Instructions

(1) Homework must be typed and answered in the order given (problem 1(a)(b)(c)(d) first, problem 2(a)(b)... second, etc...)

(2) Undergrads and grads will answer all questions.

(3) Include in each part of the homework only the answer. R code and R output (without mistakes), must be included in the appendix to the question. For example, for question 1.a, write only the answer and your comments. The code and output for that part of the question will be in the appendix (the last part of question 1).

(4) No late homework under any circumstances.

(5) Write your name and ID this way: Last name, first name, UCLA ID, date, Homework number.

(6) Do not just give a number as an answer. For example, if asked for probability that posterior proportion is larger than 0.7, write $Prob(p > 0.7) = 0.3$, say and write comments or explanations if needed.

(7) The homework must be turned in in lecture (no mail box, no e-mail).

Problem 1. Suppose we are interested in the estimation of the prevalence of a disease, for example, coronary heart disease, for men aged 30-40. Data from a prospective study of sample size equal to 1250 men have indicated 5 men experienced at least one incident. We are interested in estimating the prevalence rate of the disease for this specific age group. Do the following:

(a) Specify a non-informative Beta prior for the prevalence rate (U(0,1) written as a Beta -see lesson 3). Use R to find the prior mean, standard deviation, and probability that the prevalence rate is less than 2 incidents per 1000 males (hint: lessons 3 and 4).

(b) Specify a Binomial likelihood function that reflects what you observed in the data (hint: lesson 4).

(c) Specify the type of distribution that is your posterior distribution.

(d) Use R to find the posterior mean, the posterior standard deviation, the probability that the prevalence rate is less than 2 incidents per 1000 males and a 95% posterior interval.

(e) Make a plot that shows the prior and posterior distribution together in the same frame. Put a legend and distinguish the two curves. Discuss your results.

(f) Appendix: include documented R code (see web site for course, computing section files) and output here at the end of the problem. Use courier font for code and output. It must be clear what code is for which part of the problem. Write code in the same order that you answer the questions.

Solution 1. (a) *Prior p is distributed as Beta(1,1) = U(0,1).*

Prior mean $= 1/(1+1) = 1/2$

Prior variance$= \frac{(1\times1)}{((1+1)^2(1+1+1))} = 0.08333333$

Prior standard deviation$= 0.2886751$

Prob(p < 0.002) = 0.002

(b) *Let p = proportion of men age 30-40 that have coronary heart disease.*

Let Y = number of men age 30-40 that have coronary heart disease

Let n = number of men age 30-40 in the sample

$P(Y = 5 \mid n = 1250, p) = \binom{1250}{5} p^5 (1-p)^{(1250-5)}$

(c) Specify the type of distribution that is your posterior distribution.
It will be a Beta(1+5, 1+1245) = Beta(6,1246)

(d) *Posterior mean*= 6/(6 + 1246) = 0.004792332
Posterior Variance = $\frac{(6\times1246)}{((6+1246)^2(6+1246+1))}$ = $3.806357e-06$
Posterior standard deviation= 0.001950989
Prob(p < 0.002) = 0.04198746
Posterior 95% interval for p = (0.001762080, 0.009302367)

(e) *I rescale a little bit the likelihood so that I can show the two curves together*

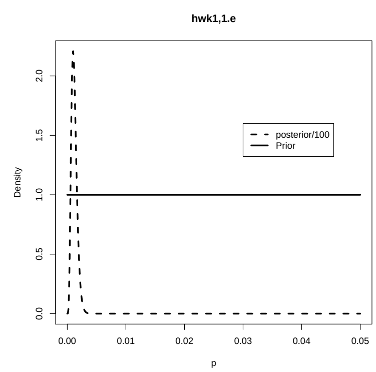

Figure 1: Prior and Posterior beta densities for prevalence of coronary heart disease in men aged 30-40

