PHP2530 HW1

Antonella Basso

2/12/2022

Problem 1

Basu's elephant example was described in class. In this example the circus statistician uses probability sampling to obtain an estimate of the total weight. Assume that the sample size is 1, and use the statistician estimator composed of the weight of the sampled elephant, multiplied by the inverse probability of an elephant to be sampled. Calculate the variance of this estimator (This estimator and its variance are also known as the Horvitz-Thompson estimators, and as we saw in class can produce some illogical results.)

Solution

Horvitz-Thompson Estimator:

Let Y be some property of which we want to estimate the total, τ, for some population (N) with a sample of size n. Then, the Horvitz-Thompson estimator for τ is:

$$\hat{\tau} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

where π_i is the probability that y_i is included in the sample. With a sample size of 1 with a 99% of being chosen, as in Basu's elephant example, the variance of the estimator is very large. We show this via simulation.

```
#arbitrary weights for elephants
weights_sorted = sort(rnorm(101, mean = 50, sd = 5))
probs = c(rep(1/10000, 50), .99, rep(1/10000, 50))

horv <- function(weights, probs){
  #Horvitz_Thompson estimator with sample size = 1
  sample = sample(weights_sorted, 1, prob = probs)
  return(sample/probs[which(weights_sorted == sample)])
}

sims = replicate(1000000, horv(weights_sorted, probs))
var(sims)
```

```
## [1] 2523405243
```

Problem 2

Conditional probability: approximately 1/125 of all births are fraternal twins and 1/300 of births are identical twins. Elvis Presley had a twin brother (who died at birth). What is the probability that Elvis was an identical twin? (You may approximate the probability of a boy or girl birth as 1/2.)

Solution

Possible Twin Combinations:

- Identical: MM, FF
- Fraternal: MM, FF, MF, FM

Events:

$$I = \text{identical twins}$$
$$F = \text{fraternal twins}$$
$$M = \text{male brothers}$$

Probability of Interest:

$$P(I|M) = \frac{P(M|I)P(I)}{P(M|I)P(I) + P(M|F)P(F)} = \frac{(1/2)(1/300)}{(1/2)(1/300) + (1/4)(1/125)} = \frac{5}{11} \approx 0.4545$$

Thus, the probability that Elvis was an identical twin is roughly 45.45%.

Problem 3

Simulation of a queuing problem: a clinic has three doctors. Patients come into the clinic at random, starting at 9 a.m., according to a Poisson process with time parameter 10 minutes: that is, the time after opening at which the first patient appears follows an exponential distribution with expectation 10 minutes and then, after each patient arrives, the waiting time until the next patient is independently exponentially distributed, also with expectation 10 minutes. When a patient arrives, he or she waits until a doctor is available. The amount of time spent by each doctor with each patient is a random variable, uniformly distributed between 15 and 20 minutes. The office stops admitting new patients at 4 p.m. and closes when the last patient is through with the doctor.

a) Simulate this process once. How many patients came to the office? How many had to wait for a doctor? What was their average wait? When did the office close?
b) Simulate the process 100 times and estimate the median and 50% interval for each of the summaries in (a).

Solution

```
#sim <- function(){
#  duration <- 420 #minutes (7 hrs.)
#  lambda <- 0.1 #arrival rate (number of patients per minute)
#  arrivals <- rpois(duration, lambda)

#  num_patients <- c(0) #all patients that have come in
#  waiting_patients <- c(0) #patients actively waiting to be seen

#  doc1 <- c(0)
#  doc2 <- c(0)
#  doc3 <- c(0)

  #service_time <- floor(runif(1, min=15, max=21)) #number of mins spend with doc

 # for (i in 1:length(arrivals)){
 #   if (arrivals[i]==1){
 #     num_patients <- c(num_patients, num_patients[i-1] + 1)
 #   }
 #   else {
 #     num_patients <- c(num_patients, num_patients[i-1])
 #   }
```

```
 # }

#  for (i in 1:length(arrivals)){
#    if (arrivals[i]>=1){
#      waiting_patients[i] <- num_patients[i-1] + 1
#    }
#    else {
#      waiting_patients[i] <- num_patients[i-1]
#    }
#  }

#  for (i in 1:length(arrivals)){
#    if (waiting_patients[i]>=1){
#      doc1[i:floor(runif(1, min=15, max=21))] <- 1
#      waiting_patients[i+1] <- waiting_patients[i] - 1
#    }
#    else{
#      waiting_patients[i] <- num_patients[i-1]
#    }
#  }

  #return(data.frame(arrivals, num_patients))
#}

#sim()
```

