1 Basu's Elephant Round 2

Please note, I used a chapter of an online textbook to guide my solution to this problem[1]. Let us call the estimator, $\hat{\theta}$. Indeed,

$$\hat{\theta} = \sum_{i=1}^{N} \frac{Y_i}{\pi_i} I_i$$

where π_i is the probability of an elephant to be sampled, Y_i is the weight of the sampled elephant, and I_i is the indicator function (1 if $i \in A$ and 0 if $i \notin A$ where A is the sample). Note that π_i is the probability that i is included in the sample. We also define π_{ij} to be the probability that both i and j are included in the sample. Therefore,

$$Var(\hat{\theta}) = \sum_{A \in \text{set of all samples}} P(A)(\hat{\theta}(A) - E(\hat{\theta}))^2$$

Sweep over set of all samples and write $(\hat{\theta}(A) - E(\hat{\theta}))^2$ as $Cov(I_i, I_j)$ to build:

$$Var(\hat{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} Cov(I_i, I_j)$$

$$Var(\hat{\theta}) = \sum_{i=1}^{N} \sum_{i=j}^{N} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} (\pi_{ij} - \pi_i \pi_j)$$

2 Chapter 1 Problem 3

2.1 Part 1

First, show that among brown-eyed children of brown-eyed parents, the expected proportion of heterozygotes is 2p/(1 + 2p). Note that in this solution I will use "Xx" to mean both "Xx" and "xX."

$$Pr(\text{Child}_{Xx}|\text{Child}_B, \text{Parents}_B) = \frac{Pr(\text{Child}_B, \text{Parents}_B|\text{Child}_{Xx})Pr(\text{Child}_{Xx})}{Pr(\text{Child}_B, \text{Parents}_B)}$$

Use joint probability instead for the numerator.

$$Pr(\text{Child}_{Xx}|\text{Child}_B, \text{Parents}_B) = \frac{Pr(\text{Child}_B, \text{Parents}_B, \text{Child}_{Xx})}{Pr(\text{Child}_B, \text{Parents}_B)}$$

[1] https://jkim.public.iastate.edu/teaching/book2.pdf

If the child is heterozygous then it will also have brown eyes.

$$Pr(\text{Child}_{Xx}|\text{Child}_B, \text{Parents}_B) = \frac{Pr(\text{Parents}_B, \text{Child}_{Xx})}{Pr(\text{Child}_B, \text{Parents}_B)}$$

Let us solve first for the numerator.

$$\text{Num} = Pr(\text{C}_{Xx}|\text{P}_{XX,XX})Pr(\text{P}_{XX,XX}) + 2Pr(\text{C}_{Xx}|\text{P}_{XX,Xx})Pr(\text{P}_{XX,Xx}) + Pr(\text{C}_{Xx}|\text{P}_{Xx,Xx})Pr(\text{P}_{Xx,Xx})$$

$$\text{Num} = (0)(1-p)^2(1-p)^2 + 2\left(\frac{1}{2}\right)(1-p)^2(2p(1-p)) + \left(\frac{1}{2}\right)(2p(1-p))(2p(1-p))$$

$$\text{Numerator} = 2p(1-p)^3 + 2p^2(1-p)^2$$

Now the denominator.

$$\text{Denom} = Pr(\text{C}_B|\text{P}_{XX,XX})Pr(\text{P}_{XX,XX}) + 2Pr(\text{C}_B|\text{P}_{XX,Xx})Pr(\text{P}_{XX,Xx}) + Pr(\text{C}_B|\text{P}_{Xx,Xx})Pr(\text{P}_{Xx,Xx})$$

$$\text{Denom} = (1)(1-p)^2(1-p)^2 + (2)(1)(1-p)^2(2p(1-p)) + \left(\frac{3}{4}\right)(2p(1-p))(2p(1-p))$$

$$\text{Denominator} = (1-p)^4 + 4p(1-p)^3 + 3p^2(1-p)^2$$

Now simplify:

$$Pr(\text{Child}_{Xx}|\text{Child}_B, \text{Parents}_B) = \frac{2p(1-p)^3 + 2p^2(1-p)^2}{(1-p)^4 + 4p(1-p)^3 + 3p^2(1-p)^2}$$

$$Pr(\text{Child}_{Xx}|\text{Child}_B, \text{Parents}_B) = \frac{2p((1-p)+p)}{(1-p)^2 + 4p(1-p) + 3p^2}$$

