

# Basic Probability

Robert Platt  
Northeastern University

Some images and slides are used from:

1. CS188 UC Berkeley
2. RN, AIMA

# Definition

- Probability theory is nothing but common sense reduced to calculation. ~Pierre Laplace
- What is probability? What does it mean when we say “the probability that a coin will land head is 0.5”

# Random variables

## What is a random variable?

Suppose that the variable  $a$  denotes the outcome of a role of a single six-sided die:

$$a \in \{1, 2, 3, 4, 5, 6\} = A$$

$a$  is a *random variable*

this is the *domain* of  $a$

Another example:

Suppose  $b$  denotes whether it is raining or clear outside:

$$b \in \{rain, clear\} = B$$

# Probability distribution

A probability distribution associates each with a probability of occurrence.

A probability table is one way to encode the distribution:

$$a \in \{1, 2, 3, 4, 5, 6\} = A \quad b \in \{rain, clear\} = B$$

a	P(a)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

b	P(b)
rain	1/4
clear	3/4

All probability distributions must satisfy the following:

1.  $\forall a \in A, a \geq 0$
2.  $\sum_{a \in A} a = 1$

# Writing probabilities

a	P(a)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

b	P(b)
rain	1/4
clear	3/4

For example:  $p(a = 2) = 1/6$   
 $p(b = \text{clear}) = 3/4$

But, sometimes we will abbreviate this as:  $p(2) = 1/6$

$$p(\text{clear}) = 3/4$$

# Joint probability distributions

Given random variables:  $X_1, X_2, \dots, X_n$

The *joint distribution* is a probability assignment to all combinations:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

or: 
$$P(x_1, x_2, \dots, x_n)$$

As with single-variate distributions, joint distributions must satisfy:

1. 
$$P(x_1, x_2, \dots, x_n) \geq 0$$

2. 
$$\sum_{x_1, \dots, x_n} P(x_1, x_2, \dots, x_n) = 1$$

# Joint probability distributions

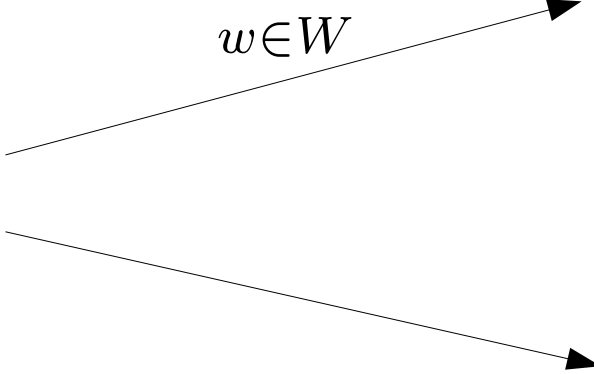
Joint distributions are typically written in table form:

T	W	$P(T,W)$
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

# Marginalization

Given  $P(T,W)$ , calculate  $P(T)$  or  $P(W)$ ...

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(T) = \sum_{w \in W} P(T, w)$$


T	P(T)
hot	0.5
cold	0.5

$$P(W) = \sum_{t \in T} P(t, W)$$

W	P(W)
sun	0.5
rain	0.4



# Conditional Probabilities


$P(\textit{sun}|\textit{hot}) \equiv$  Probability that it is sunny *given* that it is hot.

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

# Conditional Probabilities

Calculate the conditional probability using the product rule:

Product rule


$$P(a|b) = \frac{P(a, b)}{P(b)}$$

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$\begin{aligned} P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4 \\ &= P(W = s, T = c) + P(W = r, T = c) \\ &= 0.2 + 0.3 = 0.5 \end{aligned}$$

# Conditional distribution

Given  $P(T,W)$ , calculate  $P(T|w)$  or  $P(W|t)$ ...

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



W	$P(W \mid t = \text{hot})$
sun	0.8
rain	0.2

$$P(W|t) = \frac{P(W, t)}{P(t)}$$

# Conditional distribution

Given  $P(T,W)$ , calculate  $P(T|w)$  or  $P(W|t)$ ...

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



W	$P(W \mid t = \text{hot})$
sun	0.8
rain	0.2

$$P(W|t) = \frac{P(W, t)}{P(t)}$$


$$\begin{aligned} P(\text{sun}|\text{hot}) &= \frac{P(\text{sun}, \text{hot})}{P(\text{hot})} = \frac{P(\text{sun}, \text{hot})}{P(\text{sun}, \text{hot}) + P(\text{rain}, \text{hot})} \\ &= \frac{0.4}{0.4 + 0.1} \end{aligned}$$

# Conditional distribution


Given  $P(T,W)$ , calculate  $P(T|w)$  or  $P(W|t)$ ...

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W|t) = \frac{P(W, t)}{P(t)}$$



W	$P(W \mid t = \textit{hot})$
sun	0.8
rain	0.2



W	$P(W \mid t = \textit{cold})$
sun	0.4
rain	0.6

# Conditional distribution

Given  $P(T,W)$ , calculate  $P(T|w)$  or  $P(W|t)$ ...

