18.05 Problem Set 5, Spring 2014

Problem 1. (10 pts.) **Fitting a line to data using the MLE**.
Suppose you have bivariate data $(x_1, y_1), \ldots, (x_n, y_n)$. A common model is that there is a linear relationship between x and y, so in principle the data should lie exactly along a line. However since data has random noise and our model is probably not exact this will not be the case. What we can do is look for the line that best fits the data. To do this we will use a simple linear regression model.

For bivariate data the simple linear regression model assumes that the x_i are not random but that for some values of the parameters a and b the value y_i is drawn from the random variable

$$Y_i \sim ax_i + b + \varepsilon_i$$

where ε_i is a normal random variable with mean 0 and variance σ^2. We assume all of the random variables ε_i are independent.

Notes. 1. The model assumes that σ is a known constant, the same for each ε_i.
2. We think of ε_i as the measurement error, so the model says that

$$y_i = ax_i + b + \text{ random measurement error.}$$

3. Remember that (x_i, y_i) are not variables. They are values of data.

(a) The distribution of Y_i depends on a, b, σ and x_i. Of these only a and b are not known. Give the formula for the likelihood function $f(y_i \mid a, b, x_i, \sigma)$ corresponding to one random value y_i. (Hint: $y_i - ax_i - b \sim \text{N}(0, \sigma^2)$.)

(b) (i) Suppose we have data $(1, 8)$, $(3, 2)$, $(5, 1)$. Based on our model write down the likelihood and log likelihood as functions of a, b, and σ.

(ii) For general data $(x_1, y_1), \ldots, (x_n, y_n)$ give the likelihood and and log likelihood functions (again as functions of a, b, and σ).

(c) Assume σ is a constant, known value. For the data in part b(i) find the maximum likelihood estimates for a and b

(d) Use R to plot the data and the regression line you found in problem (1c). The commands `plot(x,y, pch=19)` and `abline()` will come in handy.

Print the plot and turn it in.

Problem 2. (10 pts.) **Estimating uniform parameters**
(a) Suppose we have data `1.2, 2.1, 1.3, 10.5, 5` which we know is drawn indepenedently from a uniform(a, b) distribution. Give the maximum likelihood estimate for the parameters a and b.

Hint: in this case you should not try to find the MLE by differentiating the likelihood function.

(b) Suppose we have data $x_1, x_2, \ldots, x_n$ which we know is drawn indepenedently from a uniform(a, b) distribution. Give the maximum likelihood estimate for the parameters a and b.

Problem 3. (10 pts.) **Monty Hall: Sober and drunk.**

Recall the Monty Hall problem: Monty hosts a game show. There are three doors: one hides a car and two hide goats. The contestant Shelby picks a door, which is not opened. Monty then opens another door which has nothing behind it. Finally, Shelby must decide whether to stay with her original choice or switch to the other unopened door. The problem asks which is the better strategy: staying or switching?

To be precise, let's label the door that Shelby picks first by A, and the other two doors by B and C. Hypothesis A is that the car is behind door A, and similarly for hypotheses B and C.

(a) In the usual formulation, Monty is sober and knows the locations of the car and goats. So if the contestant picks a door with a goat, Monty always opens the other door with a goat. And if the contestant picks the door with a car, Monty opens one of the other two doors at random. Suppose that sober Monty Hall opens door B, revealing a goat (this is the data). Make a Bayes' table with prior, likelihood and posterior. Use the posterior probabilities to determine the best strategy.

(b) Now suppose that Monty is drunk, i.e. he has completely forgotten where the car is and is only aware enough to open one of the doors not chosen by the contestant at random. It's entirely possible he might accidentally reveal the car, ruining the show.

Suppose that drunk Monty Hall opens door B, revealing a goat. Make a Bayes' table with prior, likelihood and posterior. Use the posterior probabilities to determine the best strategy. (Hint: the data is the same but the likelihood function is not.)

(c) Based on Monty's pre-show behavior, Shelby thinks that Monty is sober with probability .7 and drunk with probability .3. Repeat the analysis from parts (a) and (b) in this situation.

