18.05 Problem Set 6, Spring 2014 Solutions

Problem 1. (10 pts.) **(a)** Throughout this problem we will let x be the data of 140 heads out of 250 tosses. We have $140/250 = .56$. Computing the likelihoods:

$$p(x|H_0) = \binom{250}{140}(.5)^{250} \qquad p(x|H_1) = \binom{250}{140}(.56)^{140}(.44)^{110}$$
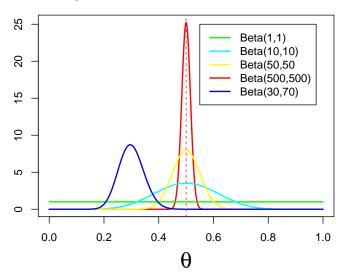
which yields Bayes factor

$$\frac{p(x|H_0)}{p(x|H_1)} = \frac{(.5)^{250}}{(.56)^{140}(.44)^{110}} = 0.16458,$$

(Actually, we computed the log Bayes factor since it is numerically more stable. Then we exponentiated to get the Bayes factor.)

Since we chose the probability $140/250$ of H_1 to exactly match the data it is not surprising that the probability of the data given H_1 is much greater than the probability given H_0. Said differently, the data will pull our prior towards one centered at $140/250$.

(b) Here are the plots of the five priors. The vertical dashed red line is at $\theta = 0.5$. The R code is posted alongside these solutions.

A priori I would want my prior centered at 0.5. This rules out Beta(30,70). Beta(500,500) seems too narrow. Beta(1,1) doesn't really match my experience with coins, but I might go with it and just let the data speak for itself. Both Beta(10,10) and Beta(50,50) seem plausible. Even if they're wrong they aren't so strong that they would cause us to ignore the evidence in the data.

(c) The prior probability of a bias in favor of heads is $P(\theta > 1/2)$. Looking at the plots of the prior pdf's in part (b) we see that (i)-(iv) are symmetric about .5, therefore they predict the probability of heads is $1/2$. That is they are all unbiased. (v) has most of it's probability below .5. So it is strongly biased against heads. Thus,

the ranking in order of bias from least to greatest is (v) followed by a four-way tie between (i)-(iv).

(d)   All of the prior pdf's are beta distributions, so they have the form

$$f(\theta) = c_1 \theta^{a-1}(1-\theta)^{b-1}.$$

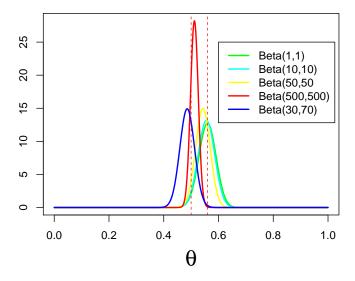For a fixed hypothesis θ the likelihood function (given the data x) is

$$p(x|\theta) = \binom{250}{140}\theta^{140}(1-\theta)^{110}.$$

Thus the posterior pdf is

$$f(\theta|x) = c_2 \theta^{140+a-1}(1-\theta)^{110+b-1} \sim \text{beta}(140+a, 110+b).$$

So the five posterior distributions (i)-(v) are beta(141, 111), beta(150, 120), beta(190, 160), beta(640, 610), and beta(170, 180).

Here are the plots of the five posteriors.

Each prior is centered on a value of θ. The sharpness of the peak is a measure of the prior 'commitment' to this value. So prior (iv) is strongly committed to $\theta = .5$, but prior (ii) is only weakly committed and (i) is essentially uncommitted. The effect of the data is to pull the center of the prior towards the data mean of .56. That is, it averages the center of the prior and the data mean. The stronger the prior belief the less the data pulls the center towards .56. So prior (iv) is only moved a little and prior (i) is moved almost all the way to .56. Priors (ii) and (iii) are intermediate. Prior (v) is centered at $\theta = .3$. The data moves the center a long way towards .56. But, since it starts so much farther from .56 than the other priors, the posterior is still centered the farthest from .56.

