STAT 886 HW 3

David Nguyen

2021-02-23

2.12 Jeffrey's prior distribution:

Suppose $y|\theta \sim Poisson(\theta)$. Find Jeffrey's prior density for θ and give α and β for which the Gamma(α, β) is a close match to the Jeffrey's prior.

The Jeffrey's prior is given by $p(\theta) \propto [J(\theta)]^{1/2}$ where $J(\theta)$ is the Fisher's information for θ:

$$J(\theta) = -E\left(\frac{d^2 logp(y|\theta)}{d\theta^2}\middle|\theta\right)$$

.

For the poisson distribution:

$$\begin{aligned}
p(y|\theta) &= \frac{\theta^y \theta^{-\theta}}{y!} \\
log(p(y|\theta)) &= ylog(\theta) - \theta - log(y!) \\
\frac{dlog(p(y|\theta))}{d\theta} &= \frac{y}{\theta} - 1 \\
\frac{d^2 log(p(y|\theta))}{d\theta^2} &= -\frac{y}{\theta^2} \\
-E\left(\frac{dlog(p(y|\theta))}{d\theta}\right) &= -E\left(-\frac{y}{\theta^2}\right) \\
&= \frac{E(y)}{\theta^2} \\
&= \frac{\theta}{\theta^2} \\
&= \frac{1}{\theta}
\end{aligned}$$

Thus the Jeffrey's prior for the Poisson distribution is $p(\theta) \propto \sqrt{\frac{1}{\theta}} = \theta^{-1/2}$.

This can be written as Gamma($\alpha = 1/2, \beta = 0$).

2.13 Discrete data:

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.

```r
flight <- data.frame(
  matrix(c(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, 0.13,
           1983, 20, 809, 0.13,
           1984, 16, 223, 0.03,
           1985, 22, 1066, 0.15),
         byrow = TRUE, nrow = 10, ncol = 4))
names(flight) <- c("Year", "Fatal_Accidents", "Deaths", "Death_rate")
knitr::kable(flight)
```

Year	Fatal_Accidents	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	0.13
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

The posterior distribution of θ using the Jeffrey's prior $p(\theta) = Gamma(1/2, 0)$ is:

$$p(\theta|y) = Gamma(1/2 + 238, 10)$$

```r
# rate estimation for number of accidents per year
# update Jeffrey's prior
total_accidents <- sum(flight$Fatal_Accidents)
alpha <- 0.5 + total_accidents
beta_inv <- 1/nrow(flight)

# simulate from Jeffrey's posterior
nsims <- 1000
theta <- rgamma(nsims, shape = alpha, scale = beta_inv)
sim_accidents <- rpois(nsims, lambda = theta)

# get 95 % CrI for annual number of fatal accidents
accident_95_CI <- sort(sim_accidents)[c(nsims*0.025, nsims*0.975)]
```

The 95% credible interval for annual fatal accidents is 14 to 34.

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.

```r
# Death_rate is given in deaths per 100 million passenger miles
flight$miles <- 1/(flight$Death_rate/flight$Deaths) # passenger miles in units of 100 million
knitr::kable(flight[,c("Year", "miles")], caption = "Passenger miles (100 million)")
```

Table 2: Passenger miles (100 million)

Year	miles
1976	3863.158
1977	4300.000
1978	5026.667
1979	5481.250
1980	5814.286
1981	6033.333
1982	5876.923
1983	6223.077
1984	7433.333
1985	7106.667

```r
# rate estimation for number of accidents per year
# update Jeffrey's prior
total_miles <- sum(flight$miles)
alpha <- 0.5 + total_accidents
beta_inv <- 1/total_miles

miles1986 <- 8 * 10^3 # in units of 100 million

# simulate from Jeffrey's posterior
nsims <- 1000
theta <- rgamma(nsims, shape = alpha, scale = beta_inv)
sim_accidents <- rpois(nsims, lambda = theta*miles1986)

# get 95 % CrI for annual number of fatal accidents
accidents_95_CI <- sort(sim_accidents)[c(nsims*0.025, nsims*0.975)]
```

Under the Jeffrey's prior the posterior distribution of the fatal accidents model given the data is:

$$p(\theta|x, y) \sim Gamma(1/2 + 238, 57158.69)$$

where passenger miles are measured in units of 100 million.

The 95% credible interval for number of fatal accidents given 8×10^{11} passenger miles flown is 22 to 45.

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

```r
# rate estimation for number of deaths per year
# update Jeffrey's prior
total_death <- sum(flight$Deaths)
alpha <- 0.5 + total_death
beta_inv <- 1/nrow(flight)

# simulate from Jeffrey's posterior
nsims <- 1000
theta <- rgamma(nsims, shape = alpha, scale = beta_inv)
sim_deaths <- rpois(nsims, lambda = theta)

# get 95 % CrI for annual number of fatal accidents
deaths_95_CI <- sort(sim_deaths)[c(nsims*0.025, nsims*0.975)]
```

Under the Jeffrey's prior for posterior distribution of of the passenger death model is:

$$p(\theta|x, y) \sim Gamma(1/2 + 6919, 10)$$

The 95% credible interval for number of passenger deaths is 636 to 748.

