AMS 207 Homework Assignment 2

Mary Silva

May 1 2018

Write the Bayes factor, BIC, DIC and Gelfand and Ghost criterion to compare a model where n observations are assumed to be sampled with a poisson distribution with a gamma prior, to a model where the observations are sampled from a binomial distribution, with a fixed, large, number of trials and beta prior for the probability of success.

1. Consider the data on deaths by horsekicks in the Prussian army. Fit the data using the two different models.
2. Perform a prior sensitivity analysis.
3. Present a model comparison analysis using the criteria mentioned above.

Posterior

$$M_1 : X \sim Poi(\theta) \quad \theta \sim Ga(\alpha, \beta)$$

The posterior is defined up to a constant by

$$f(\theta|X) \propto exp\{-n\theta\}\theta^{\sum_{i=1}^n X_i}\theta^{\alpha-1}exp\{-\beta\theta\}$$

Recognize the kernel for a Gamma distribution, the posterior is

$$\theta|X \sim Ga(\alpha + \sum_{i=1}^n X_i, \beta + n)$$

$$M_2 : X \sim Bin(N, p) \quad N = 1000 \quad p \sim Be(c, d)$$

The posterior is defined up to a constant by

$$f(p|x) \propto p^{\sum_{i=1}^n X_i}(1-p)^{Nn-\sum_{i=1}^n X_i}p^{c-1}(1-p)^{d-1}$$

Recognize the kernel for a Beta distribution, the posterior is

$$p|X \sim Be(c + \sum_{i=1}^n X_i, d + Nn - \sum_{i=1}^n X_i)$$

Bayes factor

In general the Bayes factor B_{12} is defined by the ratio of the marginal

$$B_{12} = \frac{m_1(X)}{m_2(X)}$$

$$\begin{aligned}
m_1(X) &= \int [\prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!}]\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}exp\{-\beta\theta\}d\theta \\
&= \prod_{i=1}^n \frac{1}{x_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\int exp\{-n\theta\}\theta^{\sum_{i=1}^n x_i}\theta^{\alpha-1}exp\{-\beta\theta\}d\theta \\
&= \prod_{i=1}^n \frac{1}{x_i!}\frac{\Gamma(\alpha + \sum_{i=1}^n x_i)\beta^\alpha}{\Gamma(\alpha + \sum_{i=1}^n x_i)(\beta+n)^{\alpha+\sum_{i=1}^n x_i}}
\end{aligned}$$

$$\begin{aligned} m_2(X) &= \int [\prod_{i=1}^{n} \binom{N}{x_i} p^{\sum_{i=1}^n x_i} (1-p)^{N-\sum_{i=1}^n x_i}] \frac{1}{Beta(c,d)} p^{c-1}(1-p)^{d-1} dp \\ &= [\prod_{i=1}^{n} \binom{N}{x_i}] \frac{Beta(c+\sum_{i=1}^n x_i, d+Nn-\sum_{i=1}^n x_i)}{Beta(c,d)} \end{aligned}$$

BIC

BIC is defined as

$$BIC = -2log(f(x|\hat{\theta})) + n_{param} log(n)$$

M_1 BIC

To find the MLE, set the score function to 0

$$\frac{d}{d\theta} log(f(X|\theta)) = \frac{\sum_{i=1}^n x_i}{\theta} - n = 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\begin{aligned} BIC_{M_1} &= -log[\prod_{i=1}^{n} \frac{e^{-\hat{\theta}} \hat{\theta}^{x_i}}{x_i!}] + log(n) \\ &= \sum_{i=1}^{n} log(x_i!) - \sum_{i=1}^{n} x_i log(\bar{X}) + n\bar{X} + log(n) \end{aligned}$$

Notice that $n_{param} = 1$ in M_1.