(e) For each of the five posterior distributions, we compute $p(\theta \geq 0.5|x)$:

$$P_{(i)}(\theta > 0.5|x) = 1\text{- pbeta(.5, 141,111)} = 0.9710$$
$$P_{(ii)}(\theta > 0.5|x) = 1\text{- pbeta(.5, 150,120)} = 0.9664$$
$$P_{(iii)}(\theta > 0.5|x) = 1\text{- pbeta(.5, 190,160)} = 0.9459$$
$$P_{(iv)}(\theta > 0.5|x) = 1\text{- pbeta(.5, 640,610)} = 0.8020$$
$$P_{(v)}(\theta > 0.5|x) = 1\text{- pbeta(.5, 170,180)} = 0.2963$$

This is consistent with the plot in d), as the posterior computed from the uniform prior has the most density past 0.5 while the posterior computed from prior (v) has the least.

(f) Step 1. Since the intervals are small we can use the relation

$$\text{probability} \approx \text{density} \cdot \Delta\theta.$$

So

$$P_{(i)}(H_0|x) = P_{(i)}(0.49 \leq \theta \leq 0.51|x) \approx \cdot f_{(i)}(0.5|x) \cdot .2$$

and

$$P_{(i)}(H_1|x)P_i(0.55 \leq \theta \leq 0.57) \approx \cdot f_{(i)}(0.56) \cdot .2.$$

So the the posterior odds (using prior (i)) of H_1 versus H_0 are approximately

$$\frac{P_{(i)}(H_1|x)}{P_{(i)}(H_0|x)} \approx \frac{f_{(i)}(0.56)}{f_{(i)}(0.5)} = \frac{\text{dbeta(.56,141,111)}}{\text{dbeta(.5,141,111)}} = \frac{c(0.56)^{140}(0.44)^{110}}{c(0.5)^{140}(0.5)^{110}} \approx 6.07599$$

By similar reasoning, the posterior odds (using prior (iv)) of H_1 versus H_0 is approximately 0.00437.

Problem 2. (10 pts.) Let A be the event that Alice is collecting tickets and B the event that Bob is collecting tickets. Denoting our data as D, we have the likelihoods

$$P(D|A) = \frac{10^{12+10+11+4+11}e^{-50}}{12!10!11!4!11!}$$
$$P(D|B) = \frac{15^{12+10+11+4+11}e^{-75}}{12!10!11!4!11!}.$$

Moreover, we are given prior odds, $O(A) = \frac{P(A)}{P(B)} = \frac{1}{10}$. Thus, our posterior odds are

$$O(A|D) = \frac{P(D|A)}{P(D|B)}O(A) = \left(\frac{10}{15}\right)^{48}e^{25} \cdot \frac{1}{10} \approx 25.408$$

Note that the Bayes factor is about 250.

Problem 3. (10 pts.) **(a)** We have a flat prior pdf $f(\theta) = 1$. For a single data value x, our likelihood function is:

$$f(x|\theta) = \begin{cases} 0 & \text{if } \theta < x \\ \frac{1}{\theta} & \text{if } x \leq \theta \leq 1 \end{cases}$$
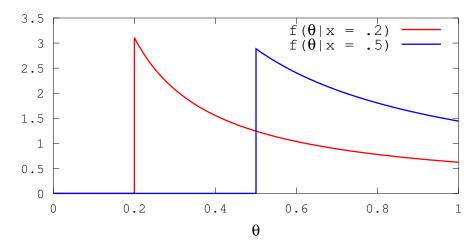
Thus our table is

hyp.	prior $f(\theta)$	likelihood $f(x|\theta)$	unnormalized posterior	posterior $f(\theta|x)$
$\theta < x$	$d\theta$	0	0	0
$x \leq \theta \leq 1$	$d\theta$	$\frac{1}{\theta}$	$\frac{d\theta}{\theta}$	$\frac{c}{\theta}\,d\theta$
Tot.	1		T	1

The normalizing constant c must make the total posterior probability 1, so

$$c \int_x^1 \frac{d\theta}{\theta} = 1 \;\Rightarrow\; c = -\frac{1}{\ln(x)}.$$

Note that since $x \leq 1$, we have $c = -1/\ln(x) > 0$.

