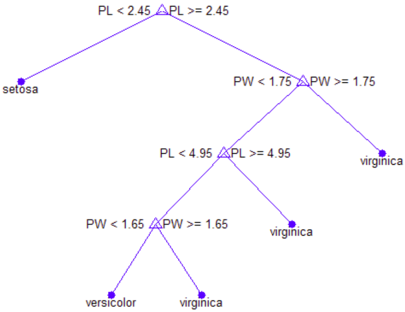
- The other rules follow accordingly

A Simple Probability Model

Let θ be “the frequency of males in Singapore over 170cm height”.

- this is a real world frequency so it exists
- however, it changes minute by minute
  - old people shrink slightly and young people grow!
  - daily, probably only changes in the 6th decimal place
- we cannot realistically measure it to 6 decimal places
- we can estimate it to a 2 or 3 decimal place accuracy by measuring enough Singaporean males
- call θ a parameter of a probability model

A Simple Classification Model



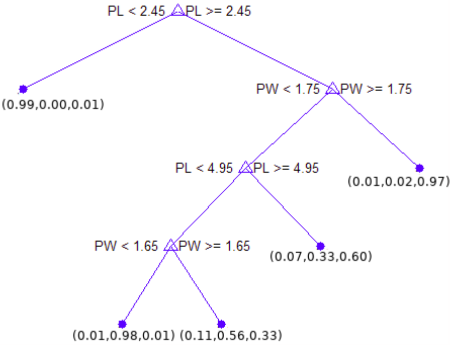
(as a tree)

```
if PL < 2.45
then class = setosa
else if PW < 1.75
    then if PL < 4.95
        then
            if PW < 1.65
            then class = versicolor
            else class = virginica
        else  #  PL < 4.95
            class = virginica
        endif  #  PL < 4.95
    else  #  PW >= 1.75
        class = virginica
    endif  #  PW >= 1.75
```

(as psuedocode)

a decision tree predicting class for the *Iris data*
it is not a probability model

A Class Probability Model



(as a class probability tree)

- each 3-dim vector is the probability of setosa, versicolor and virginica
- represents a conditional probability in the form $p(species|PL, PW, SL, SW)$
- matches a real world frequency, so can be measured

NB. this model has 12 parameters: 4 “cut-point” parameters (2.46,1.75,4.95,1.65) and 4x3-dim vectors, with 2 parameters each

Generative Probability Models

- later on we will develop probability models for regression, classification and clustering tasks
- these will be in forms like:
  - $p(species, PL, PW, SL, SW)$
  - $p(species | PL, PW, SL, SW)$
  - $p(PL, PW, SL, SW | species)$
- probability models are needed to develop sampling results and give probability predictions
- in learning we seek to estimate the model parameters from training data

Independence

Motivating Independence

- When framing a prediction problem, we often go and find out what variables are likely to influence our target.
  - e.g. suppose you want to predict whether a patient coming to your office has colon cancer, without doing an expensive biopsy
- So we list out some relevant/predictive variables:
  - e.g. regularly eat hot chillies;
    drink a lot of alcohol;
    parents had colon cancer
- But using causal reasoning we know not to include irrelevant variables:
  - e.g. the number of letters in their first name;
    the colour of their car;
    the day of the week for their first visit

Causality and relevance are dealt with in probability theory using the notion of **independence**.

Independence

Independence:
Let the random variable pair (X, Y) be from domain $\mathcal{X} \times \mathcal{Y}$. We say X and Y are **independent** if any of the following three (equivalent) conditions hold for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$
\begin{aligned}
P(X{=}x \mid Y{=}y) &= P(X{=}x) && \text{when } p(Y{=}y) > 0 \\
P(Y{=}y \mid X{=}x) &= P(Y{=}y) && \text{when } P(X{=}x) > 0 \\
P(Y{=}y \cap X{=}x) &= P(X{=}x)P(Y{=}y)
\end{aligned}
$$

- notice the three equalities are known to be equivalent (exercise: why?)

Independence, Example

Simple Booleans:
You have two Boolean valued events A and B. A full table of the joint probabilities is as follows:

	B=*true*	B=*false*
A=*true*	0.04	0.06
A=*false*	0.36	0.54

Are A and B independent?

The marginal calculations by the sum rule give $p(A{=}true) = 0.1$ and $p(B{=}true) = 0.4$. From this we can confirm that for all cases of the table $p(A \cap B) = p(A)p(B)$, so independence holds.

Independence, Example

Coin problem: You toss a biased coin ten separate times. Let the ten Boolean RVs $H_1, ..., H_{10}$ indicate whether the coin tossed head. So $H_i = \textit{true}$ if the i-th toss yielded a head.

What is the probability of $P(H_1, ..., H_{10})$?