M_2 BIC

$$\frac{d}{d\theta} log(f(X|p)) = \frac{\sum_{i=1}^n x_i}{p} - \frac{Nn - \sum_{i=1}^n x_i}{1-p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{Nn}$$

Notice that $n_{param} = 1$ in M_2.

$$\begin{aligned} BIC_{M_2} &= -log[\prod_{i=1}^{n} \binom{N}{x_i} p^{\sum_{i=1}^n x_i}] + log(n) \\ &= -\sum_{i=1}^{n} log(\binom{n}{x_i}) - \sum_{i=1}^{n} x_i log(\hat{p}) - (Nn - \sum_{i=1}^{n} x_i) log(1-\hat{p}) + log(n) \end{aligned}$$

DIC

$$DIC = -2log(y|\hat{\theta_{Bayes}}) + 2p_{DIC}$$

$$p_{DIC} = 2(logp(y|\hat{\theta_{Bayes}}) - 2E(log(y|\theta)|y))$$

$$DIC = 2logp(y|\hat{\theta_{Bayes}}) - 4E(log(y|\theta)|y)$$

For M_1

$$\hat{\theta_{Bayes}} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\beta + n}$$

For M_2

$$\hat{p_{Bayes}} = \frac{c + \sum_{i=1}^{n}}{c + d + Nn}$$

For the first part of DIC, plug the posterior expectation into the log likelihood. For the second part, get a sample of posterior distribution and calculate the log likelihood at each point, average values at all points.

Gelfand and Ghosh Criterion

$$D = G + P \quad G = \sum_{l=1}^{n}(\mu_l - x_l)^2 \quad P = \sum_{l=1}^{n}\sigma_l^2$$

$$\mu_l = E(z_l|x) \quad \sigma_l^2 = Var(z_l|x)$$

To find μ_l and σ_l^2, create S copy of replicas of the orginal data, where S is the size of posterior samples. Average over z_l at each posterior sample to find $E(z_l|x)$ and take the variance of z_l accross posterior sample to find $Var(z_l|x)$.

Prior sensitivity

To test prior sensitivity, I selected 3 priors: non-informative, less informative and informative for each model. The mean of the latter two are set to be close to the MLE solution of the two models, $\hat{\lambda} = 9, 8$, $\hat{p} = 0.098$. The non-informative priors for both Gamma and Beta distribution sets both parameters in each case to be a small number and the prior mean is not necessarily centered around the MLE solution. Table 1 and 3 present the prior mean and variance, and the three priors have increasing variance going from informative to non-informative. Figure 1 illustrates the density of 3 different priors for each model. The non-informative priors for both models are flat.

Table 2 and 4 present the posterior summary of both models under 3 prior specifications. The results within two models are similar to each other, indicating that the model is not very sensitive to prior selections. However, when we give the model a very biased prior, i.e. mean not centered around MLE and the variance small, both model tends to show more variation. This is probably due to the fact that the sample size is very small. In this sense, the model is under heavy influence of the prior. But it is not entirely reasonable to give a biased prior, so I ruled out these prior specification.

Table 1: Prior summary of Model 1

	(0.001, 0.001)	(1, 0.1)	(10, 1)
Prior mean	1	10	10
Prior var.	1000	100	10

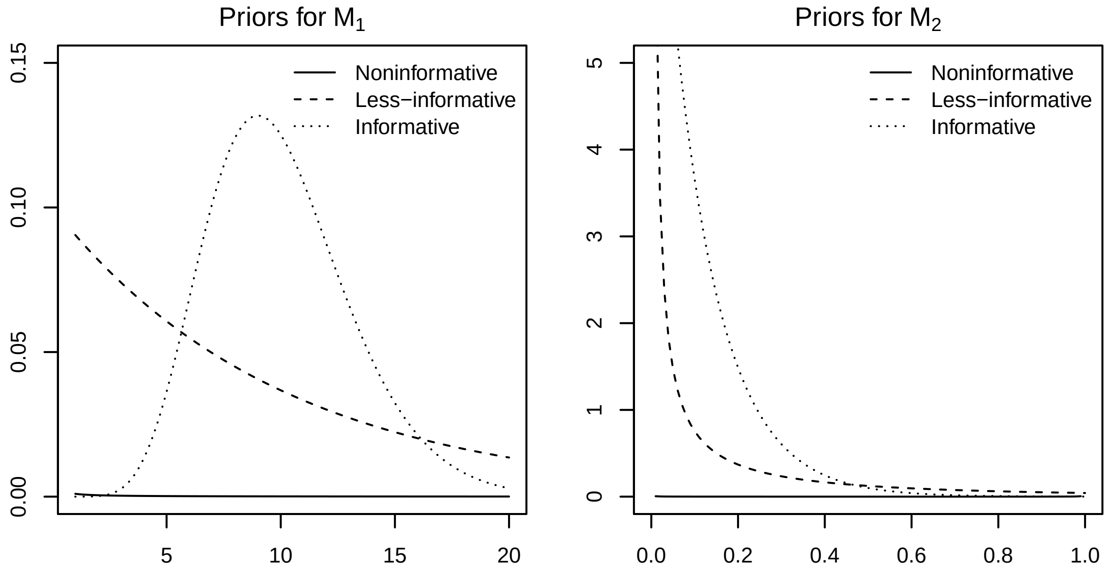

Figure 1: Priors for M_1 and M_2

Table 2: Posterior summaries of Model 1 under non-infromative, less informative, informative priors

