

Homework # 3

1. The authors of the paper I chose use two sets of priors. In both, they are simply using the prior as an adjustment to the mean performance of market benchmarks for 3 crops: corn, wheat and soybean. The authors write that "there are two levels of parameters in the normal hierarchical model for market advisory service performance. The expected performance of each advisory program $\theta = (\theta_1, \dots, \theta_n)$, in the lower level, ... and hyperparameters (μ, τ) " in the normal model where

$$\theta_j | \mu, \tau \sim N(\mu, \tau^2)$$

In the first set considered, a normal prior is used, with uninformed hyperparameters. For the mean of their prior, they use a uniform prior with no truncation. They argue this is reasonable as their sample size is large enough to rely only on the sample for estimation of this parameter. They also use a uniform prior on the variance estimate, which, they argue (citing Gelman (2006)) is supposed to perform well when the number of groups is larger than 2 or 3. They seem to do a reasonable job at arguing their choices of priors here. They write "The prior is based on the notion that performance across programs is not independent, but instead, data on pricing performance of a particular program as well as the performance of the rest of the programs are helpful in estimating expected performance."

In the second set of priors, they survey farmers (to elicit expert opinion) with two questions:

- (a) What is the most likely difference between the price obtained by following the recommendations of advisory programs and the market benchmark price?
- (b) What is the probability that an advisory program outperforms the market benchmark, on average, by more than 5%?

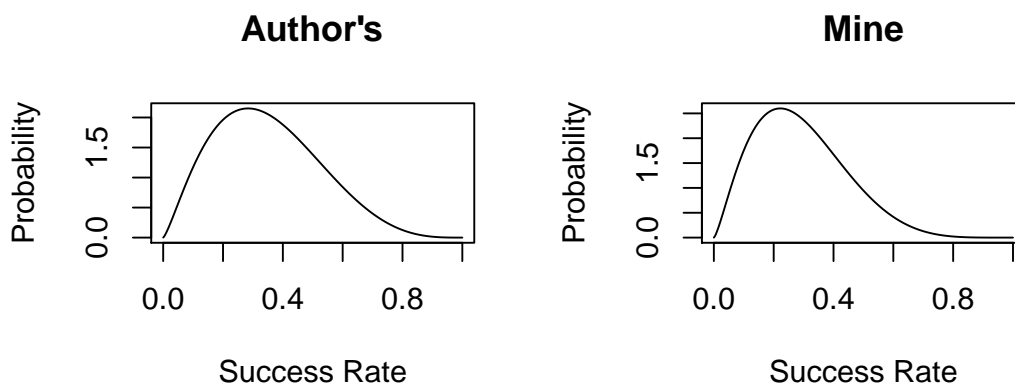
They then use the mean of the first question as the hyperparameter on a normal prior and use both the mean of the first and the answer to the second question to form the variance hyperparameter. They say that they use the normal distribution as a prior for "skeptical beliefs" as they say it simplifies computation of the posterior distribution. My main issue with this is that they then make no discussion of a sensitivity analysis to justify that this decision wouldn't greatly impact the posterior.

2. My main take-away from this section is that if there are any simplifications made in specifying a prior (i.e. going with "non-informative or reference prior densities"), then the researcher ought to at least justify their choice, check that the posterior is proper, and perform a sensitivity analysis on priors for any modeling assumptions made out of convenience.

3. (a) I do agree with Gelman's stance on weakly informative priors. To me, specifying priors has the same costs, benefits, and subjectivity as model specification, to which it is often reasonable to simplify model assumptions for a myriad of different reasons. For instance, we often may think a-priori that a quadratic model may fit the data, only to find out that we don't gain much prediction capability by this choice. We then may choose to go with the simpler model to simplify inferences made, even though the quadratic model has better fit. This is similar to the choice of not incorporating all prior information into a prior distribution. Sometimes the costs simply outweigh the benefits of incorporating model complexities.

Further, I like his advice on not using prior information heavily when it is part of the analysis. I've seen a number of studies (especially in the realm of social sciences), where it seems that the inferences made match too well to their prior beliefs. I often suspect in these cases that they are simply using data which they already have a good idea of what it should look like, to then support their argument being made. To me, I feel in these circumstances that one should "lean against the hypothesis" with prior specifications.

