COMP2610/COMP6261 - Information Theory

Lecture 4: Bayesian Inference

Mark Reid and **Aditya Menon**

Research School of Computer Science
The Australian National University

THE AUSTRALIAN NATIONAL UNIVERSITY

July 30th, 2014

Last time

- Examples of joint, marginal and conditional distributions
- When can we say that X, Y do not influence each other?
- What, if anything, does $p(X = x|Y = y)$ tell us about $p(Y = y|X = x)$?

Review Exercise

Suppose we have random variables X, Y such that

$$\begin{aligned} p(X = 1) &= 0.6 \\ p(Y = 1|X = 0) &= 0.7 \\ p(Y = 1|X = 1) &= 0.8 \end{aligned}$$

Then,

$p(X = 1|Y = 1) =$

Review Exercise

Suppose we have random variables X, Y such that

$$\begin{aligned} p(X = 1) &= 0.6 \\ p(Y = 1|X = 0) &= 0.7 \\ p(Y = 1|X = 1) &= 0.8 \end{aligned}$$

Then,

$$p(X = 1|Y = 1) = \frac{p(Y = 1|X = 1)p(X = 1)}{p(Y = 1)} \text{ Bayes' rule}$$

Review Exercise

Suppose we have random variables X, Y such that

$$
\begin{aligned}
p(X = 1) &= 0.6 \\
p(Y = 1|X = 0) &= 0.7 \\
p(Y = 1|X = 1) &= 0.8
\end{aligned}
$$

Then,

$$
\begin{aligned}
p(X = 1|Y = 1) &= \frac{p(Y = 1|X = 1)p(X = 1)}{p(Y = 1)} \text{ Bayes' rule} \\
&= \frac{p(Y = 1|X = 1)p(X = 1)}{p(Y = 1|X = 1)p(X = 1) + p(Y = 1|X = 0)p(X = 0)}
\end{aligned}
$$

Review Exercise

Suppose we have random variables X, Y such that

$$
\begin{aligned}
p(X = 1) &= 0.6 \\
p(Y = 1 | X = 0) &= 0.7 \\
p(Y = 1 | X = 1) &= 0.8
\end{aligned}
$$

Then,

$$
\begin{aligned}
p(X = 1 | Y = 1) &= \frac{p(Y = 1 | X = 1) p(X = 1)}{p(Y = 1)} \text{ Bayes' rule} \\
&= \frac{p(Y = 1 | X = 1) p(X = 1)}{p(Y = 1 | X = 1) p(X = 1) + p(Y = 1 | X = 0) p(X = 0)} \\
&= \frac{(0.8)(0.6)}{(0.8)(0.6) + (0.7)(0.4)} \\
&\approx 0.63
\end{aligned}
$$

This time

- More examples on Bayes' theorem:
  - Eating hamburgers
  - Detecting terrorists
  - The Monty Hall problem
  - Document modelling
- Are there notions of probability beyond frequency counting?

Outline

- 90% of people with McD syndrome are frequent hamburger eaters

Bayesian Inference:
Example 1 (Barber, BRML, 2011)

- 90% of people with McD syndrome are frequent hamburger eaters
- Probability of someone having McD syndrome: 1/10000

Bayesian Inference:

Example 1 (Barber, BRML, 2011)

- 90% of people with McD syndrome are frequent hamburger eaters
- Probability of someone having McD syndrome: 1/10000
- Proportion of hamburger eaters is about 50%

Bayesian Inference:

Example 1 (Barber, BRML, 2011)

- 90% of people with McD syndrome are frequent hamburger eaters
- Probability of someone having McD syndrome: 1/10000
- Proportion of hamburger eaters is about 50%

What is the probability that a hamburger eater will have McD syndrome?

