Problem Set 6, Spring 2018 Solutions

Problem 1. (20 pts.) Hypotheses and data.
(a) (i) The possible hypotheses are that the coin is fair or has probability 0.4 of landing heads.
(ii) The data \( x \) is the result of the experiment: toss the chosen coin 3 times and count the number of heads.
(iii) Let \( H_{0.5} \) be the hypothesis the coin is fair and \( H_{0.4} \) the hypothesis is has probability 0.4 of landing heads. We give the prior and likelihood in tabular form:

| Hypothesis: \( H \) | prior: \( p(H) \) | likelihood: \( p(x|H) \) |
|----------------------|-----------------|------------------|
| \( H_{0.5} \)       | 1/4             | 3/8 = 0.375      |
| \( H_{0.4} \)       | 3/4             | 3(0.4)^2(0.6) = 0.288 |

(b) (i) The unknown value is the probability of success with a random patient. We hypothesize this value: the possible hypotheses for \( \theta \) are any value between 0 and 1.
(ii) The data \( x \) is the number of successes out of the number of patients tested.
(iii) Call the unknown probability of success \( \theta \), the number of patients tested \( n \) and the number of successes \( x \). We use the same letter \( \theta \) to indicate the hypothesis. Having no belief about the effectiveness translates to a uniform prior. The likelihoods come from a binomial distribution.

| Hypothesis: \( \theta \) | prior: \( f(\theta)d\theta \) | likelihood: \( p(x|\theta,n) \) |
|--------------------------|-----------------------------|------------------|
| \( \theta \)             | \( d\theta \)              | \( \binom{n}{x}\theta^x(1-\theta)^{n-x} \) |

(c) The answers to (i) and (ii) are the same as in part (b). The data is the number of successes out of 30 patients.
(iii) There are many reasonable priors. One type of reasonable prior for \( \theta \) is a beta distribution with mean 0.75 since this matches the mean of the previous drug in the same class. The mean of beta(\( a, b \)) is \( a/(a+b) \), so we could choose beta(3\( m, m \)) where \( m \) is any positive value. The bigger \( m \) is, the more focused the prior will be around \( \theta = 0.75 \).

| Hypothesis: \( \theta \) | prior: \( f(\theta)d\theta \) | likelihood: \( p(x|\theta,30) \) |
|--------------------------|-----------------------------|------------------|
| \( \theta \)             | \( c \theta^{3m-1} \theta^{m-1}d\theta \) | \( \binom{30}{x}\theta^x(1-\theta)^{30-x} \) |

Here, \( c \) is the normalizing coefficient for the beta distribution.

d (i) The unknown value is \( \lambda \) the rate constant in the exponential distribution.
We hypothesize this value: the possible hypotheses for \( \lambda \) are any positive value.
(ii) The experiment is to give the drug to a patient and time how long it takes before the drug is not detected. The data \( x \) is this length of time.
(iii) We are given the prior explicitly:
Hypothesis: $\lambda$

| prior: $f(\lambda)d\lambda$ | likelihood: $\phi(x|\lambda)$ |
|-----------------------------|-------------------------------|
| $\frac{1}{\sqrt{2\pi \sigma_0}} e^{-\frac{(\lambda-\theta)^2}{2\sigma_0^2}}$ | $\lambda e^{-\lambda x}$ |

Problem 2. (30 pts.) Spun gold.

(a) Exact formula: $\sum_{k=140}^{250} \binom{250}{k} (0.5)^{250}$.

R calculation: $1 - \text{pbinom}(139, 250, 0.5)$. Result: 0.0332105756200217

(b) Throughout this problem we will let $x$ be the data of 140 heads out of 250 tosses. We have $140/250 = 0.56$. Computing the likelihoods:

$$p(x|H_0) = \binom{250}{140} (0.5)^{250}$$
$$p(x|H_1) = \binom{250}{140} (0.56)^{140}(0.44)^{110}$$

which yields Bayes factor

$$\frac{p(x|H_0)}{p(x|H_1)} = \frac{(0.5)^{250}}{(0.56)^{140}(0.44)^{110}} = 0.16458,$$

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 $140/250$.

(c) 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.

![Prior Plots](image)

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.

(d) 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 0.5.
Therefore they all predict the total probability of bias toward heads (that is, $\theta > 1/2$) is $1/2$. That is all the priors are unbiased. (v) has most of its probability below 0.5. So it is strongly biased against $\theta$ > $1/2$. Thus, the ranking in order of probability of bias toward heads from least to greatest is (v) followed by a four-way tie among (i)-(iv).

