While you are waiting...

- **socrative.com,** room number **SIMLANG2016**

Simulating Language
Lecture 9: Bayesian Learning

Simon Kirby

simon@ling.ed.ac.uk

Which reading are you doing / planning to do?

A. Stone, chapter 1 only

B. Stone, chapters 1-4

C. Kruschke, chapters

D. Some other really great introduction to Bayes

E. None of the above

1.1. Example 1: Poxy Diseases

The Patient's Perspective

Suppose that you wake up one day with spots all over your face, as in Figure 1.2. The doctor tells you that 90% of people who have smallpox have the same symptoms as you have. In other words, the probability of having these symptoms given that you have smallpox is 0.9 (ie 90%). As smallpox is often fatal, you are naturally terrified.

However, after a few moments of contemplation, you decide that you do not want to know the probability that you have these symptoms (after all, you already know them). Instead, what you really want to know is the probability that you have smallpox.

So you say to your doctor, "Yes, but what is the probability that I have smallpox given that I have these symptoms?". "Ah", says your doctor, "a very good question." After scribbling some equations, your doctor looks up. "The probability that you have smallpox given that you have these symptoms is 1.1%, or equivalently, 0.011." Of course,
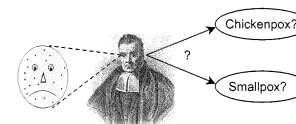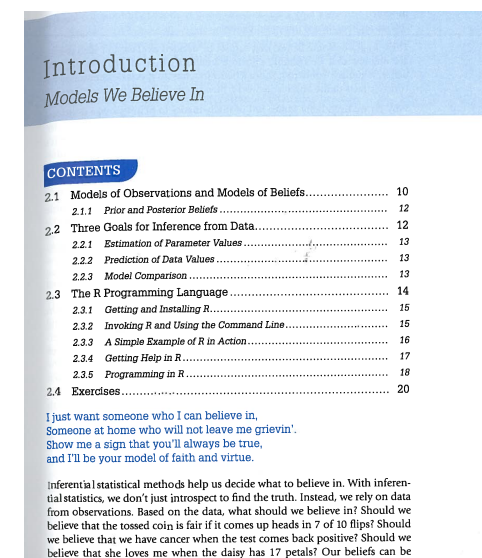
Chickenpox?

?

Smallpox?

Figure 1.2.: Thomas Bayes diagnosing a patient.

Introduction

Models We Believe In

CONTENTS

I just want someone who I can believe in,
Someone at home who will not leave me grievin'.
Show me a sign that you'll always be true,
and I'll be your model of faith and virtue.

Inferential statistical methods help us decide what to believe in. With inferential statistics, we don't just introspect to find the truth. Instead, we rely on data from observations. Based on the data, what should we believe in? Should we believe that the tossed coin is fair if it comes up heads in 7 of 10 flips? Should we believe that we have cancer when the test comes back positive? Should we believe that she loves me when the daisy has 17 petals? Our beliefs can be

Stone   Kruschke

The story so far, and what comes next

- Signalling systems (and languages) can evolve as a result of their transmission

  - We can model this

- The **biases** of learners shapes what evolves

- This potentially allows us to link findings about biases in learning at the individual level to predictions / observations about language at the population level

  - But caution (or better, a model) is required - the acquisition test here was misleading

- **Next up:** a class of models that allow us to be very clear and very precise about bias

We need a more general model

- Ideally, we'd like to be able to mix up a bunch of simple ingredients and work out what language should look like after cultural evolution has run for some time:

  - BIAS (i.e. what agents are born with)

  - LANGUAGE MODEL (i.e. set of possible languages, set of possible data)

  - POPULATION MODEL (e.g. diffusion chain, closed group etc.)

  - OTHER FEATURES OF CULTURAL TRANSMISSION (e.g. how much data a learner sees, the type of errors that occur, ...)

Towards a more general model of learning bias: a medical quiz

- Your friend coughs. Is this cough caused by:

  A. Lung cancer

  B. A cold

  C. Athlete's foot

- Resolving this question requires you to draw on two probabilities:

  - How likely is it that someone with the illness in question would exhibit that symptom?

  - How common is each illness?

Likelihood of symptoms given illnesses

Lung cancer: coughing is very likely, if you have lung cancer

A cold: coughing is very likely, if you have a cold

Athlete's foot: coughing is very very unlikely to be caused by athlete's foot

- If all we care about are the likelihood of the symptoms given each illness, we would conclude that your friend either has lung cancer or a cold

Probability of illnesses

Lung cancer: is fairly rare

A cold: the common cold is very common

Athlete's foot: is very common (let's say)

- If all we care about are the prevalence of each illness, we would conclude that your friend either has a cold or athlete's foot

- But you didn't conclude this: you brought these two quantities together in a smart way. How did you do it?