Bayesian Inference:

Example 1: Formalization

Let $McD \in \{0, 1\}$ be the variable denoting having the McD syndrome and $H \in \{0, 1\}$ be the variable denoting a hamburger eater. Therefore:

Let $McD \in \{0, 1\}$ be the variable denoting having the McD syndrome and $H \in \{0, 1\}$ be the variable denoting a hamburger eater. Therefore:

$$p(H = 1 | McD = 1) = 9/10$$

Let $McD \in \{0,1\}$ be the variable denoting having the McD syndrome and $H \in \{0,1\}$ be the variable denoting a hamburger eater. Therefore:

$$p(H = 1|McD = 1) = 9/10 \qquad p(McD = 1) = 10^{-4}$$

Let $McD \in \{0, 1\}$ be the variable denoting having the McD syndrome and $H \in \{0, 1\}$ be the variable denoting a hamburger eater. Therefore:

$$p(H = 1|McD = 1) = 9/10 \qquad p(McD = 1) = 10^{-4}$$
$$p(H = 1) = 1/2$$

Let $McD \in \{0, 1\}$ be the variable denoting having the McD syndrome and $H \in \{0, 1\}$ be the variable denoting a hamburger eater. Therefore:

$$p(H = 1|McD = 1) = 9/10 \qquad p(McD = 1) = 10^{-4}$$
$$p(H = 1) = 1/2$$

We need to compute $p(McD = 1|H = 1)$, *the probability of a hamburger eater having McD syndrome*.

Let $McD \in \{0, 1\}$ be the variable denoting having the McD syndrome and $H \in \{0, 1\}$ be the variable denoting a hamburger eater. Therefore:

$$p(H = 1|McD = 1) = 9/10 \qquad p(McD = 1) = 10^{-4}$$
$$p(H = 1) = 1/2$$

We need to compute $p(McD = 1|H = 1)$, *the probability of a hamburger eater having McD syndrome*.

Any ballpark estimates of this probability?

Bayesian Inference:

Example 1: Solution

$$
\begin{aligned}
p(McD = 1 | H = 1) &= \frac{p(H = 1 | McD = 1) p(McD = 1)}{p(H = 1)} \\
&= 1.8 \times 10^{-4}
\end{aligned}
$$

$$p(McD = 1|H = 1) = \frac{p(H = 1|McD = 1)p(McD = 1)}{p(H = 1)}$$
$$= 1.8 \times 10^{-4}$$

Repeat the above computation if the proportion of hamburger eaters is rather small: (say in France) 0.001.

Example 2: Detecting Terrorists:

From `understandinguncertainty.org`

- Scanner detects true terrorists with 95% accuracy

Example 2: Detecting Terrorists:

From `understandinguncertainty.org`

- Scanner detects true terrorists with 95% accuracy
- Scanner detects upstanding citizens with 95% accuracy

Example 2: Detecting Terrorists:

From understandinguncertainty.org

- Scanner detects true terrorists with 95% accuracy
- Scanner detects upstanding citizens with 95% accuracy
- There is 1 terrorist on your plane with 100 passengers aboard

Example 2: Detecting Terrorists:

From `understandinguncertainty.org`

- Scanner detects true terrorists with 95% accuracy
- Scanner detects upstanding citizens with 95% accuracy
- There is 1 terrorist on your plane with 100 passengers aboard
- The shifty looking man sitting next to you tests positive

Example 2: Detecting Terrorists:

From `understandinguncertainty.org`

- Scanner detects true terrorists with 95% accuracy
- Scanner detects upstanding citizens with 95% accuracy
- There is 1 terrorist on your plane with 100 passengers aboard
- The shifty looking man sitting next to you tests positive

What are the chances of this main being a terrorist?

Example 2: Detecting Terrorists:

Simple Solution Using "Natural Frequencies" (David Spiegelhalter)

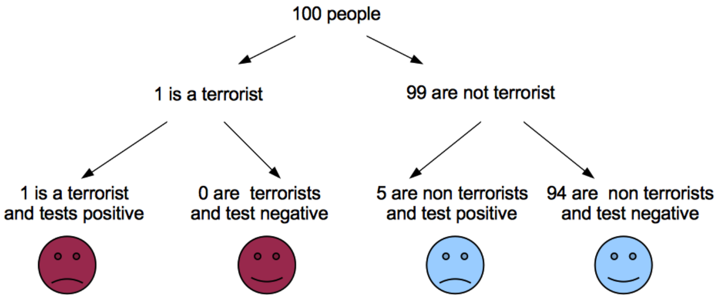


Figure: Figure reproduced from `understandinguncertainty.org`

Example 2: Detecting Terrorists:

Simple Solution Using “Natural Frequencies” (David Spiegelhalter)

