Introduction to Bayes Factors

Jim Albert

March 18, 2018

Models for Fire Calls

To motivate the discussion of plausible models, the website
`http://www.franklinvillefire.org/callstatistics.htm` gives the number
of fire calls for each month in Franklinville, NC for the last several years.

Suppose we observe the fire call counts $y_1, ..., y_N$ for N consecutive months.
Here is a general model for these data.

- $y_1, ..., y_N$ are independent $f(y|\theta)$

- θ has a prior $g(\theta)$

Also suppose we have some prior beliefs about the mean fire count $E(y)$. We
believe that this mean is about 70 and the standard deviation of this guess is
10. Given this general model structure, we have to think of possible choices for
f, the sampling density. We think of the popular distributions, say Poisson,
normal, exponential, etc. Also we should think about different choices for the
prior density. For the prior, there are many possible choices – we typically
choose one that can represent my prior information.

Once we decide on several plausible choices of sampling density and prior,
then we'll compare the models by Bayes factors. To do this, we compute the
prior predictive density of the actual data for each possible model. The Laplace
method provides a convenient and accurate approximation to the logarithm of
the predictive density and we'll use the function `laplace` from the `LearnBayes`
package.

Continuing our example, suppose our prior beliefs about the mean count of
fire calls θ is Gamma(280, 4). (Essentially this says that our prior guess at θ
is 70 and the prior standard deviation is about 4.2.) But we're unsure about
the sampling model – it could be (model M_1) Poisson(θ), (model M_2) normal
with mean θ and standard deviation 12, or (model M_3) normal with mean θ
and standard deviation 6.

To get some sense about the best sampling model, a histogram of the fire
call counts are graphed below. I have overlaid fitted Poisson and normal dis-
tributions where I estimate θ by the sample mean. The Poisson model appears
to be the best fit, followed by the Normal model with standard deviation 6,

and the Normal model with standard deviation 12. We want to formalize this comparison by computation of Bayes factors.

```
> fire.counts <- c(75, 88, 84, 99, 79, 68, 86, 109, 73, 85, 101, 85,
+                  75, 81, 64, 77, 83, 83, 88, 83, 78, 83, 78, 80,
+                  82, 90, 74, 72, 69, 72, 76, 76, 104, 86, 92, 88)
> hist(fire.counts, probability=TRUE, ylim=c(0, .08))
> x <- 60:110
> lines(x, dpois(x, lambda=mean(fire.counts)), col="red")
> lines(x, dnorm(x, mean=mean(fire.counts), sd=12), col="blue")
> lines(x, dnorm(x, mean=mean(fire.counts), sd=6), col="green")
> legend("topright", legend=c("M1: Poisson(theta)",
+                              "M2: N(theta, 12)",
+                              "M3: N(theta, 6)"),
+         col=c("red", "blue", "green"), lty=1)
```

Histogram of fire.counts
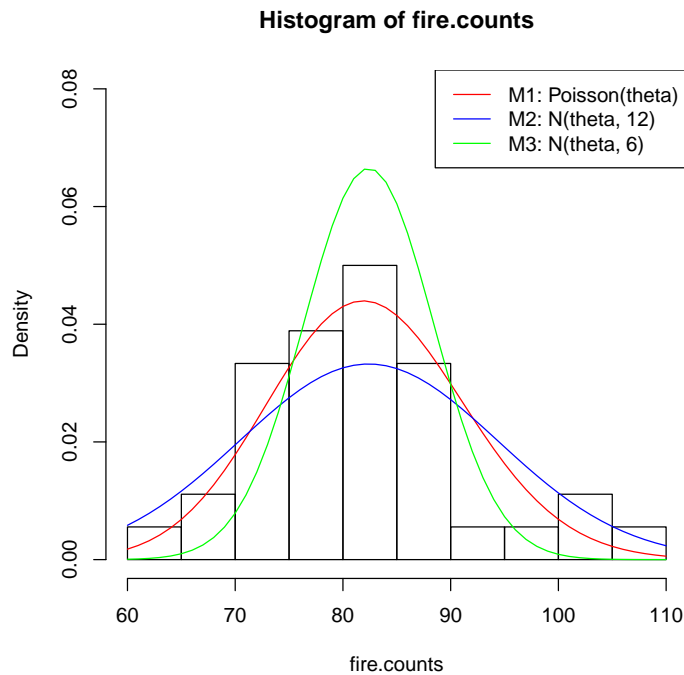
Bayesian Model Comparison

Under the general model, the predictive density of y is given by the integral

$$f(y) = \int \prod_{j=1}^{N} f(y_j|\theta)g(\theta)d\theta.$$

This density can be approximated by the Laplace method implemented in the `laplace` function.

One compares the suitability of two Bayesian models by comparing the corresponding values of the predictive density. The Bayes factor in support of model M_1 over model M_2 is given by the ratio

$$BF_{12} = \frac{f_1(y)}{f_2(y)}.$$

Computationally, it is convenient to compute the predictive densities on the log scale, so the Bayes factor can be expressed as

$$BF_{12} = \exp\left(\log f_1(y) - \log f_2(y)\right).$$

To compute the predictive density for a model, say model M_1, we initially define a function `model.1` which gives the log posterior.

```
> model.1 <- function(theta, y){
+   sum(log(dpois(y, theta))) +
+     dgamma(theta, shape=280, rate=4)
+ }
```

Then the log predictive density at y is computed by using the `laplace` function with inputs the function name, a guess at the posterior mode, and the data (vector of fire call counts). The component `int` gives the log of $f(y)$

```
> library(LearnBayes)
> log.pred.1 <- laplace(model.1, 80, fire.counts)$int
> log.pred.1
```

```
[1] -131.6253
```

We similarly find the predictive densities of the models M_2 and M_3 by defining functions for the corresponding posteriors and using `laplace`:

```
> model.2 <- function(theta, y){
+   sum(log(dnorm(y, theta, 6))) +
+     dgamma(theta, shape=280, rate=4)
+ }
> model.3 <- function(theta, y){
+   sum(log(dnorm(y, theta, 12))) +
+     dgamma(theta, shape=280, rate=4)
+ }
> log.pred.2 <- laplace(model.2, 80, fire.counts)$int
> log.pred.3 <- laplace(model.3, 80, fire.counts)$int
```

Displaying the three models and predictive densities, we see that model M_1 is preferred to M_3 which is preferred to model M_2.

```
> data.frame(Model=1:3, log.pred=c(log.pred.1, log.pred.2, log.pred.3))
```

```
   Model  log.pred
1      1 -131.6253
2      2 -144.7554
3      3 -132.9464
```

The Bayes factor in support of model M_1 over model M_3 is given by

```
> exp(log.pred.1 - log.pred.3)
```

```
[1] 3.74751
```