Problem 4

Predictive distributions: let y be the number of 6's in 1000 independent rolls of a particular real die, which may be unfair. Let θ be the probability that the die lands on 6. Suppose your prior distribution for θ is as follows:

$$P(\theta = 1/12) = 0.25$$
$$P(\theta = 1/6) = 0.5$$
$$P(\theta = 1/4) = 0.25$$

a) Using the normal approximation for the conditional distributions, $p(y|\theta)$, sketch your approximate prior predictive distribution for y.

Solution

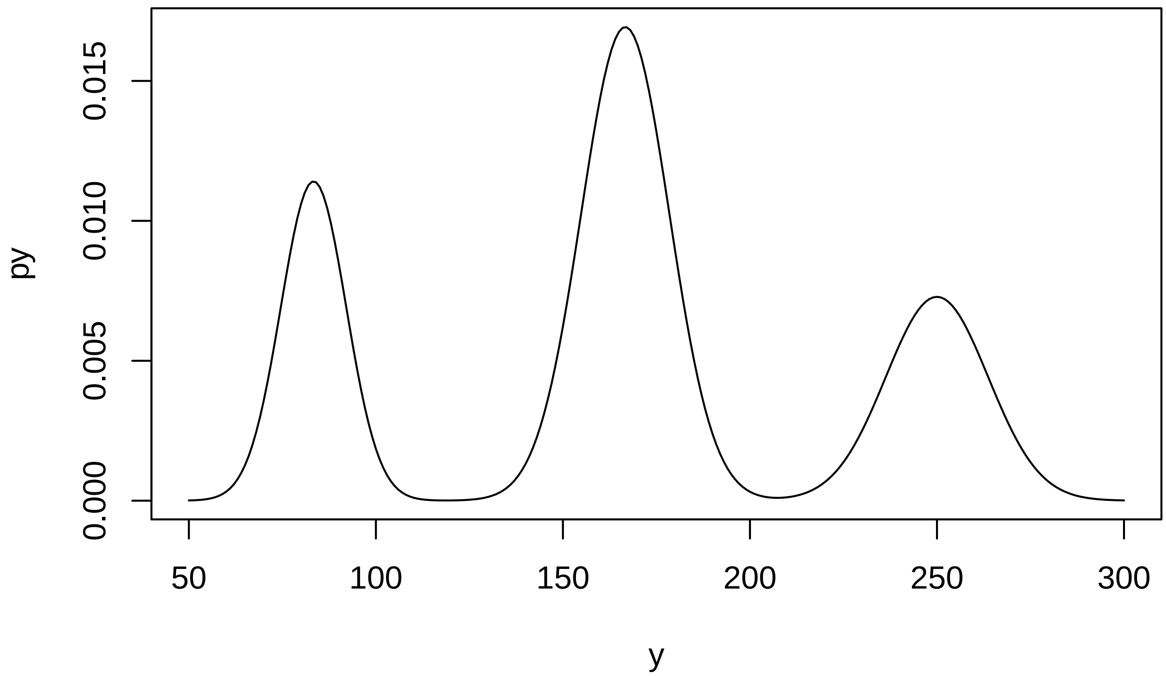
```
#Prior Predictive Distribution for Y

y <- seq(50,300,1)

p <- function (x, theta){
  dnorm (x, 1000*theta, sqrt(1000*theta*(1-theta))) #using binomial mean and sd
}

py <- 0.25*p(y,1/12) + 0.5*p(y,1/6) + 0.25*p(y,1/4)

plot(y, py, type='l')
```



b) Give approximate 5%, 25%, 50%, 75%, and 95% points for the distribution of y. (Be careful here: y does not have a normal distribution, but you can still use the normal distribution as part of your analysis.)