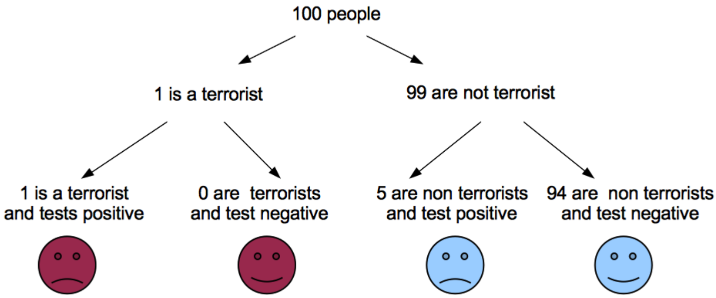


Figure: Figure reproduced from `understandinguncertainty.org`

The chances of the man being a terrorist are $\approx \frac{1}{6}$

Example 2: Detecting Terrorists:

Simple Solution Using "Natural Frequencies" (David Spiegelhalter)

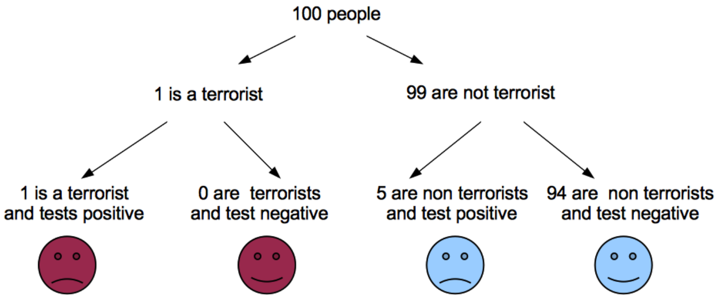


Figure: Figure reproduced from `understandinguncertainty.org`

The chances of the man being a terrorist are $\approx \frac{1}{6}$

- Relation to disease example
- Consequences when catching criminals

Example 2: Detecting Terrorists:

Formalization with Actual Probabilities

Let $T \in \{0, 1\}$ denote the variable regarding whether the person is a terrorist and $S \in \{0, 1\}$ denote the outcome of the scanner.

Example 2: Detecting Terrorists:

Formalization with Actual Probabilities

Let $T \in \{0, 1\}$ denote the variable regarding whether the person is a terrorist and $S \in \{0, 1\}$ denote the outcome of the scanner.

$$p(S = 1|T = 1) = 0.95 \qquad p(S = 0|T = 1) = 0.05$$

Example 2: Detecting Terrorists:

Formalization with Actual Probabilities

Let $T \in \{0,1\}$ denote the variable regarding whether the person is a terrorist and $S \in \{0,1\}$ denote the outcome of the scanner.

$$p(S=1|T=1) = 0.95 \qquad p(S=0|T=1) = 0.05$$
$$p(S=0|T=0) = 0.95 \qquad p(S=1|T=0) = 0.05$$

Example 2: Detecting Terrorists:

Formalization with Actual Probabilities

Let $T \in \{0, 1\}$ denote the variable regarding whether the person is a terrorist and $S \in \{0, 1\}$ denote the outcome of the scanner.

$$
\begin{aligned}
p(S = 1 | T = 1) &= 0.95 & \qquad p(S = 0 | T = 1) &= 0.05 \\
p(S = 0 | T = 0) &= 0.95 & \qquad p(S = 1 | T = 0) &= 0.05 \\
p(T = 1) &= 0.01 & \qquad p(T = 0) &= 0.99
\end{aligned}
$$

Example 2: Detecting Terrorists:
Formalization with Actual Probabilities

Let $T \in \{0, 1\}$ denote the variable regarding whether the person is a terrorist and $S \in \{0, 1\}$ denote the outcome of the scanner.

$$p(S = 1 | T = 1) = 0.95 \qquad p(S = 0 | T = 1) = 0.05$$
$$p(S = 0 | T = 0) = 0.95 \qquad p(S = 1 | T = 0) = 0.05$$
$$p(T = 1) = 0.01 \qquad p(T = 0) = 0.99$$

We want to compute $p(T = 1 | S = 1)$, *the probability of the man being a terrorist given that he has tested positive.*

