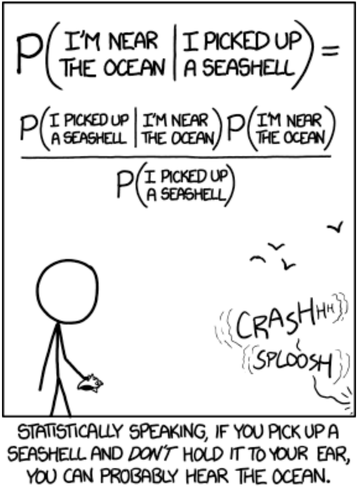
Bayesian Updating: Discrete Priors: 18.05 Spring 2022



https://xkcd.com/1236/

Announcements/Agenda

Announcements

- Pset 6 is due tomorrow. Pset 7 will be due on Monday April 4.
- R Studio 4 is graded. Good job overall.
  - A few people used the sample mean and variance to draw the normal curve – the instructions asked for the theoretical values.
  - Coding hint: Don't keep retyping values like 18/38 and 20/38. Put them in a variable and use that.

Agenda

- Bayesian updating from prior to posterior probabilities.
- Organizing the computation in tables.
- Today: Known priors (base rates) –not like in real science.

Concept question: Learning from experience

(a) Which treatment would you choose?

1. Treatment 1: cured 100% of patients in a trial.
2. Treatment 2: cured 95% of patients in a trial.
3. Treatment 3: cured 90% of patients in a trial.

Concept question: Learning from experience

(a) Which treatment would you choose?

1. Treatment 1: cured 100% of patients in a trial.
2. Treatment 2: cured 95% of patients in a trial.
3. Treatment 3: cured 90% of patients in a trial.

(b) Which treatment would you choose?

1. Treatment 1: cured 3 out of 3 patients in a trial.
2. Treatment 2: cured 19 out of 20 patients treated in a trial.
3. Standard treatment: cured 90000 out of 100000 patients in clinical practice.

Which die is it?

- I have a bag containing dice of two types: 4-sided and 6-sided.
- Suppose I pick a die at random.
- What are the hypotheses for which die I chose?
- I'll roll the die.
- Based on what I rolled which type would you guess I picked?

Which die is it?

- I have a bag containing dice of two types: 4-sided and 6-sided.
- Suppose I pick a die at random.
- What are the hypotheses for which die I chose?
- I'll roll the die.
- Based on what I rolled which type would you guess I picked?

• Suppose you find out that the bag contained one 4-sided die and 999 6-sided dice. Does this change your guess?

Bayesian updating terminology, trees, tables

Example. Bag with one 4-sided die and 999 6-sided dice. Pick one at random and roll it. Suppose I get a 3.

For this example, all of the following are given in the posted solutions to today's problems.

Terminology: data, hypotheses, likelihoods, prior probabilities, posterior probabilities.

$$\overset{\text{Posterior}}{P(\mathcal{H}_4 \mid R=3)} = \frac{\overset{\text{Likelihood}}{P(R=3 \mid \mathcal{H}_4)} \cdot \overset{\text{Prior}}{P(\mathcal{H}_4)}}{\underset{\text{Total probability of the data}}{P(R=3)}}$$

Updating using trees and tables.

Board Question: Updating from data

- A certain disease has a prevalence of 0.005.
- A screening test has 2% false positives an 1% false negatives.

Suppose a random patient is screened and has a positive test.
(a) Represent this information with a tree and use Bayes' theorem to compute the probabilities the patient does and doesn't have the disease.

(b) Identify the data, hypotheses, likelihoods, prior probabilities and posterior probabilities.

(c) Make a full likelihood table containing all hypotheses and possible test data.

(d) Redo the computation using a Bayesian update table. Match the terms in your table to the terms in your previous calculation.

Board Question: Dice

Five dice: 4-sided, 6-sided, 8-sided, 12-sided, 20-sided.

I pick one at random, roll it and report that the roll was a 13.

Goal: Find the probabilities the die is 4, 6, 8, 12 or 20 sided.

(a) Identify the hypotheses.

(b) Make a likelihood table with columns for the data 'rolled a 13', 'rolled a 5' and 'rolled a 9'.

(c) Make a Bayesian update table and compute the posterior probabilities that the chosen die is each of the five dice.

(d) Same question if I had reported a 5.

(e) Same question if I had reported a 9.

(Keep the tables for 5 and 9 handy for the next problem!)

Tabular solution

$\mathcal{D}$ = 'rolled a 13'

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\mathcal{D})$
$\mathcal{H}_4$	$1/5$	0	0	0
$\mathcal{H}_6$	$1/5$	0	0	0
$\mathcal{H}_8$	$1/5$	0	0	0
$\mathcal{H}_{12}$	$1/5$	0	0	0
$\mathcal{H}_{20}$	$1/5$	$1/20$	$1/100$	1
total	1		$1/100$	1

Tabular solution

$\mathcal{D}$ = 'rolled a 5'

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
$\mathcal{H}_4$	1/5	0	0	0
$\mathcal{H}_6$	1/5	1/6	1/30	0.392
$\mathcal{H}_8$	1/5	1/8	1/40	0.294
$\mathcal{H}_{12}$	1/5	1/12	1/60	0.196
$\mathcal{H}_{20}$	1/5	1/20	1/100	0.118
total	1		0.085	1

Tabular solution

$\mathcal{D}$ = 'rolled a 9'

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\mathcal{D})$
$\mathcal{H}_4$	$1/5$	0	0	0
$\mathcal{H}_6$	$1/5$	0	0	0
$\mathcal{H}_8$	$1/5$	0	0	0
$\mathcal{H}_{12}$	$1/5$	$1/12$	$1/60$	0.625
$\mathcal{H}_{20}$	$1/5$	$1/20$	$1/100$	0.375
total	1		0.0267	1

Iterated Updates

Suppose I rolled a 5 and then a 9.

