Bayesian Statistics HW 2

Betsy Cowdery

1. 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. (See Goodman, 1952, for a discussion and references to several versions of this problem, and Jeffreys, 1961, Lee, 1989, and Jaynes, 2003, for Bayesian treatments.)

   (a) Assume your prior distribution on N is geometric with mean 100; that is,

   $$p(N) = (1/100)(99/100)^{N-1}, \text{ for } N = 1, 2, ...$$

   What is your posterior distribution for N?

   $$P(X|N) = \begin{cases} 1/N & \text{if } N \geq 203 \\ 0 & \text{if } N < 203 \end{cases}$$

   Then for $N \geq 203$

   $$\begin{aligned} P(N|X) &\propto P(N)P(X|N) \\ &= (1/N)(1/100)(99/100)^{N-1} \\ P(N|X) &\propto \frac{(99/100)^N}{N} \\ P(N|X) &= c\frac{(99/100)^N}{N} \end{aligned}$$

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

   $$c = \frac{P(N|X)}{(1/N)(99/100)^N}$$

   We also know that

   $$1 = \sum_{N=203}^{\infty} P(N|X)$$

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

   We can combine these two equations to get:

   $$\begin{aligned} \frac{P(N|X)}{(1/N)(99/100)^N} &= \frac{1}{\sum_{N=203}^{\infty} \frac{(99/100)^N}{N}} \\ P(N|X) &= \frac{(99/100)^N}{N} \frac{1}{\sum_{N=203}^{\infty} \frac{(99/100)^N}{N}} \\ &\approx \frac{(99/100)^N}{N} \frac{1}{.04658} \\ c &\approx \frac{1}{.04658} = 21.47 \end{aligned}$$

Then

$$
\begin{aligned}
E[N|X] &= \sum_{N=203}^{\infty} (N * P(N|X)) \\
&= c \sum_{N=203}^{\infty} (99/100)^N \\
&= 21.74 \frac{(99/100)^{203}}{1-(99/100)} \\
&= 279.1
\end{aligned}
$$

$$
\begin{aligned}
Var(N|X) &= \sum_{N=203}^{\infty} (N - E[N|X])^2 * P(N|X) \\
&\approx \sum_{N=203}^{\infty} (203 - 279.1)^2 * (21.47) \frac{(99/100)^{203}}{203} \\
&\approx 6336.16 \\
\sigma(N) &\approx \sqrt(6336.16) = 79.6
\end{aligned}
$$

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

I'm going to recycle some of my answer from problem 3 and use Jeffrey's prior for a geometric distribution:
Given that $\mathrm{E}[X|N] = 1/N$

$$
\begin{aligned}
I(N) &= -\mathrm{E}\left[-\frac{1}{N^2} - \frac{X-1}{(1-N)^2} \middle| \theta \right] \\
&= -\frac{1}{N^2} + \frac{\frac{1}{N} - 1}{(1-N)^2} \\
&= \frac{1}{N^2(1-N)}
\end{aligned}
$$

Therefore

$$
P_J(N) = I(N)^{\frac{1}{2}} \propto N^{-1}(1-N)^{-\frac{1}{2}} = \text{ Beta}\left(0, \frac{1}{2}\right)
$$

This is an improper distribution, but the posterior will be proper for $X \geq 1$.

$$
P(X|N) = N^{X-1}(1-N)^{N-X-1}
$$

$$
\begin{aligned}
P(N|X) &\propto P(N)P(X|N) \\
&= (N^{-1}(1-N)^{-\frac{1}{2}})(N^{X-1}(1-N)^{N-X-1}) \\
&= N^{X-2}(1-N)^{N-X-\frac{3}{2}} \\
&= \text{ Beta}(-1, \frac{1}{2})
\end{aligned}
$$

Then

$$
\begin{aligned}
E[N|X] &= \sum_{N=0}^{203} (N * N^{X-2}(1-N)^{N-X-\frac{3}{2}}) \\
&= \sum_{N=0}^{203} (N^{X-1}(1-N)^{N-X-\frac{1}{2}}) \\
Var(N|X) &= \sum_{N=203}^{\infty} (N - E[N|X])^2 * P(N|X) \\
&\approx \sum_{N=203}^{\infty} (203 - 279.1)^2 * (21.47)\frac{(99/100)^{203}}{203} \\
&\approx 6336.16 \\
\sigma(N) &\approx \sqrt{(}6336.16) = 79.6
\end{aligned}
$$

---

2. 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.

