

# Stat 532 Assignment 3

Kenny Flagg

September 21, 2015

1. Sinding-Larsen and Xu[1] are concerned with modeling the sizes of undiscovered oil pools in an oil field off the coast of Norway. The field includes 22 pools that have so far been discovered.

The model they use is specified in Xu and Sinding-Larsen[2] as follows: pool sizes are assumed to be independent and identical draws from an infinite lognormal population, and the pools are discovered in succession as random draws with replacement from a population with finite but unknown size. The probability of discovering a particular pool is proportional to the pool's volume raised to a power.

The joint distribution is:

$$\phi(Y, S, \theta) = \binom{N}{n} g(\theta) \prod_{j=1}^n \frac{f(Y_j|\theta) Y_j^\beta}{\sum_{k=j}^n Y_k^\beta + \sum_{k=n+1}^N S_k^\beta} \prod_{j=n+1}^N f(S_j|\theta)$$

where

- $N$  is the total number of pools,
- $n$  is the number of discovered pools,
- $Y = (Y_1, \dots, Y_n)$  is a vector of the sizes of the discovered pools in the order in which they were discovered (so  $Y_i$  is the  $i$ th discovery),
- $S = (S_{n+1}, \dots, S_N)$  is a vector of the sizes of the undiscovered pools,
- $\theta = (N, \beta, \mu, \sigma^2)$  is the vector of unknown parameters,
- $g(\theta)$  is the prior distribution of  $\theta$ , and
- $f(\cdot|\theta)$  is the lognormal density function with mean  $\mu$  and variance  $\sigma^2$ .

They denote the posterior distribution as  $\phi^*(S, \theta|Y)$ .

The prior distribution of  $N$  was determined by a Monte Carlo simulation described in [2]. The assumed distribution is

$$p(N = n) = \sum_{m=n}^{\infty} \binom{m}{n} \pi^n (1 - \pi)^{m-n} Pr(M = m)$$

where  $M$  is the number of prospective pools and  $\pi$  is the “discovery probability” of a prospect being an actual pool.

The prior distribution of  $M$  should be estimated “from geological and geophysical information”[2, p. 222-223] but neither paper gives any further information about  $M$ . I do not know if this prior is meant to be informative or not. For now, I will suppose that it is based on current expert knowledge and I should do further reading to learn what models are accepted or appropriate.

The hyperprior on  $\pi$  is a point of contention. They fit a Beta distribution to historical data, but did not justify their choice of parameters. The illustration they present does not convince me that the

prior accurately represents the previous data. Since this is a hyperprior, it is not crucial to choose the single best distribution. I would have chosen something less informative.

They define a prior distribution on the exponent  $\beta$  through the “empirical equation” [1, p. 239]

$$b(\beta) = b_0(1 - e^{-\sigma\beta})$$

where  $b(\beta)$  is the slope of the linear regression of expected pool size (for a specified  $\beta$ ) against the pools’ ranks in the order of their discovery and  $b_0$  is the slope of the regression line when  $\beta = \infty$ . They examined a plot of size vs discovery order and selected a Uniform prior distribution for  $b$  based on the possible slopes that could be fit through the points.

They used a Normal prior for  $\mu$  and a  $\chi^2$  mixture prior for  $\sigma^2$ . The parameter values they chose were point estimates from size bias samples selected from the discovered pools. Their sampling and estimation process is complicated and I do not fully understand it.

The priors were chosen to be informative and to precisely reflect previous knowledge. I think their approach is unnecessarily complicated and a misuse of “Bayesian” concepts. They use  $Y$  to set prior parameter values – did they consider  $Y$  to be prior information? If so, what is the *data* in this model? I think they got too caught up in computation to think about how they were using their information. I would prefer a simpler approach with less-informative priors based on scientific consensus instead of calculations.

## References

- [1] Sinding-Larsen, R., and Xu, J., 2005, Bayesian Discovery Process Modeling of the Lower and Middle Jurassic Play of the Halten Terrace, Offshore Norway, as Compared with the Previous Modeling: Natural Resources Research, v. 14, no. 3, p. 235-248.
- [2] Xu, J., and Sinding-Larsen, R., 2005, How to choose priors for Bayesian estimation of the discovery process model: Natural Resources Research, v. 14, no. 3, p.211 -233.

- 2. If you let the data speak it does not matter much which prior you choose, but always check that the posterior is proper and do a sensitivity analysis to make sure the model is reasonable.
- 3. (a) Gelman’s main idea is that prior distributions should be meaningful in context and contain at least enough information to guarantee reasonable results. I don’t like that he relies entirely on practical arguments from his experience rather than including some theoretical arguments, but overall I agree with his point of view.

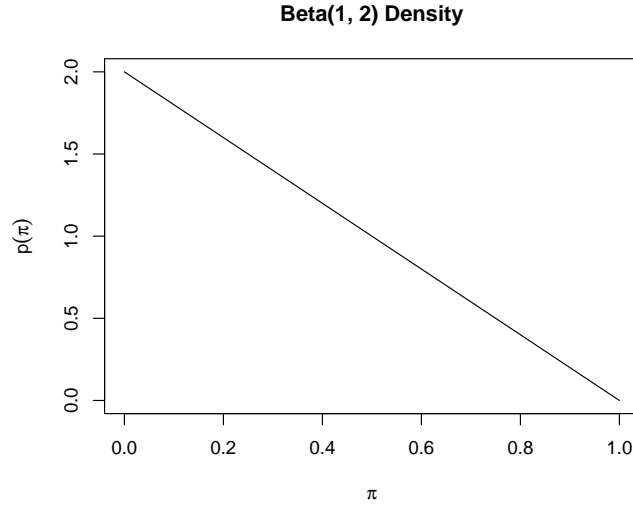
One of Gelman’s first points is that all models contain at least some prior information, such as the design used in an experiment, or notions about which predictors will be relevant. I interpret this as meaning that no Bayesian model can truly be uninformative, so there is little reason to exclude information that would help constrain results. If the goal is to let the data tell their own story then this information can be included conservatively. Simulations can be done to check that the prior does not bias the results.

The most interesting point he makes is that there is nothing wrong with using convenient or default models, as long as they are justified. The point of a weakly informative prior is to let the data dominate in the posterior, so choosing between a few standard priors is acceptable as long as a little thought is involved.

The most obvious counterargument I see would relate to the choice of measurement scale and parameterization. The argument for Jeffreys’ prior is that this decision can have implicit effects on the posterior inferences. However, if the data are supposed to dominate then this choice has little effect as long as we compare some different priors and choose one that will give a reasonable posterior.