Solution

1) The 5% point of p(y) is the 20% point of the first hump ($P(\theta = 1/12) = 0.25$): ≈ 76
2) The 25% point of p(y) is the 99.99% point (tail) of the first hump ($P(\theta = 1/12) = 0.25$): ≈ 120
3) The 50% point of p(y) is the 50% point (peak/mean) of the second hump ($P(\theta = 1/6) = 0.5$): ≈ 167
4) The 75% point of p(y) is is the 99.99% point (tail) of the second hump ($P(\theta = 1/6) = 0.5$): ≈ 117
5) The 95% point of p(y) is the 80% point of the third hump ($P(\theta = 1/4) = 0.25$): ≈ 262

```
#1
qnorm(0.2, 1000*(1/12), sqrt(1000*(1/12)*(1-(1/12))))
```

```
## [1] 75.9775
```

```
#2
qnorm(0.99999, 1000*(1/12), sqrt(1000*(1/12)*(1-(1/12))))
```

```
## [1] 120.6088
```

```
#3
qnorm(0.5, 1000*(1/6), sqrt(1000*(1/6)*(1-(1/6))))
```

```
## [1] 166.6667
```

```
#4
qnorm(0.99999, 1000*(1/6), sqrt(1000*(1/6)*(1-(1/6))))
```

```
## [1] 216.9289
```

```
#4
qnorm(0.8, 1000*(1/4), sqrt(1000*(1/4)*(1-(1/4))))
```

```
## [1] 261.5244
```

Problem 5

Noninformative prior densities:

a) For the binomial likelihood, $y \sim \text{Bin}(n, \theta)$, show that $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is the uniform prior distribution for the natural parameter of the exponential family.

Solution

Transformation of Binomial Distribution to Natural Parameter in the Exponential Family:

$$\Phi = h(\theta) = logit(\theta) = log\left(\frac{\theta}{1-\theta}\right)$$

By Jefferey's Invariance Principle:

$$p(\Phi) = p(\theta) \cdot |h'(\theta)|^{-1} \propto [\theta^{-1}(1-\theta)^{-1}] \cdot \left|\frac{1}{\theta(1-\theta)}\right|^{-1} = \frac{1}{\theta(1-\theta)} \cdot \theta(1-\theta) = 1$$

Thus, since $p(\Phi) \propto 1$, it follows that $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is the uniform prior distribution for the natural parameter of the exponential family.

b) Show that if $y = 0$ or n, the resulting posterior distribution is improper.

Solution

For a binomial distribution modeling the number of successes (y) in n trials, each with a probability θ of success, it follows that $y = n\theta$, and similarly, that $\theta = y/n$. Thus, if $y = 0$, then $\theta = 0$. Likewise, if $y = n$, then $\theta = 1$. Both such cases, give us undefined priors, resulting in a posterior that does not integrate to 1 and is hence, improper.

Problem 6

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.)

Solution

Given that $\bar{y}|\theta, \sigma^2 \sim \text{N}(\theta, \sigma^2/n)$ and $\theta \sim \text{N}(\mu_0, \tau_0^2)$, it follows that $\theta|\bar{y} \sim \text{N}(\mu_n, \tau_n^2)$, where:

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \text{ and } \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

Thus, the posterior distribution for θ is normal with the given parameters:

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

*Note:

- $\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$, so $\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$

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.)