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

Passenger Miles Flown $= \{3.836, 4.3, 5.027, 5.481, 6.033, 5.877, 6.223, 7.433, 7.106\} \times 10^{11}$

(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.

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

where

$$
\begin{aligned}
y_i &= \text{number of fatal accidents in year } i \in 1, ..., 10 \\
\theta &= \text{expected number of accidents per year}
\end{aligned}
$$

Then the prior distribution for θ is $\theta \sim \text{Gamma}(\alpha, \beta)$
And the posterior distribution is $\theta|y \sim \text{Gamma}(\alpha + 10\bar{y}, \beta + 10)$
Because we have $n = 10$ we can be comfortable using an informative prior, setting $(\alpha, \beta) = 0$
Thus $\theta|y =$ Gamma$(238, 10)$ with 95% predictive interval

If $\tilde{y}$ is the number of fatal accidents in 1986, then $\tilde{y} \sim$ Poisson(θ) with a computed 95% predictive interval of $[15, 35]$

(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 19761985. (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.

Let
m_i be the number of passenger miles flown in year i
θ be expected accident rate per passenger mile
If we assume an informative prior $\theta \sim$ Gamma$(0, 0)$
Then

$$
\begin{aligned}
y_i|m_i\theta &\sim \text{ Poisson}(m_i\theta) \\
y|m\theta &\sim \text{ Poisson}(m\theta) \\
y|\theta &\sim \text{ Gamma}(10\bar{y}, 10\bar{m}) \\
&= \text{ Gamma}(238, 5.716 \times 10^{12})
\end{aligned}
$$

Then the predictive distribution $\tilde{y} \sim$ Poisson($\tilde{m}\theta$)
and the calculated 95% predictive interval is $[22, 46]$

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

Let d_i be the number of passenger deaths in year $i \in 1, ..., 10$
θ be the expected number of accidents per year
Assume an uninformative prior $\theta \sim$ Gamma$(0, 0)$
And the posterior distribution

$$
\begin{aligned}
\theta|y &\sim \text{ Gamma}(10\bar{d}, 10) \\
&= \text{ Gamma}(6919, 10)
\end{aligned}
$$

Predictive distribution $\tilde{d} \sim$ Poisson(θ)
and the calculated 95% predictive interval is $[638, 749]$

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

Let
m_i be the number of passenger miles flown in year i
θ be expected accident rate per passenger mile
If we assume an informative prior $\theta \sim$ Gamma$(0, 0)$
Then

$$
\begin{aligned}
d_i|m_i\theta &\sim \text{ Poisson}(m_i\theta) \\
d|m\theta &\sim \text{ Poisson}(m\theta) \\
d|\theta &\sim \text{ Gamma}(10\bar{d}, 10\bar{m})
\end{aligned}
$$

Then the predictive distribution $\tilde{d} \sim$ Poisson$(\tilde{m}\theta)$
and the calculated 95% predictive interval is $[905, 1032]$

(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.

In general I think that a model in which number of deaths are dependent on the number of miles flown, since intuitively if more people are flying, then the probability of someone getting in an accident is larger and vice versa. This leaves us with (b) and (d). Furthermore, while fatal accidents are independent, passenger deaths are not - since if a passenger dies it most likely is the result of a fatal accident in which other passengers may have died. This narrows down the choices to case (b) as the most reasonable of the four models.

---

3. Suppose that you keep trying Bernoulli experiments with probability of success θ until you have r failures, and you observe $X|\theta \sim NegBin(r, \theta)$ successes.