	Mean	Sd	2.5%	25%	50%	75%	97.5%
(0.001, 0.001)	9.854067	0.6876091	8.540244	9.383859	9.836548	10.29818	11.26438
(1, 0.1)	9.819521	0.7112321	8.487589	9.334277	9.821096	10.29599	11.19251
(10, 1)	9.793167	0.6801805	8.490227	9.336800	9.783826	10.22483	11.21298

Table 3: Prior summary of Model 2

	(0.0001, 0.0001)	(0.1, 0.9)	(1, 9)
Prior mean	0.50000	0.100	0.1000000
Prior var.	0.24995	0.045	0.0081818

Table 4: Posterior summaries of Model 2 under non-infromative, less informative, informative priors

	Mean	Sd	2.5%	25%	50%	75%	97.5%
(0.0001, 0.0001)	0.0097860	0.0006923	0.0084729	0.0093131	0.0097682	0.0102811	0.0111193
(0.1, 0.9)	0.0098132	0.0006984	0.0084614	0.0093397	0.0098107	0.0102662	0.0112267
(1, 9)	0.0098198	0.0006783	0.0086143	0.0093231	0.0098203	0.0102755	0.0111474

Model selection

The Bayes factor is calculated using 6 sets of priors for both models defined. The row name vector in Table 5 shows the value of α, β, c, d that specifies the Gamma and Beta prior respectively. We can see that the Bayes factor is greater than 1 using different priors, but its value highly depends on the choice of prior (ranging from 2 to 164). A Bayes factor greater than 1 means the Poisson model is preferred. Yet it is unclear how decisive such preference is. We can use Jefferey's scale of $log_{10}(B_12)$ to evaluate the evidence in favor of M_1. From Table 5 we see that the power of such evidence range from 0.3 to 2.2, with 0.3 indicating poor and 2.2 substantial evidence in favor of M_1. Such dependence on prior specification makes Bayes factor less reliable for model comparison given such a small sample size.

From table 6 we can see that BIC agrees with Bayes factor, since the BIC for Poisson model is slightly lower than that of Binomial. DIC is also lower for Poisson model, although in both cases of BIC and DIC the difference is not drastic.

If we exclude the influence of prior on computing Bayes factor by choosing the non-informative prior, we could see that the model comparison result presented by Bayes factor agrees with that given by BIC and DIC: the two models are very similar to each other. It is not surprising since when N is very large, Poisson approximates the Binomial with N trials well.

Gelfand and Ghost criterion presents a different story: GG criterion for Binomial is lower than that of Poisson. The source of such discrepancy is two folds: difference in posterior sample and sampling distribution. The posterior inference presented by both models are actually quite similar. When N is very large in a Binomial model, we can use a Poisson model with mean Np as a decent approximation. With this reasoning, we multiply 1000 to the result in Table 4 to compare to the result in Table 2. This comparison tells us the posterior inference are quite on par in two models. The high variance in Binomial model therefore can only be due to the nature of Binomial. Notice that given $Np \approx \theta$, the Poisson sampling distribution has variance θ and Binomial sampling distribution has variance $Np(1-p) \approx \theta(1-p)$, which is smaller than θ. In this case, it is reasonable for Binomial model to have a lower GG, since the penalty term evaluated by variance is not as high as that of Poisson model.

Table 5: Bayes factor under different priors

	B_{12}	$log_{10}B_{12}$
(0.1,0.01,0.001,0.001)	164.728377	2.2167684
(1e-05,1e-05,1e-05,1e-05)	2.170625	0.3365847
(0.001,0.001,0.001,0.001)	2.151011	0.3326426
(0.001,0.001,1e-04,1e-04)	21.420550	1.3308306
(1,0.1,0.1,0.9)	6.488162	0.8121217
(10,1,1,9)	16.385775	1.2144670

Table 6: Information Criterias

	BIC	DIC	GG
Poisson	122.1529	121.0689	576.0858
Binomial	122.3268	121.3192	566.8174