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. Use material from Outline 5, section 7.

Suppose the causes of death are reviewed in detail for a city in the United States for a single year. It is found that 10 persons, out of a population of 500,000, died of skin cancer, giving an estimated cancer mortality rate in the city of 2 cases per 100,000 persons per year. A Poisson sampling model is often used for epidemiological data of this form. The Poisson model derives from an assumption of exchangeability among all small intervals of exposure. Define y as the number of deaths in a city of 500,000 in one year, and θ as the true underlying long-term cancer mortality rate in the city (measured in cases per 100,000 persons per year). The model for the data is then Poisson(5θ).

We can use knowledge about skin cancer mortality rates around the world to construct a prior distribution for θ and then combine the datum $y = 10$ with that prior distribution to obtain a posterior distribution.

(a) What is a sensible Gamma prior distribution for θ? Google a little to find mortality rates around the world, and potential information about probabilities of rates being larger or smaller than 2 cases per thousand, probabilities of being larger than 3, etc.... Enough information to determine whether there is a right tail or left tail, or whether the distribution is symmetric. Give the (a, b) of your gamma prior, the mean, and some probabilities.

(b) What is the posterior distribution for θ (specify posterior parameters of the posterior Gamma distribution.

(c) Compute the posterior probability that the long-term death rate from skin cancer in our city is more than 1.0 per 100,000 per year.

Solution 1. (a) $y = 10$ is a single observation with exposure $x = 5$

According to the World Health Organization (WHO,

```
http://en.wikipedia.org/wiki/Skin_cancer#cite_note-24
and
http://www.who.int/healthinfo/global_burden_disease/estimates_country/en/index.html
```

more than 50% of the world have rates lower than 4 cancers per 100,000 people. There are some cancer rates up to 7.7. The mean is probably around 3 cases per 100000 with a standard deviation of about 1.5. I decide that a Gamma(6,2) prior might be appropriate. This has the following summaries:

Mean= 6/2=3; Variance= 6/4=1.5; sd=1.22; Prob(rate < 4) = 0.80876

(b) Gamma(6+10, 2+5) = Gamma(16,7)

(c) Prob(rate > 1 per 10000) = 0.99759

Posterior mean is 16/7 =2.2857

Problem 2. The table below gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period.

(a) Assume that the number of fatal accidents in each year are independent with a Poisson(θ) distribution. Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 through 1985. Under this model, give a 95% predictive interval for the number of passenger deaths in 1986.

(b) Assume that the numbers of fatal accidents in each year follow independent Poisson distributions with a constant rate and an exposure in each year proportional to the number of passenger miles flown. Set a prior distribution for θ and determine the posterior distribution based on the data for 1976-1985. (Estimate the number of passenger miles flown in each year by dividing the appropriate columns of the table and ignoring round-off errors.) Give a 95% predictive interval for the number of passenger deaths in 1986 under the assumption that 8000 million passenger miles are flown that year.

(c) Repeat (a) above, replacing "fatal accidents" with "passenger deaths."

(d) Repeat (b) above, replacing "fatal accidents" with "passenger deaths."

Year	Fatal accidents	Passenger deaths	Death rate
1976	24	734	0.19
1977	25	516	0.12
1978	31	754	0.15
1979	31	877	0.16
1980	22	814	0.14
1981	21	362	0.06
1982	26	764	0.33
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

Solution 2. (a) The conjugate prior for a Poisson likelihood, $Y \sim Poi(\theta)$, is a gamma prior, $\theta \sim Ga(\alpha, \beta)$. Choose α, β to match your prior beliefs about mean and variance on fatal airline accidents worldwide per year. For example, I believe there might be around 50 give or take 40 fatal airline accidents per year., i.e, $E(Y) = 50$, $Var(Y) = 40^2$. Note that the marginal distribution of Y is negative binomial $NBin(\alpha, \beta)$. Matching mean and variance of a neg binom with 50 and 40^2, respectively, I find $\alpha = 1.6, \quad , \beta = 0.03$. Alternatively, you could have guessed an a-priori mean and var of θ and matched with moments of the gamma. It's probably easier though to think about an observable (Y), rather than an unobservable parameter (θ).