The Bayesian approach

- What you're trying to figure out is the probability that your friend has a particular illness, given the symptoms they are exhibiting. We call this quantity:

$$P(illness|symptoms)$$

- We are trying to work this out based on two quantities which we know (roughly):

  - The likelihood of exhibiting a particular symptom given that you have a certain illness

$$P(symptoms|illness)$$

  - The prior probability of each illness

$$P(illness)$$

Bayes' rule

- Bayes' rule provides a convenient way of expressing the quantity we want to know in terms of the quantities we already know:

$$P(illness|symptoms) \propto P(symptoms|illness)P(illness)$$

- Or, in full:

$$P(illness|symptoms) = \frac{P(symptoms|illness)P(illness)}{P(symptoms)}$$

Breaking it down

$$P(illness|symptoms) = \frac{P(symptoms|illness)P(illness)}{P(symptoms)}$$

$P(illness|symptoms)$

- The thing we want to know is called the **posterior**

$P(symptoms|illness)$

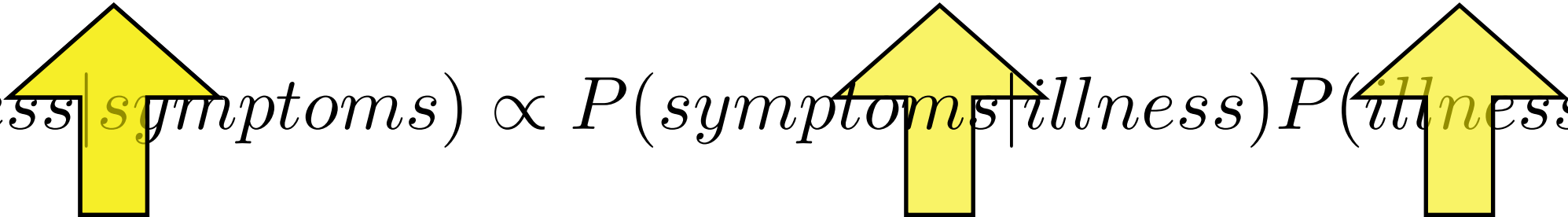- The probability of a particular set of symptoms given that you have a specific illness is called the **likelihood**

$P(illness)$

- The probability that you have a particular illness, before I have any evidence from your symptoms, is called the **prior**

$P(symptoms)$

- The term on the bottom (the probability of the symptoms independent of illness) is actually not very interesting to us, since it is the same for all illnesses.

It makes intuitive sense...

$$P(illness|symptoms) \propto P(symptoms|illness)P(illness)$$

- If the likelihood of symptoms given a certain illness is high, this will increase the posterior probability of that illness

- If the prior probability of a certain illness is high, this will increase the posterior probability of that illness

- If a particular illness has low prior probability, we need some really convincing evidence to make us believe it to be true

    - imagine if your friend was coughing blood and having seizures

Errr... hello... isn't this a course about language?

- In the medical example, we were trying to use evidence provided by symptoms to infer what underlying illness your friend had

- What if you aren't a medic, but a child hearing utterances from a parent and learning a language? You are trying to use evidence provided by utterances to infer what grammar your parent has in their head

        illnesses = languages
        symptoms = utterances
        prior for each illness = **bias in favour of particular languages**

- An ideal language learner will estimate the posterior probability of each possible language given the utterances heard

- Children probably don't calculate sums in their head while learning, but if their learning process is sensible, we can characterise it this way

Bayesian language learning

- Evaluate hypotheses about language given some prior bias (perhaps provided by your biology?) and the data that you've heard

- You want to know the **posterior** but all you have direct access to is the **prior** and the **likelihood** (assuming you know how sentences are produced from a given model of language)

- Bayes' rule provides the solution:

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

Bayesian language learning

- Evaluate hypotheses about language given some prior bias (perhaps provided by your biology?) and the data that you've heard

- You want to know the **posterior** but all you have direct access to is the **prior** and the **likelihood** (assuming you know how sentences are produced from a given model of language)