Solution

Given the Adam and Eve laws and the results from part (a), it follows that the posterior predictive distribution is $\tilde{y}|y \sim \mathrm{N}(\mu_n, \sigma_n^2 + \tau_n^2)$. Specifically,

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

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

Solution

```
#Computing 95% Posterior Interval for Theta, n=10

mu_10 <- (((1/(40^2))*180)+((10/(20^2))*150))/((1/(40^2))+(10/(20^2)))
tau2_10 <- 1/((1/(40^2))+(10/(20^2)))

theta_lower_10 <- mu_10-(1.96*sqrt(tau2_10))
theta_upper_10 <- mu_10+(1.96*sqrt(tau2_10))

c(theta_lower_10, theta_upper_10)
```

```
## [1] 138.4877 162.9757
```

```
#Computing 95% Posterior Predictive Interval for New Y, n=10

y_lower_10 <- mu_10-(1.96*sqrt(tau2_10+(20^2)))
y_upper_10 <- mu_10+(1.96*sqrt(tau2_10+(20^2)))

c(y_lower_10, y_upper_10)
```

```
## [1] 109.6640 191.7994
```

d) Do the same for $n = 100$.

Solution

```
#Computing 95% Posterior Interval for Theta, n=100

mu_100 <- (((1/(40^2))*180)+((100/(20^2))*150))/((1/(40^2))+(100/(20^2)))
tau2_100 <- 1/((1/(40^2))+(100/(20^2)))

theta_lower_100 <- mu_100-(1.96*sqrt(tau2_100))
theta_upper_100 <- mu_100+(1.96*sqrt(tau2_100))

c(theta_lower_100, theta_upper_100)
```

```
## [1] 146.1597 153.9899
```

```
#Computing 95% Posterior Predictive Interval for New Y, n=100

y_lower_100 <- mu_100-(1.96*sqrt(tau2_100+(20^2)))
y_upper_100 <- mu_100+(1.96*sqrt(tau2_100+(20^2)))

c(y_lower_100, y_upper_100)
```

```
## [1] 110.6798 189.4698
```

Problem 7

Discrete sample spaces: suppose there are N cable cars in San Francisco, numbered sequentially from 1 to N. You see a cable car at random; it is numbered 203. You wish to estimate N.

(a) Assume your prior distribution on N is geometric with mean 100; that is, $p(N) = (\frac{1}{100})(\frac{99}{100})^{N-1}$, for $N = 1, 2, ...$ What is your posterior distribution for N?

Solution

Prior Distribution:

$$p(N) = \left(\frac{1}{100}\right)\left(\frac{99}{100}\right)^{N-1}$$

Likelihood:

$$p(y|N) = \begin{cases} \frac{1}{N} & \text{if } N \geq 203 \\ 0 & \text{otherwise} \end{cases}$$

Posterior Distribution:

$$p(N|y) \propto p(N)p(y|N) = \left(\frac{1}{100}\right)\left(\frac{99}{100}\right)^{N-1} \cdot \frac{1}{N} \propto \frac{1}{N}\left(\frac{99}{100}\right)^{N-1}$$

(b) What are the posterior mean and standard deviation of N? (Sum the infinite series analytically or approximate them on the computer.)

Solution

Normalizing Constant:

$$p(N|y) \propto \frac{1}{N}\left(\frac{99}{100}\right)^{N-1}$$

$$p(N|y) = c \cdot \frac{1}{N}\left(\frac{99}{100}\right)^{N-1}$$

$$\frac{1}{c} = \sum_{N=203}^{\infty} \frac{1}{N}\left(\frac{99}{100}\right)^{N-1}$$

```r
#Normalizing Constant

N <- 203

infsum <- function(N){
  x = (1/N)*((99/100)^(N-1))
  return(x)
}

cumN <- c()

for (i in N:1000){
  cumN <- c(cumN, infsum(i))
}

tail(cumsum(cumN), n=1) #1/c
```

```
## [1] 0.04704688
```

```r
c <- 1/(tail(cumsum(cumN), n=1))

c
```

```
## [1] 21.25539
```

$$E(N|y) = \sum_{N=203}^{\infty} Np(N|y) = \sum_{N=203}^{\infty} N \cdot \frac{1}{N}\left(\frac{99}{100}\right)^{N-1} \cdot c = \sum_{N=203}^{\infty} \left(\frac{99}{100}\right)^{N-1} \cdot c$$

```r
#Posterior Mean/Expectation

infsum2 <- function(N){
  x = c*((99/100)^(N-1))
  return(x)
}

cumN2 <- c()

for (i in N:1000){
  cumN2 <- c(cumN2, infsum2(i))
}

post_exp <- tail(cumsum(cumN2), n=1)

post_exp
```

```
## [1] 279.0202
```

