# COMP2610 / COMP6261 - Information Theory

## Lecture 3: Probability Theory and Bayes' Rule



Australian National University

July 30, 2018

# Last time



- A general communication system
- Wh



- Basics of probability theory



- Joint, marginal and conditional distributions

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

$$p(A = 1, B = 0)$$
$$p(B = 1)$$
$$p(A = 0)$$
$$p(B = 1 | A = 1)$$
$$p(A = 0 | B = 0)$$

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | |
|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ |
| $p(B = 1)$ | |
| $p(A = 0)$ | |
| $p(B = 1 \vert A = 1)$ | |
| $p(A = 0 \vert B = 0)$ | |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | |
|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ |
| $p(B = 1)$ | $100/1000$ |
| $p(A = 0)$ | |
| $p(B = 1 \mid A = 1)$ | |
| $p(A = 0 \mid B = 0)$ | |

## Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | |
|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ |
| $p(B = 1)$ | $100/1000$ |
| $p(A = 0)$ | $690/1000$ |
| $p(B = 1 \vert A = 1)$ | |
| $p(A = 0 \vert B = 0)$ | |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are “brilliant” or not, $B = \{0, 1\}$



(Aside: “Brilliance” is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | |
|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ |
| $p(B = 1)$ | $100/1000$ |
| $p(A = 0)$ | $690/1000$ |
| $p(B = 1 \vert A = 1)$ | $90/310$ |
| $p(A = 0 \vert B = 0)$ | |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | |
|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ |
| $p(B = 1)$ | $100/1000$ |
| $p(A = 0)$ | $690/1000$ |
| $p(B = 1 \mid A = 1)$ | $90/310$ |
| $p(A = 0 \mid B = 0)$ | $680/900$ |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are “brilliant” or not, $B = \{0, 1\}$



(Aside: “Brilliance” is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | | |
|---|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ | Joint |
| $p(B = 1)$ | $100/1000$ | |
| $p(A = 0)$ | $690/1000$ | |
| $p(B = 1 \mid A = 1)$ | $90/310$ | |
| $p(A = 0 \mid B = 0)$ | $680/900$ | |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | | |
|---|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ | Joint |
| $p(B = 1)$ | $100/1000$ | Marginal |
| $p(A = 0)$ | $690/1000$ | |
| $p(B = 1 \mid A = 1)$ | $90/310$ | |
| $p(A = 0 \mid B = 0)$ | $680/900$ | |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | | |
|---|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ | Joint |
| $p(B = 1)$ | $100/1000$ | Marginal |
| $p(A = 0)$ | $690/1000$ | Marginal |
| $p(B = 1 \vert A = 1)$ | $90/310$ | |
| $p(A = 0 \vert B = 0)$ | $680/900$ | |

# Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are "brilliant" or not, $B = \{0, 1\}$



(Aside: "Brilliance" is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | | |
|---|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ | Joint |
| $p(B = 1)$ | $100/1000$ | Marginal |
| $p(A = 0)$ | $690/1000$ | Marginal |
| $p(B = 1 \vert A = 1)$ | $90/310$ | Conditional |
| $p(A = 0 \vert B = 0)$ | $680/900$ | |

## Review Exercise

Suppose I go through the records for $N = 1000$ students, checking their admission status, $A = \{0, 1\}$, and whether they are “brilliant” or not, $B = \{0, 1\}$



(Aside: “Brilliance” is a dodgy concept, and does not predict scientific achievement as well as persistence and combinatorial ability; see e.g. Dean Simonton, *Scientific Genius: A Psychology of Science*, Cambridge University Press, 2009; this is just a toy example!)