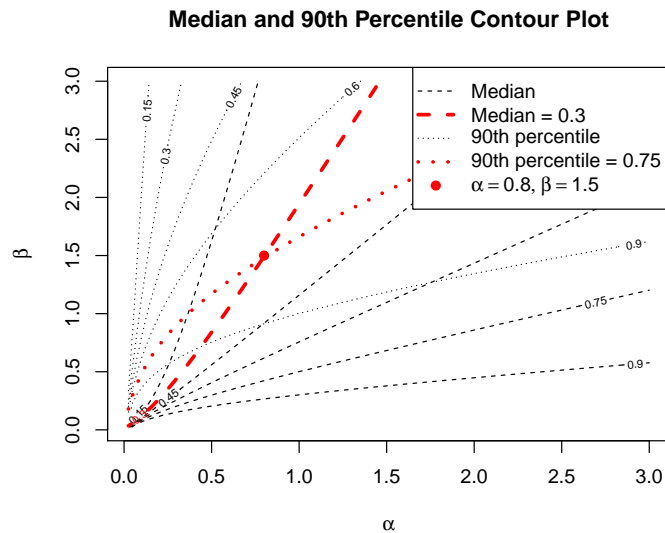
I would summarize the Gelman point of view as “be reasonable.” Justify, but do not overthink. Always do a sensitivity analysis. There is little reason to exclude all prior information. When the issue is put into these simple terms, I cannot disagree.

- (b) Looking only at the histogram and knowing nothing else, I would choose the Beta(1, 2) (triangular) distribution.

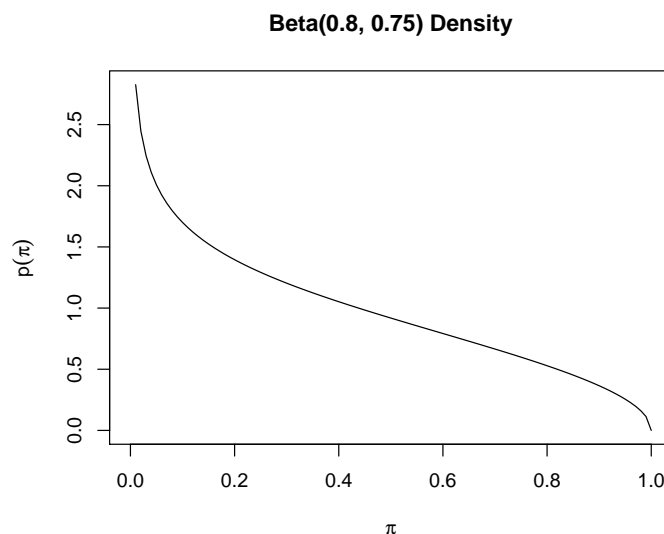


This distribution has a mean of  $\frac{1}{3}$  and standard deviation of 0.2357, so it is more variable than the prior that the authors used. It concentrates most of the mass on the cluster of observations below 0.4. It allows occasional large values, with  $Pr(\pi > 0.8) = 0.04$ . It does not allow  $\pi = 1$ , but the researchers must believe  $\pi < 1$  if they are investigating this quantity. Finally, the Beta(1, 2) distribution has the computational convenience of being a conjugate prior for the Binomial model.

4. Below I have plotted the level curves for the median and 90% percentile of a Beta( $\alpha, \beta$ ) distribution as functions of  $\alpha$  and  $\beta$ .



The curves for a median of 0.3 and a 90% percentile of 0.75 intersect near  $\alpha = 0.8$  and  $\beta = 1.5$ . The Beta(0.8, 1.5) distribution has a median of 0.2987 and  $Pr(\pi > 0.75) = 0.0974$ , which is reasonably consistent with our prior knowledge. The density curve appears below.



Below is the code to generate the level curve plot.

```
# Alpha and beta values at which to find the medians and percentiles
grids <- seq(0.025, 3, 0.025)

# Repeat these through matrices.
# Alphas values go down the columns, beta values go across the rows.
alphas <- matrix(grids, nrow = 120, ncol = 120, byrow = FALSE)
betas <- matrix(grids, nrow = 120, ncol = 120, byrow = TRUE)

# Get the medians and 90% percentiles.
meds <- qbeta(0.5, alphas, betas)
pcts <- qbeta(0.9, alphas, betas)

# Levels at which to plot curves
levels <- seq(0.15, 0.9, 0.15)

# Plot the contours, omitting the second level (0.3) for the medians
# and the fifth level (0.75) for the percentiles.
contour(x = grids, y = grids, z = meds, levels = levels[-2], lty = 2,
        lwd = 1, main = 'Median and 90th Percentile Contour Plot',
        xlab = expression(alpha), ylab = expression(beta))
contour(x = grids, y = grids, z = pct, levels = levels[-5], lty = 3,
        lwd = 1, add = TRUE)

# Now plot the previously omitted curves in red.
contour(x = grids, y = grids, z = meds, levels = 0.3, lty = 2,
        lwd = 3, col = 'red', add = TRUE)
```

```

contour(x = grids, y = grids, z = pcts, levels = 0.75, lty = 3,
        lwd = 3, col = 'red', add = TRUE)

# Plot a convenient point.
points(x = 0.8, y = 1.5, pch = 19, col = 'red')

# Add a legend.
legend('topright', bg = 'white', lty = c(2, 2, 3, 3, NA),
       pch = c(NA, NA, NA, NA, 19), lwd = c(1, 3, 1, 3, NA),
       col = c('black', 'red', 'black', 'red', 'red'),
       legend = c('Median = 0.3', '90th percentile',
                  '90th percentile = 0.75',
                  expression(list(alpha==0.8, beta==1.5))))

```

5. (a) The Poisson probability mass function is

$$p(y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \propto \lambda^y e^{-\lambda},$$

which looks very much like a Gamma kernel. The  $\text{Gamma}(\alpha, \beta)$  density is

$$p(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}.$$

These can both be written in exponential family form, with

$$p(y|\lambda) \propto g(\lambda)^1 e^{y\phi(\lambda)}$$

and

$$p(\lambda) \propto g(\lambda)^\beta e^{(\alpha-1)\phi(\lambda)}$$

where  $g(\lambda) = e^{-\lambda}$  and  $\phi(\lambda) = \log(\lambda)$ . Then the posterior density is

$$p(\lambda|y) \propto p(y|\lambda)p(\lambda) \propto g(\lambda)^{1+\beta} e^{(y+\alpha-1)\phi(\lambda)}.$$

This hints at a general exponential family result, where if

$$p(y|\lambda) = f(y)g(\lambda)^a e^{t(y)\phi(\lambda)}$$

then the conjugate prior for  $\lambda$  has the form

$$p(\lambda) \propto g(\lambda)^b e^{c\phi(\lambda)}.$$

- (b) If  $y|\lambda \sim \text{Poisson}(\lambda)$  and  $\lambda \sim \text{Gamma}(\alpha, \beta)$  then

$$\begin{aligned}
p(\lambda|y) &= \frac{p(y|\lambda)p(\lambda)}{p(y)} \\
&= \left( \frac{e^{-\lambda} \lambda^y}{y!} \right) \left( \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \right) / p(y) \\
&\propto \lambda^{\alpha+y-1} e^{-(\beta+1)\lambda}
\end{aligned}$$

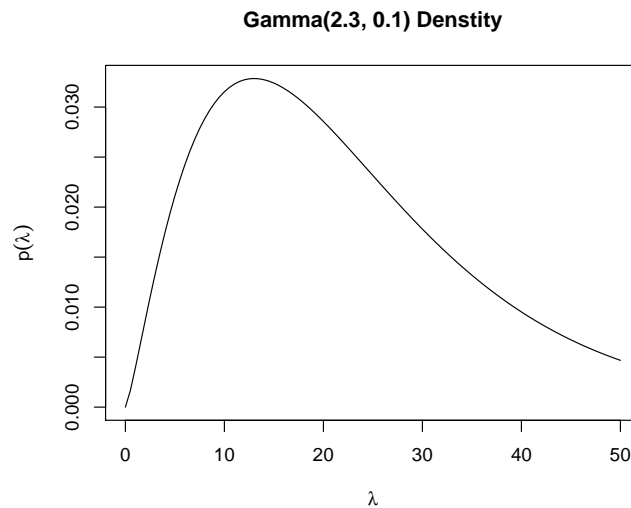
and so  $\lambda|y \sim \text{Gamma}(\alpha + y - 1, \beta + 1)$ .