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W|t) = \frac{P(W, t)}{P(t)}$$

W	P(W   $t = hot$ )
sun	0.8
rain	0.2

W	P(W   $t = cold$ )
sun	0.4
rain	0.6

$$\begin{aligned} P(sun|cold) &= \frac{P(sun, cold)}{P(cold)} = \frac{P(sun, cold)}{P(sun, cold) + P(rain, cold)} \\ &= \frac{0.2}{0.2 + 0.3} \end{aligned}$$

# Normalization

Given  $P(T,W)$ , calculate  $P(T|w)$  or  $P(W|t)$ ...

T	W	P(T,W)
---	---	--------

Can we avoid explicitly computing this?

$$P(W|t) = \frac{P(W, t)}{P(t)}$$

W	$P(W   t = \text{hot})$
sun	0.8
rain	0.2

W	$P(W   t = \text{cold})$
sun	0.4
rain	0.6

$$\begin{aligned}
 P(\text{sun}|\text{cold}) &= \frac{P(\text{sun}, \text{cold})}{P(\text{cold})} = \frac{P(\text{sun}, \text{cold})}{P(\text{sun}, \text{cold}) + P(\text{rain}, \text{cold})} \\
 &= \frac{0.2}{0.2 + 0.3}
 \end{aligned}$$

# Normalization

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



W	P(W, t=hot)
sun	0.4
rain	0.1



W	P(W   t = hot)
sun	0.8
rain	0.2

Select corresponding elts  
from the joint distribution

Scale the numbers so  
that they sum to 1.

$$P(\text{sun}|\text{cold}) = \frac{P(\text{sun}, \text{cold})}{P(\text{cold})} = \frac{P(\text{sun}, \text{cold})}{P(\text{sun}, \text{cold}) + P(\text{rain}, \text{cold})}$$



# Normalization

T	W	P(T,W)
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



W	P(W, t=hot)
sun	0.4
rain	0.1



W	P(W   t = hot)
sun	0.8
rain	0.2

Select corresponding elts  
from the joint distribution

Scale the numbers so  
that they sum to 1.

$$P(\text{sun}|\text{cold}) = \frac{P(\text{sun}, \text{cold})}{P(\text{cold})} = \frac{P(\text{sun}, \text{cold})}{P(\text{sun}, \text{cold}) + P(\text{rain}, \text{cold})}$$



The only purpose of this denominator is to make the  
distribution sum to one.

– we achieve the same thing by scaling.

# Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$



# Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

It's easy to derive from the product rule:

$$P(a, b) = P(b|a)P(a) = \underbrace{P(a|b)P(b)}$$



Solve for this

# Using Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$



$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

# Using Bayes Rule

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$



$$P(\textit{cause}|\textit{effect}) = \frac{P(\textit{effect}|\textit{cause})P(\textit{cause})}{P(\textit{effect})}$$

But harder to estimate this

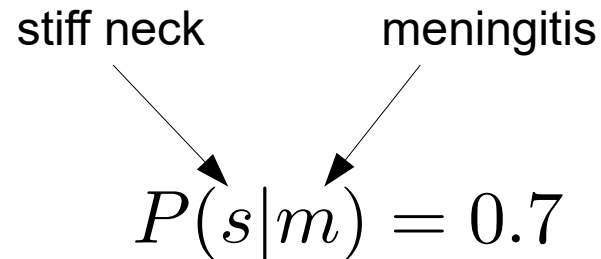
It's often easier to estimate this

# Bayes Rule Example

$$P(\textit{cause}|\textit{effect}) = \frac{P(\textit{effect}|\textit{cause})P(\textit{cause})}{P(\textit{effect})}$$

Suppose you have a stiff neck...

Suppose there is a 70% chance of meningitis if you have a stiff neck:



stiff neck                      meningitis

$$P(s|m) = 0.7$$

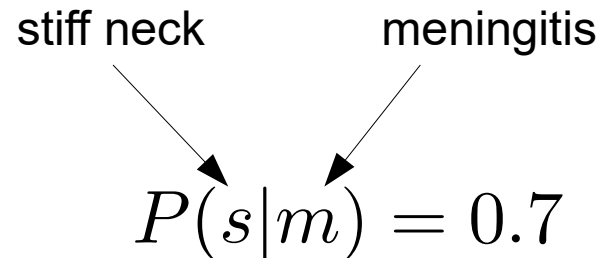
What are the chances that you have meningitis?

# Bayes Rule Example

$$P(\textit{cause}|\textit{effect}) = \frac{P(\textit{effect}|\textit{cause})P(\textit{cause})}{P(\textit{effect})}$$

Suppose you have a stiff neck...

Suppose there is a 70% chance of meningitis if you have a stiff neck:



stiff neck                      meningitis

$$P(s|m) = 0.7$$

What are the chances that you have meningitis?

We need a little more information...

# Bayes Rule Example

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

$$P(s|m) = 0.7$$

$$P(s) = 0.01$$



Prior probability of stiff neck

$$P(m) = \frac{1}{50000}$$



Prior probability of meningitis

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \times \frac{1}{50000}}{0.01} = 0.0014$$



# Bayes Rule Example

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

$$P(s|m) = 0.7$$

$$P(s) = 0.01$$



Prior probability of stiff neck

$$P(m) = \frac{1}{50000}$$



Prior probability of meningitis

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \times \frac{1}{50000}}{0.01} = 0.0014$$