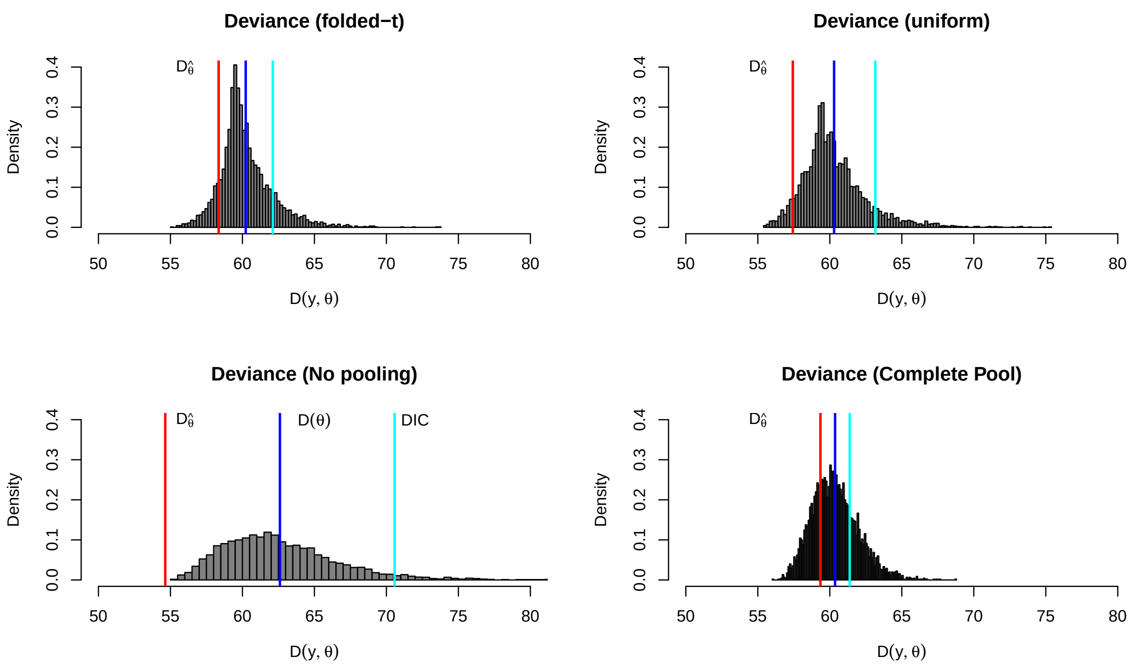
Stat 532 Assignment 10

Kenny Flagg

November 24, 2015

1. I ran through the DIC example code and included the most interesting output below. The biggest thing that I wonder about is the distribution of the deviance for the draws under the no pooling model. It makes sense that the deviance is more variable, but I was surprised that the center was not shifted compare the distributions under the other models. However, seeing that it was centered so close to the others helped me understand what the penalty does.



	D.theta.hat	D.avg.hat	pD	DIC	AIC	WAIC
Folded-t	58.35	60.23	1.88	62.11	62.35	61.40
Uniform	57.43	60.30	2.87	63.17	61.43	61.94
No pooling	54.64	62.60	7.96	70.56	70.64	
Complete pool	59.35	60.37	1.02	61.38	61.35	

2. Use the Beta-Binomial model to investigate whether people prefer Liz's cupcakes over Megan's cupcakes. 21 out of 28 tasters prefered Liz's cupcakes.

   (a) The model is

   $$y|\pi \sim \text{Binomial}(28, \pi);$$
   $$\pi \sim \text{Beta}(\alpha, \beta)$$

   where y is the number of tasters who prefer Liz's cupcakes and π is the probability that a taster prefers Liz's cupcakes.

   Choosing an informative prior in this situation is difficult because I have had 15 or 20 of Liz's cupcakes and thought each was the best cupcake ever, but I have not had a cupcake made by Megan. However, knowing the how much dilligence and attention to detail Megan puts into all of her work, I expect that she would make an excellent cupcake that might be a contender against Liz's cupcakes. I would expect Megan's cupcakes to be preferred half as often as Liz's, so I consider distributions of the form $\text{Beta}(\alpha, \alpha/2)$. Setting $\alpha = 15$ to represent my prior experience with Liz's cupcakes is unfair because I have not compared them to 7.5 cupcakes made by Megan. I chose the flatter $\text{Beta}(6, 3)$ distribution, which has a prior probability $Pr(\pi < 0.5) = 0.145$ to give Megan a fighting chance. This prior is illustrated in Figure 1.