Say that th



| | | |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |



Then:

| | | |
|---|---|---|
| $p(A = 1, B = 0)$ | $220/1000$ | Joint |
| $p(B = 1)$ | $100/1000$ | Marginal |
| $p(A = 0)$ | $690/1000$ | Marginal |
| $p(B = 1 \mid A = 1)$ | $90/310$ | Conditional |
| $p(A = 0 \mid B = 0)$ | $680/900$ | Conditional |

# This time



- More on joint, marginal and conditional distributions



- Wh

- What, if anything, does $p(X = x|Y =$
  $p(Y = y|X = x)$?

# This time



- More on joint, marginal and conditional distributions
- Wh 
- What, if anything, does $p(X = x | Y =$ $p(Y = y | X = x)$?



Philosophically related to “How do we know / learn ab

# This time



- More on joint, marginal and conditional distributions
- Wh



- What, if anything, does $p(X = x | Y =$
  $p(Y = y | X = x)$?



Philosophically related to “How do we know / learn ab

I am *not* providing a general answer; but keep it in mind!

# Outline

1. More on Joint, Marginal and Conditional Distributions
2. Stati
3. Bayes' Theorem
4. Wrapping up

1 More on Joint, Marginal and Conditional Distributions



2 Stati



3 Baye



4 Wrapping up

# Document Modelling Example

Suppose we have a large document of English text, represented as a sequence of characters:



$$X_1 X_2 X_3 \ldots X_N$$

- e.g. `hello␣how␣are␣you`



Treat eac ... variable



```
X = ‘h’, Y = ‘e’

X = ‘e’, Y = ‘l’

X = ‘l’, Y = ‘l’
```

$$\vdots$$

# Document Modelling: Marginal and Joint Distributions

| $i$ | $a_i$ | $p_i$ |
|---|---|---|
| 1 | a | 0.0575 |
| 2 | b | 0.0128 |
| 3 | c | 0.0263 |
| 4 | d | 0.0285 |
| 5 | e | [illegible] |
| 6 | f | [illegible] |
| [illegible] | g | 0.0133 |
| 8 | h | 0.0313 |
| 9 | i | 0.0599 |
| 10 | j | 0.0006 |
| 11 | k | 0.0084 |
| 26 | z | 0.0007 |
| 27 | – | 0.1928 |



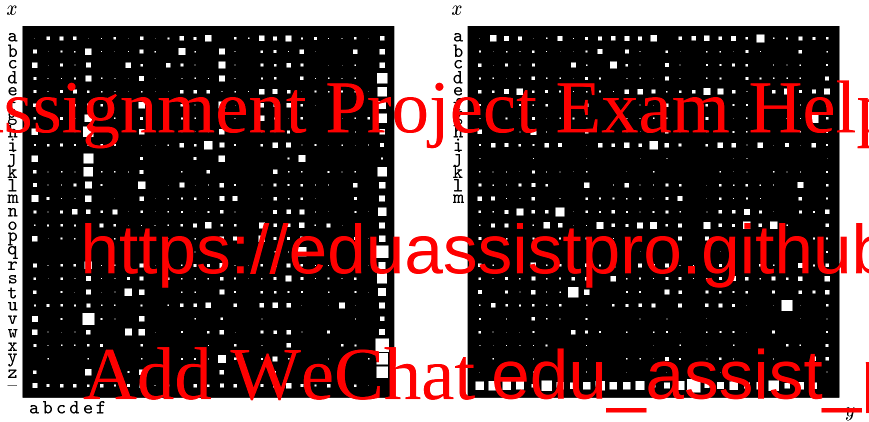


Unigram / Monogram

Bigram

Marginal and joint distributions for English alphabet, estimated from the "FAQ manual for Linux". Figure from Mackay (ITILA, 2003); areas of squares proportional to probability (the right way to do it!).

# Document Modelling: Conditional Distributions






Conditional distributions for English alphabet, estimated from the “FAQ manual for Linux”. Are these distributions “symmetric”? Figure from Mackay (ITILA, 2003)

$P(X = x|Y = y) = P(Y = y|X = x)$? $P(X = x|Y = y) = P(X = y|Y = x)$?.