Here are plots for $x = .2$ $(c = 0.621)$ and $x = .5$ $(c = 1.443)$.

(b) Notice that θ cannot be less than any one of $x_1, \ldots, x_n$. So the likelihood function is given by

$$f(x_1, \ldots, x_n|\theta) = \begin{cases} 0 & \text{if } \theta < \max\{x_1, \ldots, x_n\} \\ \frac{1}{\theta^n} & \text{if } \max\{x_1, \ldots, x_n\} \leq \theta \leq 1. \end{cases}$$
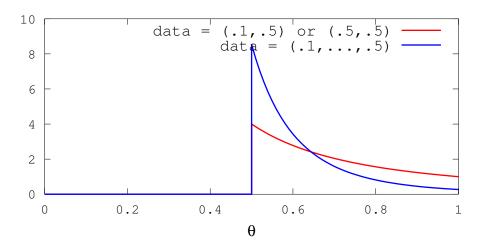
Let $x_M = \max\{x_1, \ldots, x_n\}$. So our table is

hyp.	prior $f(\theta)$	likelihood $f(\text{data}|\theta)$	unnormalized posterior	posterior $f(\theta|\text{data})$
$\theta < x_M$	$d\theta$	0	0	0
$x_M \leq \theta \leq 1$	$d\theta$	$\frac{1}{\theta^n}$	$\frac{d\theta}{\theta^n}$	$\frac{c}{\theta^n}\,d\theta$
Tot.	1		T	1

The normalizing constant c must make the total posterior probability 1, so

$$c \int_{x_M}^1 \frac{d\theta}{\theta^n} = 1 \;\Rightarrow\; c = \frac{n-1}{x_M^{1-n} - 1}.$$

The posterior pdf depends only on n and x_M, therefore the data $(.1,.5)$ and $(.5,.5)$ have the same posteriors. Here are the plots of the posteriors for the given data.

(c) We now have $x_M = 0.5$ so from part (b) the posterior density is

$$f(\theta|x_1, \ldots, x_5) = \begin{cases} 0 & \text{if } \theta < .5 \\ \frac{c}{\theta^5} & \text{if } .5 \le \theta \le 1, \end{cases}$$

where $c = \frac{4}{.5^{-4}-1} = \frac{4}{15}$.

Now, let x be amount by which Jane is late for the sixth class. The likelihood is

$$f(x|\theta) = \begin{cases} 0 & \text{if } \theta < x \\ \frac{1}{\theta} & \text{if } \theta \ge x \end{cases}$$

We have the posterior predictive probability

$$f(x|x_1, \ldots, x_5) = \int f(x|\theta) f(\theta|x_1, \ldots, x_5)\, d\theta$$

$$= \begin{cases} \int_{.5}^{1} \frac{1}{\theta} \cdot \frac{4}{15\theta^5}\, d\theta = -\frac{4}{75}\theta^{-5}\big|_{.5}^{1} = \frac{124}{75} & \text{if } 0 \le x < .5 \\ \int_{x}^{1} \frac{1}{\theta} \cdot \frac{4}{15\theta^5}\, d\theta = -\frac{4}{75}\theta^{-5}\big|_{x}^{1} = \frac{4}{75}(x^{-5} - 1) & \text{if } .5 \le x \le 1 \end{cases}$$

Thus the posterior predictive probability that $x \le 0.5$ is

$$P(x \le .5|x_1, \ldots, x_5) = \int_0^{0.5} f(x|x_1, \ldots, x_n)\, dx = \int_0^{.5} \frac{124}{75}\, dx = \frac{62}{75} = \boxed{0.82667}.$$
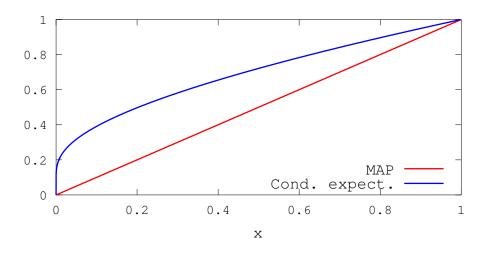
(d) The graphs or the formula in part (a) show that $f(\theta|x)$ is decreasing for $\theta \ge x$, so the MAP is when $\theta = x$.