Example 2: Detecting Terrorists:

Solution with Bayes' Rule

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

Example 2: Detecting Terrorists:

Solution with Bayes' Rule

$$
\begin{aligned}
p(T = 1|S = 1) &= \frac{p(S = 1|T = 1)p(T = 1)}{p(S = 1|T = 1)p(T = 1) + p(S = 1|T = 0)p(T = 0)} \\
&= \frac{(0.95)(0.01)}{(0.95)(0.01) + (0.05)(0.99)}
\end{aligned}
$$

Example 2: Detecting Terrorists:
Solution with Bayes' Rule

$$
\begin{aligned}
p(T=1|S=1) &= \frac{p(S=1|T=1)p(T=1)}{p(S=1|T=1)p(T=1)+p(S=1|T=0)p(T=0)} \\
&= \frac{(0.95)(0.01)}{(0.95)(0.01)+(0.05)(0.99)} \\
&\approx 0.16
\end{aligned}
$$

Example 2: Detecting Terrorists:

Solution with Bayes' Rule

$$
\begin{aligned}
p(T = 1|S = 1) &= \frac{p(S = 1|T = 1)p(T = 1)}{p(S = 1|T = 1)p(T = 1) + p(S = 1|T = 0)p(T = 0)} \\
&= \frac{(0.95)(0.01)}{(0.95)(0.01) + (0.05)(0.99)} \\
&\approx 0.16
\end{aligned}
$$

The probability of the man being a terrorist is $\approx \frac{1}{6}$

Example 2: Detecting Terrorists:
Posterior Versus Prior Belief

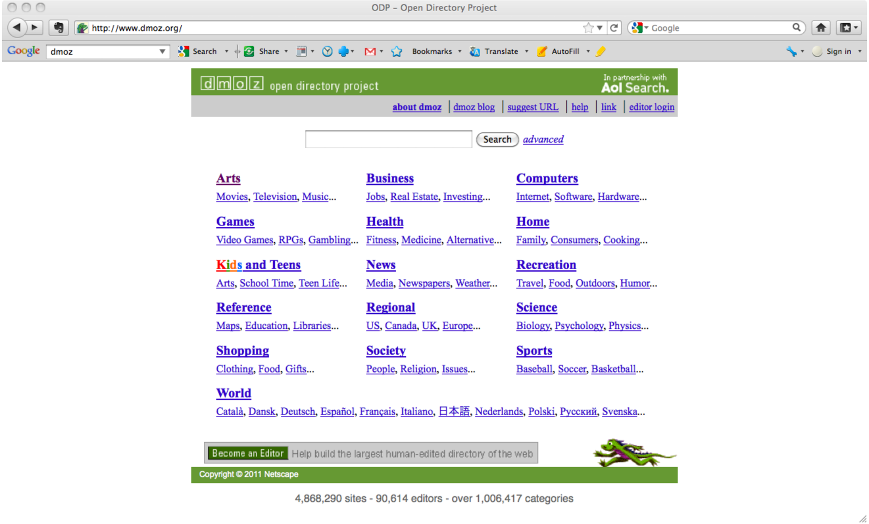
While the man has a low probability of being a terrorist, our belief has increased compared to our prior:

$$\frac{p(T = 1|S = 1)}{p(T = 1)} = \frac{0.16}{0.01} = 16$$

i.e. our belief in him being a terrorist has gone up by a factor of 16

Since terrorists are so rare, a factor of 16 does not result in a very high (absolute) probability or belief

Example 3: Document Classification



Example 3: Document Classification:

Document Modelling

Bag-of-words: Describe a document as a D-dimensional binary vector $\mathbf{x}$, indicating the presence/absence of a word in a vocabulary $\mathcal{V}$.

Example: consider the following tiny vocabulary:

$$\mathcal{V} = \{\text{football, defence, strategy, goal, office}\}$$

Then, a document

$$\mathbf{x} = (1, 0, 0, 1, 1)$$

contains only the words “football”, “goal”, and “office”

- We do not care about the order of the words

Example 3: Document Classification:

Binary Classification

We want to classify documents as being about sports ($\mathcal{C}_1$) or politics ($\mathcal{C}_2$).