$$Pr(\text{Child}_{Xx}|\text{Child}_B, \text{Parents}_B) = \frac{2p}{1+2p}$$

□

2.2 Part 2

Second, find the posterior probability that Judy is a heterozygote. So we are looking for:

$$Pr(\text{Judy}_{Xx}|\text{NChildren}_B, \text{Husband}_{Xx})$$

$$= \frac{Pr(\text{NC}_B, \text{Husb}_{Xx}|\text{J}_{Xx})Pr(\text{J}_{Xx})}{Pr(\text{NC}_B, \text{Husb}_{Xx}|\text{J}_{Xx})Pr(\text{J}_{Xx}) + Pr(\text{NC}_B, \text{Husb}_{Xx}|\text{J}_{XX})Pr(\text{J}_{XX})}$$

We can use the probability from the first part now as our prior. We know the prior (before children) probabilities of Judy being a heterozygote is what we calculated above and her not being a heterozygote (with all aforementioned conditions) is one minus the above value. The only probabilities we still need is the probability the n children are all brown-eyed, based on if Judy is Xx or XX.

If Judy is Xx:

$$Pr(\text{Child}_B|\text{J}_{Xx},\text{Husb}_{Xx}) = 1 - Pr(\text{Child}_N B|\text{J}_{Xx},\text{Husb}_{Xx})$$

$$Pr(\text{Child}_B|\text{J}_{Xx},\text{Husb}_{Xx}) = 1 - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{3}{4}$$

The genes each child inherits are independent from the others so :

$$Pr(\text{nC}_B|\text{J}_{Xx},\text{Husb}_{Xx}) = \left(\frac{3}{4}\right)^n$$

If Judy is XX we know the child will always have brown eyes due to the knowledge that the father is a heterozygote:

$$Pr(\text{Child}_B|\text{J}_{XX},\text{Husb}_{Xx}) = 1$$

Put it all together:

$$\frac{Pr(\text{nC}_B,\text{Husb}_{Xx}|\text{J}_{Xx})Pr(\text{J}_{Xx})}{Pr(\text{nC}_B,\text{Husb}_{Xx}|\text{J}_{Xx})Pr(\text{J}_{Xx}) + Pr(\text{nC}_B,\text{Husb}_{Xx}|\text{J}_{XX})Pr(\text{J}_{XX})}$$

$$= \frac{(\frac{3}{4})(\frac{2p}{1+2p})}{(\frac{3}{4})(\frac{2p}{1+2p}) + (1)(1-\frac{2p}{1+2p})}$$

$$= \frac{(\frac{3}{4})(\frac{2p}{1+2p})}{(\frac{3}{4})(\frac{2p}{1+2p}) + (1)(\frac{1}{1+2p})}$$

2.3 Part 3

Lastly, find the probability that her first grandchild has blue eyes.

This would only happen if Judy's child (i.e. the grandchild's parent) is a heterozygote. First let us calculate this probability.

$$Pr(\text{C}_{Xx}|\text{H}_{Xx},\text{J}_B,\text{C}_B) = Pr(\text{C}_{Xx}|\text{H}_{Xx},\text{J}_{Xx},\text{C}_B)Pr(\text{J}_{Xx}) + Pr(\text{C}_{Xx}|\text{H}_{Xx},\text{J}_{XX},\text{C}_B)Pr(\text{J}_{XX})$$

$$= \left(\frac{2}{3}\right)\frac{(\frac{3}{4})(\frac{2p}{1+2p})}{(\frac{3}{4})(\frac{2p}{1+2p}) + (1)(\frac{1}{1+2p})} + \left(\frac{1}{2}\right)(1 - \frac{(\frac{3}{4})(\frac{2p}{1+2p})}{(\frac{3}{4})(\frac{2p}{1+2p}) + (1)(\frac{1}{1+2p})}) = A$$

Note we denoted the final above probability A for later reference. Now, Judy's child can mate with either a XX, Xx, or xx. The probability Judy's grandchild will have blue eyes depends on her child's mate. If they are XX the probability is 0, if Xx prob 1/4, and if xx prob 1/2. We assume the random mixing from above. Note I am not going to write all of the conditioned info that we know.