## Recap: Sum and Product Rules



**Sum rule**:



**Product rule**:



$$p(X = x_i, Y = y_j) = p(Y = y$$

# Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x|Y = y)$, $p(X = x)$ and $p(Y = y)$?

## Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x|Y = y)$, $p(X = x)$ and $p(Y = y)$? Yes.

# Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x | Y = y)$, $p(X = x)$ and $p(Y = y)$? Yes.



Now supp $x, y$.
Could we c

# Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x | Y = y)$, $p(X = x)$ and $p(Y = y)$? Yes.



Now supp $x, y$.
Could we c

# Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x | Y = y)$, $p(X = x)$ and $p(Y = y)$? Yes.



Now supp $x, y$.
Could we c

# Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x|Y = y)$, $p(X = x)$ and $p(Y = y)$? Yes.



Now supp $x, y$.

Could we c



The differ

# Relating the Marginal, Conditional and Joint

Suppose we knew $p(X = x, Y = y)$ for all values of $x, y$. Could we compute all of $p(X = x|Y = y)$, $p(X = x)$ and $p(Y = y)$? Yes.



Now supp $x, y$.
Could we c



The differ



| | $B = 0$ | $B = 1$ |
|---|---|---|
| $A = 0$ | 680 | 10 |
| $A = 1$ | 220 | 90 |

| | |
|---|---|
| $A$ | |
| $A$ | |

These have the same marginals, but different joint distributions

# Joint as the "Master" Distribution



In genera
of margin



The joint di
dependence

1 More on Joint, Marginal and Conditional Distributions



2 Stati



3 Baye



4 Wrapping up

## Statistical Independence

Suppose that both boxes (red and blue) contain the same proportion of apples and oranges



If fruit is sel

# Statistical Independence

Suppose that both boxes (red and blue) contain the same proportion of apples and oranges



If fruit is sel



*The probability of selecting an apple (or an orange) i box that is chosen*



We may study the properties of $F$ and $B$ separately: this often simplifies analysis

# Statistical Independence: Definition



**Definition: Independent Variables**

Two varia ⊥⊥ $Y$, if

and only if t

marginal



This definition generalises to more than two variabl

# Statistical Independence: Definition



**Definition: Independent Variables**

Two varia ⊥ $Y$, if
and only if t
marginal



This definition generalises to more than two variabl



Are the variables in the language example statistic

# A Note on Notation

When we write

$$p(X, Y) = p(X)p(Y)$$



we have n

This state



$$p(X = x, Y = y) = p(X$$

for every possible $x$ and $y$



This notation is sometimes called implied universality

## Conditional independence

We may also consider random variables that are conditionally independent given some other variable



**Definition: Conditionally Independent Variables**

Two varia enoted
$X \perp\!\!\!\perp Y|$



$$p(X, Y|Z) = p(X|Z)p(Y|Z)$$

Intuitively, $Z$ is a common cause for $X$ and



Example: $X$ = whether I have a cold
$Y$ = whether I have a headache
$Z$ = whether I have the flu

1 More on Joint, Marginal and Conditional Distributions



2 Stati



3 Baye



4 Wrapping up

# Revisiting the Product Rule

The product rule tells us:



$$p(X, Y) = p(Y|X)p(X)$$

This can e



probability:

$$p(Y|X) = \frac{p(X}{p(}$$



Can we use these to relate $p(X|Y)$ and $p(Y|X)$?