(f) Appendix:

```
############## Part (a) Prior Beta(1,1) #################
1/(1+1)  #  formula for prior mean of p
0.5       # prior mean of p
(1*1)/((1+1)^2*(1+1+1)) # formula for prior variance of p
 0.08333333  #   prior variance
 pbeta(0.002,1,1)   # find priorProb(p < 0.002)
[1] 0.002     # Prior prob(p<0.002)
##################################################


############## Part (d) Posterior Beta(6,1251) #################
6/(6+1251)  #  formula for posterior mean of p
0.00477327  # posterior mean of p
(6*1251)/((6+1251)^2*(6+1251+1)) # formula for posterior variance of p
 3.776221e-06  #   posterior variance
  pbeta(0.002,6,1251)   # find posterior Prob(p < 0.002)
[1] 0.04265915    # Posterior prob(p<0.002)
 qbeta(c(0.025,0.975),6,1251)   # find posterior 95 percent interval
   0.001755057 0.009265433  #  posterior 95 percent probability interval
##################################################
```

```
################Part (e) Plot ########################
p=seq(0,0.05,length=500)
 plot(p,scale.posterior,type="l",ylab="Density",lty=2,lwd=3, main="hwk1,1.e")
 lines(p,prior,lty=1,lwd=3)
 legend(0.03,1.6,c("posterior/100","Prior"),lty=c(2,1),lwd=c(3,3))
####################################################
```

Problem 2. Borrowing owls sometimes build their nests in holes that were dug by prairie dogs, coyotes, or badgers and have been since abandoned. They sometimes line the nests with cattle or horse dung. Why? Perhaps it insulates the nest from temperature extremes. A possibility suggested by biologist Dennis Martin is that the owls use the dung to keep predators away. To test this theory, biologist Gregory Green observed lined and unlined owl nests in the Columbia River basin. He recorded whether the nest was raided or not in both samples. We will deal with both samples later in the quarter, but for now consider only the lined nests. Green identified $n = 25$ nests lined with dung and observed that only $s = 2$ were raided. Assume exchangeability of the 25 nests.

Green's prior probabilities of the proportion of nest lined with dung being raided are given below

Proportion of Raided nests (p)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Prior probabilities	0.08	0.12	0.08	0.04	0.12	0.28	0.12	0.04	0.04	0.04	0.04

(a) Summarize in one sentence Green's prior probability distribution and enter it into R.

(b) Write down the likelihood function and enter it into R.

(c) Find the posterior distribution with R. Write it in a table that also contains the prior probabilities. Has Green changed his views very much after observing the data? Explain your answer.

(d) Plot the posterior and prior distributions.

(e) Appendix: include documented R code (see web site for course, computing section files) and output here at the end of the problem. Use courier font for code and output. It must be clear what code is for which part of the problem. Write code in the same order that you answer the questions.

Solution 2. (a) *Green seems to believe that the proportion of nests lined with dung being raided is most likely to be between 40 and 60%, and is certainly lower than 60%.*

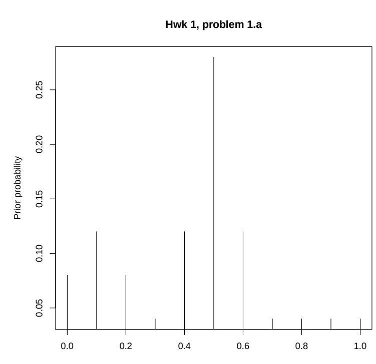

Figure 2: Prior probability of proportion of nest lined dung being raided

(b) $l(p) = p^2(1-p)^{23}$

For entering it into R, see the Appendix, part e of this question

(c) *The posterior distribution is the product of the likelihood and the prior at each p. According to the table below, we can see that after seeing the data, the distribution for p has shifted to the left. That is, it is more likely that 1 or 2 % of nests are raided. Definitely, that proportion is less than 0.3. Thus, Green should change his opinion after seeing the data.*

	p	prior.prob	likelihood	posterior
[1,]	0.0	0.08	0.000000e+00	0.000000e+00
[2,]	0.1	0.12	8.862938e-04	8.414847e-01
[3,]	0.2	0.08	2.361183e-04	1.494538e-01
[4,]	0.3	0.04	2.463187e-05	7.795513e-03
[5,]	0.4	0.12	1.263568e-06	1.199685e-03
[6,]	0.5	0.28	2.980232e-08	6.602303e-05
[7,]	0.6	0.12	2.533275e-10	2.405198e-07
[8,]	0.7	0.04	4.613016e-13	1.459931e-10
[9,]	0.8	0.04	5.368709e-17	1.699093e-14
[10,]	0.9	0.04	8.100000e-24	2.563494e-21
[11,]	1.0	0.04	0.000000e+00	0.000000e+00