$$
\begin{aligned}
Pr(\mathrm{G}_{xx}) = Pr(\mathrm{G}_{xx}|\mathrm{C}_{Xx}, \mathrm{M}_{XX})Pr(\mathrm{C}_{Xx}, \mathrm{M}_{XX}) + Pr(\mathrm{G}_{xx}|\mathrm{C}_{Xx}, \mathrm{M}_{Xx})Pr(\mathrm{C}_{Xx}, \mathrm{M}_{Xx}) + \\
Pr(\mathrm{G}_{xx}|\mathrm{C}_{Xx}, \mathrm{M}_{xx})Pr(\mathrm{C}_{Xx}, \mathrm{M}_{xx}) \\
= (0)((1-p)^2)A + \left(\frac{1}{4}\right)(2p(1-p))A + \left(\frac{1}{2}\right)(p^2))A \\
= \left(\frac{1}{4}\right)(2p(1-p))A + \left(\frac{1}{2}\right)(p^2))A \\
= \left(\left(\frac{1}{4}\right)(2p(1-p)\right) + \left(\frac{1}{2}\right)(p^2))A
\end{aligned}
$$

3 Chapter 1 Problem 9

3.1 Part A

1. 165 patients came to the office
2. 162 patients had to wait for a doctor
3. Their average wait was 292 minutes.
4. The office closed 535 minutes past 4pm, or 12:55am.

3.2 Part B

1. 155 patients patients came to the office, [147, 162].
2. 152 patients patients had to wait for a doctor, [144, 159].
3. Their average wait was 275 minutes, [257, 300].
4. The office closed 481 minutes past 4pm, or 12:01am, [432, 520].

My simulation code:

```
#initializations for storage to find median / interval
sims <- 100
i.list <- c(rep(0,sims))
total.num.patients.list <- c(rep(0,sims))
patients.who.waited.list <- c(rep(0,sims))
virtual.line <- c(NA)
time.waited.list <- c(rep(0,sims))
i <- 0

#simulation
for (run in (1:sims)) {
  #initializations
  i <- 0 #time (minutes)
  num.wait <- 0 #the number of patients waiting
  busy <- c(0,0,0) #are or arent the doctors available, 0 if avail, 1 if busy
  free <- c(0,0,0) #when each of the doctors will be free again
  total.num.patients <- 0 #counter for total number of patients who visit
  patients.who.waited <- 0 #counter for number of patients who had to wait
  time.waited <- c(0) #counter for amount of time waited by each patient

  # from 9am to 4pm
  for (i in (1:420)) {
    #if a doctor finishs with a patient at time i, then set corresponding busy to
       0 for free
    busy[free==i] <- 0
    #if a doctor finishs with a patient at time i
    if (busy[1]==0 || busy[2]==0 || busy[3]==0) {
      #count the amount of time patient has been waiting ...
      time.waited[i] <- virtual.line[which(!is.na(virtual.line))[1]]
      #... and then remove from the line
      virtual.line[which(!is.na(virtual.line))[1]] <- NA
    }
    #does a new patient arrive? if yes add one to the waiting line
    if (rexp(1, 10)>=0.10) {
      num.wait <- num.wait+1
      total.num.patients <- total.num.patients+1
      #and add patient to virtual line
      virtual.line[i] <- 1
      #check if patient will have to wait
      if (length(busy[which(busy==0)]) < num.wait){
        patients.who.waited <- patients.who.waited+1
      }
    }
    #check which / how many of the doctors ...
```

```
    for (j in (1:3)) {
      #... are not busy, AND check if there is a patient waiting
      if (busy[j]==0 && num.wait > 0){
        #take a patient off the waiting line
        num.wait <- num.wait-1
        #note that doctor is now busy
        busy[j] <- 1
        #set the time when the doctor will no longer be busy to i + pulled value
        free[j] <- i + round(runif(1,15,20),0)
      }
    }
    #if no doctor is free, mark that patients in line had to wait
    for (mark in (which(!is.na(virtual.line)))) {
      virtual.line[mark] <- virtual.line[mark] +1
      }
  }
  # post 4pm
  while (num.wait>0) {
    #if a doctor finishes with a patient at time i, then set corresponding busy to
      0
    busy[free==i] <- 0
    #if a doctor finishes with a patient at time i
    if (busy[1]==0 || busy[2]==0 || busy[3]==0) {
      #count the amount of time patient has been waiting ...
      time.waited[i] <- virtual.line[which(!is.na(virtual.line))[1]]
      #... and then remove from the line
      virtual.line[which(!is.na(virtual.line))[1]] <- NA
    }
    #check which / how many of the doctors ...
    for (j in (1:3)) {
      #... are not busy, AND check if there is a patient waiting
      if (busy[j]==0 && num.wait > 0){
        #take a patient off the waiting line
        num.wait <- num.wait-1
        #note that that doctor is now busy
        busy[j] <- 1
        #set the time when the doctor will no longer be busy to i + pulled value
        free[j] <- i + round(runif(1,15,20),0)
      }
    }
    #if no doctor is free, mark that patients in line had to wait
    for (mark in (which(!is.na(virtual.line)))) {
      virtual.line[mark] <- virtual.line[mark] +1
    }
```

```
    #count elapsed minute
    i <- i + 1
    virtual.line[i] <- NA
  }
  i.list[run] <- i
  total.num.patients.list[run] <- total.num.patients
  patients.who.waited.list[run] <- patients.who.waited
  time.waited.list[run] <- mean(time.waited,na.rm=T)
}

#How many patients came to the office?
median(total.num.patients.list)
total.num.patients.list.ordered <- (sort(total.num.patients.list))
total.num.patients.interval <- total.num.patients.list.ordered[c(25,75)]
total.num.patients.interval

#How many had to wait for a doctor?
median(patients.who.waited.list)
patients.who.waited.list.ordered <- (sort(patients.who.waited.list))
patients.who.waited.list.interval <- patients.who.waited.list.ordered[c(25,75)]
patients.who.waited.list.interval

#What was their average wait?
median(time.waited.list)
time.waited.list.ordered <- (sort(time.waited.list))
time.waited.list.interval <- time.waited.list.ordered[c(25,75)]
time.waited.list.interval

#When did the office close?
i.list <- i.list-420
median(i.list)
i.list.ordered <- (sort(i.list))
i.list.interval <- i.list.ordered[c(25,75)]
i.list.interval
```