$$Var(N|y) = E(N^2|y) - E(N|y)^2 = \left(\sum_{N=203}^{\infty} N^2 p(N|y)\right) - E(N|y)^2$$

$$= \left(\sum_{N=203}^{\infty} N^2 \cdot \frac{1}{N}\left(\frac{99}{100}\right)^{N-1} \cdot c\right) - E(N|y)^2 = \left(\sum_{N=203}^{\infty} N\left(\frac{99}{100}\right)^{N-1} \cdot c\right) - E(N|y)^2$$

```r
#Posterior Standard Deviation

infsum3 <- function(N){
  x = N*c*((99/100)^(N-1))
  return(x)
}

cumN3 <- c()

for (i in N:1000){
  cumN3 <- c(cumN3, infsum3(i))
}

post_var <- tail(cumsum(cumN3), n=1) - (post_exp^2)
post_sd <- sqrt(post_var)

post_sd
```

```
## [1] 79.61534
```

(c) Choose a reasonable 'noninformative' prior distribution for N and give the resulting posterior distribution, mean, and standard deviation for N.

Solution

Prior Distribution:

$$p(N) = \begin{cases} \frac{1}{N} & \text{if } N \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Likelihood:

$$p(y|N) = \begin{cases} \frac{1}{N} & \text{if } N \geq 203 \\ 0 & \text{otherwise} \end{cases}$$

Posterior Distribution:

$$p(N|y) \propto p(N)p(y|N) = \frac{1}{N} \cdot \frac{1}{N} = \frac{1}{N^2}$$

Normalizing Constant:

$$p(N|y) = c \cdot \frac{1}{N^2}$$

$$\frac{1}{c} = \sum_{N=203}^{\infty} \frac{1}{N^2}$$

```
#Normalizing Constant

infsum4 <- function(N){
  x = 1/(N^2)
  return(x)
}

cumN4 <- c()

for (i in N:100000){
  cumN4 <- c(cumN4, infsum4(i))
}

tail(cumsum(cumN4), n=1) #1/c
```

```
## [1] 0.004928262
```

```
c2 <- 1/(tail(cumsum(cumN4), n=1))

c2
```

```
## [1] 202.9113
```

$$E(N|y) = \sum_{N=203}^{\infty} Np(N|y) = \sum_{N=203}^{\infty} N \cdot \frac{1}{N^2} \cdot c = \sum_{N=203}^{\infty} \frac{c}{N}$$

```
#Posterior Mean/Expectation

infsum5 <- function(N){
  x = c2/N
  return(x)
}
```

```r
cumN5 <- c()

for (i in N:100){
  cumN5 <- c(cumN5, infsum5(i))
}

post_exp2 <- tail(cumsum(cumN5), n=1)

post_exp2
```

```
## [1] 145.1841
```

Problem 8

Discrete data: Table 2.2 gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period. We use these data as a numerical example for fitting discrete data models.

(a) Assume that the numbers 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 fatal accidents in 1986. You can use the normal approximation to the gamma and Poisson or compute using simulation.

Solution

Let y_i be the number of fatal accidents in year i, for $i = 1, 2, ..., n$ (years in the data, $n = 10$), and θ be the expected number of accidents in a year.

Prior Distribution:

Using Gamma(0,0) as a noninformative conjugate prior for the Poisson likelihood,

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} = \theta^{-1}$$

Likelihood: $y_i|\theta \sim \text{Poisson}(\theta)$

$$p(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}e^{-\theta}}{y_i!} = \frac{\theta^{\sum y_i}e^{-n\theta}}{\prod_{i=1}^{n} y_i!} \propto \theta^{n\bar{y}}e^{-n\theta}$$

Posterior Distribution:

$$p(\theta|y) \propto p(\theta)p(y|\theta) = \theta^{-1} \cdot \theta^{n\bar{y}}e^{-n\theta} = \theta^{n\bar{y}-1}e^{-n\theta}$$

$$\theta|y \sim \text{Gamma}(n\bar{y}, n)$$

$$\theta|y \sim \text{Gamma}(238, 10)$$

*Note that:

