Prediction and Odds

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

Probabilistic Prediction, or Probabilistic Forecasting

Assign a **probability** to each outcome of a future experiment.

Prediction: “It will rain tomorrow.”

WEP Prediction: “It is likely to 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, ...

Why quantify predictions using probability?

Bin Laden Determined to Strike in US

"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."

http://en.wikipedia.org/wiki/Words_of_Estimative_Probability

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

Example: Three types of coins

- Type A coins are fair, with probability .5 of heads
- Type B coins have probability .6 of heads
- Type C coins have probability .9 of heads

A drawer contains one coin of each type. You pick one at random.

1. *Prior* predictive probability: Before taking any data, what is the probability a toss will land heads? Tails?

2. *Posterior* predictive probability: First toss lands heads. What is the probability the next toss lands heads? Tails?

Solution 1

1. Use the law of total probability:

Let $D_{1,H}$ = 'toss 1 is heads', $D_{1,T}$ = 'toss 1 is tails'.

$$\begin{aligned} P(D_{1,H}) &= P(D_{1,H}|A)P(A) + P(D_{1,H}|B)P(B) + P(D_{1,H}|C)P(C) \\ &= .5 \cdot .3333 + .6 \cdot .3333 + .9 \cdot .3333 \\ &= .6667 \end{aligned}$$

$P(D_{1,T}) = 1 - P(D_{1,H}) = .3333$

Solution 2

2. We are given the data $D_{1,H}$. First update the probabilities for the type of coin.

Let $D_{2,H}$ = 'toss 2 is heads', $D_{2,T}$ = 'toss 2 is tails'.

hypothesis	prior	likelihood	unnormalized posterior	posterior
H	$P(H)$	$P(D_{1,H}\|H)$	$P(D_{1,H}\|H)P(H)$	$P(H\|D_{1,H})$
A	$1/3$	.5	.1667	.25
B	$1/3$	.6	.2	.3
C	$1/3$	.9	.3	.45
total	1		.6667	1

Next use the law of total probability:

$$\begin{aligned} P(D_{2,H}|D_{1,H}) &= P(D_{2,H}|A)P(A|D_{1,H}) + P(D_{2,H}|B)P(B|D_{1,H}) \\ &\quad + P(D_{2,H}|C)P(C|D_{1,H}) \\ &= .71 \end{aligned}$$

$P(D_{2,T}|D_{1,H}) = .29.$

Three coins, continued.

As before: 3 coins with probability .5, .6 and .9 of heads.

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

Concept question: What's your best guess for the probability of heads on the next roll?

(1) .1 (2) .2 (3) .3 (4) .4 (5) .5
(6) .6 (7) .7 (8) .8 (9) .9

Three coins, continued.

As before: 3 coins with probability .5, .6 and .9 of heads.

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

Concept question: What's your best guess for the probability of heads on the next roll?

(1) .1 (2) .2 (3) .3 (4) .4 (5) .5
(6) .6 (7) .7 (8) .8 (9) .9

Board question: For this example:

1. Specify clearly the set of hypotheses and the prior probabilities.

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

answer: *See next slide.*

Solution

Data = '1 head and 4 tails'

hypothesis	prior	likelihood	unnormalized posterior	posterior
H	$P(H)$	$P(D\|H)$	$P(D\|H)P(H)$	$P(H\|D)$
A	$1/3$	$\binom{5}{1}.5^5$	.0521	.669
B	$1/3$	$\binom{5}{1}.6 \cdot .4^4$	.0256	.329
C	$1/3$	$\binom{5}{1}.9 \cdot .1^4$	.00015	.002
total	1		.0778	1

So,

$$P(\text{heads}|D) = .669 \cdot .5 + .329 \cdot .6 + .002 \cdot .9 = 0.53366$$
$$P(\text{tails}|D) = 1 - P(\text{heads}|D) = .46634.$$

Concept Question

Does the order of the 1 head and 4 tails affect the posterior distribution of the coin type?

1. Yes   2. No

Does the order of the 1 head and 4 tails affect the posterior predictive distribution of the next flip?

1. Yes   2. No

answer: No for both questions.