- (c) If we have  $n$  independent observations from the same Poisson distribution, then the posterior distribution is

$$p(\lambda|y) \propto p(y|\lambda)p(\lambda) \propto (e^{-n\lambda}\lambda^{y_1+\dots+y_n}) (\lambda^{\alpha-1}e^{-\beta\lambda}) = \lambda^{y_1+\dots+y_n+\alpha-1}e^{-(n+\beta)\lambda}$$

so  $\lambda|y \sim \text{Gamma}(\sum y_i + \alpha - 1, n + \beta)$ . In the posterior, the prior distribution is equivalent to  $\alpha - 1$  Poisson events in  $\beta$  observations.

- (d) I would use a  $\text{Gamma}(2.3, 0.1)$  prior in this situation.



Features of this distribution include:

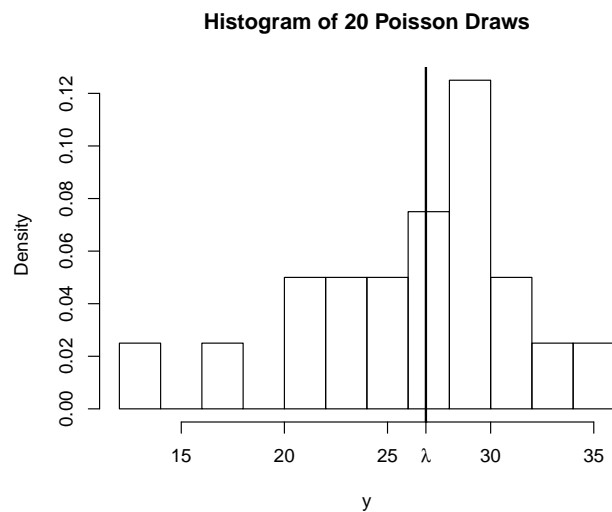
- $E(\lambda) = 23$ ,  $Var(\lambda) = 230$ , Median = 19.77
- $Pr(15 < \lambda < 25) = 0.2831$
- $Pr(10 < \lambda < 35) = 0.6250$
- $Pr(\lambda < 1) = 0.0017$ ,  $Pr(\lambda < 10) = 0.1905$
- $Pr(\lambda > 50) = 0.0596$

The mean and median are in the interval thought to be most likely, and it is highly right-skewed so values above 50 are somewhat likely to occur. Very small values less than 1 are unlikely, but it does have about a fifth of its mass below 10. Since we don't have more specific information, like exact values for the mean or median, I prefer to use a prior with a large variance and let the data vouch for themselves.

- (e) I got  $\lambda = 26.86$ .

```
set.seed(9231346)
lambda <- rgamma(1, 2.3, 0.1)
obs <- rpois(20, lambda)
```

- (f) My observations have mean 26.45 and variance 29.41.

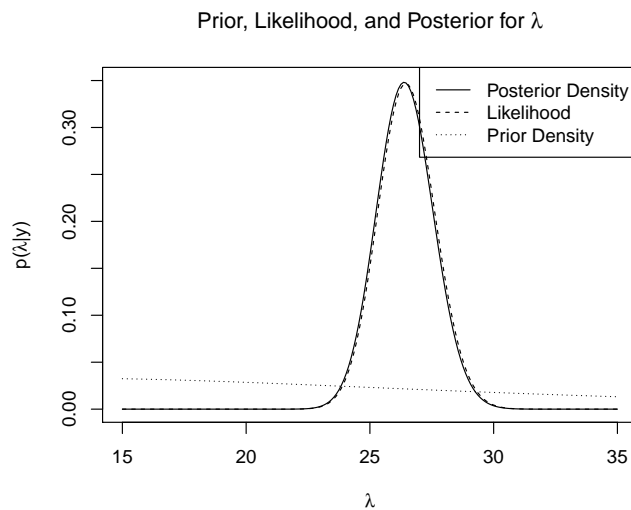


(g) The posterior density is

$$\begin{aligned}
 p(\lambda|y) &= p(y|\lambda)p(\lambda)/p(y) \\
 &\propto \left(e^{-n\lambda}\lambda^{\sum y_i}\right) (\lambda^{1.3}e^{-0.1\lambda}) \\
 &= \lambda^{\sum y_i + 1.3} e^{-(n+0.1)\lambda}
 \end{aligned}$$

which is a  $\text{Gamma}(\sum y_i + 2.3, n + 0.1)$  density. Inserting the observed values, this becomes  $\lambda|y \sim \text{Gamma}(531.3, 20.1)$ .

The posterior density, prior density, and normalized likelihood appear on the plot below.



(h) `post.int <- qgamma(c(0.05, 0.95), 531.3, 20.1)`

A 90% posterior interval for  $\lambda$  is  $[24.58, 28.35]$ .

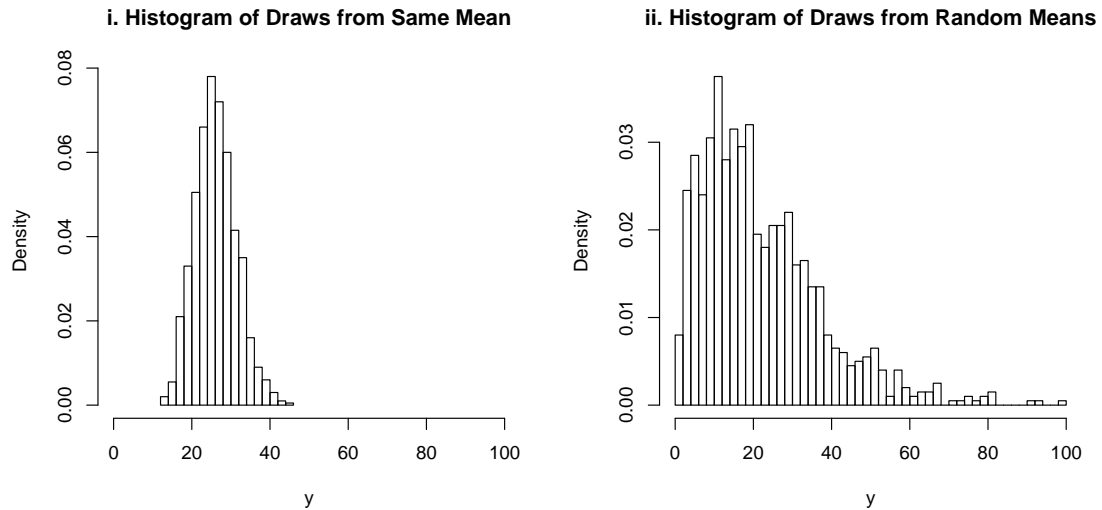
(i) `post.prob1 <- pgamma(20, 531.3, 20.1) - pgamma(10, 531.3, 20.1)`

$$Pr(10 < \lambda < 20|y) = 4.4528192\text{e-}10$$

(j) `post.prob2 <- pgamma(5, 531.3, 20.1)`

$$Pr(\lambda < 5|y) = 1.5879948\text{e-}199$$

(k) Here I've plotted histograms of random draws on the same horizontal scale.



