Prediction and Odds
18.05 Spring 2022

Announcements/Agenda

Announcements

- R studio is posted, so you can do it ahead of time.
- We made the second problem optional so it will be short.
- Exam solutions are posted in the 'Review Materials' section on MITx.

Agenda

- Probabilistic prediction in words
- Prior and posterior predictive probabilities
- Odds
- Bayesian updating using odds
- Bayes' factors and the weight of evidence

Probabilistic Prediction

Also called probabilistic forecasting.
Assign a probability to each outcome of a future experiment.

Prediction: "It will rain tomorrow."

Probabilistic prediction: "Tomorrow it will rain with probability 60% (and not rain with probability 40%)."

Examples: medical treatment outcomes, weather forecasting, climate change, sports betting, elections, ...

Words of estimative probability (WEP)

WEP Prediction: "It is likely to rain tomorrow."

Memo: *Bin Laden Determined to Strike in US*

See https://en.wikipedia.org/wiki/Words_of_Estimative_Probability

"The language used in the [Bin Laden] memo lacks words of estimative probability (WEP) that reduce uncertainty, thus preventing the President and his decision makers from implementing measures directed at stopping al Qaeda's actions."

"Intelligence analysts would rather use words than numbers to describe how confident we are in our analysis," a senior CIA officer who's served for more than 20 years told me. Moreover, "most consumers of intelligence aren't particularly sophisticated when it comes to probabilistic analysis. They like words and pictures, too. My experience is that [they] prefer briefings that don't center on numerical calculation."

WEP versus Probabilities: medical consent

No common standard for converting WEP to numbers.

Suggestion for potential risks of a medical procedure:

Word	Probability
Likely	Will happen to more than 50% of patients
Frequent	Will happen to 10-50% of patients
Occasional	Will happen to 1-10% of patients
Rare	Will happen to less than 1% of patients

From same Wikipedia article

Predictive probabilities

Prior predictive probability: Probability of an outcome **before taking data**.

Posterior predictive probability: Probability of an outcome **after taking data**.

This is different from the prior and posterior probabilities of hypotheses.

Example. Two types of dice: 4 and 12-sided. Choose one at random. First roll: $R_1 = 3$. Second roll: R_2.

- Prior probabilities (of hypotheses): $P(\text{4-sided})$.
- Posterior probabilities (of hypotheses): $P(\text{12-sided}|R_1 = 3)$.
- Prior predictive probability (of outcome): $P(R_1 = 3)$.
- Posterior predictive probability (of outcome): $P(R_2 = 8|R_1 = 3)$.

Concept question: Three coins

- Type $C_{0.5}$ coins are fair, with probability 0.5 of heads
- Type $C_{0.6}$ coins have probability 0.6 of heads
- Type $C_{0.9}$ coins have probability 0.9 of heads

A drawer has one of each.

Pick one; toss 5 times; Suppose you get 1 head out of 5 tosses.

What's your best guess for the probability of heads on the next toss?

(a) 0.1 **(b)** 0.2 **(c)** 0.3 **(d)** 0.4

(e) 0.5 **(f)** 0.6 **(g)** 0.7 **(h)** 0.8

(i) 0.9 **(j)** 1.0

(This is answered in the next board question.)

Board question: Three coins

- We have 3 coins with probabilities 0.5, 0.6, and 0.9 of heads.
- Pick one at random; toss 5 times.
- Suppose you get 1 head out of 5 tosses.

Compute the posterior probabilities for the type of coin and the posterior predictive probabilities for the results of the next toss.

(a) Specify clearly the set of hypotheses and the prior probabilities.

(b) Compute the prior and posterior predictive distributions, i.e. give the probabilities of all possible outcomes.

Concept Question: Does order matter?

Suppose instead of saying 1 head in 4 tails, we told you the tosses, in order, were THTTT. Does this affect posterior distribution of the coin type?

(a) Yes **(b)** No

Odds

Definition The odds of an event are

$$O(E) = \frac{P(E)}{P(E^c)}.$$

- Usually for two choices: E and *not* E.
- Can split multiple outcomes into two groups.
- Bayesian focus: Updating the odds of a hypothesis H given data D.

Key formula: posterior odds $=$ likelihood $times$ prior odds

This simple formula can be a reason to prefer odds to probabilities.

Odds examples

- A fair coin has $O(\text{heads}) = \dfrac{0.5}{0.5} = 1$.
  We say '1 to 1' or 'fifty-fifty'.

- The odds of rolling a 4 with a six-sided die are $\dfrac{1/6}{5/6} = \dfrac{1}{5}$.
  We say '1 to 5 for' or '5 to 1 against'

- For event E, if $P(E) = p$ then $O(E) = \dfrac{p}{1-p}$.

- If an event is rare, then $P(E) \approx O(E)$ (since $1 - p \approx 1$).

Bayesian framework: Marfan's Syndrome

Marfan's syndrome (M) is a genetic disease of connective tissue. The main ocular features (F) of Marfan's syndrome include bilateral ectopia lentis (lens dislocation), myopia and retinal detachment.