- $\alpha = \beta = 0$
- $\frac{\sum y_i}{n} = \bar{y}$, so $\sum y_i = n\bar{y}$
- $n = 10$
- $\bar{y} = 23.8$

```r
#Computing 95% Predictive Interval for New Y (via Simulation)

post_theta <- rgamma(10000, 238, 10)
```

```r
y_pred <- rpois(10000, post_theta)

#hist(y_pred)

lower <- mean(y_pred)-(1.96*sd(y_pred))
upper <- mean(y_pred)+(1.96*sd(y_pred))

c(lower, upper)
```

```
## [1] 13.81175 33.81565
```

(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 Table 2.2 and ignoring round-off errors.) Give a 95% predictive interval for the number of fatal accidents in 1986 under the assumption that 8×10^{11} passenger miles are flown that year.

Solution

Let x_i be the number of passenger miles flown in year i, for $i = 1, 2, ..., n$ (years in the data, $n = 10$), and and θ be the expected accident rate per passenger mile.

Calculating x_i:

Year	x_i
1976	$734/0.19 \times 100$ million $= 3.863 \times 10^{11}$
1977	$516/0.12 \times 100$ million $= 4.300 \times 10^{11}$
1978	$754/0.15 \times 100$ million $= 5.027 \times 10^{11}$
1979	$877/0.16 \times 100$ million $= 5.481 \times 10^{11}$
1980	$814/0.14 \times 100$ million $= 5.814 \times 10^{11}$
1981	$362/0.06 \times 100$ million $= 6.033 \times 10^{11}$
1982	$764/0.06 \times 100$ million $= 5.877 \times 10^{11}$
1983	$809/0.13 \times 100$ million $= 6.223 \times 10^{11}$
1984	$223/0.03 \times 100$ million $= 7.433 \times 10^{11}$
1985	$1066/0.15 \times 100$ million $= 7.106 \times 10^{11}$

Prior Distribution:

Using Gamma(0,0) as a noninformative conjugate prior for the Poisson likelihood,

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} = \theta^{-1}$$

Likelihood: $y_i|x_i, \theta \sim \text{Poisson}(\theta x_i)$

$$p(y|x,\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} x_i^{y_i} e^{-\theta x_i}}{y_i!} = \frac{\theta^{\sum y_i} \prod_{i=1}^{n} x_i^{y_i} e^{-\theta \sum x_i}}{\prod_{i=1}^{n} y_i!} \propto \theta^{n\bar{y}} e^{-n\bar{x}\theta}$$

Posterior Distribution:

$$p(\theta|y,x) \propto p(\theta)p(y|x,\theta) = \theta^{-1} \cdot \theta^{n\bar{y}} e^{-n\bar{x}\theta} = \theta^{n\bar{y}-1} e^{-n\bar{x}\theta}$$

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

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

*Note that:

- $\alpha = \beta = 0$
- $\frac{\sum y_i}{n} = \bar{y}$, so $\sum y_i = n\bar{y}$
- $\frac{\sum x_i}{n} = \bar{x}$, so $\sum x_i = n\bar{x}$
- $n = 10$
- $\bar{y} = 23.8$
- $\bar{x} = 5.716 \times 10^{11}$

```
#Computing 95% Predictive Interval for New Y (via Simulation)

post_theta2 <- rgamma(10000, 238, 5.716e12)
y_pred2 <- rpois(10000, post_theta2*8e11)

#hist(y_pred2)

lower2 <- mean(y_pred2)-(1.96*sd(y_pred2))
upper2 <- mean(y_pred2)+(1.96*sd(y_pred2))

c(lower2, upper2)
```

```
## [1] 21.18796 45.42544
```

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

Solution

Let y_i be the number of passenger deaths in year i, for $i = 1, 2, ..., n$ (years in the data, $n = 10$), and θ be the expected number of deaths in a year.

