Statistics 700 Homework 2

Bayesian Parametric Models

Due date: 6:00 pm (EST) Oct. 3, 2017

Fly Safe. The table below 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

Table 1: Worldwide airline fatalities. 1976-1985. Death rate is passenger deaths per 100 million passenger miles. Source: Statistical Abstract of the United States.

1. 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.

2. 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 roundi-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.

3. Repeat 1 above, replacing 'fatal accidents' with 'passenger deaths'.

4. Repeat 2 above, replacing 'fatal accidents' with 'passenger deaths'.

5. In which of the cases 1-4 above does the Poisson model seem more or less reasonable? Why? Discuss based on general principles, without specific reference to the numbers in the table.

Now we expand the model by assuming that the number of fatal accidents in year t follows a Poisson distribution with mean $\alpha + \beta t$.

1. Discuss various choices for a 'noninformative' prior for (α, β). Choose one.

2. Discuss what would be a realistic informative prior distribution for (α, β). Sketch its contours and then put it aside. Do parts 3-8 of this problem using your noninformative prior distribution from 1.

3. Write the posterior density for (α, β). What are the sufficient statistics?

4. Check that the posterior density is proper.

5. Calculate crude estimates and uncertainties for (α, β) using linear regression.

6. Plot the contours and take 1000 draws from the posterior density of (α, β).

7. Using your samples of (α, β), plot a histogram of the posterior density for the expected number of fatal accidents in 1986, $\alpha + 1986\beta$.

8. Create simulation draws and obtain a 95% predictive interval for the number of fatal accidents in 1986.

9. How does your hypothetical informative prior distribution in 2 differ from the posterior distribution in 6 and 7, obtained from the noninformative prior distribution and the data? If they disagree, discuss.

Incidentally, in 1986, there were 22 fatal accidents, 546 passenger deaths, and a deathrate of 0.06 per 100 million miles flown.

Make some progress on your project. What are the parameters of interest in your project? What kind of priors do you want to use? Why?

Optional Reading. Read one of the following papers and post your summary and thoughts on Canvas. Bonus points up to 5 will be rewarded.

1. Invariant prior distributions, Hartigan, J. (1964), Annals of Mathematical Statistics 35, 836-845.
2. The selection of prior distributions by formal rules, Kass, R. E, and Wasserman, L. (1996), Journal of the American Statistical Association 91, 1343-1370.
3. Marginalization paradoxes in Bayesian and structural inferences (with discussion), Dawid, A. P., Stone, M, and Zidek, J. V. (1973), Journal of the Royal Statistical Society B 35, 189-233.
4. Parameterization and Bayesian modeling, Gelman, A. (2004), Journal of the American Statistical Association 99, 537-545.
5. Prior distributions for variance parameters in hierarchical models, Gelman, A. (2006), Bayesian Analysis 1, 515-533.