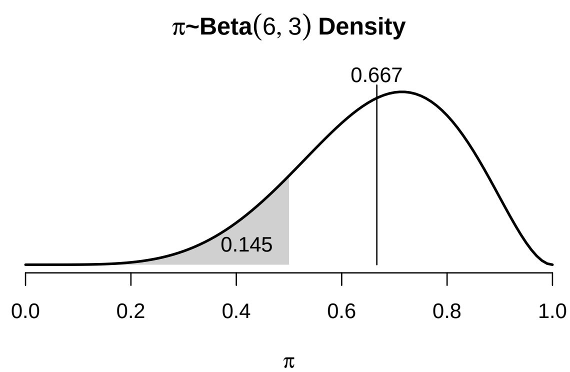


Figure 1: The informative prior distibution. The shaded area shows the prior probability of 0.145 that Megan's cupcakes are prefered more often than Liz's cupcakes. The vertical line marks the prior mean of 0.667.

   For comparison, I also used $\text{Beta}(0, 0)$, $\text{Beta}(1, 1)$, and $\text{Beta}(2, 2)$ priors. These all have prior means of 0.5 and are condisered default uninformative priors.

   In the double-blind taste test, 21 of 28 tasters prefered Liz's cupcakes. The resulting posterior distributions appear in Figure 2.

   The informative prior distribution resulted in a $\pi|y = 21 \sim \text{Beta}(27, 10)$ posterior distribution. On average, tasters prefer Liz's cupcakes over Megan's cupcakes in 73.0% of all trials, and there is a posterior 99.8% chance that Liz's cupcakes are prefered at least half of the time.