4 Chapter 2 Problem 4

4.1 Part A

$$p(Y|\theta = \frac{1}{12}) \sim N(83.33, 8.74)$$
$$p(Y|\theta = \frac{1}{6}) \sim N(166.67, 11.79)$$

$$p(Y|\theta = \frac{1}{4}) \sim N(250, 13.69)$$

Where I calculated the mean as $1000 * \theta$ and the standard deviation as $\sqrt{1000 * \theta * (1 - \theta)}$.

4.2 Part B

Because we are given the prior distribution for the value of θ, we are able to approximate the distribution of y. Notably, by determining the means and spreads of each of the conditional distributions, we see that for the most part they do not overlap. The range (I am defining for this exercise as 3 SDs out in either direction) of the first is [57.11, 109.55], the second is [131.3, 202.04], and the third is [208.93, 291.07]. Due to the little overlap we are able to say that $\frac{1}{4}$ of the distribution is in the first curve, $\frac{1}{2}$ is in the second, and $\frac{1}{4}$ is in the third.

1. 5% would be 20% into the first bell curve. We will use the standard normal distribution to approximate where that 20% would fall (-0.84 in the standard normal curve).
$$83.3 - (0.84)8.7 = 75.9$$
2. 25% would be directly in between the first two curves. It appears to be around 120 (between 109.55 and 131.3).
3. 50% would be in the middle/mean of the second bell curve: 166.67.
4. 75% would be directly in between the second and third curves. It appears to be around 205 (between 202.04 and 208.93).
5. 95% would be 80% into the last bell curve. We will again use the standard normal distribution to approximate where that 80% would fall (0.84 in the standard normal curve).
$$250 + (0.84)13.7 = 261.5$$

5 Chapter 2 Problem 7

5.1 Part A

We have the binomial likelihood: $y \sim \text{Bin}(n, \theta)$. We can write the PDF as

$$p(y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

We can put it in the form of the exponential family by using

$$p(y) = \binom{n}{y} \left(\exp(y \log \left(\frac{\theta}{1 - \theta} \right) + n\log(1 - \theta) \right)$$