Causality in this case implies joint independence of the tosses. So for all outcomes of the RVs

$$P(H_1, ..., H_{10}) = P(H_1) \cdot ... \cdot P(H_{10}) = \prod_{i=1}^{10} P(H_i)$$

Bayes Theorem

Bayes Theorem

Bayes Theorem:
on domain $\mathcal{X} \times \mathcal{Y}$ for $A \subseteq \mathcal{X} \times \mathcal{Y}$ and $x \in \mathcal{X}$

$$P(x|A) = \frac{P(A|x)P(x)}{P(A)} = \frac{P(A|x)P(x)}{\sum_{x \in \mathcal{X}} P(A|x)P(x)}$$

Someone tells you a regular die has rolled odd. What is the probability it will be a 3:

$$P(3|odd) = \frac{P(odd|3)P(3)}{P(odd)} = \frac{1\ \frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Bayes Theorem: Example

Cancer problem: You have been referred to a speciality clinic. You have been told that 1/100 who go to the clinic have cancer X. Tests are positive 80% of the time for those with cancer X, and negative 80% of the time for those without.

What is the probability you have cancer X?

Since you haven't been tested yet, it is 1/100, just 1%.

Bayes Theorem: Example

Cancer problem: You have been referred to a speciality clinic. You have been told that 1/100 who go to the clinic have cancer X. Tests are positive 80% of the time for those with cancer X, and negative 80% of the time for those without.

Now you test positive. What is the probability you have cancer X?

$P(D) = 0.01$, $P(T+|D) = 0.8$, $P(T-|N) = 0.8$, $P(T+|N) = 0.2$

We apply Bayes theorem,

$$
\begin{aligned}
P(D|T+) &= \frac{P(T+|D)P(D)}{P(T+)} = \frac{P(T+|D)P(D)}{P(T+|D)p(D) + P(T+|N)P(N)} \\
&= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.2 \times 0.99} = \frac{8}{8 + 198} \approx 0.039
\end{aligned}
$$

Bayes Theorem: Example

Cancer problem: You have been referred to a speciality clinic. You have been told that 1/100 who go to the clinic have cancer X. Tests are positive 80% of the time for those with cancer X, and negative 80% of the time for those without.

With the test 80% accurate, then $P(D|T+) \approx 0.039$ and $p(D|T-) \approx 0.0025$

- after the test you are roughly 4 times more/less likely to have cancer X then before the test
- the low test accuracy means things don't change that much
- with a positive test still only 4/100 chance of cancer X
- medical screening tests often work this way!

Beliefs

Kinds of Probabilities

Consider probabilities of:

1. *a virus warning by Microsoft will be announced for Windows 10 in the next week*
2. *the Euro will go above AU$1.50 at some time in 09-12/2020*
3. *the height of a Singaporean male is between 180–185cm*
4. *an Iris flower with petal length less than 2.45mm is species setosa*

- items (1) and (2) are not about events with a practical sample space
  - for (1) next week (for Windows) is unique in history
  - for (2) the later part of 2020 is also unique
- we can rethrow a dice, but we cannot replay history
- therefore we cannot meaningfully talk about probabilities for items (1) and (2) as long term frequencies

One-off Events

Decision making about one-off events in complex dynamic contexts has this problem all the time:

- betting on a specific horse in a given race
- betting on candidates in the 2020 USA election
- USD versus JPY currency trading
- etc.

We can talk about probabilities for these sorts of events:

- it no longer falls under the context of random variables: variables with **known probability** of outcomes
- the probabilities cannot be measured or estimated; they do not correspond to real world frequencies
- we refer to them as beliefs or subjective probabilities

Caution

- subjectivity and science are usually considered incompatible
- but if you must make a decision in a one-off context, you have no choice but to be subjective
- subjectivity is therefore acceptable in many intelligence or robotics contexts
- but not if you are advising the Federal Government on the safety of a drug

Three Kinds of Probabilities

Proportions: *e.g.*, the height of a Singaporean male is between 180–185cm

- true proportions about the world
- measurable but sometimes approximated
- Kolmogorov Axioms hold

Beliefs: *e.g.*, Euro will go above AU$2.00 in 09-12/2020

- (subjective) beliefs, particularly on one-off events
- not practically measurable on real world data
- need to be elicited

Beliefs about Proportions: *e.g.*, Warren Buffet's belief that the "Euro will go above AU$2.00 in 09-12/2020"

- (subjective) beliefs about true proportions (or other parameters of a probability distribution)
- not practically measurable
- realm of Bayesian Statistics or Full Probability Modelling

Beliefs as Probabilities

- beliefs are hard to justify in an objective sense
- but if we are going to work with them, then using a probability calculus to manipulate them makes sense
- historically, this is controversial in statistics and science

End of Week 2