A simple *model* for $p(\mathbf{x}|\mathcal{C}_j)$ is:

$$p(\mathbf{x}|\mathcal{C}_j) = \prod_{i=1}^{D} p(x_i|\mathcal{C}_j)$$

This is called Naive Bayes due to its unrealistic assumption of conditional independence of words given the class label

Assume the vocabulary:

$$\mathcal{V} = \{\text{football, defence, strategy, goal, office}\}$$

and the conditional probability tables (CPTs) are given by:

$$
\begin{array}{ll}
p(\mathcal{C}_1) = 0.5 & p(\mathcal{C}_2) = 0.5 \\
p(f = 1|\mathcal{C}_1) = 0.8 & p(f = 1|\mathcal{C}_2) = 0.1 \\
p(d = 1|\mathcal{C}_1) = 0.7 & p(d = 1|\mathcal{C}_2) = 0.7 \\
p(s = 1|\mathcal{C}_1) = 0.2 & p(s = 1|\mathcal{C}_2) = 0.8 \\
p(g = 1|\mathcal{C}_1) = 0.7 & p(g = 1|\mathcal{C}_2) = 0.3 \\
p(o = 1|\mathcal{C}_1) = 0.2 & p(o = 1|\mathcal{C}_2) = 0.7
\end{array}
$$

Example 3: Document Classification:

Conditional Probability Tables

Assume the vocabulary:

$$\mathcal{V} = \{\text{football, defence, strategy, goal, office}\}$$

and the conditional probability tables (CPTs) are given by:

$$\begin{array}{ll}
p(\mathcal{C}_1) = 0.5 & p(\mathcal{C}_2) = 0.5 \\
p(f = 1|\mathcal{C}_1) = 0.8 & p(f = 1|\mathcal{C}_2) = 0.1 \\
p(d = 1|\mathcal{C}_1) = 0.7 & p(d = 1|\mathcal{C}_2) = 0.7 \\
p(s = 1|\mathcal{C}_1) = 0.2 & p(s = 1|\mathcal{C}_2) = 0.8 \\
p(g = 1|\mathcal{C}_1) = 0.7 & p(g = 1|\mathcal{C}_2) = 0.3 \\
p(o = 1|\mathcal{C}_1) = 0.2 & p(o = 1|\mathcal{C}_2) = 0.7
\end{array}$$

A new document arrives and is described by $\mathbf{x} = (0, 1, 1, 1, 0)$.

Example 3: Document Classification:
Conditional Probability Tables

Assume the vocabulary:

$$\mathcal{V} = \{\text{football, defence, strategy, goal, office}\}$$

and the conditional probability tables (CPTs) are given by:

$$\begin{array}{ll}
p(\mathcal{C}_1) = 0.5 & p(\mathcal{C}_2) = 0.5 \\
p(f = 1|\mathcal{C}_1) = 0.8 & p(f = 1|\mathcal{C}_2) = 0.1 \\
p(d = 1|\mathcal{C}_1) = 0.7 & p(d = 1|\mathcal{C}_2) = 0.7 \\
p(s = 1|\mathcal{C}_1) = 0.2 & p(s = 1|\mathcal{C}_2) = 0.8 \\
p(g = 1|\mathcal{C}_1) = 0.7 & p(g = 1|\mathcal{C}_2) = 0.3 \\
p(o = 1|\mathcal{C}_1) = 0.2 & p(o = 1|\mathcal{C}_2) = 0.7
\end{array}$$

A new document arrives and is described by $\mathbf{x} = (0, 1, 1, 1, 0)$.

What is the probability of this document being about sports?

Example 3: Document Classification:
Solution

$$
\begin{aligned}
p(\mathcal{C}_1|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) \cdot p(\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) \cdot p(\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2) \cdot p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2)}
\end{aligned}
$$

Example 3: Document Classification:

Solution

$$
\begin{aligned}
p(\mathcal{C}_1|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) \cdot p(\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) \cdot p(\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2) \cdot p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2)} \\
&= \frac{(0.2)(0.7)(0.2)(0.7)(0.8)}{(0.2)(0.7)(0.2)(0.7)(0.8) + (0.9)(0.7)(0.8)(0.3)(0.3)}
\end{aligned}
$$

Example 3: Document Classification:

Solution

$$
\begin{aligned}
p(\mathcal{C}_1|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) \cdot p(\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) \cdot p(\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2) \cdot p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2)} \\
&= \frac{(0.2)(0.7)(0.2)(0.7)(0.8)}{(0.2)(0.7)(0.2)(0.7)(0.8) + (0.9)(0.7)(0.8)(0.3)(0.3)} \\
&\approx 0.26.
\end{aligned}
$$

Example 3: Document Classification:

Solution

$$\begin{aligned}
p(\mathcal{C}_1|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)p(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1)\cdot p(\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1)\cdot p(\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2)\cdot p(\mathcal{C}_2)} \\
&= \frac{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1)}{\prod_{d=1}^{D} p(x_d|\mathcal{C}_1) + \prod_{d=1}^{D} p(x_d|\mathcal{C}_2)} \\
&= \frac{(0.2)(0.7)(0.2)(0.7)(0.8)}{(0.2)(0.7)(0.2)(0.7)(0.8) + (0.9)(0.7)(0.8)(0.3)(0.3)} \\
&\approx 0.26.
\end{aligned}$$

We would classify this document as politics as $p(\mathcal{C}_2|\mathbf{x}) \approx 0.74$.

Example 4: The Monty Hall Problem

Problem Statement

- Three boxes, one with a prize and the other two are empty

Example 4: The Monty Hall Problem

Problem Statement

- Three boxes, one with a prize and the other two are empty
- Each box has equal probability of having the prize

Example 4: The Monty Hall Problem

Problem Statement

- Three boxes, one with a prize and the other two are empty
- Each box has equal probability of having the prize
- Your goal is to pick up the box with the prize in it

Example 4: The Monty Hall Problem

Problem Statement

- Three boxes, one with a prize and the other two are empty
- Each box has equal probability of having the prize
- Your goal is to pick up the box with the prize in it
- You select one of the boxes

Example 4: The Monty Hall Problem

Problem Statement

- Three boxes, one with a prize and the other two are empty
- Each box has equal probability of having the prize
- Your goal is to pick up the box with the prize in it
- You select one of the boxes
- The host, who knows the location of the prize, opens the empty box out of the other two boxes

Example 4: The Monty Hall Problem

Problem Statement

- Three boxes, one with a prize and the other two are empty
- Each box has equal probability of having the prize
- Your goal is to pick up the box with the prize in it
- You select one of the boxes
- The host, who knows the location of the prize, opens the empty box out of the other two boxes

Should you switch to the other box? Would that increase your chances of winning the prize?

Let $C \in \{r, g, b\}$ denote the box that contains the prize where r, g, b refer to the identity of each box.

WLOG assume the following:

Example 4: The Monty Hall Problem:
Formalization

Let $C \in \{r, g, b\}$ denote the box that contains the prize where r, g, b refer to the identity of each box.

WLOG assume the following:

- You have selected box r

Example 4: The Monty Hall Problem:
Formalization

Let $C \in \{r, g, b\}$ denote the box that contains the prize where r, g, b refer to the identity of each box.

WLOG assume the following:

- You have selected box r
- Denote the event: "the host opens box b" with $H=\mathrm{b}$

Example 4: The Monty Hall Problem:

Formalization

Let $C \in \{r, g, b\}$ denote the box that contains the prize where r, g, b refer to the identity of each box.

WLOG assume the following:

- You have selected box r
- Denote the event: "the host opens box b" with H=b

$$P(C = r) = \frac{1}{3} \qquad p(C = g) = \frac{1}{3} \qquad p(C = b) = \frac{1}{3}$$

Example 4: The Monty Hall Problem:
Formalization

Let $C \in \{r, g, b\}$ denote the box that contains the prize where r, g, b refer to the identity of each box.

WLOG assume the following:

- You have selected box r
- Denote the event: "the host opens box b" with H=b

$$P(C = r) = \frac{1}{3} \qquad p(C = g) = \frac{1}{3} \qquad p(C = b) = \frac{1}{3}$$

$$p(H = b|C = r) = \frac{1}{2} \quad p(H = b|C = g) = 1 \quad p(H = b|C = b) = 0$$

Example 4: The Monty Hall Problem:

Formalization

Let $C \in \{r, g, b\}$ denote the box that contains the prize where r, g, b refer to the identity of each box.