Both distributions are centered near the mean of the original  $\text{Gamma}(2.3, 0.1)$  distribution. The single  $\lambda$  used to generate the data on the left is a bit larger than the mean of the prior distribution, so the distribution is shifted right. The draws from a common  $\lambda$  and tightly packed around their mean because their common Poisson distribution is their only source of variation. In contrast, the draws from distributions with random  $\lambda$  are much more spread out since they include additional variation from the distribution of  $\lambda$ .

6. (a) The likelihood is

$$p(y|\pi) = \binom{30}{y} \pi^y (1 - \pi)^{30-y}.$$

I don't immediately see a helpful way to put this in terms of  $\eta = \text{logit}(\pi)$  so I will find Jeffreys'



prior in terms of  $\pi$  and then transform.

$$\begin{aligned}
\log(p(y|\pi)) &= \log \binom{30}{y} + y \log(\pi) + (30 - y) \log(1 - \pi) \\
\frac{d}{d\pi} \log(p(y|\pi)) &= \frac{y}{\pi} - \frac{30 - y}{1 - \pi} \\
\frac{d^2}{d\pi^2} \log(p(y|\pi)) &= -\frac{y}{\pi^2} - \frac{30 - y}{(1 - \pi)^2} \\
J(\pi) &= -E \left( \frac{d^2}{d\pi^2} \log(p(y|\pi)) \right) \\
&= E \left( \frac{y}{\pi^2} + \frac{30 - y}{(1 - \pi)^2} \right) \\
&= \frac{30}{\pi} + \frac{30}{1 - \pi} \\
&= \frac{30}{\pi(1 - \pi)}
\end{aligned}$$

So Jeffreys' prior for  $\pi$  is

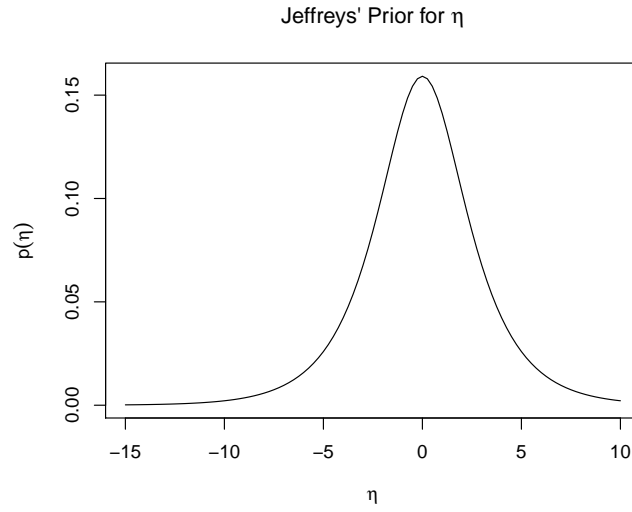
$$p(\pi) \propto (J(\pi))^{\frac{1}{2}} = \sqrt{30} \pi^{-\frac{1}{2}} (1 - \pi)^{-\frac{1}{2}}$$

which is  $Beta\left(\frac{1}{2}, \frac{1}{2}\right)$ .

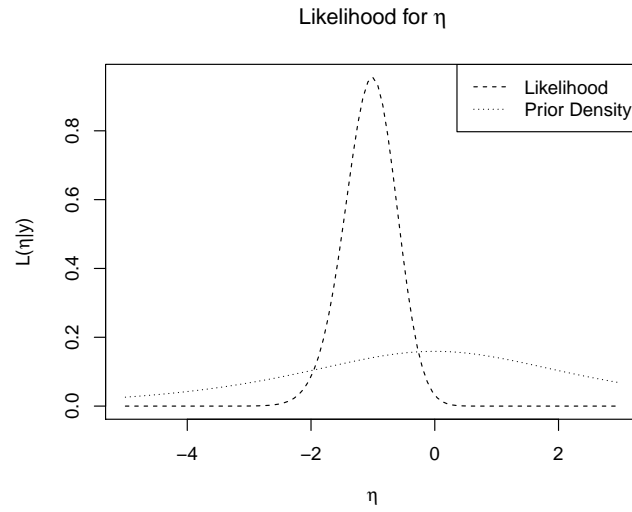
Since  $\pi = \frac{e^\eta}{1 + e^\eta}$  and  $\frac{d\pi}{d\eta} = \frac{e^\eta}{(1 + e^\eta)^2}$ , the prior for  $\eta$  is

$$\begin{aligned}
p(\eta) &= p \left( \pi = \frac{e^\eta}{1 + e^\eta} \right) \left| \frac{d\pi}{d\eta} \right| \\
&\propto \left( \frac{e^\eta}{1 + e^\eta} \right)^{-\frac{1}{2}} \left( \frac{1}{1 + e^\eta} \right)^{-\frac{1}{2}} \left( \frac{e^\eta}{(1 + e^\eta)^2} \right) \\
&= \frac{(e^\eta)^{\frac{1}{2}}}{(1 + e^\eta)^{-1} (1 + e^\eta)^2} \\
&= \frac{e^{\frac{\eta}{2}}}{1 + e^\eta}
\end{aligned}$$

for  $-\infty < \eta < \infty$ .



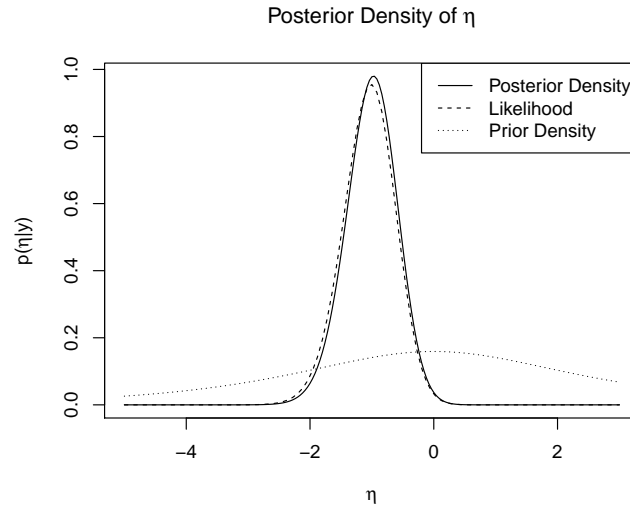
(b) The plot below shows the normalized likelihood and the prior density.