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.
- Can do odds of A vs. $B = P(A)/P(B)$.
- Our Bayesian focus:
  Updating the odds of a hypothesis H given data D.

Examples

- A fair coin has $O(\text{heads}) = \dfrac{.5}{.5} = 1$.
  We say '1 to 1' or 'fifty-fifty'.
- The odds of rolling a 4 with a 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 $P(E) \approx O(E)$.

Bayesian framework: Marfan's Syndrome

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

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

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

hypothesis	prior	likelihood	unnormalized posterior	posterior
H	$P(H)$	$P(D\|H)$	$P(D\|H)P(H)$	$P(H\|D)$
M	.000067	.7	.0000467	.00066
M^c	.999933	.07	.069995	.99933
total	1		.07004	1

Odds form

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

Prior odds:

$$O(M) = \frac{P(M)}{P(M^c)} = \frac{1/15000}{14999/15000} = \frac{1}{14999} = .000067.$$

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

Posterior odds: can use the unnormalized posterior!

$$O(M|D) = \frac{P(M|D)}{P(M^c|D)} = \frac{P(D|M)P(M)}{P(D|M^c)P(M^c)} = .000667.$$

The posterior odds is a product of factors:

$$O(M|D) = \frac{P(D|M)}{P(D|M^c)} \cdot \frac{P(M)}{P(M^c)} = \frac{.7}{.07} \cdot O(M)$$

Bayes factors

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

posterior odds = Bayes factor · prior odds

- The Bayes factor is the ratio of the likelihoods.
- 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 $.7/.07 = 10$.

Board Question: screening tests

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.

1. If a patient tests positive what are the odds they have the disease?

2. What is the Bayes factor for this data?

3. Based on you answers to (1) and (2) would you say the data provides strong or weak evidence for the presence of the disease.

answer: See next slide

Solution

Likelihood table:

		Possible data	
		Positive	negative
Hypotheses	Healthy	.05	.95
	Sick	.98	.02

Bayesian update table following a positive test (likelihood column taken from likelihood table)

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\text{Positive}\mid\mathcal{H})$	$P(\text{Positive}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\text{Positive})$
Healthy	.995	.05	.04975	.909
Sick	.005	.98	.00499	.091
total	1		.05474	1

Posterior odds of being sick = .091/.909 = .04975/.00499 = .100
Bayes factor: P(sick—positive)/P(healthy—positive) = .98/.05 = 19.6
This is strong evidence, nonetheless the small prior makes it unlikely the patient has the disease.

Board Question: CSI Blood Types*

- 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

1. 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?

2. Same question for suspect Alberto who has type AB blood.

Show helpful hint on next slide.

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

Helpful hint

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'

Solution to CSI Blood Types

For Oliver:

$$\text{Bayes factor} = \frac{P(D|S)}{P(D|S^c)} = \frac{.01}{2 \cdot .6 \cdot .01} = .83.$$

Therefore the posterior odds $= .83 \times$ prior odds $(O(S|D) = .83 \cdot O(S))$
Since the odds of his presence decreased this is (weak) evidence of his innocence.

For Alberto:

$$\text{Bayes factor} = \frac{P(D|S)}{P(D|S^c)} = \frac{.6}{2 \cdot .6 \cdot .01} = 50.$$

Therefore the posterior odds $= 50 \times$ prior odds $(O(S|D) = 50 \cdot O(S))$
Since the odds of his presence increased this is (strong) evidence of his presence at the scene.

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."

Updating again and again

Collect data: D_1, D_2, ...

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

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

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

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

Marfan's Symptoms

The Bayes factor for ocular features (F) is

$$BF_F = \frac{P(F|M)}{P(F|M^c)} = \frac{.7}{.07} = 10$$

The wrist sign (W) is the ability to wrap one hand around your other wrist to cover your pinky nail with your thumb. In our class, 5 or 50 students have the wrist sign and so we estimate $P(W|M^c) = .1$. So:

$$BF_W = \frac{P(W|M)}{P(W|M^c)} = \frac{.9}{.1} = 9$$

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

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

$$P(M|F,W) \approx \frac{6}{1000+6} \approx .596\%$$