Where its natural parameter is: $\eta(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$. We are told to use a uniform prior density, i.e. $p(\eta) \propto 1$ on the real line, on $\eta(\theta)$. We can transform this using the variable transformation technique we learned in class:

$$p_v(v) = |J| p_u(f^{-1}(v))$$

Let u $p(\eta) \propto 1$ and $v = \log \frac{\theta}{1-\theta}$. Therefore,

$$p_v(v) = \left| \frac{d}{d\theta} \log\left(\frac{\theta}{1-\theta}\right) \right| p_u\left(\exp\left(\frac{\theta}{1-\theta}\right)\right)$$

$$p_v(v) \propto \left| \frac{1}{\theta(1-\theta)} \right| (1)$$

$$p_v(v) \propto \frac{1}{\theta(1-\theta)} = \theta^{-1}(1-\theta)^{-1}$$

□

5.2 Part B

To find posterior distribution:

$$p(\theta|y) = p(y|\theta)p(\theta)$$

$$p(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}\theta^{-1}(1-\theta)^{-1}$$

$$p(\theta|y) \propto \theta^{y-1}(1-\theta)^{n-y-1}$$

Therefore, when $y = 0$, $p(\theta|y) \propto \theta^{-1}(1-\theta)^{n-1}$, which is improper when $\theta \to 0$. Lastly, when $y = n$, $p(\theta|y) \propto \theta^{n-1}(1-\theta)^{-1}$, which is improper when $\theta \to 1$.

6 Chapter 2 Problem 8

6.1 Part A

Our prior: $p(\theta) \sim N(180, 40^2)$. Our sampling distribution: $p(y|\theta) \sim N(\theta, 20^2)$ We are looking for the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$$p(y|\theta) \propto \prod_{i=1}^{n} \exp\left(\frac{-1}{2\sigma_0^2}(y-\theta)^2\right), p(\theta) \propto \exp\left(\frac{-1}{2\sigma_1^2}(\theta-\mu)^2\right)$$

$$p(\theta|y) \propto \exp\left[\left(\frac{-1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i-\theta)^2\right) + \left(\frac{-1}{2\sigma_1^2}(\theta-\mu)^2\right)\right]$$

$$p(\theta|y) \propto \exp -\frac{1}{2}\left[\left(\frac{\sum_{i=1}^{n}(y_i-\theta)^2}{\sigma_0^2}\right)+\left(\frac{(\theta-\mu)^2}{\sigma_1^2}\right)\right]$$
$$p(\theta|y) \propto \exp -\frac{1}{2}\left[\left(\frac{\sum_{i=1}^{n}(y_i-\theta)^2}{\sigma_0^2}\right)+\left(\frac{(\theta-\mu)^2}{\sigma_1^2}\right)\right]$$

Which tells us that:

$$p(\theta|y) \sim N\left(\frac{\frac{1}{\sigma_1^2}\mu+\frac{n}{\sigma_0^2}y}{\frac{1}{\sigma_1^2}+\frac{n}{\sigma_0^2}}, \frac{1}{\frac{1}{\sigma_1^2}+\frac{n}{\sigma_0^2}}\right)$$

As seen on page 40 of the textbook and in class. Therefore, plug in and we have:

$$p(\theta|y) \propto N\left(\frac{\frac{1}{40^2}180+\frac{n}{20^2}150}{\frac{1}{40^2}+\frac{n}{20^2}}, \frac{1}{\frac{1}{40^2}+\frac{n}{20^2}}\right)$$

6.2 Part B

As seen on page 41 of the textbook and in class:

$$E[\tilde{y}|y] = E[E[\tilde{y}|\theta, y]|y] = E[\theta|y] = \text{found above}$$

$$\text{var}[\tilde{y}|y] = E[\text{var}[\tilde{y}|\theta, y]|y] + \text{var}[E[\tilde{y}|\theta, y]|y]$$
$$\text{var}[\tilde{y}|y] = E[\sigma^2|y] + \text{var}[\theta|y] = \sigma^2_{sample} + \sigma^2_{above}$$

Therefore,

$$p(\tilde{y}|y) \propto N\left(\frac{\frac{1}{40^2}180+\frac{n}{20^2}150}{\frac{1}{40^2}+\frac{n}{20^2}}, \frac{1}{\frac{1}{40^2}+\frac{n}{20^2}}+20^2\right)$$

6.3 Part C

Find the 95% posterior interval for θ when $n = 10$. When $n = 10$ then $p(\theta|y) \propto N(150.7, 6.25)$. Therefore, $150.7 \pm 1.96 * 6.25 = [138, 163]$. Find the 95% posterior interval for $\tilde{y}$ when $n = 10$. When $n = 10$ then $p(\theta|y) \propto N(150.7, 20.95)$. Therefore, $150.7 \pm 1.96 * 20.95 = [110, 192]$.