- (b) I think that Kenny's comments were reasonable in that it may be possible that the researchers made some more informed decisions about the prior specification than was lead on. Perhaps, as suggested by Gelman, they may be trying to "lean against the hypothesis" or that some simplification was made as the costs of incorporating this information would outweigh the benefits. However, in absence of access to the reasearchers, I think it would be reasonable to try to fit the data given better than the authors have done. After looking into this further, I actually came up with a fairly similar decision as far as parameters on the beta distribution. I think that the scaling on the y-axis of their graph just seems to be off, as I wouldn't see how either distribution (histogram or beta) sums to one. In looking at the distribution though, and making some rough estimates on quantiles, I came up with a $\text{beta}(2.370455, 5.800109)$ distribution, which isn't all that dissimilar to what the authors of Kenny's paper came up with. Here's a plot of the author's choice versus mine. Code used is provided in the appendix.



I think that so long as there is enough data collected, these priors will work well enough. However, in the case of small sample sizes, we may need to take greater care and move away from the beta distribution.

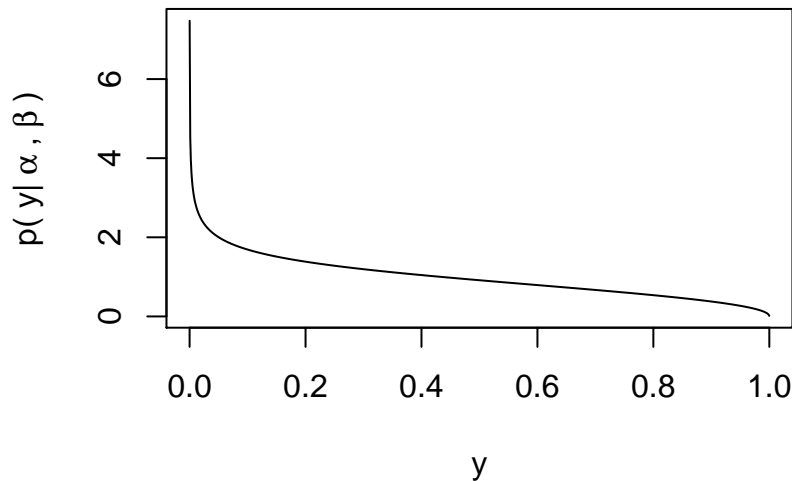
4. We use the following code to computationally approximate the parameters of the beta distribution which have the given characteristics.

```
error_func <- function(v){
  abs(qbeta(.5,v[1],v[2]) - .3) + abs((1 - pbeta(.75, v[1], v[2])) - .1)
}

nlm(error_func, c(2,2), stepmax = 1, iterlim = 1000, gradtol = .000000001)
```

We find that a $\text{beta}(0.7920481, 1.4748258)$ distribution reflects the two criterion given. A plot of this distribution is given below.

Beta dist. with $\alpha = 0.7920481$, $\beta = 1.4748258$



5. (a) We have $y|\lambda \sim \text{Pois}(\lambda)$, which gives

$$\begin{aligned}
 P(y|\lambda) &= \frac{\lambda^y e^{-\lambda}}{y!} \\
 &\propto \lambda^y e^{-\lambda} \\
 &= e^{-\lambda} e^{y \log \lambda} \\
 &= g(\lambda)^n e^{\phi(\lambda)^T t(y)}
 \end{aligned}$$

and so, a prior density of the form

$$\begin{aligned}
 p(\lambda) &\propto g(\lambda)^\beta e^{\phi(\lambda)^T(\alpha-1)} \\
 &= (e^{-\lambda})^\beta e^{(\log \lambda)(\alpha-1)} \\
 &= \underbrace{e^{-\lambda\beta} \lambda^{\alpha-1}}_{\text{Kernel of } \textit{Gamma}(\alpha, \beta)}
 \end{aligned}$$

is conjugate to the poisson pmf.