Update in two steps:

First for the 5

Then update the update for the 9.

Tabular solution

$\mathcal{D}_1$ = 'rolled a 5'
$\mathcal{D}_2$ = 'rolled a 9'
Bayes numerator$_1$ = likelihood$_1 \times$ prior.
Bayes numerator$_2$ = likelihood$_2 \times$ Bayes numerator$_1$

hyp.	prior	likel. 1	Bayes num. 1	likel. 2	Bayes num. 2	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}_1 \mid \mathcal{H})$	$***$	$P(\mathcal{D}_2 \mid \mathcal{H})$	$***$	$P(\mathcal{H} \mid \mathcal{D}_1, \mathcal{D}_2)$
$\mathcal{H}_4$	1/5	0	0	0	0	0
$\mathcal{H}_6$	1/5	1/6	1/30	0	0	0
$\mathcal{H}_8$	1/5	1/8	1/40	0	0	0
$\mathcal{H}_{12}$	1/5	1/12	1/60	1/12	1/720	0.735
$\mathcal{H}_{20}$	1/5	1/20	1/100	1/20	1/2000	0.265
total	1				0.0019	1

Board Question: Iterated updates

Suppose I rolled a 9 and then a 5.

(a) Do the Bayesian update in two steps:
Step 1: First update for the 9.
Step 2: Then update the update for the 5.

(b) Do the Bayesian update in one step.
That is, the data is $\mathcal{D} =$ '9 followed by 5'

Tabular solution: two steps

$\mathcal{D}_1$ = 'rolled a 9'
$\mathcal{D}_2$ = 'rolled a 5'
Bayes numerator$_1$ = likelihood$_1 \times$ prior.
Bayes numerator$_2$ = likelihood$_2 \times$ Bayes numerator$_1$

hyp.	prior	likel. 1	Bayes num. 1	likel. 2	Bayes num. 2	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}_1\|\mathcal{H})$	$***$	$P(\mathcal{D}_2\|\mathcal{H})$	$***$	$P(\mathcal{H}\|\mathcal{D}_1, \mathcal{D}_2)$
$\mathcal{H}_4$	1/5	0	0	0	0	0
$\mathcal{H}_6$	1/5	0	0	1/6	0	0
$\mathcal{H}_8$	1/5	0	0	1/8	0	0
$\mathcal{H}_{12}$	1/5	1/12	1/60	1/12	1/720	0.735
$\mathcal{H}_{20}$	1/5	1/20	1/100	1/20	1/2000	0.265
total	1				0.0019	1

Tabular solution: one step

$\mathcal{D}$ = 'rolled a 9 then a 5'

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
$\mathcal{H}_4$	$1/5$	0	0	0
$\mathcal{H}_6$	$1/5$	0	0	0
$\mathcal{H}_8$	$1/5$	0	0	0
$\mathcal{H}_{12}$	$1/5$	$1/144$	$1/720$	0.735
$\mathcal{H}_{20}$	$1/5$	$1/400$	$1/2000$	0.265
total	1		0.0019	1

Board Question: probabilistic prediction

Along with finding posterior probabilities of hypotheses. We might want to make posterior predictions about the next roll.

With the same setup as before let:
$\mathcal{D}_1$ = result of first roll, $\quad \mathcal{D}_2$ = result of second roll

(a) Find $P(\mathcal{D}_1 = 5)$.

(b) Find $P(\mathcal{D}_2 = 4 | \mathcal{D}_1 = 5)$.

Solution

$\mathcal{D}_1 =$ 'rolled a 5', $\quad \mathcal{D}_2 =$ 'rolled a 4'

			Bayes			
hyp.	prior	likel. 1	num. 1	post. 1	likel. 2	post. 1 × likel. 2
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}_1 \mid \mathcal{H})$	$* * *$	$P(\mathcal{H} \mid \mathcal{D}_1)$	$P(\mathcal{D}_2 \mid \mathcal{H}, \mathcal{D}_1)$	$P(\mathcal{D}_2 \mid \mathcal{H}, \mathcal{D}_1) P(\mathcal{H} \mid \mathcal{D}_1)$
$\mathcal{H}_4$	1/5	0	0	0	$*$	0
$\mathcal{H}_6$	1/5	1/6	1/30	0.392	1/6	$0.392 \cdot 1/6$
$\mathcal{H}_8$	1/5	1/8	1/40	0.294	1/8	$0.294 \cdot 1/40$
$\mathcal{H}_{12}$	1/5	1/12	1/60	0.196	1/12	$0.196 \cdot 1/12$
$\mathcal{H}_{20}$	1/5	1/20	1/100	0.118	1/20	$0.118 \cdot 1/20$
total	1		0.085	1		0.124

The law of total probability tells us $P(\mathcal{D}_1)$ is the sum of the Bayes numerator 1 column in the table: $\boxed{P(\mathcal{D}_1) = 0.085}$.

The law of total probability tells us $P(\mathcal{D}_2 \mid \mathcal{D}_1)$ is the sum of the last column in the table: $\boxed{P(\mathcal{D}_2 \mid \mathcal{D}_1) = 0.124}$