

## 18.05 Problem Set 5, Spring 2014 Solutions

**Problem 1.** (10 pts.) (a) We know that  $y_i - ax_i - b = \varepsilon_i \sim N(0, \sigma^2)$ . Therefore  $y_i = \varepsilon_i + ax_i + b \sim N(ax_i + b, \sigma^2)$ . That is,

$$f(y_i | a, b, x_i, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y_i - ax_i - b)^2}{2\sigma^2}}.$$

(b) (i) The  $y$  values are 8, 2, 1. The likelihood function is a product of the densities found in part (a)

$$f(8, 3, 2 | a, b, \sigma) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^3 e^{-((8-a-b)^2 + (2-3a-b)^2 + (1-5a-b)^2)/2\sigma^2}$$

$$\ln(f(8, 3, 2 | a, b, \sigma)) = -3 \ln(\sigma) - \frac{3}{2} \ln(2\pi) - \frac{(8-a-b)^2 + (2-3a-b)^2 + (1-5a-b)^2}{2\sigma^2}$$

(ii) We just copy our answer in part (i) replacing the explicit values of  $x_i$  and  $y_i$  by their symbols

$$f(y_1, \dots, y_n | a, b, \sigma) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-\sum_{j=1}^n (y_j - ax_j - b)^2 / 2\sigma^2}$$

$$\ln(f(8, 3, 2 | a, b, \sigma)) = -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \sum_{j=1}^n (y_j - ax_j - b)^2 / 2\sigma^2$$

(c) We set partial derivatives to 0 to try and find the MLE. (Don't forget that  $\sigma$  is a constant.)

$$\begin{aligned} \frac{\partial}{\partial a} \ln(f(8, 3, 2 | a, b, \sigma)) &= -\frac{-2(8-a-b) - 6(2-3a-b) - 10(1-5a-b)}{2\sigma^2} \\ &= \frac{-84a - 18b + 38}{2\sigma^2} \\ &= 0 \\ &\Rightarrow 70a + 18b = 38 \end{aligned}$$

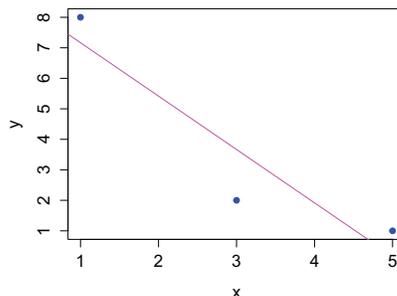
$$\begin{aligned} \frac{\partial}{\partial b} \ln(f(8, 3, 2 | a, b, \sigma)) &= -\frac{-2(8-a-b) - 2(2-3a-b) - 2(1-5a-b)}{2\sigma^2} \\ &= \frac{-18a - 6b + 22}{2\sigma^2} \\ &= 0 \\ &\Rightarrow 18a + 6b = 22 \end{aligned}$$

We have two simultaneous equations:  $70a + 18b = 38$ ,  $18a + 6b = 22$ . These are easy to solve, e.g. first eliminate  $b$  and solve for  $a$ . We get

$$a = -\frac{7}{4} \quad b = \frac{107}{12}$$

(d) Here's the R code I used to make the plot

```
x = c(1,3,5)
y = c(8,2,1)
a = -7/4
b = 107/12
plot(x,y,pch=19,col="blue")
abline(a=b,b=a, col="magenta")
```



### Problem 2. (10 pts.) Estimating uniform parameters

(a) The pdf for uniform( $a, b$ ) one data value is  $f(x_i | a, b) = 1/(b-a)$  if  $x_i$  is in the interval  $[a, b]$  and 0 if it is not. So the likelihood function for our 5 data values is

$$f(\text{data} | a, b) = \begin{cases} 1/(b-a)^5 & \text{if all data is in } [a, b] \\ 0 & \text{if not} \end{cases}$$

This is maximized when  $(b-a)$  is as small as possible. Since all the data has to be in the interval  $[a, b]$  we minimize  $(b-a)$  by taking  $a = \text{minimum of data}$  and  $b = \text{maximum of data}$ .

**answer:**  $a = 1.2, b = 10.5$ .

(b) The same logic as in part (a) shows  $a = \min(x_1, \dots, x_n)$  and  $b = \max(x_1, \dots, x_n)$ .

**Problem 3.** (10 pts.) In all three parts to this problem we have 3 hypotheses:

$H_A$  = 'the car is behind door A'

$H_B$  = 'the car is behind door B'

$H_C$  = 'the car is behind door C'.

In all three parts the data is  $D$  = 'Monty opens door B and reveals a goat'.

