18.05 Problem Set 6, Spring 2014

Problem 1. (10 pts.) **Beta try again.** (Adapted from 'Information Theory, Inference, and Learning Algorithms' by David J. C. Mackay.)

A statistical statement appeared in *The Guardian* on Friday January 4, 2002:

> When spun on edge 250 times, a Belgian one-euro coin came up heads 140 times and tails 110. 'It looks very suspicious to me', said Barry Blight, a statistics lecturer at the London School of Economics. 'If the coin were unbiased the chance of getting a result as extreme as that would be less than 7%'.

The figure 7% is a frequentist p-value. We will study that in the next unit and return to this problem. Here we will do a Bayesian analysis of the data.

In class and in the notes we considered the odds and Bayes factor for a hypothesis H vs. H^c. Given data D we have the posterior odds of H vs. H^c are

$$\text{odds} = \frac{P(H|D)}{P(H^c|D)} = \frac{P(D|H)P(H)}{P(D|H^c)P(H^c)}$$

The factor $\frac{P(D|H)}{P(D|H^c)}$ is the *Bayes factor* and the factor $\frac{P(H)}{P(H^c)}$ is the *prior odds*. We have the same definitions for comparing any two hypotheses H_1 vs. H_2

$$\text{odds} = \frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)P(H_1)}{P(D|H_2)P(H_2)}; \quad \text{Bayes factor} = \frac{P(D|H_1)}{P(D|H_2)}; \quad \text{prior odds} = \frac{P(H_1)}{P(H_2)}.$$

(a) Consider the hypotheses H_0 = 'probability of heads is 1/2' and H_1 = 'probability of heads is $\frac{140}{250}$'. Using the data of 140 heads and 110 tails, what is the Bayes factor for H_0 versus H_1?

It shouldn't surprise you that the Bayes factor is much less than 1. Give a short explanation of why you would expect this.

(b) Let θ be the probability that the spun coin comes up heads. Below are five priors, use R to graph them all on a single plot using a different color for each prior. Sample R plotting code is on our website in the file sampleplotting.r

Which of the priors do you find most reasonable and why? There is no one right answer.

i) $f(\theta)$ is Beta(1,1), or equivalently uniform(0, 1).

ii) $f(\theta)$ is Beta(10, 10).

iii) $f(\theta)$ is Beta(50, 50).

iv) $f(\theta)$ is Beta(500, 500).

v) $f(\theta)$ is Beta(30, 70).

(c) Without computing, rank these priors according to the prior probability that the coin is biased toward heads (from highest to lowest). Explain your ranking, including any ties.

(d) Let x be the data of 140 heads and 110 tails. The posterior distribution $f(\theta|x)$ depends on our choice for prior distribution $f(\theta)$. Find the posterior for each of the above priors and graph all them on a single plot using the same colors as in part (b).

Considering the mean and shape of each of the priors, give a short explanation for the positions of each of the posteriors in in graph.

(e) Use the R function `pbeta(x,a,b)` to compute the posterior probability of a bias in favor of heads for each of the priors. Rank the posteriors from most biased to least biased towards heads. Is this consistent with your plot in part (d)?

(f) Note that $140/250 = .56$. Let $H_1 = $ '$.55 \leq \theta \leq .57$' and let $H_0 = $ '$.49 \leq \theta \leq .51$'. Use the following steps to *estimate* the posterior odds of H_1 versus H_0, first for prior (i) and then for prior (iv).

Step 1. Use the posterior densities and the fact that H_0 and H_1 cover small intervals to estimate the posterior probabilities of H_0 and H_1. (No integral is needed. Leave the normalizing constant uncomputed.)

Step 2. Use the posterior probabilities to give the posterior odds.

Explain the difference between the results for priors (i) and (iv).

Problem 2. (10 pts.) **Bayes at the movies.**: A local theater employs two ticket collectors, Alice and Bob, although only one of them works on any given day. The number of tickets X that a ticket collector can collect in an hour is modeled by a distribution which has mean λ, and probability mass function

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

for $k = 0, 1, 2, \ldots$. (This distribution is called a *Poisson distribution*. It is an important *discrete* distribution in biology and physics.)

Suppose that Alice collects, on average, 10 tickets an hour and Bob collects, on average, 15 tickets an hour. One day the manager stays home sick. He knows Bob is supposed to work that day, but thinks there are 1 to 10 odds that Alice is filling in for Bob (based on Bob's prior history of taking advantage of Alice's generous nature when the manager is away). The suspicious manager monitors ticket sales online and observes that over the span of 5 hours there are 12, 10, 11, 4, and 11 tickets collected. What are the manager's posterior odds that Alice is filling in for Bob?