(e) Extra credit: 5 points
(i) From part (a) we have the posterior pdf $f(\theta|x)$. The conditional expectation is

$$E(\theta|x) = \int \theta f(\theta|x)\, d\theta = \int_x^1 \theta \cdot \frac{c}{\theta} = c(1 - x) = \boxed{-\frac{1-x}{\ln(x)}}.$$

(ii) After observing x, we know that $\theta \geq x$, and as a result the conditional expectation $E[\theta|x] \geq x$. So the conditional expectation estimator is always at least as big as the MAP estimator. However, the MAP estimator is precisely x, the amount by which Jane is late on the first class. In this context, the MAP is not reasonable as it suggests that on the first class, Jane arrived as late as possible and that in the future, she will arrive less than x hours late.

Problem 4. (10 pts.) **(a)** Leaving the scale factors as letters our table is

	prior $f(\theta) \sim N(5,9)$	likelihood $f(x|\theta) \sim N(\theta, 4)$	posterior $f(\theta|x)$
hyp.			
θ	$c_1 e^{-(\theta-5)^2/18}\, d\theta$	$c_2 e^{(6-\theta)^2/8}$	$c\exp\left(-\dfrac{(\theta-5)^2}{18} - \dfrac{(6-\theta)^2}{8}\right)$
Tot.	1		1

All we need is some algebraic manipulations of the exponent in the posterior:

$$-\frac{(\theta-5)^2}{18} - \frac{(6-\theta)^2}{8} = -\frac{1}{2}\left(\frac{\theta^2 - 12\theta + 36}{4} + \frac{\theta^2 - 10\theta + 25}{9}\right)$$

$$= -\frac{1}{2}\left(\frac{13\theta^2 - 148\theta + 424}{36}\right)$$

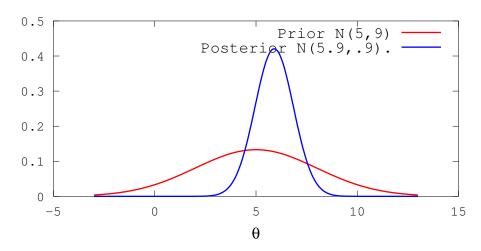$$= -\frac{1}{2}\left(\frac{(\theta - 74/13)^2}{36/13} + k\right)$$

where k is a constant. Thus the posterior

$$f(\theta|x) \; \propto \; \exp\left(-\frac{(\theta - 74/13)^2}{2 \cdot 36/13}\right)$$

This has the form of a pdf for $N\left(\frac{74}{13}, \frac{36}{13}\right)$.   QED

(b)   We have $\mu_{prior} = 5, \sigma^2_{prior} = 9, \bar{x} = 6, \sigma^2 = 4, n = 4$ So we have

$$a = \frac{1}{9}, \quad b = 1, \quad a + b = \frac{10}{9} \Rightarrow \mu_{post} = \frac{5/9 + 6}{10/9} = 5.9, \quad \sigma^2_{post} = \frac{1}{10/9} = .9.$$

After observing $x_1, \ldots, x_4$, we see that the posterior mean is close to $\bar{x}$ and the posterior variance is much smaller than the prior variance. The data has made us more certain about the location of θ.

(c)   As more data is received n increases, so b increases, so the mean of the posterior is closer to the data mean and the variance of the posterior decreases. Since the variance goes down, we gain more certainty about the true value of θ.