(c) The posterior distribution of  $\eta$  is

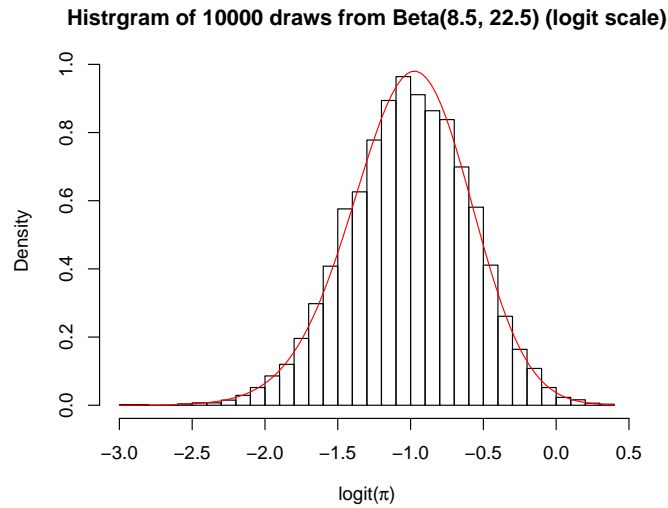
$$\begin{aligned}
 p(\eta|y) &\propto p(y|\eta)p(\eta) \\
 &\propto \left( \left( \frac{e^\eta}{1+e^\eta} \right)^y \left( \frac{1}{1+e^\eta} \right)^{30-y} \right) \left( \frac{e^{\frac{\eta}{2}}}{1+e^\eta} \right) \\
 &= e^{(y+\frac{1}{2})\eta} (1+e^\eta)^{-31}.
 \end{aligned}$$

The posterior density, likelihood, and prior density are plotted together in the plot below.



The maximum of the posterior density is slightly to the left of the maximum of the likelihood. I expected it to be to the right, between the maximum of the likelihood and the maximum of the prior density.

- (d) Jeffreys' prior for  $\pi$  is  $\text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$  so if we observe  $y = 8$  successes and  $n - y = 22$  failures the posterior distribution is  $\pi|y \sim \text{Beta}(8.5, 22.5)$ .

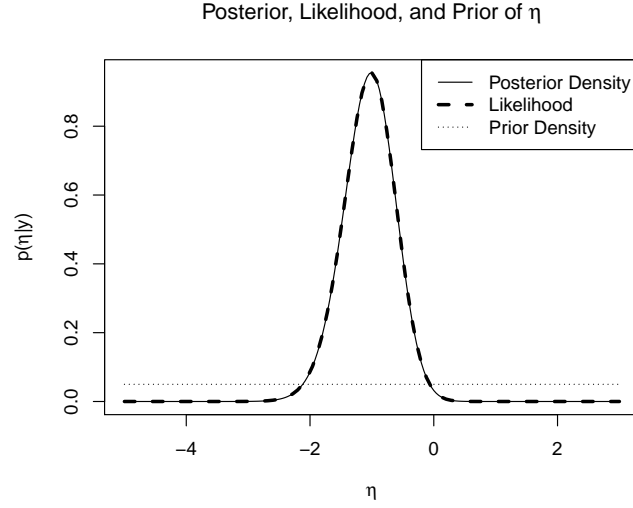


I drew 10000 draws from the posterior distribution of  $\pi$ . After transforming them to the logit scale, the distribution of the draws is a perfect match for the posterior distribution of  $\eta$ . It looks like these draws could have come from either distribution.

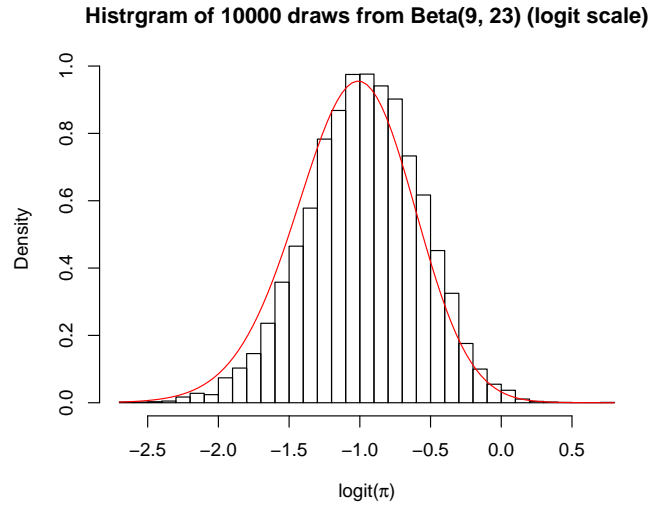
- (e) Using the prior  $p(\eta) \propto 1$ , the posterior distribution of  $\eta$  is

$$\begin{aligned} p(\eta|y) &\propto \left(\frac{e^\eta}{1+e^\eta}\right)^y \left(\frac{1}{1+e^\eta}\right)^{30-y} \\ &= e^{y\eta} (1+e^\eta)^{-30}. \end{aligned}$$

The posterior is identical to the normalized likelihood.



Using  $\pi \sim \text{Unif}(0, 1)$ , the posterior is  $\pi|y \sim \text{Beta}(9, 23)$ .

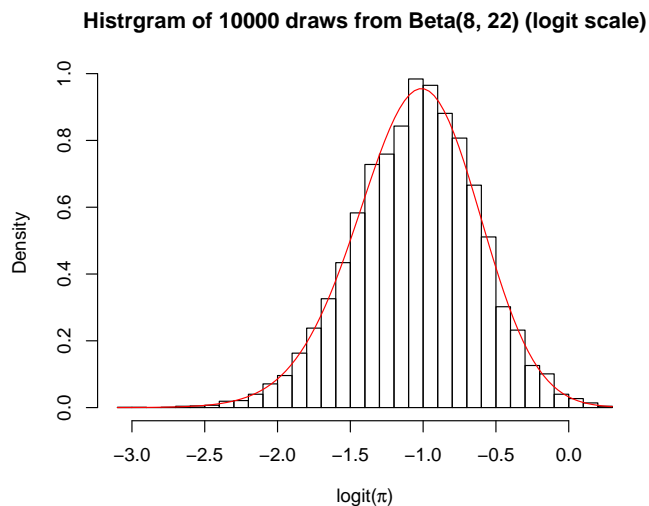


In the case where we use a uniform prior for  $\pi$ , draws from the posterior distribution of  $\pi$  do not match the posterior distribution that we would get for  $\eta$  by using a uniform prior for  $\eta$ . The transformed distribution of  $\pi|y$  is shifted slightly to the right compared to the distribution of  $\eta|y$ . This is a problem if we want to truly represent a lack of prior knowledge. In this case, using a uniform prior changed the results based on which parameterization the prior was applied to. Thus, saying “all values are equally likely” on one scale does not imply that all values are equally likely on another scale. In the absence of information, we may wish to avoid choosing between different scales.