Prior Distribution:

Using Gamma(0,0) as a noninformative conjugate prior for the Poisson likelihood,

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} = \theta^{-1}$$

Likelihood: $y_i|\theta \sim \text{Poisson}(\theta)$

$$p(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}e^{-\theta}}{y_i!} = \frac{\theta^{\sum y_i}e^{-n\theta}}{\prod_{i=1}^{n} y_i!} \propto \theta^{n\bar{y}}e^{-n\theta}$$

Posterior Distribution:

$$p(\theta|y) \propto p(\theta)p(y|\theta) = \theta^{-1} \cdot \theta^{n\bar{y}}e^{-n\theta} = \theta^{n\bar{y}-1}e^{-n\theta}$$

$$\theta|y \sim \text{Gamma}(n\bar{y}, n)$$

$$\theta|y \sim \text{Gamma}(6919, 10)$$

*Note that:

- $\alpha = \beta = 0$
- $\frac{\sum y_i}{n} = \bar{y}$, so $\sum y_i = n\bar{y}$
- $n = 10$
- $\bar{y} = 691.9$

```r
#Computing 95% Predictive Interval for New Y (via Simulation)

post_theta3 <- rgamma(10000, 6919, 10)
y_pred3 <- rpois(10000, post_theta3)

#hist(y_pred3)

lower3 <- mean(y_pred3)-(1.96*sd(y_pred3))
upper3 <- mean(y_pred3)+(1.96*sd(y_pred3))

c(lower3, upper3)
```

```
## [1] 638.0961 746.4653
```

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

Solution

Let x_i be the number of passenger miles flown in year i, for $i = 1, 2, ..., n$ (years in the data, $n = 10$), and θ be the expected death rate per passenger mile.

Prior Distribution:

Using Gamma(0,0) as a noninformative conjugate prior for the Poisson likelihood,

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} = \theta^{-1}$$

Likelihood: $y_i|x_i, \theta \sim \text{Poisson}(\theta x_i)$

$$p(y|x,\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} x_i^{y_i} e^{-\theta x_i}}{y_i!} = \frac{\theta^{\sum y_i} \prod_{i=1}^{n} x_i^{y_i} e^{-\theta \sum x_i}}{\prod_{i=1}^{n} y_i!} \propto \theta^{n\bar{y}} e^{-n\bar{x}\theta}$$

Posterior Distribution:

$$p(\theta|y,x) \propto p(\theta)p(y|x,\theta) = \theta^{-1} \cdot \theta^{n\bar{y}} e^{-n\bar{x}\theta} = \theta^{n\bar{y}-1} e^{-n\bar{x}\theta}$$

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

$$\theta|y,x \sim \text{Gamma}(6919, 5.716 \times 10^{12})$$

*Note that:

- $\alpha = \beta = 0$
- $\frac{\sum y_i}{n} = \bar{y}$, so $\sum y_i = n\bar{y}$
- $\frac{\sum x_i}{n} = \bar{x}$, so $\sum x_i = n\bar{x}$
- $n = 10$
- $\bar{y} = 691.9$
- $\bar{x} = 5.716 \times 10^{11}$

```r
#Computing 95% Predictive Interval for New Y (via Simulation)

post_theta4 <- rgamma(10000, 6919, 5.716e12)
y_pred4 <- rpois(10000, post_theta4*8e11)

#hist(y_pred4)
```

```r
lower4 <- mean(y_pred4)-(1.96*sd(y_pred4))
upper4 <- mean(y_pred4)+(1.96*sd(y_pred4))

c(lower4, upper4)
```

```
## [1]  902.5166 1034.7578
```

(e) In which of the cases (a)-(d) above does the Poisson model seem more or less reasonable? Why? Discuss based on general principles, without specific reference to the numbers in Table 2.2.

Solution

The most reasonable of the four models above is that from part (d). This is because it generally makes more sense to model passenger deaths using a Poisson distribution whose unknown parameter is precisely a death rate, which we may reasonably assume is proportional to the number of passenger miles flown in one unit of time (that is, we expect to see more deaths in years where more passenger miles were flown). However, as we may not assume that deaths are independent events in this case, a Poisson model may be more suitable for modeling fatal accidents, as in part (a), which does not violate this assumption.