The sample mean is $\bar{y} = 23.8$ and $n = 10$.

$$\theta \sim Ga(1.6, 0.03) \quad (1)$$
$$y_i \mid \theta \sim Pois(\theta) \quad (2)$$
$$\theta \mid y \sim Ga(1.6 + 10(23.8), 0.03 + 10) = Ga(239.6, 10.03) \quad (3)$$
$$y_{n+1} \mid y_i, y_2, \ldots.., y_n) \sim Negbin(239.6, 10.03) \quad (4)$$

The predictive distribution $p(y_{n+1} \mid y) = \int p(y_{n+1} \mid \theta) p(\theta \mid y) d\theta = \int Poi(y_{n+1} \mid \theta) Ga(\theta \mid \alpha_1, \beta_1) d\theta$. The posterior density is $p(y_{n+1} \mid y) = NBin(239.6, 10.03)$. A 95% predictive interval is (14, 34), obtained with the following R code

```
 ######## hwk2.p2.a #########
mu1=239.6/10.03
size1=239.6
qnbinom(c(0.025,0.975),mu=mu1,size=size1)
```

Thus, there is 95% probability that the number of fatal accidents in 1986 will be between 14 and 34.

(b) Let X equal passenger miles. In the table, X=pass deaths/death rate (in 100 mil miles).

$$
\begin{aligned}
p(\theta \mid y) &= Ga(1.6 + 10(23.8), 0.03 + 53596.92) = Ga(239.6, 53596.95) && (5)\\
p(y_{n+1} \mid y) &= \int p(y_{n+1} \mid \theta) p(\theta \mid y) d\theta && (6)\\
&= \int Poi(y_{n+1} \mid x_{n+1}\theta) Ga(\theta \mid \alpha_2, \beta_2) d\theta && (7)\\
&= \int Poi(y_{n+1} \mid \eta) Ga(\eta \mid \alpha_2, \beta_2/x_{n+1}) d\theta && (8)\\
&= Nbin(y_{n+1} \mid \alpha_2, \beta_2/x_{n+1} && (9)\\
&= NBin(y_{n+1} \mid 239.6, 6.699619) && (10)
\end{aligned}
$$

where 6.699619 comes from dividing 53596.95 by 8000 The third equation comes from a change of variables from θ to $\eta = x_{n+1}\theta$. Note that if $\theta \sim Ga(a, b)$, then $\eta = r\theta \sim Ga(a, b/r)$. The fourth equation is true by theoretical results on the Poisson. The desired predictive interval is the central 95% interval for a $NBin(\alpha_2, \beta_2/x_{n+1})$. We can find that it is (24,49), from the following R code

```
#########  hwk2.p2.b ###########
mu2=239.6/6.699619
 size2=239.6
 qnbinom(c(0.025,0.975),mu=mu2,size=size2)
[1] 24 49
###############################
```

Thus there is 95% probability that the number of fatal accidents in 1986 is between 24 and 49 in this year where 8000 million miles were travelled.

(c) y_i is now the number of passenger deaths per year. To set the prior distribution, I checked some aviation accident statistics and noticed that there is a lot of variability. An average of 500 accidents give or take a standard deviation of 100 is an approximation. I will use a negative binomial with this mean and standard deviation to find the parameters. A $NegBin(\alpha, beta)$ has mean α/β and variance $(\alpha/beta^2)(\beta + 1)$. So I make $500 = \alpha/\beta$ and $10000 = (\alpha/\beta^2)(\beta + 1)$. I obtain this way $\alpha = 26.3$ and $\beta = 0.0526$.

```
http://www.ntsb.gov/aviation/Table10.htm
```

The sample mean is now $\bar{y} = 691.9$.

$$
\begin{aligned}
\theta &\sim Ga(26.3, 0.0526) && (11)\\
y_i \mid \theta &\sim Pois(\theta) && (12)\\
\theta \mid y &\sim Ga(26.3 + 10(691.9), 0.0526 + 10) = Ga(6945.3, 10.0526) && (13)\\
y_{n+1} \mid y_i, y_2, \ldots, y_n) &\sim Negbin(6945.3, 10.0526) && (14)
\end{aligned}
$$

A 95% prediction interval is now (637, 745). Thus there is 95% probability that in 1986 there will be between 637 and 745 passenger deaths.

```
########## hwk2.p2.c ###########
mu3=6945.3/10.0526
size3=6945.3
qnbinom(c(0.025,0.975),mu=mu3,size=size3)
##################################
```