## Posterior Inference:
### Example 1 (Mackay, 2003)

- Dicksy Sick had a test for a rare disease
  - Only 1% people of Dicksy's background have the disease

# Posterior Inference:

## Example 1 (Mackay, 2003)

- Dicksy Sick had a test for a rare disease
  - Only 1% people of Dicksy's background have the disease
- The test simply classifies a person as having the disease, or not

# Posterior Inference:

## Example 1 (Mackay, 2003)

- Dicksy Sick had a test for a rare disease
  - Only 1% people of Dicksy's background have the disease
- The test simply classifies a person as having the disease, or not
- The t
  - $p(\text{identifies sick} \mid \text{sick}) = 95\%$.
  - It correctly identifies a healthy individual 96
    $p(\text{identifies healthy} \mid \text{healthy}) = 96$

# Posterior Inference:

## Example 1 (Mackay, 2003)

- Dicksy Sick had a test for a rare disease
  - Only 1% people of Dicksy's background have the disease



- The test simply classifies a person as having the disease, or not
- The t



  - $p(\text{identifies sick} \mid \text{sick}) = 95\%$.
  - It correctly identifies a healthy individual 96



    $p(\text{identifies healthy} \mid \text{healthy}) = 96$
- Dicksy has tested positive (apparently sick)

# Posterior Inference:

## Example 1 (Mackay, 2003)

- Dicksy Sick had a test for a rare disease
  - Only 1% people of Dicksy's background have the disease
- The test simply classifies a person as having the disease, or not
- The t
  - $p(\text{identifies sick} \mid \text{sick}) = 95\%$.
  - It correctly identifies a healthy individual 96
    $p(\text{identifies healthy} \mid \text{healthy}) = 96$
- Dicksy has tested positive (apparently sick)
- What is the probability of Dicksy having the disease?

# Posterior Inference:

## Example 1: Formalization



Let $D \in \{0, 1\}$ denote whether Dicksy has the disease, and $T \in \{0, 1\}$ the outcome

# Posterior Inference:

## Example 1: Formalization



Let $D \in \{0, 1\}$ denote whether Dicksy has the disease, and $T \in \{0, 1\}$ the outcome



.99

.04

$p(T = 0|D = 1) = 0.05$

# Posterior Inference:

## Example 1: Formalization



Let $D \in \{0, 1\}$ denote whether Dicksy has the disease, and $T \in \{0, 1\}$ the outcome



.99

.04

$$p(T = 0|D = 1) = 0.05$$



We need to compute $p(D = 1|T = 1)$, *th*

*the disease given that the test has resulted positive.*

## Posterior Inference:

### Example 1: Solution

$$p(D = 1 | T = 1) = \frac{p(D = 1, T = 1)}{p(T = 1)}$$ Def. conditional prob.

# Posterior Inference:

## Example 1: Solution

$$p(D = 1 | T = 1) = \frac{p(D = 1, T = 1)}{p(T = 1)} \quad \text{Def. conditional prob.}$$

$$= \frac{p(T = 1, D = 1)}{\phantom{x}}$$

# Posterior Inference:

## Example 1: Solution

$$p(D=1|T=1) = \frac{p(D=1,T=1)}{p(T=1)}$$ Def. conditional prob.

$$\frac{p(T=1,D=1)}{\quad}$$

## Posterior Inference:

### Example 1: Solution

$$p(D = 1|T = 1) = \frac{p(D = 1, T = 1)}{p(T = 1)} \quad \text{Def. conditional prob.}$$

$$\frac{p(T = 1, D = 1)}{}$$

$$= \frac{p(T = 1|D = 1)p(D}{\sum_d p(T = 1|D = d)p}$$

# Posterior Inference:

## Example 1: Solution



$$p(D = 1 | T = 1) = \frac{p(D = 1, T = 1)}{p(T = 1)} \quad \text{Def. conditional prob.}$$

$$= \frac{p(T = 1, D = 1)}{\quad}$$



$$= \frac{p(T = 1 | D = 1) p(D}{\sum_d p(T = 1 | D = d) p}$$



$$= \frac{p(T =}{p(T = 1 | D = 1) p(D = 1) + p(T = 1 | D = 0) p(D = 0)}$$

# Posterior Inference:

## Example 1: Solution



$$
\begin{aligned}
p(D=1|T=1) &= \frac{p(D=1,T=1)}{p(T=1)} \quad \text{Def. conditional prob.} \\
&= \frac{p(T=1,D=1)}{\quad} \\
&= \frac{\quad}{\quad} \\
&= \frac{p(T=1|D=1)p(D}{\sum_d p(T=1|D=d)p} \\
&= \frac{p(T=}{p(T=1|D=1)p(D=1)+p(T=1|D=0)p(D=0)} \\
&\approx 0.19.
\end{aligned}
$$