The strength of Jeffreys’ prior is that, if we do not wish to decide on one parameterization, we can use any parameterization that is convenient and the results will be the same. In binomial problem, the posterior distributions were identical when transformed between the probability scale and the

logit scale, so any conclusion we make would be the same whether the inference was based on  $\pi$  or  $\eta$ .

- (f) The prior  $\pi \sim \text{Beta}(0, 0)$  results in the posterior  $\pi|y = 8 \sim \text{Beta}(8, 22)$  when 8 successes and 22 failures are observed.



The distribution of the transformed posterior draws appears to match the posterior distribution of  $\eta$ . The posterior density of  $\pi$  is

$$p(\pi|y = 8) \propto \pi^7(1 - \pi)^{21}.$$

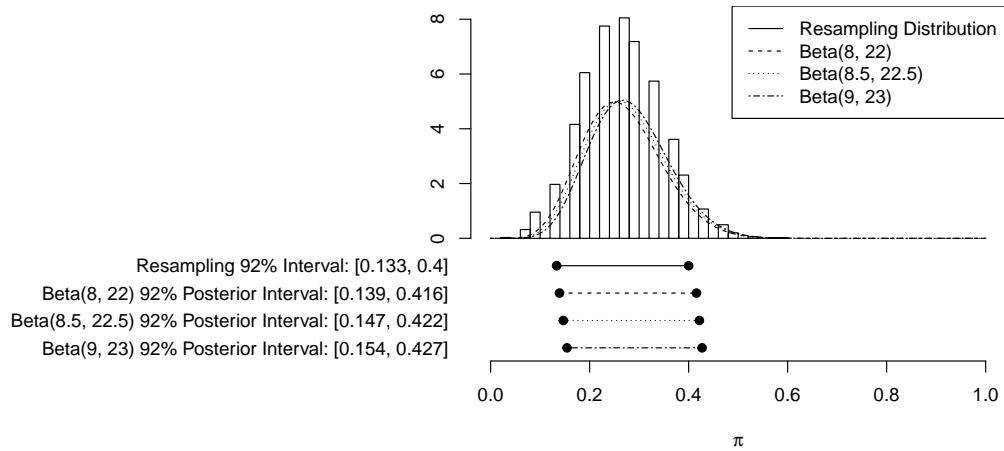
Transforming this to  $\eta$ ,

$$\begin{aligned} p(\eta|y = 8) &= p\left(\pi = \frac{e^\eta}{1 + e^\eta} | y = 8\right) \left| \frac{d\pi}{d\eta} \right| \\ &\propto \left( \frac{e^\eta}{1 + e^\eta} \right)^7 \left( \frac{1}{1 + e^\eta} \right)^{21} \frac{e^\eta}{(1 + e^\eta)^2} \\ &= e^{8\eta}(1 + e^\eta)^{-30}, \end{aligned}$$

which is the distribution found in part e. This result makes sense because the Beta(0, 0) prior has its mass concentrated at 0 and 1, and the constant prior on the logit scale has infinite mass in the tails which would be concentrated at 0 and 1 when transformed to the probability scale.

- (g) With a sample proportion of  $\frac{8}{30}$  and a sample size of only 30 the sampling distribution will be skewed, so I will bootstrap. The frequentist interval is based on 10000 resamples.

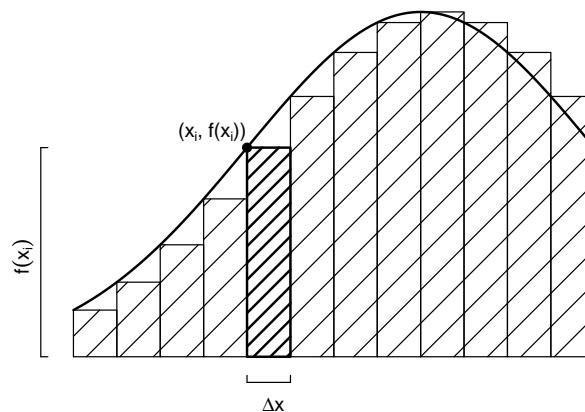
Relative Frequency Histogram of Resamples with Density Curves



All three Bayesian intervals are similar to the frequentist interval. However, each prior is symmetric with a median of 0.5 and so all the of the posterior intervals are pulled a little bit towards 0.5. The Beta(1, 1) prior had the biggest shift, but a small effect can be seen even the Beta(0, 0) prior. This is interesting because the Beta(8, 22) posterior mean is equal to the sample proportion, but the posterior distribution is slightly more skewed than the resampling distribution.

7. The labeled rectangle has area  $f(x_i)\Delta x$ , where  $\Delta x = x_{i+1} - x_i$ . Numerical integration is done by summing the areas of all the rectangles in the interval being integrated over.

Illustration of Numerical Integration



8. (a) The Beta(0, 0) density is

$$p(\pi) \propto \frac{1}{\pi} \frac{1}{1 - \pi}$$

and the likelihood is

$$p(y = 10|\pi) \propto \pi^{10}(1 - \pi)^{30}$$

so the posterior density is

$$\begin{aligned} p(\pi|y = 10) &\propto (\propto \pi^{10}(1 - \pi)^{30}) \left( \frac{1}{\pi} \frac{1}{1 - \pi} \right) \\ &= \pi^9(1 - \pi)^{29} \end{aligned}$$

and thus  $\pi|y = 8 \sim \text{Beta}(8, 30)$  which is a proper distribution.

(b) If we observe  $y = 0$  successes then the posterior density is

$$\begin{aligned} p(\pi|y = 0) &\propto (\propto \pi^0(1 - \pi)^{40}) \left( \frac{1}{\pi} \frac{1}{1 - \pi} \right) \\ &= \frac{1}{\pi}(1 - \pi)^{39} \end{aligned}$$

This is an improper distribution because

$$\begin{aligned} \int_0^1 \frac{1}{\pi}(1 - \pi)^{39} d\pi &\geq \int_0^1 \frac{1}{\pi} d\pi \\ &= \log(\pi)|_{\pi=0}^1 \\ &= \infty. \end{aligned}$$

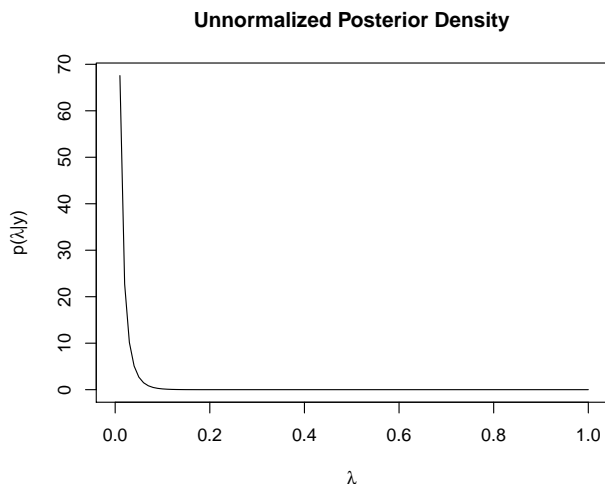
Similarly, if we observe  $y = 40$  successes then the posterior density is

$$\begin{aligned} p(\pi|y = 40) &\propto (\propto \pi^{40}(1 - \pi)^0) \left( \frac{1}{\pi} \frac{1}{1 - \pi} \right) \\ &= \pi^{39} \frac{1}{1 - \pi} \end{aligned}$$

which is also improper because

$$\begin{aligned} \int_0^1 \pi^{39} \frac{1}{1 - \pi} d\pi &\geq \int_0^1 \frac{1}{1 - \pi} d\pi \\ &= -\log(1 - \pi)|_{\pi=0}^1 \\ &= \infty. \end{aligned}$$

- (c) One way to use computation to check if whether the posterior is proper is to examine numerical integration near potentially problematic points and infer whether the integral would converge.
- i. First, examine the unnormalized posterior density.