(a) The key to our Bayesian update table is the likelihoods: Since Monty is sober he always reveals a goat.

$P(D|H_A)$ :  $H_A$  says the car is behind A. So Monty is equally likely to pick B or C and reveal a goat. Thus  $P(D|H_A) = 1/2$ .

$P(D|H_B)$ : Since  $H_B$  says the car is behind B, sober Monty will never choose B (and if he did it would not reveal a car). So  $P(D|H_B) = 0$ .

$P(D|H_C)$ :  $H_C$  says the car is behind C. Since sober Monty doesn't make mistakes he will open door B and reveal a goat. So  $P(D|H_C) = 1$ .

Here is the table for this situation.

| $H$    | $P(H)$ | $P(D H)$ | Unnorm. Post. | Posterior |
|--------|--------|----------|---------------|-----------|
| $H_A$  | 1/3    | 1/2      | 1/6           | 1/3       |
| $H_B$  | 1/3    | 0        | 0             | 0         |
| $H_C$  | 1/3    | 1        | 1/3           | 2/3       |
| Total: | 1      | –        | 1/2           | 1         |

Therefore, Shelby should switch, as her chance of winning the car after switching is double that had she stayed with her initial choice.

b) Some of the likelihoods change in this setting.

$P(D|H_A)$ :  $H_A$  says the car is behind  $A$ . So Monty is equally likely to show  $B$  or  $C$  and reveal a goat. Thus  $P(D|H_A) = 1/2$ .

$P(D|H_B)$ : Since  $H_B$  says the car is behind  $B$ , drunk Monty might show  $B$ , but if he does we won't reveal a goat. (He will ruin the game.) So  $P(D|H_B) = 0$ .

$P(D|H_C)$ :  $H_C$  says the car is behind  $C$ . Drunk Monty is equally likely to  $B$  or  $C$ . If he chooses  $B$  he'll reveal a goat. So  $P(D|H_C) = 1/2$ .

Our table is now:

| $H$    | $P(H)$ | $P(D H)$ | Unnorm. Post. | Posterior |
|--------|--------|----------|---------------|-----------|
| $H_A$  | 1/3    | 1/2      | 1/6           | 1/2       |
| $H_B$  | 1/3    | 0        | 0             | 0         |
| $H_C$  | 1/3    | 1/2      | 1/6           | 1/2       |
| Total: | 1      | –        | 1/3           | 1         |

So in this case switching is just as good (or as bad) as staying with the original choice.

(c) We have to recompute the likelihoods.

$P(D|H_A)$ : If the car is behind  $A$  then sober or drunk Monty is equally likely to choose door  $B$  and reveal a goat. Thus  $P(D|H_A) = 1/2$ .

$P(D|H_B)$ : If the car is behind door  $B$  then whether he chooses it or not Monty can't reveal a goat behind it. So  $P(D|H_B) = 0$ .

$P(D|H_C)$ : Let  $S$  be the event that Monty is sober and  $S^c$  the event he is drunk. From the table in (a), we see that  $P(D|H_C, S) = 1$  and from the table in (b), we see that  $P(D|H_C, S^c) = 1/2$ . Thus, by the law of total probability

$$P(D|H_C) = P(D|H_C, S)P(S) + P(D|H_C, S^c)P(S^c) = 0.7 + \frac{1}{2}(0.3) = .85 = \frac{17}{20}.$$

| $H$    | $P(H)$ | $P(D H)$ | Unnorm. Post. | Posterior |
|--------|--------|----------|---------------|-----------|
| $H_A$  | 1/3    | 1/2      | 1/6           | 10/27     |
| $H_B$  | 1/3    | 0        | 0             | 0         |
| $H_C$  | 1/3    | 17/20    | 17/60         | 17/27     |
| Total: | 1      | –        | 9/20          | 1         |

Thus switching gives a probability of 17/27 of winning. So switching is the best strategy.

**Problem 4.** (10 pts.) (a) Let  $H_4, H_6, H_8, H_{12}$ , and  $H_{20}$  are the hypotheses that we have selected the 4, 6, 8, 12, or 20 sided die respectively.