WLOG assume the following:

- You have selected box r
- Denote the event: "the host opens box b" with H=b

$$P(C = r) = \frac{1}{3} \qquad p(C = g) = \frac{1}{3} \qquad p(C = b) = \frac{1}{3}$$

$$p(H = b|C = r) = \frac{1}{2} \quad p(H = b|C = g) = 1 \quad p(H = b|C = b) = 0$$

We want to compute $p(C = r|H = b)$ and $p(C = g|H = b)$ to decide if we should switch from our initial choice.

Example 4: The Monty Hall Problem:
Solution

We have that:

$$p(H = b) = \sum_{c \in \{r, g, b\}} p(H = b | C = c) p(C = c)$$

Example 4: The Monty Hall Problem:

Solution

We have that:

$$
\begin{aligned}
p(H = b) &= \sum_{c \in \{r, g, b\}} p(H = b | C = c) p(C = c) \\
&= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3)
\end{aligned}
$$

Example 4: The Monty Hall Problem:

Solution

We have that:

$$\begin{aligned} p(H = b) &= \sum_{c \in \{r,g,b\}} p(H = b | C = c) p(C = c) \\ &= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3) \\ &= 1/2 \end{aligned}$$

Example 4: The Monty Hall Problem:
Solution

We have that:

$$
\begin{aligned}
p(H = b) &= \sum_{c \in \{r, g, b\}} p(H = b | C = c) p(C = c) \\
&= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3) \\
&= 1/2
\end{aligned}
$$

Therefore:

Example 4: The Monty Hall Problem:
Solution

We have that:

$$\begin{aligned} p(H = b) &= \sum_{c \in \{r,g,b\}} p(H = b | C = c) p(C = c) \\ &= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3) \\ &= 1/2 \end{aligned}$$

Therefore:

$$p(C = r | H = b) = \frac{p(H = b | C = r) p(C = r)}{p(H = b)}$$

Example 4: The Monty Hall Problem:
Solution

We have that:

$$p(H = b) = \sum_{c \in \{r, g, b\}} p(H = b | C = c) p(C = c)$$
$$= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3)$$
$$= 1/2$$

Therefore:

$$p(C = r | H = b) = \frac{p(H = b | C = r) p(C = r)}{p(H = b)} = \frac{(1/2)(1/3)}{(1/2)} = 1/3$$

We have that:

$$
\begin{aligned}
p(H = b) &= \sum_{c \in \{r, g, b\}} p(H = b | C = c) p(C = c) \\
&= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3) \\
&= 1/2
\end{aligned}
$$

Therefore:

$$
p(C = r | H = b) = \frac{p(H = b | C = r) p(C = r)}{p(H = b)} = \frac{(1/2)(1/3)}{(1/2)} = 1/3
$$

Similarly, $p(C = g | H = b) = 2/3$.

Example 4: The Monty Hall Problem:
Solution

We have that:

$$\begin{aligned} p(H = b) &= \sum_{c \in \{r,g,b\}} p(H = b | C = c) p(C = c) \\ &= (1/2)\,(1/3) + (1)\,(1/3) + (0)\,(1/3) \\ &= 1/2 \end{aligned}$$

Therefore:

$$p(C = r | H = b) = \frac{p(H = b | C = r) p(C = r)}{p(H = b)} = \frac{(1/2)(1/3)}{(1/2)} = 1/3$$

Similarly, $p(C = g | H = b) = 2/3$.

You should switch from your initial choice to the other box in order to increase your chances of winning the prize!

Example 4: The Monty Hall Problem:

Illustration of the Solution

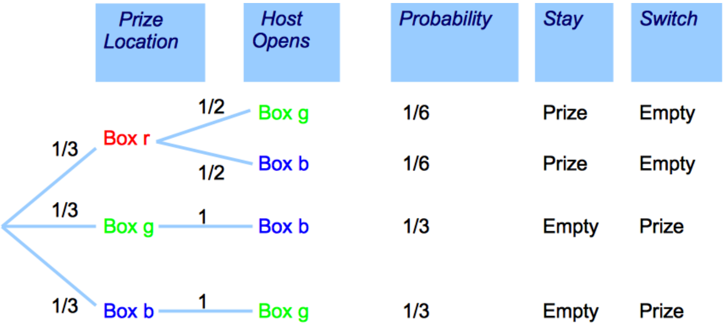


Figure: Illustration of the solution when you have initially selected box r.