(e) This is a situation with conjugate priors: 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 $\theta$ 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 $\theta$. The sharpness of the peak is a measure of the prior ‘commitment’ to this value. So prior (iv) is strongly committed to $\theta = 0.5$, 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 0.56. 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 0.56. So prior (iv) is only moved a little and prior (i) is moved almost all the way to 0.56. Priors (ii) and (iii) are intermediate. Prior (v) is centered at $\theta = 0.3$. The data moves the center a long way towards 0.56.
But, since it starts so much farther from 0.56 than the other priors, the posterior is still centered the farthest from 0.56.

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

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

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.

(g)

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

\[
\text{probability} \approx \text{density} \cdot \Delta \theta.
\]

So

\[
P_{(i)}(H_0|x) = P_{(i)}(0.49 \leq \theta \leq 0.51|x) \approx f_{(i)}(0.5|x) \cdot 0.02
\]

and

\[
P_{(i)}(H_1|x)P_{(i)}(0.55 \leq \theta \leq 0.57) \approx f_{(i)}(0.56) \cdot 0.02.
\]

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

\[
\frac{P_{(i)}(H_1|x)}{P_{(i)}(H_0|x)} \approx \frac{f_{(i)}(0.56)}{f_{(i)}(0.5)} \approx \frac{c(0.56)^{140}(0.44)^{110}}{c(0.5)^{140}(0.5)^{110}} \approx 6.07599
\]

(In reality we made these computations using R:

\[
dbeta(.56,141,111)/dbeta(.5,141,111).
\]

By similar reasoning, the posterior odds (using prior (iv)) of $H_1$ versus $H_0$ is approximately $\text{dbeta}(0.56,640,610)/\text{dbeta}(0.5,640,610) = 0.00437$.

Problem 3. (20 pts.) Bayes at the movies.

(a) Let $A$ be the event that Alice is selling tickets and $B$ the event that Bob is selling 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, \( O(A) = \frac{P(A)}{P(B)} = \frac{1}{16} \). Thus, our posterior odds are

\[
O(A|D) = \frac{P(D|A)}{P(D|B)} O(A) = \left( \frac{10}{15} \right)^{48} e^{25} \cdot \frac{1}{10} \approx 25.409
\]

Note that the Bayes factor is about 250.

(b) I believe that the decision to attend a movie is usually made long before meeting the ticket seller. The number of tickets sold in an hour is determined by the outcomes of random processes taking place earlier. A Poisson distribution is probably a good model for ticket sales, but the rate constant is not much affected by the ticket collector.

A better job for the problem could have been not selling tickets but issuing them: parking meter enforcement. The tickets are written only when an illegally parked car is encountered, and again a Poisson distribution is reasonable; but the rate constant this time depends on how fast the officer walks, and perhaps also on the choice of streets for walking on.

Problem 4. (20 pts.) Normal is the new normal.

(a) Leaving the scale factors as letters our table is

| Hyp. | Prior \( f(\theta) \sim N(5, 4^2) \) | Likelihood \( \phi(x|\theta) \sim N(\theta, 3^2) \) | Posterior \( f(\theta|x) \) |
|------|-------------------------------------|-----------------------------------------------|---------------------------------|
| \( \theta \) | \( c_1 e^{-(\theta-5)^2/32} d\theta \) | \( c_2 e^{(6-\theta)^2/18} \) | \( c \exp \left( -\frac{(\theta - 5)^2}{32} - \frac{(6 - \theta)^2}{18} \right) \) |
| Tot. | 1 | 1 | 1 |

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

\[
-\frac{(\theta - 5)^2}{32} - \frac{(6 - \theta)^2}{18} = -\frac{1}{2} \left( \frac{\theta^2 - 10\theta + 25}{16} + \frac{\theta^2 - 12\theta + 36}{9} \right) \\
= -\frac{1}{2} \left( \frac{25\theta^2 - 282\theta + 801}{144} \right) \\
= -\frac{1}{2} \left( \frac{(\theta - 141/25)^2}{144/25} + k \right)
\]

where \( k \) is a constant. Thus the posterior

\[
f(\theta|x) \propto \exp \left( -\frac{(\theta - 141/25)^2}{2 \cdot 144/25} \right)
\]

This has the form of a pdf for \( N \left( \frac{141}{25}, \frac{144}{25} \right) \). QED

(b) We have \( \mu_{\text{prior}} = 5, \sigma_{\text{prior}}^2 = 16, \bar{x} = 6, \sigma^2 = 9, n = 4 \) So we have

\[
a = \frac{1}{16}, \quad b = 4/9, \quad a+b = \frac{73}{144} \Rightarrow \mu_{\text{post}} = \frac{5/16 + 24/9}{73/144} = 5.88, \quad \sigma_{\text{post}}^2 = \frac{1}{73/144} = 1.97.
\]
After observing $x_1, \ldots, x_4$, we see that the posterior mean is close to $\pi$ and the posterior variance is much smaller than the prior variance. The data has made us more certain about the location of $\theta$.

(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 $\theta$.

Problem 5. (10 pts.) Censored data.

(a) 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

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

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.