We compute

| Hyp.<br>$H$ | Prior<br>$P(H)$ | Likelihood<br>$P(\text{data} H)$ | Unnorm. Post.   | Posterior<br>$P(H \text{data})$ |
|-------------|-----------------|----------------------------------|---|---------------------------------|
| $H_4$       | 1/5             | 0                                | 0   | 0                               |
| $H_6$       | 1/5             | 0                                | 0   | 0                               |
| $H_8$       | 1/5             | $(1/8)^n$                        | $\frac{1}{5} \cdot (1/8)^n$                             | $\frac{1}{5T} (1/8)^n$          |
| $H_{12}$    | 1/5             | $(1/12)^n$                       | $\frac{1}{5} \cdot (1/12)^n$                            | $\frac{1}{5T} (1/12)^n$         |
| $H_{20}$    | 1/5             | $(1/20)^n$                       | $\frac{1}{5} \cdot (1/20)^n$                            | $\frac{1}{5T} (1/20)^n$         |
| Total:      | 1               | –                                | $T = \frac{1}{5} \cdot ((1/8)^n + (1/12)^n + (1/20)^n)$ | 1                               |

The posterior probabilities are given in the table. To find what happens as  $n$  grows large, we rewrite the posterior probabilities by multiplying numerator and denominator by  $8^n$ :

$$P(H_8|\text{data}) = \frac{1}{1 + \left(\frac{2}{3}\right)^n + \left(\frac{2}{5}\right)^n}$$

$$P(H_{12}|\text{data}) = \frac{\left(\frac{2}{3}\right)^n}{1 + \left(\frac{2}{3}\right)^n + \left(\frac{2}{5}\right)^n}$$

$$P(H_{20}|\text{data}) = \frac{\left(\frac{2}{5}\right)^n}{1 + \left(\frac{2}{3}\right)^n + \left(\frac{2}{5}\right)^n}$$

As  $n \rightarrow \infty$ , we know that  $\left(\frac{2}{3}\right)^n \rightarrow 0$  and  $\left(\frac{2}{5}\right)^n \rightarrow 0$ . Thus, as  $n$  grows to infinity,  $P(H_8|\text{data})$  approaches 1 and the posterior probability of all the other hypotheses goes to 0.

(b) Having observed  $n$  7's already, we know that we could not have selected the 4-sided or the 6-sided die. We have three different groups of numbers: we can roll 1 to 8 with all three remaining dice; 9 to 12 with the 12 and 20-sided dice; and 13 to 20 with only the 20-sided die. Thus, rolling 1 to 8 are all equally likely, likewise 9 to 12 and 13 to 20. Since we can get 1 to 8 from all three dice each of these values is in the most likely group. The next most likely values are 9 to 12 which can happen on two dice. Least likely values are 13 to 20.

(c) By the law of total probability, for  $x_{n+1} = 1, 2, \dots, 8$ , we have

$$P(x_{n+1}|\text{data}) = \frac{1}{5T} \cdot \left(\frac{1}{8}\right)^n \cdot \frac{1}{8} + \frac{1}{5T} \cdot \left(\frac{1}{12}\right)^n \cdot \frac{1}{12} + \frac{1}{5T} \cdot \left(\frac{1}{20}\right)^n \cdot \frac{1}{20}.$$

For  $x_{n+1} = 9, 10, 11, 12$ , we have

$$P(x_{n+1}|\text{data}) = \frac{1}{5T} \cdot \left(\frac{1}{12}\right)^n \cdot \frac{1}{12} + \frac{1}{5T} \cdot \left(\frac{1}{20}\right)^n \cdot \frac{1}{20}.$$

Finally, for  $x_{n+1} = 13, 14, \dots, 20$ , we have

$$P(x_{n+1}|\text{data}) = \frac{1}{5T} \cdot \left(\frac{1}{20}\right)^n \cdot \frac{1}{20}.$$

(d) As  $n \rightarrow \infty$ , we see that  $P(D_{n+1} = x|\text{data} = \text{all sevens}) \rightarrow 1/8$  for  $x = 1, 2, \dots, 8$ , and 0 for  $9 \leq x \leq 20$ .

**Problem 5.** (10 pts.) **Odds.**

(a) Odds of  $A$  are  $P(A)/P(A^c)$ . So both types of coin have odds 10/40.