(d)

$$p(\theta \mid y) = Ga(6945.3, 0.0526 + 53596.92) = Ga(6945.3, 53596.97) \quad (15)$$

$$p(y_{n+1} \mid y) = \int p(y_{n+1} \mid \theta) p(\theta \mid y) d\theta \quad (16)$$

$$= \int Poi(y_{n+1} \mid x_{n+1}\theta) Ga(\theta \mid \alpha_2, \beta_2) d\theta \quad (17)$$

$$= \int Poi(y_{n+1} \mid \eta) Ga(\eta \mid \alpha_2, \beta_2 / x_{n+1}) d\theta \quad (18)$$

$$= Nbin(y_{n+1} \mid \alpha_2, \beta_2 / x_{n+1} \quad (19)$$

$$= NBin(y_{n+1} \mid 6945.3, 6.699619) \quad (20)$$

where 6.69962 comes from dividing 53596.97 by 8000 The third equation comes from a change of variables from θ to $\eta = x_{n+1}\theta$. Note that if $\theta \sim Ga(a, b)$, then $\eta = r\theta \sim Ga(a, b/r)$. The fourth equation is true by theoretical results on the Poisson. The desired predictive interval is the central 95% interval for a $NBin(\alpha_2, \beta_2/x_{n+1})$. We can find that it is (970,1105), from the following R code

```
#########  hwk2.p2.b ###########
mu4=6945/6.699619
 size4=6945.3
 qnbinom(c(0.025,0.975),mu=mu4,size=size4)
[1] 970 1105
################################
```

Problem 3. In section 5.9, we conclude the problem with the following statements:

Notice that there is much more overlap between these two distributions than between the posterior distributions of θ_1 and θ_2. For example, $Pr(\tilde{Y}_1 > \tilde{Y}_2 \mid \sum Y_{i,1} = 217, \sum Y_{i,2}) = 0.22$. The distinction between the events $\{\theta_1 > \theta_2\}$ and $\tilde{Y}_1 > \tilde{Y}_2$ is extremely important: Strong evidence of a difference between two populations does not mean that the difference itself is large.

Use R to check whether those statements are true. The way to do that is to first generate observations from the distribution of Y_1, then generate observations from the distribution of Y_2 and after that, count the proportion of one larger than the other. It is done in the textbook, so you may use the same commands.

Solution 3. In this problem, you just have to run the program given in the notes (also in Hoff's book) until the end, and interpret the output. A plot of the the two predictive distributions also helps see the overlap. I include both here. The output from R that we need is also here. The probability that a woman randomly chosen among the uneducated has more kinds than an educated one is close to 5%. The two distributions overlap a lot, even though the educated have higher peaks at lower values.

```
a<-2 ; b<-1           # prior parameters
n1<-111 ; s1<-217     # data in group 1
n2<-44  ; s2<-66      # data in group 2


(a+s1)/(b+n1)         # posterior mean
```

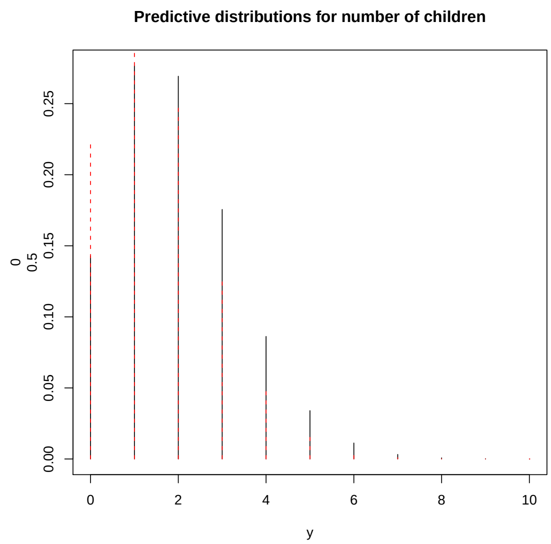

Figure 1: Predicted number of children of a randomly chosen educated and another uneducated woman. The distribution in red is the one for the educated,

```
(a+s1-1)/(b+n1)      # posterior mode
qgamma( c(.025,.975),a+s1,b+n1)   # posterior 95\% CI

(a+s2)/(b+n2)
(a+s2-1)/(b+n2)
qgamma( c(.025,.975),a+s2,b+n2)

-------
th1_mc<-rgamma(100000,a+s1,b+n1)

th2_mc<-rgamma(100000,a+s2,b+n2)

mean(th1_mc>th2_mc)

y1_mc<-rpois(1000000,th1_mc)   # to compare the predictive distributions, rn
y2_mc<-rpois(1000000,th2_mc)



######### Homework 2.p.3 ############

 mean(y1_mc>y2_mc)   # proportion of y1> y2
[1] 0.482124
 mean(y1_mc>=y2_mc)
[1] 0.699361
mean(y1_mc==y2_mc)
[1] 0.217237
```

```
### plots ######
y =0:10
without=dnbinom(y, size=(a+s1), mu=(a+s1)/(b+n1))
with= dnbinom(y, size=(a+s2), mu=(a+s2)/(b+n2))
plot(y, without, type="h", lty=1,main="Predictive distributions for number of children",ylab=c(0,0.5))
lines(y,with,type="h", lty=2,col="red")

###########################
```

Problem 4. Normal distribution with unknown mean: a random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose your prior distribution for θ is normal with mean 180 and standard deviation 40.

