COMP2610 / COMP6261 - Information Theory

Lecture 4: Bayesian Inference

Robert C. Williamson

Research School of Computer Science

Australian National University

July 31, 2018

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 binary 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 binary 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,

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

Review Exercise

Suppose we have binary 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,

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

Review Exercise

Suppose we have binary 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,

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

This time

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

Outline

1. Bayes' Rule: Examples
   - Eating Hamburgers
   - Detecting Terrorists
   - The Monty Hall Problem

2. Moments for functions of two discrete Random Variables

3. The meaning of Probability

4. Wrapping Up

Bayesian Inference:

Example 1 (Barber, BRML, 2011)

- 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:

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:

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

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:

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

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:

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

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:

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

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

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:

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

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

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

Bayesian Inference:

Example 1: Solution

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

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 (terrorist)

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 (terrorist)

What are the chances of this man being a terrorist?

Example 2: Detecting Terrorists:

Simple Solution Using “Natural Frequencies” (David Spiegelhalter)

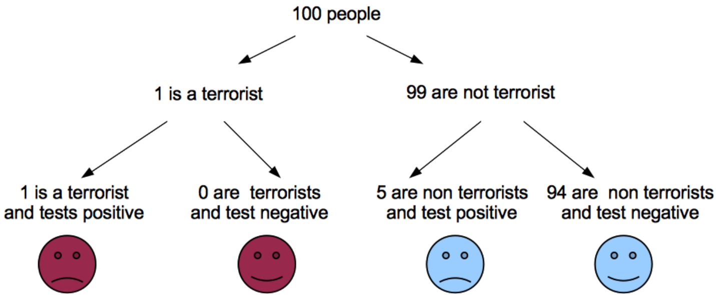

Figure reproduced from `understandinguncertainty.org`

Example 2: Detecting Terrorists:

Simple Solution Using “Natural Frequencies” (David Spiegelhalter)

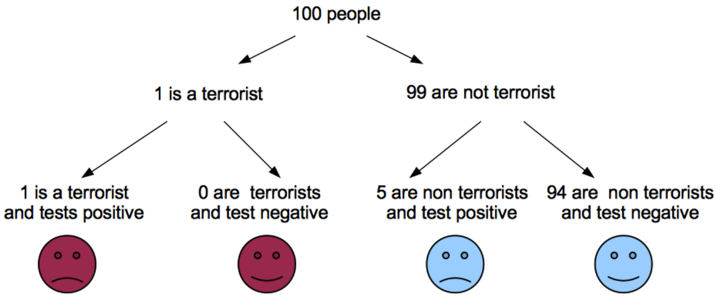

Figure reproduced from `understandinguncertainty.org`

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

Example 2: Detecting Terrorists:

Simple Solution Using “Natural Frequencies” (David Spiegelhalter)

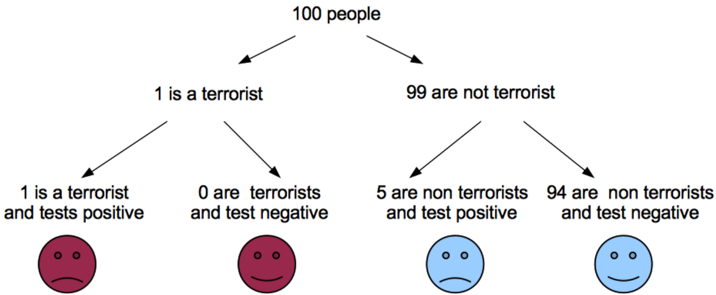

Figure reproduced from `understandinguncertainty.org`

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

(Aside: They are indeed very rare. For an intruiging (and surprising) example of the implications of inability to take account of actual base rates (in the example above we made the numbers up), and the effect on people's subsequent decisions, see Gerd Gigerenzer, Dread Risk, September 11, and Fatal Traffic Accidents, *Psychological Science* 15(4), 286–287, (2004); Gerd Gigerenzer, Out of the Frying Pan into the Fire: Behavioural Reactions to Terrorist Attacks, *Risk Analysis* 26(2), 347–351 (2006). His calculation (which of course is based on some assumptions) is that in the year following 9/11, 6 times the number of people who were killed as passengers *additionally* died on roads (that is the increase in road deaths due to people chosing to drive instead of flying)! He calls the reaction to very low probability events with a bad outcome "dread risk".)

Example 3: The Monty Hall Problem

Problem Statement

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

Example 3: 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 3: 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 3: 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 3: 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 3: 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?

Example 3: 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:

Example 3: 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 3: 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

Example 3: 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 3: 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 3: 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 3: 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 3: 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 3: 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 3: 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 3: 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 3: 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:

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

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

Example 3: 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 3: The Monty Hall Problem:

Illustration of the Solution

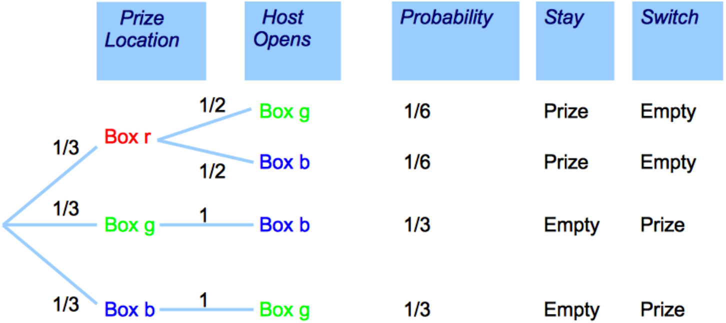

Illustration of the solution when you have initially selected box r.