Problem 4. (10 pts.) We are going to explore the dice problem from class further. I pick one of the five dice uniformly at random (4, 6, 8, 12, or 20 sides). I then roll this die n times and tell you that, miraculously, every roll resulted in the value 7. As I am in a position of authority, assume that I am telling the truth!

We write the data as $x_1 = 7$, $x_2 = 7$, $\ldots x_n = 7$, where x_i is the result of the ith roll.

(a) Find the posterior probability $P(H|\text{data})$ for each die given the data of all n rolls (your answers should involve n). What is the limit of each of these probabilities as n grows to infinity? Explain why this makes sense.

(b) Given that my first 10 rolls resulted in 7 (i.e., $n = 10$), rank the possible values for my next roll from most likely to least likely. Note any ties in rank and explain your reasoning carefully. You need not do any computations to solve this problem.

(c) Find the posterior predictive pmf for the $(n + 1)^{\text{st}}$ roll given the data. That is, find $P(x_{n+1}|x_1 = 7, \cdots, x_n = 7)$ for $x_{n+1} = 1, \ldots, 20$. (Hint: use (a) and the law of total probability. Many values of the pmf coincide, so you do not need to do 20 separate computations. You should check that your answer is consistent with your ranking in (b) for $n = 10$).

(d) What function does the pmf in part (c) converge to as n grows to infinity? Explain why this makes sense.

Problem 5. (10 pts.) **Odds.**
You have a drawer that contains 50 coins. 10 coins have probability $p = 0.3$ of heads, 30

coins have probability $p = 0.5$ and 10 coins have probability $p = 0.7$. You pick one coin at random from the drawer and flip it.

(a) What are the (prior) odds you chose a 0.3 coin? A 0.7 coin?

(b) What are the (prior) odds of flipping a heads?

(c) Suppose the flip lands heads.

   (i) What are the posterior odds the coin is a 0.3 coin?

   (ii) A 0.7 coin?

(d) What are the posterior predictive odds of heads on the next (second) flip?

Problem 6. (10 pts.) **Courtroom fallacies**.
(a) [Mackay, *Information Theory, Inference, and Learning Algorithms*, and OJ Simpson trial] Mrs S is found stabbed in her family garden. Mr S behaves strangely after her death and is considered as a suspect. On investigation of police and social records it is found that Mr S had beaten up his wife on at least nine previous occasions. The prosecution advances this data as evidence in favor of the hypothesis that Mr S is guilty of the murder. 'Ah no,' says Mr S's highly paid lawyer, '*statistically*, only one in a thousand wife-beaters actually goes on to murder his wife. So the wife-beating is not strong evidence at all. In fact, given the wife beating evidence alone, it's extremely *unlikely* that he would be the murderer of his wife – only a 1/1000 chance. You should therefore find him innocent.'

Is the lawyer right to imply that the history of wife-beating does not point to Mr S's being the murderer? Or is the lawyer a slimy trickster? If the latter, what is wrong with his argument?

Use the following scaffolding to reason precisely:
Hypothesis: M = 'Mr S murdered Mrs S'
Data: K = 'Mrs S was killed',   B = 'Mr S had a history of beating Mrs S'
How is the above probability 1/1000 expressed in these terms? How is the (posterior) probability of guilt expressed in these terms? How are these two probabilities related? Hint: Bayes' theorem, conditioning on K throughout.

b) [True story] In 1999 in Great Britain, Sally Clark was convicted of murdering her two sons after each child died weeks after birth (the first in 1996, the second in 1998). Her conviction was largely based on the testimony of the pediatrician Professor Sir Roy Meadow. He claimed that, for an affluent non-smoking family like the Clarks, the probability of a single cot death (SIDS) was 1 in 8543, so the probability of two cot deaths in the same family was around "1 in 73 million." Given that there are around 700,000 live births in Britain each year, Meadow argued that a double cot death would be expected to occur once every hundred years. Finally, he reasoned that given this vanishingly small rate, the far more likely scenario is that Sally Clark murdered her children.

Carefully explain at least two errors in Meadow's argument.

18.05 Introduction to Probability and Statistics

Spring 2014