- Bayes' rule provides the solution:
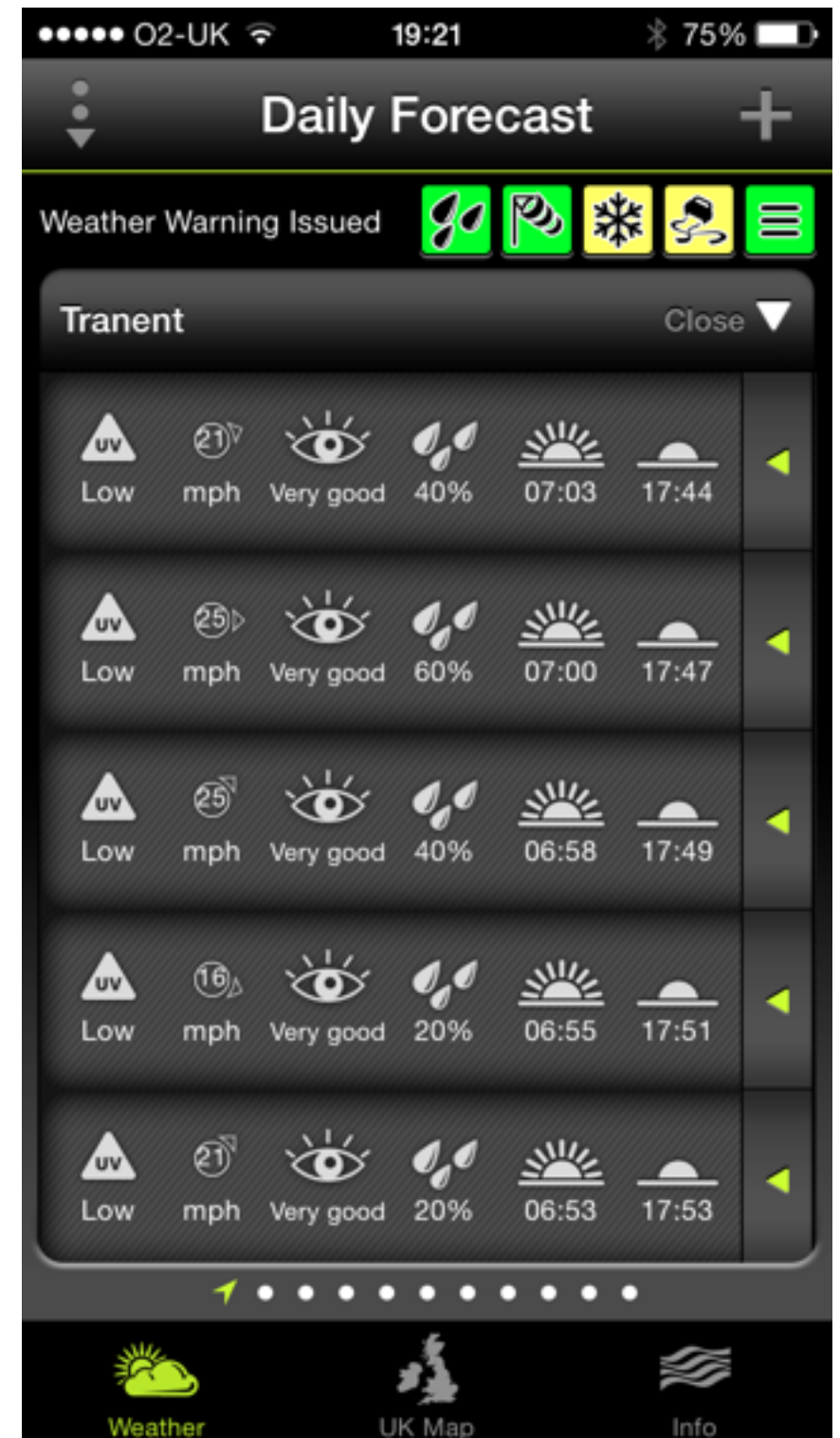
$$P(h|d) \propto P(d|h)P(h)$$