(b) We have

$$\begin{aligned}
 p(\mathbf{y}|\lambda) &= \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \\
 &= \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!}
 \end{aligned}$$

and so

$$\begin{aligned}
 p(\lambda|\mathbf{y}) &\propto \lambda^{\sum_{i=1}^n y_i} e^{-n\lambda} \lambda^{\alpha-1} e^{-\lambda\beta} \\
 &= \underbrace{\lambda^{\sum_{i=1}^n y_i + \alpha - 1} e^{-\lambda(n+\beta)}}_{\text{Kernel of } \textit{gamma}(\sum_{i=1}^n y_i + \alpha, n + \beta)}
 \end{aligned}$$

Thus, we have $\lambda|\mathbf{y} \sim \textit{gamma}(\sum_{i=1}^n y_i + \alpha, n + \beta)$

(c)

(d) I set the following criterion in the search for a gamma distribution:

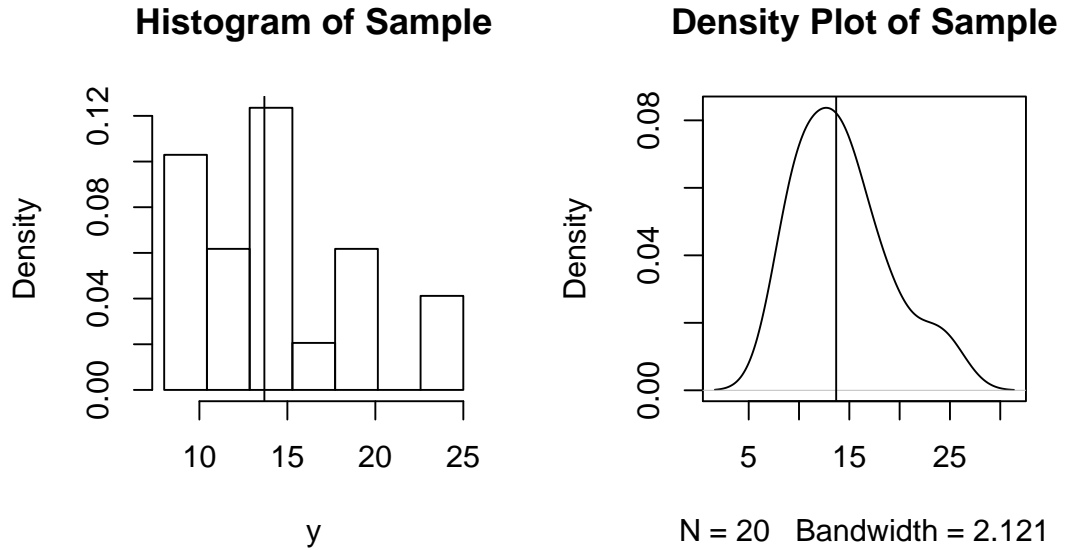
- The probability of λ being between 15 and 25 as .80.
- The probability of λ being between 10 and 35 as .95.
- The probability of λ being between 15 and 25 as .99.

Then, I searched the parameter space to find a gamma distribution that closely matched those criterion. I obtained a *gamma*(21.753974, 1.075465) distribution.

(e) I obtained the sample with the following R-code:

```
set.seed(2015)
samp1 <- rpois(20, rgamma(1, 21.753974 , 1.075465))
```

(f) Below, a histogram and a density plot are given for the data with a verticle line specifying the value of the sample drawn from the prior distribution.



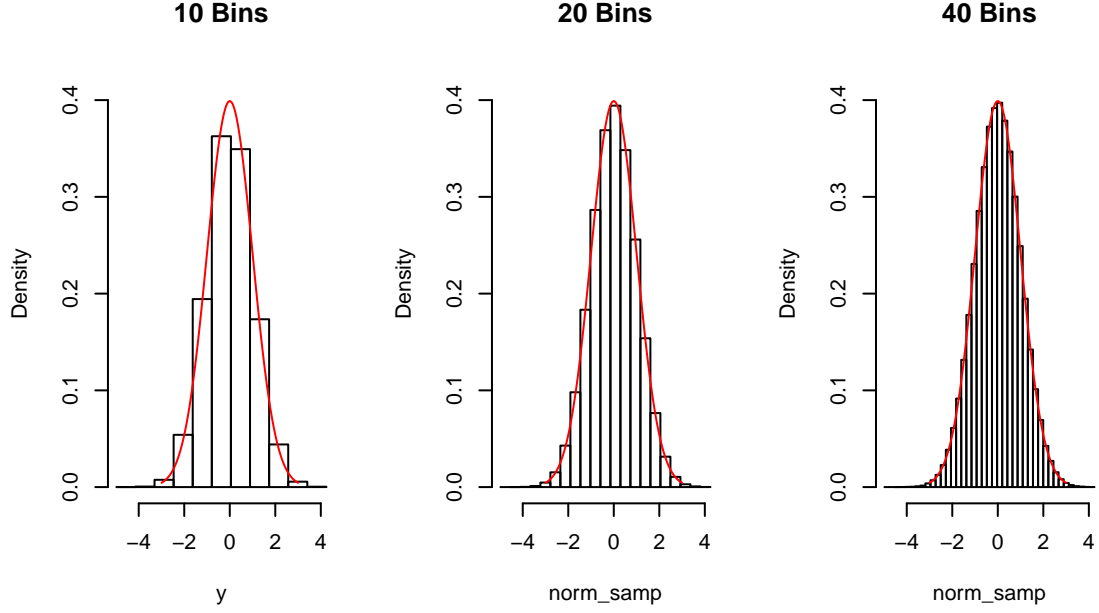
- (g) From our sample, we have that $\sum_{i=1}^{20} y_i = 286$ Thus, from the work in part (b), we know that