Here are some known statistics
$P(M) = 1/15000, \quad P(F|M) = 0.7, \quad P(F|M^c) = 0.07$

Problem. If a person has the main ocular features F what is the probability they have Marfan's syndrome?

Solution: Bayesian updating: $P(M|F) = 0.00066$

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(F\|\mathcal{H})$	$P(F\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|F)$
M	0.000067	0.7	0.0000467	0.00066
M^c	0.999933	0.07	0.069995	0.99933
total	1		0.07004	1

Same problem using odds

$P(M) = 1/15000$, $\quad P(F|M) = 0.7$, $\quad P(F|M^c) = 0.07$

Prior odds:

$$O(M) = \frac{P(M)}{P(M^c)} = \frac{1/15000}{14999/15000} = \frac{1}{14999} = 0.000067.$$

Note: $O(M) \approx P(M)$ since $P(M)$ is small.

Posterior odds:

$$\begin{aligned} O(M|F) &= \frac{P(M|F)}{P(M^c|F)} = \frac{\frac{P(F|M)P(M)}{P(F)}}{\frac{P(F|M^c)P(M^c)}{P(F)}} \\ &= \frac{P(F|M)}{P(F|M^c)} \cdot \frac{P(M)}{P(M^c)} \\ &= \frac{0.7}{0.07} \cdot 0.000067 = 0.000667. \end{aligned}$$

Bayes' factors (likelihood ratios)

$$O(M|F) = \frac{P(F|M)}{P(F|M^c)} \cdot \frac{P(M)}{P(M^c)} = \frac{P(F|M)}{P(F|M^c)} \cdot O(M)$$

$$= \frac{0.7}{0.07} \cdot O(M)$$

posterior odds $=$ Bayes' factor $\cdot$ prior odds

- The Bayes' factor is the ratio of the likelihoods. (Also called the likelihood ratio.)
- The Bayes' factor gives the strength of the 'evidence' provided by the data.
- A large Bayes' factor times small prior odds can be small (or large or in between).
- The Bayes' factor for ocular features is $0.7/0.07 = 10$.

Board Question: screening test odds

A disease is present in 0.005 of the population.

A screening test has a 0.05 false positive rate and a 0.02 false negative rate.

(a) Give the prior odds a patient has the disease

Assume the patient tests positive

(b) What is the Bayes factor for this data?

(c) What are the posterior odds they have the disease?

(d) Based on your answers to (a) and (b) would you say a positive test (the data) provides strong or weak evidence for the presence of the disease.

Board Question: CSI Blood Types*

We need to weigh the evidence at a crime scene:

- Crime scene: the two perpetrators left blood: one of type O and one of type AB
- In population 60% are type O and 1% are type AB

(a) Suspect Oliver is tested and has type O blood.

Compute the Bayes factor and posterior odds that Oliver was one of the perpetrators.

Is the data evidence for or against the hypothesis that Oliver is guilty?

(b) Same question for suspect Alberto who has type AB blood.

See helpful hints on next slide.

*From 'Information Theory, Inference, and Learning Algorithms' by David J. C. Mackay.

Hopefully Helpful Hints

Population: 60% type O; 1% type AB

For the question about Oliver we have

Hypotheses:

$S =$ 'Oliver and another unknown person were at the scene'
$S^c =$ 'two unknown people were at the scene'

Data: $D =$ 'type 'O' and 'AB' blood were found; Oliver is type O'

Prior odds: These are unknown. We can just say $O(S)$.

Updating again and again

Collect data: D_1, D_2, ...

Posterior odds to D_1 become prior odds to D_2. So,

$$O(H|D_1, D_2) = O(H) \cdot \frac{P(D_1|H)}{P(D_1|H^c)} \cdot \frac{P(D_2|H)}{P(D_2|H^c)}$$
$$= O(H) \cdot BF_1 \cdot BF_2.$$

Independence assumption:
D_1 and D_2 are *conditionally* independent.

$$P(D_1, D_2|H) = P(D_1|H)P(D_2|H).$$

Legal Thoughts

David Mackay:

"In my view, a jury's task should generally be to multiply together carefully evaluated likelihood ratios from each independent piece of admissible evidence with an equally carefully reasoned prior probability. This view is shared by many statisticians but learned British appeal judges recently disagreed and actually overturned the verdict of a trial because the jurors *had* been taught to use Bayes' theorem to handle complicated DNA evidence."

Marfan's Symptoms

The Bayes factor for ocular features (F) is

$$BF_F = \frac{P(F|M)}{P(F|M^c)} = \frac{0.7}{0.07} = 10$$

The wrist sign (W) is the ability to wrap one hand around your other wrist to cover your pinky nail with your thumb. Assume 10% of the population have the wrist sign, while 90% of people with Marfan's have it. So,

$$BF_W = \frac{P(W|M)}{P(W|M^c)} = \frac{0.9}{0.1} = 9.$$

$$O(M|F,W) = O(M) \cdot BF_F \cdot BF_W = \frac{1}{14999} \cdot 10 \cdot 9 \approx \frac{6}{1000}.$$

We can convert posterior odds back to probability, but since the odds are so small the result is nearly the same:

$$P(M|F,W) \approx \frac{6}{1000+6} \approx 0.596\%$$