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

```r
# rate estimation for number of accidents per year
# update Jeffrey's prior
total_miles <- sum(flight$miles)
total_deaths <- sum(flight$Deaths)
alpha <- 0.5 + total_deaths
beta_inv <- 1/total_miles

miles1986 <- 8 * 10^3 # in units of 100 million

# simulate from Jeffrey's posterior
nsims <- 1000
theta <- rgamma(nsims, shape = alpha, scale = beta_inv)
sim_deaths <- rpois(nsims, lambda = theta*miles1986)

# get 95 % CrI for annual number of fatal accidents
deaths_95_CI <- sort(sim_deaths)[c(nsims*0.025, nsims*0.975)]
```

Under the Jeffrey's prior the posterior distribution of the fatal accidents model given the data is:

$$p(\theta|x, y) \sim Gamma(1/2 + 6919, 57158.69)$$

where passenger miles are measured in units of 100 million.

The 95% credible interval for number of passenger deaths given 8×10^{11} passenger miles flown is 902 to 1038.

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.

The Poisson model is reasonable in the sense that the support of the Poisson is consistent with the possible observations of fatal accidents and numbers of passenger deaths, i.e., $0, 1, 2, 3, \ldots$. A binomial model could potentially be justified, since there is an upper bound on the total number of fatal accidents or number of passenger deaths. The models which include total passenger miles flown are more reasonable than the models that do not include this covariate since it is reasonable to believe that accidents and deaths are more likely when more aircraft and people are exposed.

2.19 Exponential model with conjugate prior distribution:

a. Show that if $y|\theta \sim exp(\theta)$ then the gamma prior is conjugate for an iid exponential sample.

$$
\begin{aligned}
p(\theta|y) &\propto p(\theta) \prod_{i=1}^{n} p(y|\theta) \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta} \theta^{n} e^{-n\bar{y}\theta} \\
&\propto \theta^{(\alpha+n)-1} e^{-(\beta+n\bar{Y})\theta}
\end{aligned}
$$

Which is the kernal of gamma$(\alpha + n, \beta + n\bar{Y})$. Therefore, the gamma distribution is conjugate for θ by the definition of conjugacy.

b. Show the inverse gamma distribution is a conjugate prior for the mean, $\phi = 1/\theta$.

$$
\begin{aligned}
p(\phi|y) &\propto p(\phi) \prod_{i=1}^{n} p(y|\phi) \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \phi^{-(\alpha+1)} e^{-\beta/\phi} (1/\phi)^{n} e^{-n\bar{y}(1/\phi)} \\
&\propto \theta^{-((\alpha+n)+1)} e^{\frac{-(\beta+n\bar{Y})}{\phi}}
\end{aligned}
$$

Which is the kernal of the inverse-gamma$(\alpha + n, \beta + n\bar{Y})$. Therefore, the inverse-gamma distribution is conjugate for ϕ by the definition of conjugacy.

c. The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5. (The coefficient of variation is defined as the standard deviation divided by the mean.) A random sample of light bulbs is to be tested and the lifetime of each obtained. If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

Define the prior $p(\theta) \sim gamma(\alpha_0, \beta_0)$ such that $CV(\theta) = 0.5$. We wish to solve for the prior parameters α_0, β_0 to determine the necessary sample size n.

$$
\begin{aligned}
0.5 &= \frac{\sqrt{Var(\theta_0)}}{E(\theta_0)} \\
&= \frac{\sqrt{\frac{\alpha}{\beta^2}}}{\frac{\alpha_0}{\beta_0}} \\
&= \alpha_0^{-1/2} \\
\implies \alpha_0 &= 4
\end{aligned}
$$

To get $CV(gamma(\alpha_n, \beta_n)) = 0.1$ we need $\alpha_n = 100 = \alpha_0 + 96$.

We need to test $n = 96$ more light bulbs to get $CV(\theta_n) = 0.1$

d. Same as c but with ϕ.

$$
\begin{aligned}
0.5 &= \frac{\sqrt{Var(\phi_0)}}{E(\phi_0)} \\
&= \frac{\sqrt{\frac{\beta_0^2}{(\alpha_0-1)^2(\alpha_0-2)}}}{\frac{\beta_0}{\alpha_0-1}} \\
&= (\alpha_0 - 2)^{-1/2} \\
\implies \alpha_0 &= 6
\end{aligned}
$$

To get $CV(\phi_n) = 0.1$ we need

$$
\begin{aligned}
0.1 &= (\alpha_n - 2)^{-1/2} \\
\implies 10 &= \sqrt{\alpha_n - 2} \\
\implies \alpha_n &= 102
\end{aligned}
$$

From the previously derived conjugate posterior distribution of ϕ we can conclude that $n = \alpha_n - \alpha_0 = 102 - 6 = 94$. Hence, our answer is unchanged.

2.20 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.

First, recall the memoryless property of the exponential distribution:

$$
P(X = x + a | x > a) = P(X = x)
$$

.

This implies that the likelihood is given by $p(\theta|y \geq 100) = e^{-100\theta}$.

$$
\begin{aligned}
p(\theta|y \geq 100) &\propto p(\theta)p(\theta|y \geq 100) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\theta\beta}e^{-100\theta} \\
&\propto \theta^{\alpha-1}e^{-(\beta+100)\theta}
\end{aligned}
$$

So,

- $\theta|y \geq 100 \sim gamma(\alpha, \beta + 100)$.
- $E(\theta|y \geq 100) = \alpha/(\beta + 100)$
- $Var(\theta|y \geq 100) = \alpha/(\beta + 100)^2$

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?

From 2.19 a we know that

- $\theta|y = 100 \sim gamma(\alpha + 1, \beta + 100)$
- $E(\theta|y = 100) = (\alpha + 1)/(\beta + 100)$
- $Var(\theta|y = 100) = (\alpha + 1)/(\beta + 100)^2$

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?

I'm not sure why.