Example 4: The Monty Hall Problem:

Another Perspective

Switching is bad if, and only if, we initially picked the prize box (because if not, the other remaining box must contain the prize)

We picked the prize box with probability $1/3$. This is independent of the host's action

Hence, with probability $2/3$, switching will reveal the prize box

Example 4: The Monty Hall Problem:

Variants to Ponder

Would switching be rational if:

- The host only revealed a box when he knew we picked the right one?
- The host only revealed a box when he knew we picked the wrong one?
- The host is himself unaware of the prize box, and reveals a box at random, which by chance does not have the prize?

The meaning of Probability

Frequentist : Frequencies of random repeatable experiments

The meaning of Probability

Frequentist : Frequencies of random repeatable experiments

- E.g. Prob. of biased coin landing "Heads"

The meaning of Probability

Frequentist : Frequencies of random repeatable experiments

- E.g. Prob. of biased coin landing "Heads"

Bayesian : Degrees of Belief

The meaning of Probability

Frequentist : Frequencies of random repeatable experiments

- E.g. Prob. of biased coin landing "Heads"

Bayesian : Degrees of Belief

- E.g. Prob. of Tasmanian Devil disappearing by the end of this decade

The meaning of Probability

Frequentist : Frequencies of random repeatable experiments

- E.g. Prob. of biased coin landing "Heads"

Bayesian : Degrees of Belief

- E.g. Prob. of Tasmanian Devil disappearing by the end of this decade

Cox Axioms

Given $B(x)$, $B(\bar{x})$, $B(x, y)$, $B(x|y)$, $B(y)$:

1. Degrees of belief can be ordered
2. $B(x) = f[B(\bar{x})]$
3. $B(x, y) = g[B(x|y), B(y)]$

The meaning of Probability

Frequentist : Frequencies of random repeatable experiments

- E.g. Prob. of biased coin landing "Heads"

Bayesian : Degrees of Belief

- E.g. Prob. of Tasmanian Devil disappearing by the end of this decade

Cox Axioms

Given $B(x)$, $B(\bar{x})$, $B(x, y)$, $B(x|y)$, $B(y)$:

1. Degrees of belief can be ordered
2. $B(x) = f[B(\bar{x})]$
3. $B(x, y) = g[B(x|y), B(y)]$

If a set of Beliefs satisfy these axioms they can be mapped onto probabilities satisfying the rules of probability.

Frequentists versus Bayesians: Round I

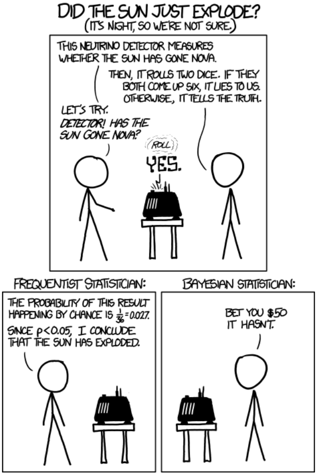


Image from `http://xkcd.com/1132/`

Frequentists versus Bayesians: Round II

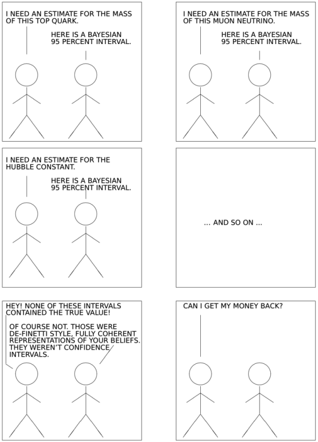


Image from
`http://normaldeviate.wordpress.com/2012/11/09/anti-xkcd/`

Summary

- Examples of application of Bayes' rule
  - Formalization
  - Solution by applying Bayes' theorem
- Intuition is usually helpful although it may sometimes deceive us
- Interesting application to document classification
- Frequentist v Bayesian probabilities
- Cox axioms

Next time

- Working through some useful probability distributions
- More on Bayesian inference