# Probability and (Bayesian) Data Analysis

Brendon J. Brewer

Department of Statistics
The University of Auckland

https://www.stat.auckland.ac.nz/~brewer/

# Where to get everything

To get all of the material (slides, code, exercises):

```
git clone --recursive
      https://github.com/eggplantbren/Madrid
```

# Book recommendations etc.

See this web page which I wrote for my incoming research students.

https://www.stat.auckland.ac.nz/~brewer/student-resources.html

# Probability

Probability is a mathematical framework that has two main applications:

(1) Describing proportions of sets.

(2) Describing the plausibility of statements.

(1) is associated with 'frequentist' statistics, and (2) is 'Bayesian'. Both are valid. The kind of 'frequentism' I disagree with is the denial of (2), not the acceptance of (1).

# The two rules of probability — general versions

For any propositions/statements $X$, $Y$, and $Z$, we have the **sum rule**:

$$P(X \vee Y|Z) = P(X|Z) + P(Y|Z) - P(X, Y|Z) \tag{1}$$

and the **product rule**:

$$P(X, Y|Z) = P(X|Z)P(Y|X, Z). \tag{2}$$

# Easier versions

The easy sum rule:

$$P(X \vee Y) = P(X) + P(Y) \tag{3}$$

when $X$ and $Y$ are mutually exclusive (they cannot both be true, i.e., they are two *alternative* hypotheses).

The easy product rule:

$$P(X, Y) = P(X)P(Y|X). \tag{4}$$

for any statements $X$, $Y$.

# Bayes' rule

From the product rule and commutativity of logical *and*:

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)} \tag{5}$$

# Special properties

Sometimes probability assignments make pairs of statements *independent*. In this special case, the product rule reduces to:

$$P(X, Y) = P(X)P(Y) \tag{6}$$

Sometimes probability assignments make pairs of statements *mutually exclusive*. In this special case, the sum rule reduces to:

$$P(X \vee Y) = P(X) + P(Y). \tag{7}$$

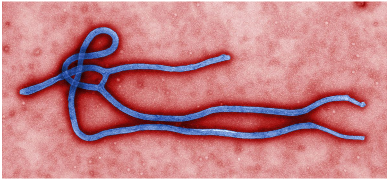
# Bayes' rule — most useful form

For a set of *mutually exclusive and exhaustive* (i.e., they're *alternatives*) hypotheses $\{H_i\}$,

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{\sum_i P(H_i)P(D|H_i)}. \tag{8}$$

- $P(H_i)$ are the prior probabilities
- $P(D|H_i)$ are the likelihoods
- The denominator, $P(D)$ is the 'marginal likelihood' or 'evidence'.

# Updating Probabilities: Example



A patient goes to the doctor because he as a fever. Define

$$\begin{aligned} H &\equiv \text{“The patient has Ebola”} \\ \neg H &\equiv \text{“The patient does not have Ebola”.} \end{aligned}$$

# Updating Probabilities: Example

Based on all of her knowledge, the doctor assigns probabilities to the two hypotheses.

$$
\begin{aligned}
P(H) &= 0.01 \\
P(\neg H) &= 0.99
\end{aligned}
$$

But she wants to test the patient to make sure.

# Updating Probabilities: Example

The patient is tested. Define

$$D \equiv \text{“The \textbf{test says} the patient has Ebola”}$$
$$\neg D \equiv \text{“The \textbf{test says} the patient does not have Ebola”.}$$

If the test were perfect, we'd have $P(D|H) = 1$, $P(\neg D|H) = 0$, $P(D|\neg H) = 0$, and $P(\neg D|\neg H) = 1$.

# Updating Probabilities: Example

The Ebola test isn't perfect. Suppose there's a 5% probability it simply gives the wrong answer. Then we have:

$$
\begin{aligned}
P(D|H) &= 0.95 \\
P(\neg D|H) &= 0.05 \\
P(D|\neg H) &= 0.05 \\
P(\neg D|\neg H) &= 0.95
\end{aligned}
$$

# Updating Probabilities: Example

Overall, there are four possibilities, considering whether the patient has Ebola or not, and what the test says.

$$(H, D)$$
$$(\neg H, D)$$
$$(H, \neg D)$$
$$(\neg H, \neg D)$$

# Updating Probabilities: Example

The probabilities for these four possibilities can be found using the product rule.

$$
\begin{aligned}
P(H, D) &= 0.01 \times 0.95 \\
P(\neg H, D) &= 0.99 \times 0.05 \\
P(H, \neg D) &= 0.01 \times 0.05 \\
P(\neg H, \neg D) &= 0.99 \times 0.95
\end{aligned}
$$

These four possibilities are **mutually exclusive** (only one of them is true) and exhaustive (it's not “something else”), so the probabilities add up to 1.

# Updating Probabilities: Example

The test results come back and say that the patient has Ebola. That is, we've learned that $D$ is true. So we can confidently rule out those possibilities where $D$ is false:

$$
\begin{aligned}
P(H, D) &= 0.01 \times 0.95 \\
P(\neg H, D) &= 0.99 \times 0.05 \\
\color{red}{P(H, \neg D)} &\color{red}{= 0.01 \times 0.05} \\
\color{red}{P(\neg H, \neg D)} &\color{red}{= 0.99 \times 0.95}
\end{aligned}
$$

# Updating Probabilities: Example

We are left with these two possibilities.

$$
\begin{aligned}
P(H, D) &= 0.01 \times 0.95 \\
P(\neg H, D) &= 0.99 \times 0.05
\end{aligned}
$$

It would be strange to modify these probabilities just because we deleted the other two. The only thing we have to do is renormalise them, by dividing by the total, so they sum to 1 again.

# Updating Probabilities: Example

Normalising, we get

$$
\begin{aligned}
P(H|D) &= (0.01 \times 0.95)/(0.01 \times 0.95 + 0.99 \times 0.05) = 0.161 \\
P(\neg H|D) &= (0.99 \times 0.05)/(0.01 \times 0.95 + 0.99 \times 0.05) = 0.839
\end{aligned}
$$

# Moral

Bayesian updating is completely equivalent to:

- Writing a list of possible answers to your question
- Giving a probability to each
- Deleting the ones that you discover are false.

It just seems more complicated than this because we often apply it to more complex sets of hypotheses.

# Basic Bayesian exercises

Do exercise set 1.