(d)   With no new data we are given the prior $f(\theta) \sim N(100, 15^2)$. For data $x = $ score on the IQ test we have the likelihood $f(x|\theta) \sim N(\theta, 10^2)$. Using the update formulas we have $\mu_{\text{prior}} = 100$, $\sigma^2_{\text{prior}} = 15^2$, $\sigma^2 = 10^2$, $n = 1$. So $a = 1/225$, $b = 1/100$ and

(i) Randall, $x = 80$: $\mu_{\text{post}} = \dfrac{a \cdot 100 + b \cdot 80}{a + b} = 86.15$

(ii) Mary, $x = 150$: $\mu_{\text{post}} = \dfrac{a \cdot 100 + b \cdot 150}{a + b} = 134.62$

Regression towards the mean!

(e) Extra credit: 5 points.   This is essentially the same manipulation as in part (a). First suppose we have one data value x_1 then

	prior $f(\theta) \sim N(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$	likelihood $f(x_1|\theta) \sim N(\theta, \sigma^2)$	posterior $f(\theta|x_1)$
hyp.			
θ	$c_1 e^{-(\theta - \mu_{\text{prior}})^2/2\sigma^2_{\text{prior}}} \, d\theta$	$c_2 e^{(x_1 - \theta)^2/2\sigma^2}$	$c \exp\left(-\dfrac{(\theta - \mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}} - \dfrac{(x_1 - \theta)^2}{2\sigma^2}\right)$
Tot.	1		1

All we need is some algebraic manipulations of the exponent in the posterior. When-

ever we get a term not involving θ we just absorb it into the constant k_1, k_2 etc.

$$-\frac{(\theta - \mu_{\text{prior}})^2}{2\sigma_{\text{prior}}^2} - \frac{(x_1 - \theta)^2}{2\sigma^2} = -\frac{1}{2}\left(\frac{\theta^2 - 2\mu_{\text{prior}}\theta + \mu_{\text{prior}}^2}{\sigma_{\text{prior}}^2} + \frac{\theta^2 - 2x_1\theta + x_1^2}{\sigma^2}\right)$$

$$= -\frac{1}{2}\left((a+b)\theta^2 - 2(a\mu_{\text{prior}} + bx_1)\theta + k_1\right)$$

$$= -\frac{1}{2}\frac{\left(\theta - \frac{a\mu_{\text{prior}}+bx_1}{a+b}\right)^2}{\frac{1}{a+b}} + k_2$$

Thus the posterior

$$f(\theta|x_1) \propto \exp\left(-\frac{1}{2}\frac{\left(\theta - \frac{a\mu_{\text{prior}}+bx_1}{a+b}\right)^2}{\frac{1}{a+b}}\right)$$

This has the form of a pdf for $N\left(\frac{a\mu_{\text{prior}}+bx_1}{a+b}, \frac{1}{a+b}\right)$. This proves the formulas (1) when $n = 1$. The formulas when $n > 1$ are a simple consequence of updating one data point at a time using the formulas when $n = 1$.

Problem 5. (10 pts.) **Censored data.** We note that we assume that, given a particular dice, the rolls are independent. Let x be the censored value on one roll. The Bayes factor for x is

$$\text{Bayes factor} = \frac{p(x|4\text{-sided})}{p(x|6\text{-sided})} = \begin{cases} \frac{3/4}{5/6} = 9/10 & \text{if } x = 0 \\ \frac{1/4}{1/6} = 3/2 & \text{if } x = 1 \end{cases}$$

Starting from the prior odds of 1, we multiply by the appropriate Bayes factor and get the posterior odds after rolls 1-5 are

$$\frac{3}{2} = 1.5, \quad \frac{27}{20} = 1.35, \quad \frac{81}{40} = 2.025, \quad \frac{243}{80} = 3.0375, \quad \frac{729}{160} = 4.5562$$

(b) In part (a) we saw the Bayes factor when $x = 1$ is $3/2$. Since this is more than 1 it is evidence in favor of the 4-sided die. When $x = 0$ the Bayes factor is $9/10$, which is evidence in favor of the 6-sided die.

We saw this in part (a) because after every value of 1 the odds for the 4-sided die went up and after the value of 0 the odds went down.

18.05 Introduction to Probability and Statistics

Spring 2014