# Posterior Inference:

## Example 1: Solution



$$p(D = 1|T = 1) = \frac{p(D = 1, T = 1)}{p(T = 1)} \quad \text{Def. conditional prob.}$$

$$\frac{p(T = 1, D = 1)}{\ }$$



$$= \frac{p(T = 1|D = 1)p(D}{\sum_d p(T = 1|D = d)p}$$



$$= \frac{p(T =}{p(T = 1|D = 1)p(D = 1) + p(T = 1|D = 0)p(D = 0)}$$

$$\approx 0.19.$$

Despite testing positive and the high accuracy of the test, the probability of Dicksy having the disease is only 0.19!

# Why is the Probability So Low?

## A “Natural Frequency” Approach

In 100 people, only 1 is expected to have the disease ($p(D = 1) = 0.01$)



This sick person will most likely test positive ($p(T = 1\ D = 1) = 0.95$)

But aroun

($p(T =$



So when the test is positive, the chance of being sick is

# Why is the Probability So Low?

## A "Natural Frequency" Approach

In 100 people, only 1 is expected to have the disease ($p(D = 1) = 0.01$)



This sick person will most likely test positive ($p(T = 1\ D = 1) = 0.95$)

But aroun
($p(T =$



So when the test is positive, the chance of being sick is



(Aside: If you can correctly perform the calculation on the previous slide
most medical doctors! See Gird Gigerenzer and Adrian Edwards, Si
from innumeracy to insight, *British Medical Journal*, 327(7417), 741–744, 27 September 2003; Gird Gigerenzer, *Reckoning with risk: Learning to live with uncertainty*, Penguin, 2002.

Moral of the story — if you get sick, don't delegate conditional probability computations to your doctor!)

## Bayes' Theorem

We have implicitly used the following (at first glance remarkable) fact:

Bayes' Theorem:



$$
\begin{aligned}
p(Z|X) &= \frac{p(Z,X)}{p(X)} \\
&= \frac{p(X|Z)p(Z)}{p(X)} \\
&= \frac{p(X|Z)p(Z)}{\sum_{Z'} p(X|Z')p(Z')}
\end{aligned}
$$



If we can express what knowledge of $X$ (test) tells us about $Z$ (disease), then we can express what knowledge of $Z$ tells us about $X$

# The Bayesian Inference Framework

### Bayesian Inference

Bayesian inference provides a mathematical framework explaining how to change our (prior) beliefs in the light of new evidence.





$$\underbrace{\qquad}_{\text{posterior}} = \frac{\qquad}{\underbrace{p(X)}_{\text{evid}}}$$



**Prior**: Belief that someone is sick

**Likelihood**: Probability of testing positive given you are sick

**Posterior**: Probability of being sick given you test positive

## Posterior Inference:

### Example 2 (Bishop, 2006)

Recall our fruit-box example:

- The proportion of oranges and apples are given by

# Posterior Inference:

## Example 2 (Bishop, 2006)

Recall our fruit-box example:

- The proportion of oranges and apples are given by



- Someone told us that in a previous experiment they ended up picking up the red box 40% of the time and the blue box 60% of the time.

## Posterior Inference:

### Example 2 (Bishop, 2006)

Recall our fruit-box example:

- The proportion of oranges and apples are given by



- Someone told us that in a previous experiment they ended up picking up the red box 40% of the time and the blue box 60% of the time.
- A piece of fruit has been picked up and it turned out to be an orange.