$r > 0$ is the number of failures until the experiment is stopped (success)
$x \geq r$ is the number of successes (failure)
$0 < \theta < 1$ is the probability of success

$$X|\theta \sim NegBin(r, \theta)$$
$$X|\theta \propto \left(\begin{array}{c} x+r-1 \\ x \end{array} \right) \theta^r (1-\theta)^x$$

(a) Find the conjugate prior of θ for this likelihood. How does it compare to the conjugate prior for a binomial likelihood?
Suppose we define $\theta \sim$ Beta(α, β)
Then

$$\begin{aligned} P(\theta|X) &\propto \left(\begin{array}{c} x+r-1 \\ x \end{array} \right) \theta^r (1-\theta)^x \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\text{Beta}(\alpha,\beta)} \\ &\propto \theta^{\alpha+r-1}(1-\theta)^{\beta+x-1} \end{aligned}$$

And thus $\theta|X \sim$ Beta$(\alpha + r, \beta + x)$
We can generalize this by realizing that the negative binomial and the binomial have likelihood functions of the form $\theta|X \propto c\theta^a(1-\theta)^b$ where c is a normalizing constant. Then

$$\begin{aligned} P(\theta|X) &\propto c\theta^a(1-\theta)^b \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\text{Beta}(\alpha,\beta)} \\ &\propto \theta^{\alpha+a-1}(1-\theta)^{\beta+b-1} \end{aligned}$$

And once again $\theta|X \sim$ Beta$(\alpha + a, \beta + b)$.
Thus the conjugate prior will be the same in both cases.

(b) Find Jeffreys prior for this likelihood. Is it proper? How does it compare to Jeffreys prior for a binomial likelihood?
We know that the Fisher information for θ is:

$$I(\theta) = -\text{E}\left[\frac{d^2 log P(X|\theta)}{d\theta^2} \middle| \theta \right]$$

Again, using the generalized likelihood function for the binomial and negative binomial cases,

$$\begin{aligned} I(\theta) &= -\mathrm{E}\left[\frac{d^2}{d\theta^2} \log(c(1-\theta)^a\theta^b) \middle| \theta \right] \\ &= -\mathrm{E}\left[\frac{-((\theta-1)^2 a + \theta^2 b)}{(\theta-1)^2\theta^2} \middle| \theta \right] \\ &= -\mathrm{E}\left[-\frac{a}{\theta^2} - \frac{b}{(1-\theta)^2} \middle| \theta \right] \end{aligned}$$

Then Jeffrey's prior
for Binomial:
$a = x$
$b = n - x$
Since $\mathrm{E}[X] = n\theta$

$$\begin{aligned} I(\theta) &= -\mathrm{E}\left[-\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \middle| \theta \right] \\ &= -\mathrm{E}\left[-\frac{n\theta}{\theta^2} - \frac{n-n\theta}{(1-\theta)^2} \middle| \theta \right] \\ &= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} \\ &= \frac{n}{\theta(1-\theta)} \end{aligned}$$

Therefore

$$P_J(\theta) = I(\theta)^{\frac{1}{2}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} = \text{ Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

This is a proper distribution.
for Negative Binomial:
$a = x$
$b = r$
Since $\mathrm{E}[X] = \frac{r\theta}{1-\theta}$

$$\begin{aligned} I(\theta) &= -\mathrm{E}\left[-\frac{x}{\theta^2} - \frac{r}{(1-\theta)^2} \middle| \theta \right] \\ &= -\mathrm{E}\left[-\frac{r\theta}{1-\theta} * \frac{1}{\theta^2} - \frac{r}{(1-\theta)^2} \middle| \theta \right] \\ &= \frac{r\theta}{(1-\theta)\theta^2} + \frac{r}{(1-\theta)^2} \\ &= \frac{r}{\theta^2(1-\theta)} \end{aligned}$$

Therefore

$$P_J(\theta) = I(\theta)^{\frac{1}{2}} \propto \theta^{-1}(1-\theta)^{-\frac{1}{2}} = \text{ Beta}\left(0, \frac{1}{2}\right)$$

This is a uniform distribution and therefore not proper.