We see that  $\lim_{\lambda \rightarrow 0^+} = \infty$ , which is reason to suspect that the integral is infinite. Since an integral is the sum of areas of small rectangles, we can show that this diverges by showing that the limit of the areas of these rectangles is positive.

To be conservative, I will use a right-sum so that the rectangles are below the curve.

- ii. The following code finds the areas of rectangles near  $\lambda = 0$ . We only need to know what happens near the asymptote, so I use a manageable number of rectangles. Smaller rectangles could be used if this is not enough precision.

```
# Define the number of rectangles and integration bounds.
nrects = 10
left = 0
right = 0.0001

# Define the endpoints of our rectangles (from right to left).
endpoints <- seq(right, left, (left - right) / nrects)
deltas <- endpoints[-length(endpoints)] - endpoints[-1]

# Find the areas of the rectangles.
heights <- posterior(endpoints[-length(endpoints)])
areas <- heights * deltas
```

- iii. Running the above produces these results:

Right Endpoint	Width	f(Right Endpoint)	Area
0.00010	0.00001	9961.07	0.09961
0.00009	0.00001	11072.18	0.11072
0.00008	0.00001	12461.06	0.12461
0.00007	0.00001	14246.77	0.14247
0.00006	0.00001	16627.71	0.16628
0.00005	0.00001	19961.04	0.19961
0.00004	0.00001	24961.03	0.24961
0.00003	0.00001	33294.36	0.33294
0.00002	0.00001	49961.01	0.49961
0.00001	0.00001	99961.01	0.99961

As the endpoint approaches 0, the area increases. Since the individual terms of the sum have a nonzero limit, the sum must diverge. Therefore, the posterior distribution is improper.



This method works when dealing with an asymptote. If we need to investigate the end behavior on an unbounded support, the method will need to be adjusted. If we suspect that the integral converges, we need to use a different method based on a different convergence test. This is why the first step of plotting the density is crucial.

## R Code Appendix

### Problem 3(b)

```
curve(dbeta(x, 1, 2), from = 0, to = 1, main = 'Beta(1, 2) Density',
      xlab = expression(pi), ylab = expression(p(pi)))
```

### Problem 4

```
curve(dbeta(x, 0.8, 1.5), from = 0, to = 1, main = 'Beta(0.8, 0.75) Density',
      xlab = expression(pi), ylab = expression(p(pi)))
```

### Problem 5(d)

```
curve(dgamma(x, 2.3, 0.1), from = 0, to = 50,
      main = 'Gamma(2.3, 0.1) Density',
      xlab = expression(lambda), ylab = expression(p(lambda)))
```

### Problem 5(f)

```
hist(obs, freq = FALSE, breaks = 15,
     main = 'Histogram of 20 Poisson Draws', xlab = 'y')
abline(v = lambda, lwd = 2)
axis(1, at = lambda, labels = expression(lambda))
```

### Problem 5(g)

```
# Find the normalizing constant for the likelihood
lambda.grid <- seq(0.005, 50, 0.005)
delta <- lambda.grid[-1] - lambda.grid[-length(lambda.grid)]
areas.L.pois <- dpois(sum(obs), 20*lambda.grid[-length(lambda.grid)]) * delta
L.pois.const <- sum(areas.L.pois)

curve(dgamma(x, 2.3, 0.1), from = 15, to = 35, lty = 3, ylim = c(0,0.35),
      main = expression(paste('Prior, Likelihood, and Posterior for ',lambda)),
      xlab = expression(lambda), ylab = expression(paste('p(',lambda,',ly)'))))
curve(dpois(sum(obs), 20*x) / L.pois.const, lty = 2, n = 200, add = TRUE)
curve(dgamma(x, 531.3, 20.1), lty = 1, n = 200, add = TRUE)
legend('topright', lty = c(1, 2, 3),
      legend = c('Posterior Density', 'Likelihood', 'Prior Density'))
```

### Problem 5(k)

```
samelambda <- rpois(1000, lambda)
lambdas <- rgamma(1000, 2.3, 0.1)
difflambda <- rpois(1000, lambdas)

par(mfrow = c(1, 2))
hist(samelambda, breaks = 12, freq = FALSE, xlim = c(0, 100), xlab = 'y',
     main = 'i. Histogram of Draws from Same Mean')
hist(difflambda, breaks = 36, freq = FALSE, xlim = c(0, 100), xlab = 'y',
     main = 'ii. Histogram of Draws from Random Means')
```

### Problem 6(a)

```

# The kernelly bits of the prior
eta.prior.unnorm <- function(x){
  return(exp(x/2)/(1+exp(x)))
}

# It's pretty close to 0 outside of the interval [-15, 5] but
# I'll integrate from -50 to 50 just to be safe (with more resolution
# in that interval).
eta.grid <- c(seq(-50, -15.01, 0.01), seq(-15, 5, 0.001), seq(5.01, 50, 0.01))

# Get the widths of the rectangles.
delta <- eta.grid[-1] - eta.grid[-length(eta.grid)]

# Now get the rectangle areas.
areas.prior <- eta.prior.unnorm(eta.grid[-length(eta.grid)]) * delta

# Find the normalizing constant.
const.prior <- sum(areas.prior)

# Here's the normalized prior!
eta.prior <- function(x){
  return(exp(x/2)/(const.prior * (1+exp(x))))
}

# Plot it
curve(eta.prior(x), from = -15, to = 10,
      main = expression(paste('Jeffreys\' Prior for ', eta)),
      xlab = expression(eta), ylab = expression(p(eta)))

```

## Problem 6(b)

```

eta.L.unnorm <- function(eta, y){
  return(dbinom(y, 30, exp(eta)/(1+exp(eta))))
}

# It's very close to 0 except in [-4, 2].
eta.grid <- seq(-5, 3, 0.001)

# Get the widths of the rectangles.
delta <- eta.grid[-1] - eta.grid[-length(eta.grid)]

# Now get the rectangle areas.
areas.L <- eta.L.unnorm(eta.grid[-length(eta.grid)], 8) * delta

# Find the normalizing constant.
const.L <- sum(areas.L)

# Normalized likelihood
eta.L <- function(eta, y){
  return(dbinom(y, 30, exp(eta)/(1+exp(eta))) / const.L)
}

# Plot
curve(eta.L(x, 8), from = -5, to = 3, lty = 2,
      main = expression(paste('Likelihood for ', eta)),
      xlab = expression(eta), ylab = expression(paste('L(', eta, '|y)'))
curve(eta.prior(x), lty = 3, add = TRUE)
legend('topright', lty = c(2, 3), legend = c('Likelihood', 'Prior Density'))