(a) Give your posterior distribution for θ. (Your answer will be a function of n.)

(b) A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$. (Your answer will still be a function of n.)

(c) For *n=10*, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$.

Solution 4. (a) The prior distribution of θ: $\theta \sim N(180, 40^2)$.

$$f(\theta \mid \tau_0^2) = \frac{1}{\tau_0\sqrt{2\pi}} \exp\left\{-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right\} \tag{21}$$

$$= \frac{1}{40\sqrt{2\pi}} \exp\left\{-\frac{1}{3200}\left(\theta-180)^2\right)\right\} \tag{22}$$

The distribution of y given θ and σ^2 is: $Y \mid \theta, \sigma^2 \sim N(\theta, 20^2)$.

$$f(y \mid \theta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta)^2\right\} \tag{23}$$

$$= \frac{1}{20\sqrt{2\pi}} \exp\left\{-\frac{1}{800}(y-\theta)^2\right\} \tag{24}$$

The posterior distribution is approximately

$$f(\theta \mid y_1, y_2, ..., y_n, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\} \exp\left\{-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right\}$$

Adding the terms in the exponents and ignoring the $-\frac{1}{2}$ for the moment, we have

$$\frac{1}{\tau_0^2}\left(\theta^2 - 2\theta\mu_0 + \mu_0^2\right) + \frac{1}{\sigma^2}\left(\sum y_i^2 - 2\theta\sum y_i + n\theta^2\right) = a\theta^2 - 2b\theta + c,$$

where

$$a = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}, \quad b = \frac{\mu_0}{\tau_0^2} + \frac{\sum y_i}{\sigma^2}, \quad \text{and } c = c(\mu_0, \tau_0^2, \sigma^2, y_1, ..., y_n).$$

Now let's see if $f(\theta \mid y_1, y_2, ..., y_n, \sigma^2)$ takes the form of a normal density:

$$f(\theta \mid \sigma^2, y_1, y_2, ..., y_n) \propto \exp\left\{-\frac{1}{2}(a\theta^2 - 2b\theta)\right\} \tag{25}$$

$$= \exp\left\{-\frac{1}{2}a\left(\theta^2 - \frac{2b\theta}{a} + \frac{b^2}{a^2}\right) + \frac{1}{2}\frac{b^2}{a}\right\} \tag{26}$$

$$\propto \exp\left\{-\frac{1}{2}a\left(\theta - \frac{b}{a}\right)^2\right\} \tag{27}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{\theta - b/a}{1/\sqrt{a}}\right)\right\}. \tag{28}$$

This function has exactly the same shape as a normal density curve, with $1/\sqrt{a}$ playing the role of the standard deviation and b/a playing the role of the mean. Since probability distributions are determined by their shape, this means that $f(\theta \mid \sigma^2, y_1, y_2, ..., y_n)$ is indeed a normal density. We refer to the mean and variance of this density as μ_n and τ_n^2, where

$$\tau_n^2 = \frac{1}{a} = \frac{1}{1/\tau_0^2 + n/\sigma^2} \quad \text{and} \quad \mu_n = \frac{b}{a} = \frac{\mu_0/\tau_0^2 + \sum y_i/\sigma^2}{1/\tau_0^2 + n/\sigma^2} = \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2}$$

So the posterior distribution of θ for our particular example is Normal, withe the following values for the mean and the variance

$$\tau_n^2 = \frac{1}{1/1600 + n/400} \quad \text{and} \quad \mu_n = \frac{180/1600 + n(150)/400}{1/1600 + n/400} = (180/1600 + n(150)/400)\tau_n^2$$

(b) We consider now predicting a new observation $\tilde{y}$ from the population after having observed $(y_1, y_2, ..., y_n)$. To find the predictive distribution, let's use the following fact:

$$\left\{\tilde{y} \mid \theta, \sigma^2\right\} \sim \text{normal}\left(\theta, \sigma^2\right) \rightarrow \tilde{y} = \theta + \tilde{\epsilon}, \quad \left\{\tilde{\epsilon} \mid \theta, \sigma^2\right\} \sim \text{normal}\left(0, \sigma^2\right)$$