Example 3: 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 3: 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 Expected Value of a Function of Two Discrete Random Variables

(Assuming you have met Expectation $E[X]$ and Variance $\mathrm{Var}(X)$ before...)
The expected value of a function $g(X, Y)$ of two discrete random variables is defined as

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) p(X = x, Y = y). \tag{1}$$

In particular, the expected value of X is given by

$$E[X] = \sum_x \sum_y x p(X = x, Y = y). \tag{2}$$

It should be noted that if we have already calculated the marginal distribution of X, then it is simpler to calculate $E[X]$ using this.

Covariance and the Correlation Coefficient

The covariance between X and Y, $\mathrm{Cov}(X, Y)$ is given by

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) \tag{3}$$

Note that by definition $\mathrm{Cov}(X, X) = E(XX) - E(X)E(X) = \mathrm{Var}(X)$.
The coefficient of correlation between X and Y is given by

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{4}$$

Always in $[-1, 1]$.

Example

Discrete random variables X and Y have the following joint distribution:

	$Y=-1$	$Y=0$	$Y=1$
$X=0$	0	$\frac{1}{3}$	0
$X=1$	$\frac{1}{3}$	0	$\frac{1}{3}$

Calculate

1. $p(X>Y)$
2. marginal distributions of X and Y
3. expected values and variances of X and Y
4. coefficient of correlation between X and Y

Are X and Y independent?

Example

To calculated the probability of such an event, note that we sum over all the cells which correspond to that event. Hence,

$$p(X > Y) = p(X = 0, Y = -1) + p(X = 1, Y = -1) \\ + p(X = 1, Y = 0) = \frac{1}{3}$$

Example

Recall that

$$p(X = x) = \sum_y p(X = x, Y = y).$$

Hence,

$$p(X = 0) = \sum_{y=-1}^{1} p(X = 0, Y = y) = 0 + \frac{1}{3} + 0 = \frac{1}{3}$$

$$p(X = 1) = \sum_{y=-1}^{1} p(X = 1, Y = y) = \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}$$

Note that after obtaining $p(X = 0)$, we could calculate $p(X = 1)$ by using the fact that

$$p(X = 1) = 1 - p(X = 0), \tag{5}$$

since X only takes the values 0 and 1.

Example

Similarly,

$$p(Y=-1)=\sum_{x=0}^{1} p(X=x, Y=-1)=0+\frac{1}{3}=\frac{1}{3}$$

$$p(Y=0)=\sum_{x=0}^{1} p(X=x, Y=0)=\frac{1}{3}+0=\frac{1}{3}$$

$$p(Y=1)=1-p(Y=-1)-p(Y=0)=\frac{1}{3}$$

Example

We then calculate the expected values and variances of X and Y from these marginal distributions.

$$E(X) = \sum_{x=0}^{1} x\, p(X = x) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$

$$E(Y) = \sum_{y=-1}^{1} y\, p(Y = y) = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0.$$

Example

To calculate the variances of X and Y, $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$, we use the formula

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2.$$

$$E(X^2) = \sum_{x=0}^{1} x^2 p(X = x) = 0^2 \times \frac{1}{3} + 1^2 \times \frac{2}{3} = \frac{2}{3}$$

$$E(Y^2) = \sum_{y=-1}^{1} y^2 p(Y = y) = (-1)^2 \times \frac{1}{3} + 0^2 \times \frac{1}{3} + 1^2 \times \frac{1}{3} = \frac{2}{3}.$$

Thus we get

$$\mathrm{Var}(X) = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}$$

$$\mathrm{Var}(Y) = \frac{2}{3} - (0)^2 = \frac{2}{3}$$

Example

To calculate the correlation coefficient, we first calculate the covariance between X and Y. We have

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

where

$$\begin{aligned} E(XY) &= \sum_{x=0}^{1} \sum_{y=-1}^{1} xy\, p(X = x, Y = y) \\ &= 0(-1)0 + 0(0)\frac{1}{3} + 0(1)0 + 1(-1)\frac{1}{3} + 1(0)0 + 1(1)\frac{1}{3} = 0 \end{aligned}$$

Thus we get

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0 - \frac{2}{3} \times 0 = 0.$$

From the definition of the correlation coefficient,

$$\rho(X, Y) = 0.$$

Example - is X and Y independent

We have that

$$p(X = 0, Y = -1) = 0 \neq p(X = 0)p(Y = -1) = \left(\frac{1}{3}\right)^2$$

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

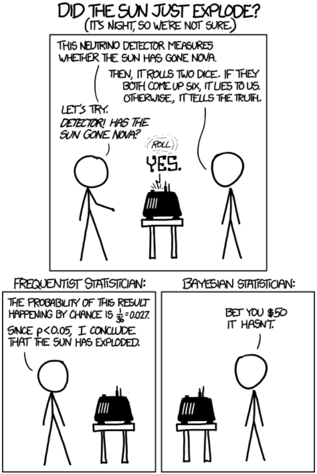

Image from `http://xkcd.com/1132/`

Frequentists versus Bayesians: Round II

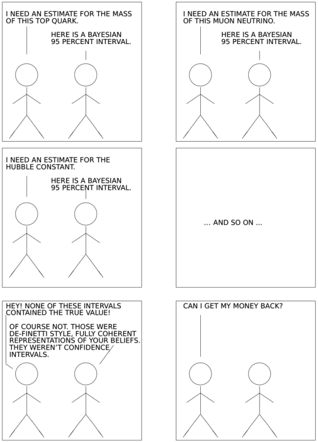

Image from `http://normaldeviate.wordpress.com/2012/11/09/anti-xkcd/`

In practice one needs to make use of both interpretations. Wise to be open to both. This is a huge topic which we can not get into further here. Note that Mackay was firmly in the Bayesian camp. . .

Summary

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

Next time

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