Problem 3. (10 pts.) **Uniformly late.** (Continuous x, continuous θ.)
When Jane started 18.05, she warned Jon that she tends to run late. Not just late, but uniformly late. That is, Jane will be late to a random class by X hours, uniformly distributed over the interval $[0, \theta]$. Since the class is one hour long (fractions be banished!), Jon decides to model his initial belief about the unknown parameter θ by a uniform density from zero to one hour.

(a) Jane was x hours late to the first class. How should Jon update his belief about θ? In other words, find the posterior $f(\theta|x)$. Roughly sketch the result for $x = .2$ and for $x = .5$.

Hints:
1. Remember that x is a fixed number, e.g. .4.
2. The likelihood will be 0 if $\theta < x$.

(b) Now suppose that after $n \geq 2$ classes, Jane has been late by amounts $x_1, \ldots, x_n$ (assume these times are conditionally independent). Find the posterior $f(\theta|x_1, \ldots, x_n)$.

Roughly sketch the posterior for the data (.1, .5), the data (.5, .5), the data (.1, .2, .3, .4, .5). Hint: Some of these will coincide.

(c) Suppose that the latest Jane arrived during the first 5 classes was 30 minutes. Find the posterior predictive probability that Jane will arrive less than 30 minutes late to the next class.

(d) The *Maximum a Posteriori Probability* (MAP) estimate for θ is the value of θ that maximizes the posterior distribution $f(\theta|x)$. For a single value x (as in part (a)) find the MAP estimate for θ. (This should be obvious from your graphs.)

(e) Extra credit: 5 points
(i) The MAP is not always the best point estimate of θ. Another estimate is the *Conditional Expectation*. This is the expected value of the posterior distribution $f(\theta|x)$. Using the posterior pdf in part (a) find the Conditional Expectation estimate for θ.

(ii) Graph the estimates in (d) and (e(i)) for $0 \leq x \leq 1$. How do these estimates compare? Which do you think is most likely to be closer to the true value of θ?

Problem 4. (10 pts.) **Normal is the new normal.** Your friend transmits an unknown value θ to you over a noisy channel. The noise is normally distributed with mean 0 and a known variance 4. so the value x that you receive is modeled by $N(\theta, 4)$. Based on previous communications, your prior on θ is $N(5, 9)$.

(a) Suppose your friend transmits a value to you that you receive as $x = 6$. Show that the posterior pdf for θ is $N(74/13, 36/13)$. The reading gave formulas for normal-normal updating, but for this problem carry out the calculations from scratch using our usual Bayesian update table and a good dose of algebra.

(b) Suppose your friend transmits the same value θ to you n times. You receive these signals plus noise as $x_1, \ldots, x_n$ with sample mean $\bar{x}$. In the reading we gave normal-normal Bayesian update formulas which apply in this case. For this problem you can use these formulas.

$$a = \frac{1}{\sigma_{\text{prior}}^2}, \qquad b = \frac{n}{\sigma^2}, \qquad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a + b}, \qquad \sigma_{\text{post}}^2 = \frac{1}{a + b}. \qquad (1)$$

Suppose your friend transmits the same value θ to you 4 times. You receive these signals plus noise as $x_1, \ldots, x_4$ with sample mean $\bar{x} = 6$. Using the same prior and know variance σ^2 as in part (a), show that the posterior on θ is $N(5.9, 0.9)$. Plot the prior and posterior on the same graph. Describe how the data changes your belief about the true value of θ.

(c) How do the mean and variance of the posterior change as more data is received? What is gained by sending the same signal multiple times? Here we want you to interpret the equations (1).

(d) IQ in the general population follows a $N(100, 15^2)$ distribution. An IQ test is unbiased with a known normal variance of 10^2; that is, if the same person is tested multiple times, their measured IQ will differ from their true IQ according to a normal distribution with mean 0 and variance 100.

i) Randall Vard scored an 80 on the test. What is the expected value of his true IQ?

ii) Mary I. Taft scored a 150 on the test. What is the expected value of her true IQ?

(e) Extra credit: 5 points. Prove the formulas in equation (1).

Problem 5. (10 pts.) **Censored data.** Sometimes data is not reported in full. This can mean only reporting values in a certain range or not reporting exact values. We call such data *censored*.

We have a 4-sided die and a 6-sided die. One of them is chosen at random and rolled 5 times. Instead of reporting the number of spots on a roll we censor the data and just report

1 if the roll is a 1;   0 if the roll is not a 1.

Suppose the data for the five rolls is 1, 0, 1, 1, 1. Starting from a flat prior on the choice of die, update in sequence and report, after each roll, the posterior odds that the chosen die is 4-sided die.

(b) A censored value of 1 is evidence in favor of which die? What about 0? How is this reflected in the posterior odds in part (a)

18.05 Introduction to Probability and Statistics

Spring 2014