6.4 Part D

Find the 95% posterior interval for θ when $n = 100$. When $n = 100$ then $p(\theta|y) \propto N(150.7, 1.99)$. Therefore, $150.7 \pm 1.96 * 1.99 = [146, 154]$. Find the 95% posterior interval for $\tilde{y}$ when $n = 100$. When $n = 100$ then $p(\theta|y) \propto N(150.7, 20.10)$. Therefore, $150.7 \pm 1.96 * 20.10 = [111, 189]$.

7 Chapter 2 Problem 10

7.1 Part A

Indeed, our prior distribution on N is geometric with mean 100: $p(N) = (1/100)(99/100)^{N-1}$ for $N = 1, 2, ...$ We wish to find the posterior distribution for N.

$$p(\text{study}|N) = 1/N$$

$$p(N|\text{study}) \propto p(N)p(\text{study}|N) = (1/100)(99/100)^{N-1}(1/N)$$

7.2 Part B

From above we can write and simplify a bit with c as a normalizing constant:

$$p(N|\text{study}) = c(99/100)^N(1/N)$$

We need to calculate c in order to determine the mean and standard deviation of the distribution. We know that

$$\sum_N p(N|\text{study}) = 1$$

So we can write:

$$\sum_N \frac{c}{N}(99/100)^N = 1$$

$$\sum_{N=203}^{\text{inf}} \frac{1}{N}(99/100)^N = \frac{1}{c}$$

I used wolfram alpha to approximate this sum. The result was an approximation of $\frac{1}{c} = 0.0466$ which equates to $c = 21.47$. Now we investigate for the mean (expectation) and standard deviation (square root of the variance).

$$E(N|\text{study}) = \sum_{N=203}^{\text{inf}} Np(N|\text{study}) = c\sum_{N=203}^{\text{inf}}(99/100)^N$$

$$E(N|\text{study}) = 21.47\sum_{N=203}^{\text{inf}}(99/100)^N$$

$$E(N|\text{study}) \approx 279.1$$

$$\mathrm{SD}(N|\mathrm{study}) = \sqrt{\sum_{N=203}^{\mathrm{inf}} (N-279.1)^2 \frac{c}{N} (99/100)^N}$$

$$\mathrm{SD}(N|\mathrm{study}) = \sqrt{\sum_{N=203}^{\mathrm{inf}} (N-279.1)^2 \frac{21.47}{N} (99/100)^N}$$

$$\mathrm{SD}(N|\mathrm{study}) \approx 79.6$$

7.3 Part C

Let us try a non-informative prior distribution of $p(N) \propto \frac{1}{N}$.

$$p(\mathrm{study}|N) = 1/N$$

$$p(N|\mathrm{study}) \propto p(N)p(\mathrm{study}|N) = \left(\frac{1}{N}\right)^2$$

From above we can write and simplify with c as a normalizing constant:

$$p(N|\mathrm{study}) = c\left(\frac{1}{N}\right)^2$$

We need to calculate c in order to determine the mean and standard deviation of the distribution. We know that

$$\sum_N p(N|\mathrm{study}) = 1$$

So we can write:

$$\sum_N c\left(\frac{1}{N}\right)^2 = 1$$

$$\sum_{N=203}^{\mathrm{inf}} \left(\frac{1}{N}\right)^2 = \frac{1}{c}$$

I also used wolfram alpha to approximate this sum. The result was an approximation of $\frac{1}{c} = 0.0049$ which equates to $c = 202.84$.
Now we investigate for the mean (expectation) and standard deviation (square root of the variance).

$$E(N|\mathrm{study}) = \sum_{N=203}^{\mathrm{inf}} Np(N|\mathrm{study}) = c\sum_{N=203}^{\mathrm{inf}} \left(\frac{1}{N}\right)^2$$

$$E(N|\text{study}) = 202.84 \sum_{N=203}^{\text{inf}} \left(\frac{1}{N}\right)^2$$
$$E(N|\text{study}) \approx 1.00168$$

$$\text{SD}(N|\text{study}) = \sqrt{\sum_{N=203}^{\text{inf}} (N - 1.00168)^2 \frac{c}{N} \left(\frac{1}{N}\right)^2}$$
$$\text{SD}(N|\text{study}) = \sqrt{\sum_{N=203}^{\text{inf}} (N - 1.00168)^2 \frac{202.84}{N} \left(\frac{1}{N}\right)^2}$$