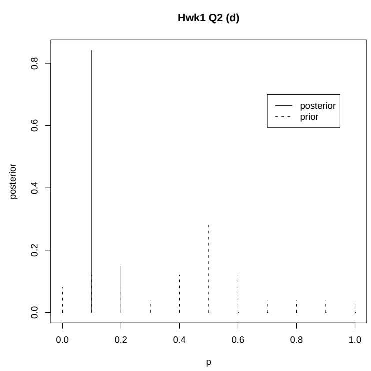

Figure 3: Prior and posterior probability of proportion of nest lined dung being raided

(d) *The image confirms what we said. Certainly opinion has shifted towards lower proportions. There are much less nests being raided that we thought a priori, in view of the evidence in the data.*

(e)
```
############### Hwk 1, problem 2(a) ###############
 p=seq(0,1,by=0.1)
prior.prob=c(0.08,0.12,0.08,0.04,0.12,0.28,0.12,0.04,0.04,0.04,0.04)
 plot(p,prior.prob,type="h",ylab="Prior probability",main="Hwk 1, problem 1.a")

############### Hwk 1, problem 2(b) ################
likelihood=(p^2)*(1-p)^(23)

###############  Hwk 1, problem 2(c) ##################
posterior=(prior.prob*likelihood)/sum(prior.prob*likelihood)
cbind(p,prior.prob,likelihood,posterior)

############## Hwk 1, problem 2(d) ##################

plot(p,posterior,type="h",lty=1,main="Hwk1 Q2 (d)")
```

```
lines(p,prior.prob,type="h",lty=2)
legend(0.7,0.7,c("posterior","prior"),lty=c(1,2))
```

Problem 3. In the matter of why burrowing owls line their nests with dung, suppose biologist Green's prior opinion concerning the proportion of dung-lined nests that are raided by badgers is given by the Beta(2,2) density.

(a) Specify the posterior density.

(b) Find his predictive probability that the next lined owl nest he observes will be raided. Show work.

Solution 3. (a) *Beta(2+2, 23+2)*

(b) *The predictive probability of a success is the posterior average probability of a success, that is, the average of p in the posterior distribution. This will be:*

$P(Y = 1) = 4/(29) = 0.1379$

Problem 4. Suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $Pr(\theta = 1) = 0.5$ and $Pr(\theta = 2) = 0.5$

(a) For $\sigma = 2$, write the formula for the marginal probability density for y and plot it using R.

(b) What is $Pr(\theta = 1 \mid y = 1)$, again supposing $\sigma = 2$? Show work.

(c) Describe how the posterior density of θ changes in shape as σ is increased and as it is decreased.

Solution 4. Prior distribution: $Pr(\theta = 1) = \frac{1}{2}$ $\quad Pr(\theta = 2) = 0.5$

$$Pr(y \mid \theta = 1) = \frac{1}{\sqrt{2\pi}}\sigma e^{-\frac{1}{2\sigma^2}(y-1)^2} \tag{1}$$

$$Pr(y \mid \theta = 2) = \frac{1}{\sqrt{2\pi}}\sigma e^{-\frac{1}{2\sigma^2}(y-2)^2} \tag{2}$$

(a)

$$Pr(y) = Pr(y|\theta = 1)Pr(\theta = 1) + Pr(y \mid \theta = 2)Pr(\theta = 2) \tag{3}$$

$$= \frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{8}(y-1)^2}\frac{1}{2} + \frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{8}(y-1)^2}\frac{1}{2} \tag{4}$$

$$= \frac{1}{4\sqrt{2\pi}}\left[e^{-\frac{1}{8}(y-1)^2} + e^{-\frac{1}{8}(y-2)^2}\right] \tag{5}$$

The following R code will generate the marginal posterior density for y

```
y = seq(-7,10,0.02)
dens = 0.5*dnorm(y,1,2) + 0.5*dnorm(y,2,2)
plot (y, dens, ylim=c(0,1.1*max(dens)),
type="l", xlab="y", ylab="f(y)", xaxs="i",
yaxs="i", yaxt="n", bty="n", cex=2)
}
```