# Posterior Inference:

## Example 2 (Bishop, 2006)

Recall our fruit-box example:

- The proportion of oranges and apples are given by



- Someone told us that in a previous experiment they ended up picking up the red box 40% of the time and the blue box 60% of the time.
- A piece of fruit has been picked up and it turned out to be an orange.
- What is the probability that it came from the red box?

# Posterior Inference:

## Example 2: Formalization



Let $B \in \{r, b\}$ denote the selected box and $F \in \{a, o\}$ the selected fruit.

# Posterior Inference:

## Example 2: Formalization



Let $B \in \{r, b$ denote the selected box and $F$ $a, o$ the selected fruit.



$10$

$/1$

$p(F = a|B = b) = 3/4$

# Posterior Inference:

## Example 2: Formalization



Let $B \in \{r, b\}$ denote the selected box and $F \quad a, o$ the selected fruit.

$10$



$/1$

$$p(F = a|B = b) = 3/4$$



We need to compute $p(B = r|F = o)$, *th*
*orange came from the red box.*

# Posterior Inference:

## Example 2: Solution



We simply use Bayes' rule:

# Posterior Inference:

## Example 2: Solution



We simply use Bayes' rule:

$p(B = r$

# Posterior Inference:

## Example 2: Solution

We simply use Bayes' rule:



$$p(B = r \ldots = \frac{\ldots}{p(F = o | B = r) p(B = \ldots = b)}$$

# Posterior Inference:

## Example 2: Solution



We simply use Bayes' rule:

$p(B = r$



$p(F = o | B = r) p(B =$ … $= b)$

$= \frac{2}{3}$

We simply use Bayes' rule:

$$p(B = r \ldots$$



$$p(F = o|B = r)p(B = \ldots = b)$$

$$= \frac{2}{3}$$



and therefore $p(B = b|F = o) = 1/3$.

# Posterior Inference:

## Example 2: Interpretation of the Solution



- If we hadn't been told any information about the fruit picked, the blue box is more likely to be selected than the red box

# Posterior Inference:

## Example 2: Interpretation of the Solution



- If we hadn't been told any information about the fruit picked, the blue box is more likely to be selected than the red box
  - 

# Posterior Inference:

## Example 2: Interpretation of the Solution



- If we hadn't been told any information about the fruit picked, the blue box is more likely to be selected than the red box
  - 
- On
incr

# Posterior Inference:

## Example 2: Interpretation of the Solution



- If we hadn't been told any information about the fruit picked, the blue box is more likely to be selected than the red box
  -
- On

incr
  - *Because the red box contains more orange*

# Posterior Inference:

## Example 2: Interpretation of the Solution



- If we hadn't been told any information about the fruit picked, the blue box is more likely to be selected than the red box
  - 
- On 
  incr
  - *Because the red box contains more orange*
- In fact, the proportion of oranges is so much hig
  this is strong evidence that the orange came fro

# Posterior Inference:

## Example 2: Interpretation of the Solution



- If we hadn't been told any information about the fruit picked, the blue box is more likely to be selected than the red box
  -
- On incr



  - *Because the red box contains more orange*
- In fact, the proportion of oranges is so much hig this is strong evidence that the orange came fro



  - *So after picking up the orange the red box is mu been selected than the blue one*

# Summary

- Recap on joint, marginal and conditional distributions



- Inte



- Stat



- Bayes rule: combination of prior, likelihood to

- Reading: Mackay § 2.1, § 2.2 and § 2.3

# Homework Exercise

Suppose we know that random variables $X, Y$ satisfy



$$p(X|Y) = p(Y|X)$$

What can y $Y$?



If $X$ and ?

Repeat the above questions for the statement



$$\frac{p(X|Y)}{p(Y|X)} = \frac{p}{p(Y)}$$

# Next time

- More examples on Bayes' theorem:



  - Eating hamburgers
  - 
  - The Monty Hall problem



  - Document modelling
- Are there notions of probability beyond frequency counting?