Bayesian language learning

$$d \quad P(h|d) \propto P(d|h)P(h)$$

$$h \quad P(d|h) \quad d$$
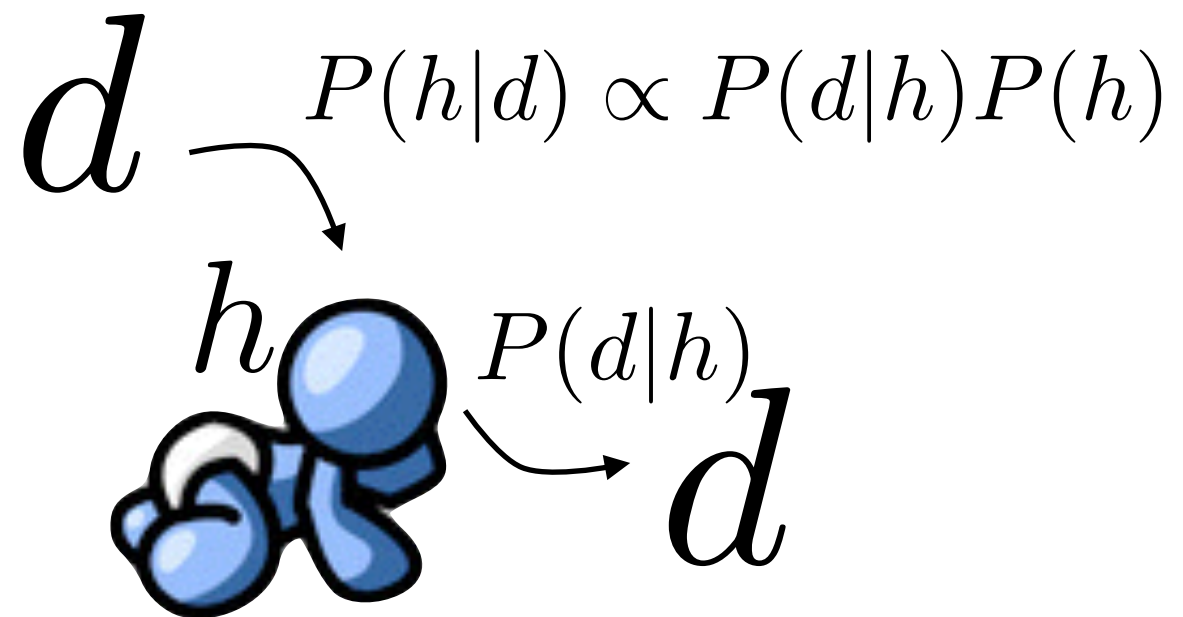
How to put numbers on uncertainty: probability

$$P(h|d) \propto P(d|h)P(h)$$

- Bayes Rule is about a relationship between probabilities. What are probabilities?

- How likely is it to rain tomorrow?

Chance of rain

Rain	No rain

- It will definitely either rain or not rain ("precipitate or not precipitate" - whatever) tomorrow: 100% chance of one of those two outcomes

Chance of rain

Rain	No rain

- It will definitely either rain or not rain ("precipitate or not precipitate" - whatever) tomorrow: 100% chance of one of those two outcomes

- The met office reckons it's slightly more likely to rain than not: estimate a 60% chance of rain

Chance of rain

Rain   No rain
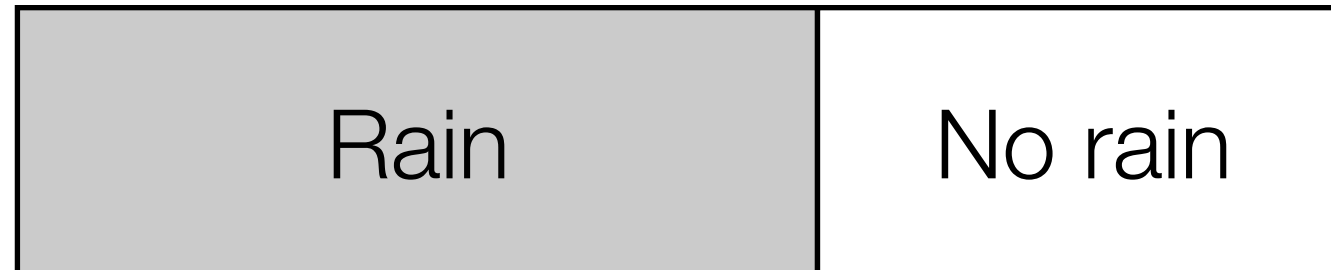
If there's a 60% chance it will rain, what chance do the Met Office assign to it not raining?

A. 0% chance of no rain

B. 40% chance of no rain

C. 60% chance of no rain

D. 100% chance of no rain

From chance to probability

Rain	No rain

- The Met Office are in fact stating their estimate of the **probability** of it raining tomorrow.

- They express is as a %

- But probabilities are conventionally expressed as a proportion of possible events

  - i.e. they lie between 0 and 1

- "60% chance" corresponds to probability of 60/100 or 6/10 or 0.6

Assigning probabilities

Consider <u>this scenario </u>Roughly what probability does the heroic forecaster assign to it raining heavily?

A. probability of heavy rain ≈ 0

B. probability of heavy rain ≈ 0.5

C. probability of heavy rain ≈ 1

D. probability of heavy rain = 1

E. probability of heavy rain ≈ 2

Assigning probabilities

Consider <u>this scenario </u>What probability does the meteorological board assign to it raining heavily?