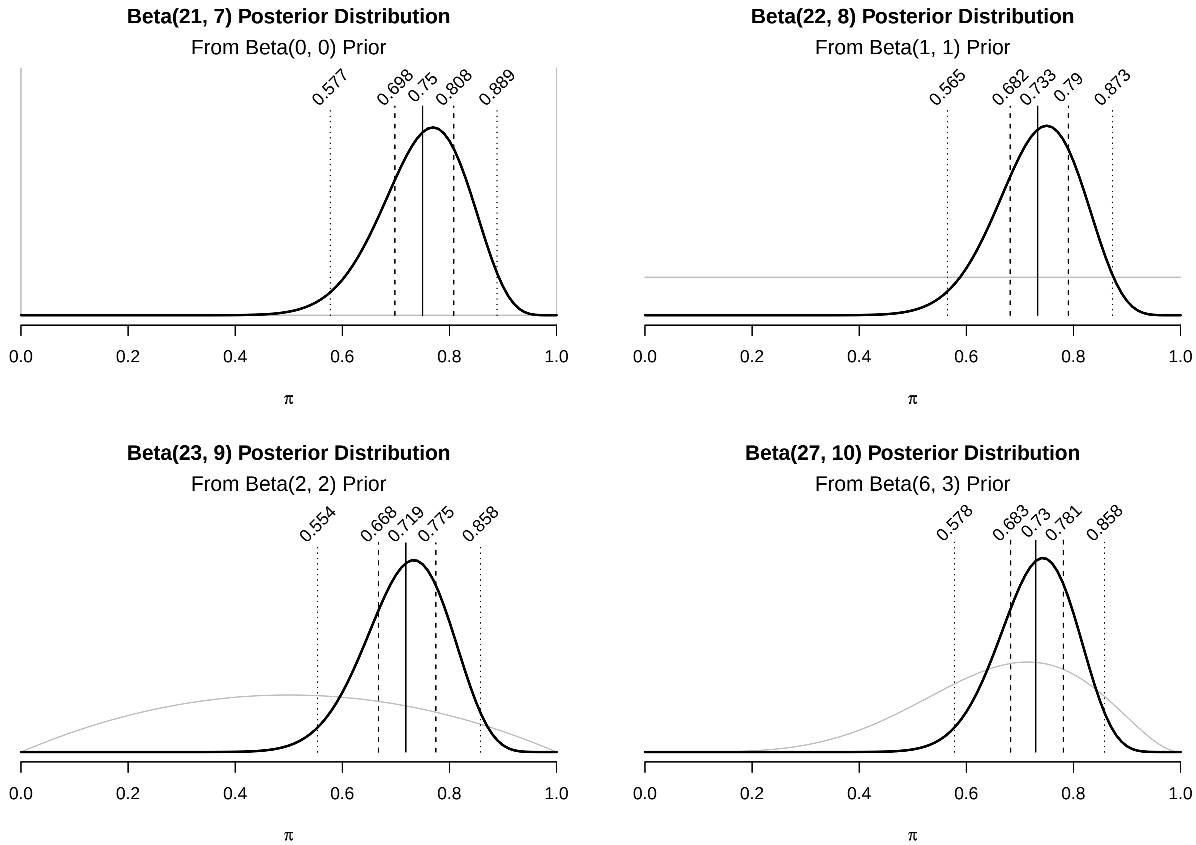


Figure 2: Posterior distributions of the probability of prefering Liz's cupcakes from four different prior distributions. The vertical lines denote posterior means, and 95% and 50% posterior intervals. The light grey curves are the prior densities.

(b) For any prior distribution, the prior probability of M_1 is $p(M_1) = Pr(\pi \geq 0.7)$ and the posterior probability of M_1 is $p(M_1|y) = Pr(\pi < 0.7|y)$. These values, as well as the prior and posterior odds and the Bayes factors, are tabulated for each of the four priors in Table 1.

	$\pi \sim$ Beta$(0,0)$	$\pi \sim$ Beta$(1,1)$	$\pi \sim$ Beta$(2,2)$	$\pi \sim$ Beta$(6,3)$
$Pr(\pi \geq 0.7)$	0.500	0.300	0.216	0.448
$Pr(\pi < 0.7)$	0.500	0.700	0.784	0.552
Prior Odds	1.000	0.429	0.276	0.812
$Pr(\pi \geq 0.7|y = 21)$	0.744	0.679	0.614	0.675
$Pr(\pi < 0.7|y = 21)$	0.256	0.321	0.386	0.325
Posterior Odds	2.901	2.111	1.587	2.073
Bayes Factor	2.901	4.926	5.762	2.552

Table 1: Probabilities, odds, and Bayes factors comparing M_1: $\pi \geq 0.7$ and M_2: $\pi < 0.7$.

3. Bayes factors are a common way to choose between different Bayesian models, but they should be used with caution. In particular, problems can arise when the models use highly dispersed prior distributions. For example, consider the model $y \sim \mathrm{N}(\mu, 1)$ for the data, where μ is the mean. Suppose we compare Model 1, where $\mu = 0$, to Model 2, where μ is unknown and has a prior distribution $\mu \sim \mu_0, \tau^2$ with τ specified. By following an example from Link and Barker (2006), the Bayes factor for Model 1 is

$$\mathrm{BF}_{1,2} = \sqrt{1 + n\tau^2} e^{-\mu_0 n\bar{y} + \frac{\mu_0^2}{2} - \frac{n^2\tau^2}{2(n\tau^2+1)}(\bar{y}-\mu_0)^2}.$$

Often, τ is set to a large value to create a very wide prior distribution that is meant not to influence the inferences. However, for very large values of τ, the Bayes factor will approach infinity and will always prefer Model 1 even if the data make $\mu = 0$ appear unlikely. Often, Bayes factors are unhelpful when uninformative priors are used.

4. Bayes factors are one way to asses the strength of evidence in support of a particular model or hypothesis over another in a Bayesian analysis. There are parallels between Bayes factors and likelihood ratios in both form and purpose. The definition of a Bayes factor is

$$\mathrm{BF}_{1,2} = \frac{\text{Posterior odds}}{\text{Prior odds}}.$$

When comparing two models M_1 and M_2, this can be rewritten as

$$\mathrm{BF}_{1,2} = \frac{p(y|M_1)}{p(y|M_2)}$$

where $p(y|M_1)$ and $p(y|M_2)$ are the marginal likelihoods under each model. That is, they are the results of averaging the likelihood function of the parameters over all possible values of the parameters in each model, Thus, the marginal likelihoods can be thought of as likelihood functions for the models themselves, and so the Bayes factor is a ratio of likelihoods.

The Bayes factor gives a relative measurement of how strongly a particular set of data y supports a one model over another, with large values of the Bayes factor indicating that the model in the numerator is more likely. This is analogous to the use of a likelihood ratio, where large values imply that the model in the denominator should not be preferred over the model in the numerator. Unlike likelihood ratios, Bayes factors cannot easily be transformed to follow known distributions, but some guidelines have been suggested to aid interpretation.