$$\lambda|\mathbf{y} \sim \text{gamma}(307.754, 21.075465)$$

We have the following plots of the likelihood, the posterior, and the prior.

- (h)
 - (i)
 - (j)
 - (k)
 - i.
 - ii.
6. (a)
- (b)
- (c)
- (d)
- (e)
- (f)
- (g)
7. With the plot below, we illustrate the process of numerical integration. They are histograms of a very large sample drawn from the normal pdf with the associated pdf superimposed. Although these plots were made with simply using the `hist()` and `dnorm()` functions in R, they are done in such a way to approximate what the graphs of a riemann sum would look like using the midpoint rule. We see that as our mesh grid is refined, our error associated with the process goes to zero. Thus, we can make our mesh

grid so fine that our error is negligible. Note, that using the midpoint rule is superior to the left and right hand rules when the function we are numerically integrating is monotonic increasing or decreasing. There are also many other more efficient forms of numerical quadrature.



8. (a) Yes, the resulting distribution would be proper. With $y|\theta \sim \text{Bin}(40, \theta)$, and $\theta \sim \text{beta}(0, 0)$, we have

$$\begin{aligned} p(\theta|y) &\propto \binom{n}{y} \theta^y (1-\theta)^{40-y} \theta^{-1} (1-\theta)^{-1} \\ &\propto \underbrace{\theta^{y-1} (1-\theta)^{40-y-1}}_{\text{Kernel of } \text{beta}(y, 40-y)} \end{aligned}$$

Thus, if we observe $y = 10$, we have

$$\theta|y \sim \text{beta}(10, 30)$$

Which is a proper distribution.

- (b) No. If we observe $y = 0$, we have $\theta|y \sim \text{beta}(0, 40)$, which is not a proper

distribution. We have

$$\begin{aligned}
\int_0^1 (\theta)^{-1} (1 - \theta)^{39} (B(0, 40))^{-1} d\theta &= \int_0^1 (\theta)^{-1} (1 - \theta)^{39} d\theta \\
&= \int_0^1 \sum_{i=0}^{38} \frac{\binom{39}{i} (-\theta)^{39-i}}{\theta} d\theta + \int_0^1 \frac{1}{\theta} d\theta \\
&= \sum_{i=0}^{38} \frac{\binom{39}{i} (-1)^{39-i}}{39-i} + \lim_{b \rightarrow 0^+} -\ln(b) \\
&= -\frac{2066035355155033}{485721041551200} + \infty \\
&= \infty
\end{aligned}$$

Thus, since the integral is not equal to one, $\text{beta}(0, 40)$ is not a proper distribution.

(c) Yes. We can simply use the following code:

```
f_x_8 <- function(x) ((1 - x)^39)/(x)
f_x_8_vec <- Vectorize(f_x_8)
integrate(f_x_8_vec, 10^(-15), 1)
```

Which gives us that the integral is equal to "30.28523 with absolute error < 0.0021" for the given bounds. Since this doesn't integrate to 1, we know that this is an improper prior. Note that I knew ahead of time that the problem was at zero and the second argument in the `integrate()` function, I used 10^{-15} . If I hadn't known this ahead of time and put 0 in for this argument, I would have received the following error in R.

```
Error in integrate(f_x_8_vec, 0, 1) :
  maximum number of subdivisions reached
```

This error would also have told me that there may be a problem in the integral being equal to 1, as the function `integrate()` uses more subdivisions when the change in y is relatively large compared to the change in x.

9.

R-Code

R-Code Here