8 Chapter 2 Problem 13

8.1 Part A

Indeed, $p(y_i|\theta) \sim \text{Poisson}(\theta)$ where y_i is the number of fatal accidents in year i and θ is the expected number a year. Let's set the prior distribution for θ as $p(\theta) \sim \text{Gamma}(\alpha, \beta)$. Using the fact that it has a conjugate family, we know that the posterior is:

$$p(\theta|y) \sim \text{Gamma}(\alpha + n\bar{y}, \beta + n)$$

(demonstrated on page 44 of the textbook). Now let us find the posterior distribution using the discrete data in Table 2.2. Since we have 10 samples, let's keep it simple and start with a prior where $\alpha = \beta = 0$. Therefore, the posterior would be $\theta|y \sim \text{Gamma}(238, 10)$. We wish to predict the number of fatal accidents in 1986, i.e. $\tilde{y}$. Given θ, the predictive distribution for $\tilde{y}$ is the original $\text{Poisson}(\theta)$.

I chose to compute using simulation (see code below). This resulted in a 95% predictive interval of [15, 35] (sometimes produced 14 or 34) for the number of fatal accidents in 1986.

```
ytilde <- rpois(10000,rgamma(10000,238)/10)
ytilde.ordered <- (sort(ytilde))
interval <- ytilde.ordered[c(250,9750)]
```

8.2 Part B

See table of the calculated number of passenger miles per year on the next page. These numbers were calculated in the following manner:

$$\text{No. of passenger miles} = \frac{\text{Passenger deaths}}{\text{death rate} * 100 \text{ million miles}}$$

Now, let $p(y_i|x_i, \theta) \sim \text{Poisson}(x_i, \theta)$ where x_i is the number of passenger miles flown in year i. Let the prior distribution for θ again be $p(\theta) \sim \text{Gamma}(0, 0)$. Then the posterior will be:

$$p(\theta|y, x) \sim \text{Gamma}(\alpha + n\bar{y}, \beta + n\bar{x})$$

$$\theta|y, x \sim \text{Gamma}(238, 5.716 * 10^{12})$$

Given θ, the predictive distribution for $\tilde{y}$ is now $\text{Poisson}(\bar{x}\theta) = \text{Poisson}(8 * 10^{11}\theta)$.

I chose to compute using simulation (see code below). This resulted in a 95% predictive interval of [22, 46] for the number of fatal accidents in 1986.

```
ytilde <- rpois(10000,rgamma(10000,238)/5.716e12*8e11)
ytilde.ordered <- (sort(ytilde))
interval <- ytilde.ordered[c(250,9750)]
```

	No. of passenger miles
1976	3.863 * 10^11
1977	4.300 * 10^11
1978	5.027 * 10^11
1979	5.481 * 10^11
1980	5.814 * 10^11
1981	6.033 * 10^11
1982	5.877 * 10^11
1983	6.223 * 10^11
1984	7.433 * 10^11
1985	7.106 * 10^11

8.3 Part C

Everything is the exact same from part A except the substitution of the value of $\bar{y} = 6919$, where 6919 is the sum of the number of deaths in the data. This results in a 95% predictive interval of [639, 747] for the number of passenger deaths in 1986.

```
ytilde <- rpois(10000,rgamma(10000,6919)/10)
ytilde.ordered <- (sort(ytilde))
interval <- ytilde.ordered[c(250,9750)]
```

8.4 Part D

Everything is the exact same from part B except the substitution of the value of $\bar{y} = 6919$, where 6919 is the sum of the number of deaths in the data (the rate stays the same). This results in a 95% predictive interval of [903, 1034] for the number of passenger deaths in 1986.

```
ytilde <- rpois(10000,rgamma(10000,6919)/5.716e12*8e11)
ytilde.ordered <- (sort(ytilde))
interval <- ytilde.ordered[c(250,9750)]
```

8.5 Part E

I think that it is poor to assume that passenger deaths would be independent, as passengers deaths happen in clusters (the groups of people who are on the malfunctioning planes together). Therefore I think it is inappropriate to be using the Poisson model for parts C and D. I think using it for parts A and B is much more appropriate, as we do expect the accidents themselves to be independent.