Bayes factors have the advantages that the incorporate uncertainty about the parameters by averaging over all possible values, and that they can be interpreted as directly favoring the numerator rather than showing only a lack of evidence for the denominator. However, the lack of a known distribution can make interpretation difficult, and rigidly following guidlines leads to the same issues as clinging closely to p-value cutoffs. Like hypothesis tests, Bayes factors should be used with caution.

5. The data are $y = (1, 3, 5, 7, 7)$. Use Bayes factors to compare a Poisson model and a Geometric model.

   (a) Let M_1 be the model $y \sim$ Poisson(5) and let M_2 be the model $y \sim$ Geometric(0.15).

      i. If both models are equally likely, the prior odds for M_2 are

$$\frac{p(M_2)}{p(M_1)} = \frac{0.5}{0.5} = 1.$$

      ii. Then the posterior odds for M_2 are

$$\begin{aligned}
\frac{p(M_2|y)}{p(M_1|y)} &= \frac{p(M_2)p(y|M_2)/p(y)}{p(M_1)p(y|M_1)/p(y)} \\
&= \frac{(0.5)\left((0.15)^5(0.85)^{\sum y_i}\right)}{(0.5)\left(e^{-25}\frac{5^{\sum y_i}}{\prod(y_i!)}\right)} \\
&= e^{25}(0.15)^5(0.85/5)^{\sum y_i}\prod(y_i!) \\
&= e^{25}(0.15)^5(0.17)^{23}(1!)(3!)(5!)(7!)(7!) \\
&= 0.1997
\end{aligned}$$

      or about 5 to 1 against.

      iii. The Bayes factor is

$$\mathrm{BF}_{2,1} = \frac{\text{Posterior odds}}{\text{Prior odds}} = \frac{0.1997}{1} = 0.1997.$$

   (b) Now let M_1 be the model $y|\lambda \sim$ Poisson(λ), $\lambda \sim$ Unif(0, 30), and let M_2 be the model $y|\pi \sim$ Geometric(π), $\frac{1-\pi}{\pi} \sim$ Uniform(0, 30). This implies a prior density of

$$p(\pi) = \frac{1}{30\pi^2}, \ \frac{1}{31} < \pi < 1.$$

      i. The prior odds for M_2 are

$$\frac{p(M_2)}{p(M_1)} = \frac{0.5}{0.5} = 1.$$

      ii. To find the posterior odds, we first need the marginal distributions of y under each model. For M_1,

$$\begin{aligned}
p(y|M_1) &= \int_0^{30} p(\lambda|M_1)p(y|\lambda, M_1)d\lambda \\
&= \int_0^{30} \frac{1}{30}\frac{e^{-5\lambda}\lambda^{23}}{(1!)(3!)(5!)(7!)(7!)}d\lambda \\
&= \frac{\Gamma(24)}{30(1!)(3!)(5!)(7!)(7!)5^{24}}Pr(G < 30)
\end{aligned}$$

where G is a Gamma$(24, 5)$ random variable (using Gelman's parameterization), so

$$p(y|M_1) = \frac{10}{12{,}650{,}291} \ .$$

Then for M_2,

$$\begin{aligned}
p(y|M_2) &= \int_{\frac{1}{31}}^{1} p(\pi|M_2)p(y|\pi, M_2)d\pi \\
&= \int_{\frac{1}{31}}^{1} \frac{1}{30\pi^2}\pi^5(1-\pi)^{23}d\pi \\
&= \frac{1}{30}\int_{\frac{1}{31}}^{1} \pi^3(1-\pi)^{23}d\pi \\
&= \frac{1}{30}\frac{\Gamma(4)\Gamma(24)}{\Gamma(28)}Pr\left(B > \frac{1}{31}\right)
\end{aligned}$$

where $B \sim$ Beta$(4, 24)$ random variable, so

$$p(y|M_2) = \frac{5}{10{,}641{,}692} \ .$$

Finally, the posterior odds for M_2 are

$$\begin{aligned}
\frac{p(M_2|y)}{p(M_1|y)} &= \frac{p(M_2)p(y|M_2)/p(y)}{p(M_1)p(y|M_1)/p(y)} \\
&= \frac{(0.5)(10/12{,}650{,}291)}{(0.5)(5/10{,}641{,}692)} \\
&= 0.5944,
\end{aligned}$$

about 10 to 6 against.

iii. The Bayes factor is

$$\text{BF}_{2,1} = \frac{\text{Posterior odds}}{\text{Prior odds}} = \frac{0.5944}{1} = 0.5944.$$