  A. substantially lower probability

  B. about the same probability

  C. substantially higher probability

Discrete probability distributions

p(rain)=0.6	p(no-rain)= 0.4

- A limited set of known possible scenarios

  - Rain, no rain

- We can assign a probability to each event

  - Since the events are mutually exclusive, the probabilities sum to 1

Continuous probability distributions
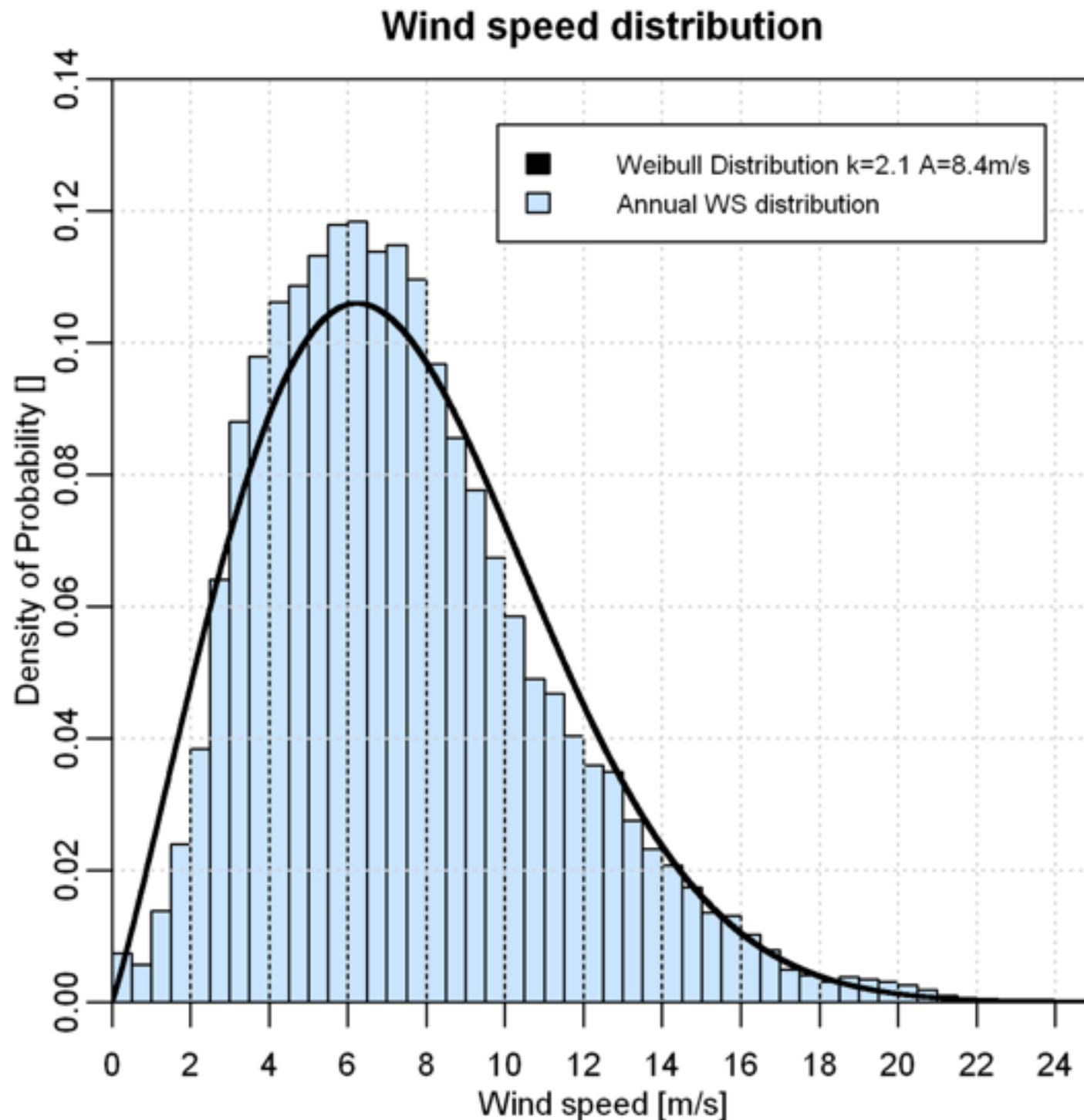
Summary and next up $P(h|d) \propto P(d|h)P(h)$

- Bayesian learning: a nice simple way to model learning

- Make the bias of learners beautifully explicit

- Involves probabilities:

  - For each possible hypothesis, what is its prior probability? What is the likelihood of the data under that hypothesis?

  - For each possible language, what is its prior probability? What is the likelihood of the linguistic data if people are using that language?

- Monday's lecture: a linguistic case study, and iterated Bayesian learning

- Thursday: lab on iterated Bayesian learning

- Friday: Jennifer Culbertson