(b) $Pr(\theta = 1 \mid y = 1)$, again supposing $\sigma = 2$ is given by

$$
\begin{aligned}
Pr(\theta = 1 \mid y = 1) &= \frac{Pr(y = 1 \mid \theta = 1)Pr(\theta = 1)}{Pr(y = 1)} && (6)\\
&= \frac{\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(1-1)^2}(1/2)}{\frac{1}{4\sqrt{2\pi}}\left[e^{-\frac{1}{8}(1-1)^2} + e^{-\frac{1}{8}(1-2)^2}\right]} && (7)\\
&= \frac{1}{1 + e^{-\frac{1}{8}}} && (8)\\
&= 0.5312 && (9)
\end{aligned}
$$

One can write R code to compute this, since the dnorm can be used everytime we have a normal density in the calculation.

```
y=seq(-7,10,0.02);
f.y = 0.5*dnorm(y,1,2)+0.5*dnorm(y,2,2)
plot(y,f.y,ylim=c(0,1.1*max

(dnorm(1,1,2))*(1/2)/
```

(c) For any σ, $Pr(\theta = 1 \mid y = 1) = \frac{1}{1+e^{-\frac{1}{2\sigma^2}}}$ is monotone decreasing in σ^2. And $p(\theta = 1 \mid y = 1) \to 1$ as $\sigma^2 \to 0$ and $p(\theta = 1 \mid y = 1) \to 0.5$ as $\sigma^2 \to \infty$. In this last case, the posterior approaches the prior, the data contains no information.

Problem 5. Since the posterior distribution is a compromise between prior information and the information provided by the new data, then it is interesting to compare the relative strengths. Perform an experiment where you flip a coin 10 times, recoding the data as zeros and ones. Produce the posterior expected value (mean) for two priors on p (the probability of a head): a uniform distribution between zero and one, and a beta distribution, with parameters [10,1]. Which prior appears to influence the posterior mean more than the other? Explain and show your work.

Solution 5. *I used the following command in R to generate 10 tosses of a coin. As we can see, I got 5 successes (the number 1).*

```
sample(c(1,0), 10,prob=c(0.5,0.5), replace=TRUE)
[1] 0 0 1 1 0 0 1 0 1 1
```

Under a Beta(1,1) (which is the same as a U(0,1)), the posterior is Beta(5+1, 5+1)= Beta(6,6). The prior mean is 0.5, and the posterior mean is 0.5

Under a Beta(10,1), the posterior is Beta(15, 6), the prior mean is 10/11=0.909 and the posterior mean 15/21 = 0.814.

Thus, for the same data, a Beta(10,1) prior has resulted in a posterior opinion that departs more from the data information than when we used a noninformative uniform prior.

Problem 6. Assume that 2% of the population of the United States are members of some extremist Militia group (Prob(M)=0.02), a fact that some members might attempt to hide and therefore not readily admit to an interviewer. A survey is 95% accurate on positive classification, $P(C \mid M) = 0.95$, and the unconditional probability of classification (i.e, regardless of actual militia status) is given by $P(C) = 0.05$. The survey is 97% accurate on negative classification ($P(C^c \mid M^c) = 0.97$).

(a) Use the law of total probability (lesson 2), and information given in the problem, to show that P(C) is the normalizing constant obtained by accumulating over all possible events. (Hint: $P(C \mid M^c) = 1 - P(C^c \mid M^c)$.

(b) Find the probability that someone positively classified by the survey as being a militia member really is a militia member

Solution 6. (a) $P(C) = P(C \cap M) + P(C \cap M^c) = P(C \mid M)P(M) + P(C \mid M^c)P(M^c) = P(C \mid M)P(M) + [1 - P(C^c \mid M^c)]P(M^c) = (0.95)(0.02) + (0.03)(0.98) = 0.0484$ As we can see, we are averaging the probability of event C over all possible states of the world.

(b) $P(M \mid C) = \frac{P(C|M)P(M)}{P(C)} = \frac{0.95(0.02)}{0.0484} = 0.392562$ The startling