9 Chapter 2 Problem 15***

Indeed, $Z \sim Beta(\alpha, \beta)$. Therefore:

$$p(Z) \propto Z^{\alpha-1}(1-Z)^{\beta-1}$$

We wish to find $E[Z^m(1-Z)^n]$ and we will make use of the law of the unconscious statistician.

$$E[Z^m(1-Z)^n] = \int Z^n(1-Z)^m Z^{\alpha-1}(1-Z)^{\beta-1}dZ$$

$$E[Z^m(1-Z)^n] = \int Z^{n+\alpha-1}(1-Z)^{m+\beta-1}dZ$$

If we make use of the rule stated in the problem statement:

$$E[Z^m(1-Z)^n] = \int_0^1 Z^{n+\alpha-1}(1-Z)^{m+\beta-1}dZ$$

$$E[Z^m(1-Z)^n] = \frac{\Gamma(n+\alpha)\Gamma(m+\beta)}{\Gamma(n+\alpha+m+\beta)}$$

Then, we note that if $m = 1$ and $n = 0$ we can use our above rule to solve for:

$$E[Z] = E[Z^1(1-Z)^0]$$

$$E[Z] = E[Z^1(1-Z)^0] = \frac{\Gamma(0+\alpha)\Gamma(\beta+1)}{\Gamma(0+\alpha+\beta+1)}$$

$$E[Z] = \frac{\Gamma(\alpha)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)}$$

Lastly, we want the variance, $Var(Z)$. Recall:

$$Var(Z) = E[Z^2] - E[Z]^2$$

To acquire $E[Z^2]$, we can set $m = 2$ and $n = 0$ as follows:

$$E[Z^2] = E[Z^2(1-Z)^0]$$

$$E[Z^2] = E[Z^2(1-Z)^0] = \frac{\Gamma(0+\alpha)\Gamma(\beta+2)}{\Gamma(0+\alpha+\beta+2)}$$

$$E[Z^2] = \frac{\Gamma(\alpha)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+2)}$$

Therefore:

$$Var(Z) = \frac{\Gamma(\alpha)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+2)} - \left(\frac{\Gamma(\alpha)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)}\right)^2$$

10 Chapter 2 Problem 20

10.1 Part A

We know from the problem statement that $p(\theta) \sim \text{Gamma}(\alpha, \beta)$ and $p(y|\theta) \sim \exp(\theta)$. We can expand these to write $p(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta}$ and $p(y|\theta) = \theta e^{-\theta y}$. Therefore, $p(y \geq 100|\theta) = \int_{100}^{\infty} \theta e^{-y\theta} d\theta \propto e^{-100\theta}$. To find the posterior:

$$p(\theta|y \geq 100) = p(y \geq 100|\theta)p(\theta)$$

$$p(\theta|y \geq 100) \propto e^{-100\theta}\theta^{\alpha-1}e^{-\beta\theta}$$

$$p(\theta|y \geq 100) \propto e^{(-100\theta-\beta\theta)}\theta^{\alpha-1}$$

$$p(\theta|y \geq 100) \propto e^{(-(100+\beta)\theta)}\theta^{\alpha-1}$$

$$p(\theta|y \geq 100) \propto \text{Gamma}(\alpha, \beta+100)$$

Therefore the posterior mean of θ is $\frac{\alpha}{\beta+100}$ and the posterior variance of θ is $\frac{\alpha}{(\beta+100)^2}$ per the definition of the Gamma distribution.

10.2 Part B

Now we know $y = 100$. Therefore, $p(y = 100|\theta) = \int_{100}^{100} \theta e^{-y\theta} d\theta \propto \theta e^{-100\theta}$. To find the posterior:

$$p(\theta|y = 100) \propto \theta e^{(-(100+\beta)\theta)}\theta^{\alpha-1}$$

$$p(\theta|y = 100) \propto e^{(-(100+\beta)\theta)}\theta^{\alpha}$$

$$p(\theta|y = 100) \propto \text{Gamma}(\alpha + 1, \beta + 100)$$

Therefore the posterior mean of θ is $\frac{\alpha+1}{\beta+100}$ and the posterior variance of θ is $\frac{\alpha+1}{(\beta+100)^2}$ per the definition of the Gamma distribution.

10.3 Part C

The referred to identity says that **on average** the variance decreases given more information. Simply looking at the variance at exactly $y = 100$ is not the same as averaging over the variance of the entire distribution, i.e. $y|y = 100$.