(b) To answer parts b-d we make a likelihood table and a Bayesian update table. We label our hypothesis 0.3, 0.5 and 0.7 meaning the chosen has that probability of heads. Our

data from the first flip is  $\mathcal{D}_1$  is the event 'heads on the first flip'.

|            |     | outcomes |       |
|------------|-----|----------|-------|
|            |     | Heads    | Tails |
| hypotheses | 0.3 | 0.3      | 0.7   |
|            | 0.5 | 0.5      | 0.5   |
|            | 0.7 | 0.7      | 0.3   |

| Hypoth.<br>$\mathcal{H}$ | Prior<br>$P(\mathcal{H})$ | likelihood<br>$P(\mathcal{D}_1 \mathcal{H})$ | unnorm. post<br>$P(\mathcal{H})P(\mathcal{D}_1 \mathcal{H})$ | posterior<br>$P(\mathcal{H} \mathcal{D}_1)$ |
|--------------------------|---------------------------|--|--|---|
| 0.3                      | 0.2                       | 0.3  | 0.06   | 0.12  |
| 0.5                      | 0.6                       | 0.5  | 0.30   | 0.60  |
| 0.7                      | 0.2                       | 0.7  | 0.14   | 0.28  |
| Total:                   | 1                         | –  | 0.50   | 1   |

The prior probability of heads is just the total in the unnormalized posterior column:  $P(\text{heads}) = 0.50$

So the prior probability of tails is  $1 - P(\text{heads}) = 0.50$

So the prior odds of heads are  $O(\text{heads}) = 1$ , i.e. 50-50 odds.

(c) (i) From the table we see the posterior probability the coin is the 0.3 coin is 0.12 so

the posterior odds are  $\frac{0.12}{0.88} = \frac{12}{88} = 0.136$ .

(ii) Likewise the posterior odds it's the 0.7 coin are  $\frac{0.28}{0.72} = \frac{28}{72} = 0.389$ .

(d) The posterior predictive probability of heads is found by summing the product of the posterior column in the Bayesian update table and the heads column in the likelihood table. We get  $P(\text{heads}|\mathcal{D}_1) = .12 \cdot .3 + .60 \cdot .5 + .28 \cdot .7 = 0.532$ .

The posterior predictive probability of tails  $P(\text{tails}|\mathcal{D}_1) = 1 - 0.532 = 0.468$ . So the posterior

predictive odds of heads are  $O(\text{heads}|\mathcal{D}_1) = \frac{0.532}{0.468} = 1.1368$ .

**Problem 6.** (10 pts.) (a) The lawyer may correctly state that  $P(M|B) = 1/1000$ , but the lawyer then conflates this with the probability of guilt given all the relevant data, which is really  $P(M|B, K)$ . From Bayes' theorem, conditioning on  $B$  throughout, we have:

$$P(M|K, B) = \frac{P(K|M, B)P(M|B)}{P(K|B)} = \frac{P(M|B)}{P(K|B)}$$

since  $P(K|M, B) = 1$ . If we let  $N$  be the event that Mrs S was murdered by someone other than her husband, then  $M$  and  $N$  partition  $K$ , so and the odds of Mr S's guilt are

$$O(M|K, B) = \frac{P(M|B)}{P(N|B)} = \frac{P(\text{Mr S murdered Mrs S given that he beat her and she was killed})}{P(\text{Someone else murdered Mrs S given that Mr S beat her and she was killed})}$$

At this point, the relevant question is clear: is a battered wife more likely to be murdered by her husband or by someone else? I would guess the odds strongly favor the former.

(b) Here are four errors in the argument

18.05 Problem Set 5, Spring 2014 Solutions

1. The prosecutor arrived at “1 in 73 million” as follows: The probability that 1 child from an affluent non-smoking family dies of SIDS is  $1/8543$ , so the probability that 2 children die is  $(1/8543)^2$ . However, this assumes that the SIDS death among siblings are independent. Due to genetic or environmental factors, we suspect that this assumption is invalid.
2. The use of the figure “700,000 live births in Britain each year.” The prosecutor had restricted attention only to affluent non-smoking families when (erroneously) computing the probability of two SIDS deaths. However, he does not similarly restrict his attention when considering the number of births.
3. The rate “once every hundred years” is not valid: The prosecutor arrived at this by multiplying the number of live births by the probability that two children die from SIDS. The result is a non-sensical rate.
4. While double SIDS is very unlikely, double infanticide may be even more unlikely. It is the odds of one explanation relative to the other given the deaths that matters, and not just how unlikely one possibility is.

The Sally Clark case is an example of the “Prosecutor’s Fallacy.” You can read about it at [https://en.wikipedia.org/w/index.php?title=Sally\\_Clark&oldid=629645024](https://en.wikipedia.org/w/index.php?title=Sally_Clark&oldid=629645024)

There is also a video that discusses legal fallacies at

[http://www.ted.com/talks/peter\\_donnelly\\_shows\\_how\\_stats\\_fool\\_juries](http://www.ted.com/talks/peter_donnelly_shows_how_stats_fool_juries)

MIT OpenCourseWare  
<https://ocw.mit.edu>

18.05 Introduction to Probability and Statistics  
Spring 2014

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.