So the posterior mean of $\tilde{y}$ is μ_n and the posterior variance of $\tilde{y}$ is $\tau_n^2 + \sigma^2$. The predictive distribution is therefore

$$\tilde{y} \mid \sigma^2, y_1, ..., y_n \sim \text{normal}\left(\mu_n, \tau_n^2 + \sigma^2\right)$$

In the particular case considered in this problem,

$$\tilde{y} \mid \sigma^2, y_1, ..., y_n \sim \text{normal}\left((180/1600 + n(150)/400)\tau_n^2, \frac{1}{1/1600 + n/400} + 400\right) = \text{Normal}\left(\frac{180 + 600n}{1 + 4n}, \frac{1600}{1 + 4n} + 400\right)$$

(c) For $n = 10$, we have

$$\tau_n^2 = \frac{1}{1/1600 + 10/400} = 39.024 \quad \text{and} \quad \mu_n = \frac{180/1600 + 10(150)/400}{1/1600 + n/400} = (180/1600 + n(150)/400)\tau_n^2 = 150.7317$$

A 95% posterior interval for θ is then $150.7317 \pm 1.96\sqrt{(39.024)}$ or (138.4877,162.9757).

We also have the mean of the posterior predictive distribution to be $\frac{180+600(10)}{1+4(10)} = 150.7317$ and the variance $\frac{1600}{1+4(10)} + 400 = 439.0244$

A 95% posterior predictive interval for $\tilde{y}$ is then $150.7317 \pm 1.96\sqrt{(439.0244)}$ or (109.664,191.7994).

Problem 5. Computing with a nonconjugate single-parameter model. suppose that $y_1, y_2, ..., y_n$ are independent samples from a Cauchy distribution with unknown center θ and known scale 1.:

$$p(y_i \mid \theta) \propto 1/(1 + (y_i - \theta)^2)$$

. Assume, for simplicity, that the prior distribution for θ is uniform on $[0, 1]$. Given the observations $(y_1, y_2, ..., y_5) = (-2, -1, 0, 1.5, 2.5)$:

(a) Compute the unnormalized posterior density function $p(\theta)p(y \mid \theta)$, on a grid of points $\theta = 0, 1/m, 2/m, ..., 1$, for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta \mid y)$ as a function of θ.

(b) Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.

(c) Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the predictive draws.

```
######## Will this program do the job?   #####

f = function(theta) {  # evaluates the likelihood
  likelihood= (1/(1+(-2-theta)^2))*(1/(1+(-1-theta)^2))*(1/(1+(0-theta)^2)) *(1/(1+(1.5-theta)^2))*(1/(
}

grid = seq(from=0, to=1, length=100)   # make a grid of theta values -step=0.01
fgrid=rep(0,100)   # initialize a vector for the function evaluations

for(i in 1:100){     # evaluate f on the grid
fgrid[i] = f(grid[i])
}

c=0.01*sum(fgrid) #compute the normalization constant
 # dens norm = dens.unn/step x (sum(dens unnorma))
px=fgrid/c

par(mfrow=c(2,1))
plot(grid, fgrid, xlab="theta",ylab="P(theta|y)", type="l",
main="un-normalized posterior",ylim=c(0,max(fgrid)), xlim=c(0,1) )

plot(grid, px, xlab="theta",ylab="P(theta|y)", type="l",
main="normalized posterior",ylim=c(0,max(px)), xlim=c(0,1) )  #plot it
```

Solution 5.

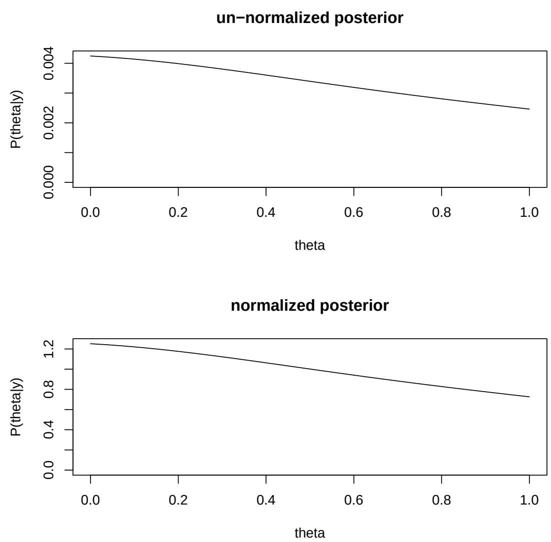

Figure 2: Unnormalized and normalized posterior distributions for a non-conjugate single-parameter model: Cauchy sampling model and U(0,1) prior.