Problem 9

Posterior intervals: unlike the central posterior interval, the highest posterior interval is **not** invariant to transformation. For example, suppose that, given σ^2, the quantity nv/σ^2 is distributed as χ^2_n, and that σ has the (improper) noninformative prior density $p(\sigma) \propto \sigma^{-1}, \sigma > 0$.

a) Prove that the corresponding prior density for σ^2 is $p(\sigma^2) \propto \sigma^{-2}$.

Solution

Let $\psi = h(\sigma) = \sigma^2$, and hence, $\sigma = \sqrt{\psi}$. Then,

$$p(\psi) = p(\sigma) \cdot \left|\frac{d\sigma}{d\psi}\right| = p(\sigma) \cdot \left|\frac{d}{d\psi}\sqrt{\psi}\right| = p(\sigma) \cdot \left(\frac{1}{2}\psi^{-\frac{1}{2}}\right) = p(\sigma) \cdot \left(\frac{1}{2}(\sigma^2)^{-\frac{1}{2}}\right)$$

$$\propto \sigma^{-1} \cdot \frac{1}{2}\sigma^{-1} = \frac{1}{2\sigma^2}$$

Therefore, $p(\sigma^2) \propto \sigma^{-2}$.

b) Show that the 95% highest posterior density region for σ^2 is not the same as the region obtained by squaring the endpoints of a posterior interval for σ.

Solution

Problem 10

Censored and uncensored data in the exponential model:

a) Suppose $y|\theta$ is exponentially distributed with rate θ, and the marginal (prior) distribution of θ is Gamma(α, β). Suppose we observe that $y \geq 100$, but do not observe the exact value of y. What is the posterior distribution, $p(\theta|y \geq 100)$, as a function of α and β? Write down the posterior mean and variance of θ.

Solution

Prior Distribution: $\theta \sim \text{Gamma}(\alpha, \beta)$

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

Likelihood: $y|\theta \sim \text{Exp}(\theta)$

*Using the sampling distribution for an outcome y, given θ, (assuming $n = 0$).

$$p(y|\theta) = \theta e^{-y\theta} \text{ for } y > 0$$

$$p(y \geq 100|\theta) = e^{-100\theta}$$

Posterior Distribution:

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

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

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

Posterior Mean & Variance of θ:

$$E(\theta|y \geq 100) = \frac{\alpha}{\beta + 100}$$

$$Var(\theta|y \geq 100) = \frac{\alpha}{(\beta + 100)^2}$$

b) In the above problem, suppose that we are now told that y is exactly 100. Now what are the posterior mean and variance of θ?

Solution

Prior Distribution: $\theta \sim \text{Gamma}(\alpha, \beta)$

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

Likelihood: $y|\theta \sim \text{Exp}(\theta)$

*Using the sampling distribution for an outcome y, given θ, (assuming $n = 1$).

$$p(y|\theta) = \theta e^{-y\theta} \text{ for } y > 0$$

$$p(y = 100|\theta) = \theta e^{-100\theta}$$

Posterior Distribution:

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

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

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

Posterior Mean & Variance of θ:

$$E(\theta|y = 100) = \frac{\alpha + 1}{\beta + 100}$$

$$Var(\theta|y = 100) = \frac{\alpha + 1}{(\beta + 100)^2}$$

c) Explain why the posterior variance of θ is higher in part (b) even though more information has been observed. Why does this not contradict identity (2.8) on page 32?

Solution

Although more information has been observed in part (b), the posterior variance we obtain in part (a) is not in fact the prior variance of part (b). The Law of Total Variance implies that the prior variance is greater than or at least equal to the posterior variance (if there is no variation in posterior means). However, since $Var(\theta|y \geq 100)$ is not actually the prior variance of θ when we have posterior $Var(\theta|y = 100)$ (and is also not averaging over all possible values of y), it is safe to say that these results do not contradict the identity (2.8) on page 32.