```

## Problem 6(c)

```

# The kernelly bits of the posterior
eta.post.unnorm <- function(eta,y){
  return(exp((y+0.5)*eta)/(1+exp(eta))^31)
}

# I'll integrate from -4 to 2.
eta.grid <- seq(-40, 20, 0.001)

# Get the widths of the rectangles.
delta <- eta.grid[-1] - eta.grid[-length(eta.grid)]

# Now get the rectangle areas.
areas.post <- eta.post.unnorm(eta.grid[-length(eta.grid)], 8) * delta

# Find the normalizing constant.
const.post <- sum(areas.post)

# Normalized posterior
eta.post <- function(eta, y){
  return(exp((y+0.5)*eta)/(const.post*(1+exp(eta))^31))
}

# Plot them
curve(eta.post(x, 8), from = -5, to = 3, lty = 1, n = 201,
      main = expression(paste('Posterior Density of ', eta)),
      xlab = expression(eta), ylab = expression(paste('p(', eta, '|y)'))
curve(eta.L(x, 8), lty = 2, add = TRUE)
curve(eta.prior(x), lty = 3, add = TRUE)
legend('topright', lty = c(1, 2, 3),
      legend = c('Posterior Density', 'Likelihood', 'Prior Density'))

```

## Problem 6(d)

```

draws <- rbeta(10000, 8.5, 22.5)
logitdraws <- log(draws) - log(1-draws)

hist(logitdraws, freq = FALSE, breaks = 30,
     main = 'Histogram of 10000 draws from Beta(8.5, 22.5) (logit scale)',
     xlab = expression(paste('logit(', pi, ')')))
curve(eta.post(x, 8), col = 'red', add = TRUE)

```

## Problem 6(e)

```

# The kernelly bits of the posterior
eta.u.post.unnorm <- function(eta, y){
  return(exp(y*eta)/(1+exp(eta))^30)
}

# I'll integrate from -4 to 2.
eta.grid <- seq(-4, 2, 0.001)

# Get the widths of the rectangles.
delta <- eta.grid[-1] - eta.grid[-length(eta.grid)]

# Now get the rectangle areas.
areas.eta.u <- eta.u.post.unnorm(eta.grid[-length(eta.grid)], 8) * delta

# Find the normalizing constant.
const.eta.u <- sum(areas.eta.u)

# Normalized posterior
eta.u.post <- function(eta, y){
  return(exp(y*eta)/(const.eta.u*(1+exp(eta))^30))
}

```

```

}

# Plot them
curve(eta.u.post(x, 8), from = -5, to = 3, lwd = 1, lty = 1, n = 201,
      main = expression(paste('Posterior, Likelihood, and Prior of ', eta)),
      xlab = expression(eta), ylab = expression(paste('p(', eta, '|y)'))
curve(eta.L(x, 8), lwd = 3, lty = 2, add = TRUE)
segments(x0 = -5, y0 = 0.05, x1 = 3, lwd = 1, lty = 3)
legend('topright', lty = c(1, 2, 3), lwd = c(1, 3, 1),
      legend = c('Posterior Density', 'Likelihood', 'Prior Density'))

draws <- rbeta(10000, 9, 23)
logitdraws <- log(draws) - log(1-draws)

hist(logitdraws, freq = FALSE, breaks = 30,
      main = 'Histogram of 10000 draws from Beta(9, 23) (logit scale)',
      xlab = expression(paste('logit(', pi, ')')))
curve(eta.u.post(x, 8), col = 'red', add = TRUE)

```

## Problem 6(f)

```

draws <- rbeta(10000, 8, 22)
logitdraws <- log(draws) - log(1-draws)

hist(logitdraws, freq = FALSE, breaks = 30,
      main = 'Histogram of 10000 draws from Beta(8, 22) (logit scale)',
      xlab = expression(paste('logit(', pi, ')')))
curve(eta.u.post(x, 8), col = 'red', add = TRUE)

```

## Problem 6(g)

```

set.seed(78634)

# Bootstrap interval
boot <- rbinom(10000, 30, 8/30) / 30
int.freq <- quantile(boot, c(0.04, 0.96))

# Posterior intervals
int.beta0 <- qbeta(c(0.04, 0.96), 8, 22)
int.beta5 <- qbeta(c(0.04, 0.96), 8.5, 22.5)
int.beta1 <- qbeta(c(0.04, 0.96), 9, 23)

# Plot them fancy-like
par(mar = c(5, 21, 4, 5) + 0.1)
hist(boot, freq = FALSE, breaks = 30, xlim = c(0, 1), ylim = c(-4, 8),
      main = 'Relative Frequency Histogram of Resamples with Density Curves',
      xlab = expression(pi), ylab = '', yaxt = 'n')
curve(dbeta(x, 8, 22), lty = 2, add = TRUE)
curve(dbeta(x, 8.5, 22.5), lty = 3, add = TRUE)
curve(dbeta(x, 9, 23), lty = 6, add = TRUE)

# Draw the intervals
points(x = int.freq, y = c(-1, -1), type = 'o', pch = 19, lty = 1)
points(x = int.beta0, y = c(-2, -2), type = 'o', pch = 19, lty = 2)
points(x = int.beta5, y = c(-3, -3), type = 'o', pch = 19, lty = 3)
points(x = int.beta1, y = c(-4, -4), type = 'o', pch = 19, lty = 6)

# Add an axis
axis(2, at = seq(0, 8, 2), labels = seq(0, 8, 2))
axis(2, las = 2, at = -1:-4, tick = FALSE, labels = c(
  paste0('Resampling 92% Interval: [',
    round(int.freq[1], 3), ', ', round(int.freq[2], 3), ']'),

```

```

paste0('Beta(8, 22) 92% Posterior Interval: [' ,
       round(int.beta0[1], 3), ' , ' , round(int.beta0[2], 3), ']''),
paste0('Beta(8.5, 22.5) 92% Posterior Interval: [' ,
       round(int.beta5[1], 3), ' , ' , round(int.beta5[2], 3), ']''),
paste0('Beta(9, 23) 92% Posterior Interval: [' ,
       round(int.beta1[1], 3), ' , ' , round(int.beta1[2], 3), ']''))

# Legen--wait for it--d.
legend('topright', lty = c(1, 2, 3, 6),
       legend = c('Resampling Distribution',
                  'Beta(8, 22)', 'Beta(8.5, 22.5)', 'Beta(9, 23)'))

```

## Problem 7

```

grids <- seq(-2, 1, 0.25)
heights <- dnorm(grids)
curve(dnorm(x), from = -2, to = 1, lwd = 2, axes = FALSE,
      main = 'Illustration of Numerical Integration',
      ylim = c(0, 0.4), xlab = '', ylab = '')
rect(grids[-c(5, 13)], 0, grids[-c(1, 6)], heights[-c(5, 13)],
     lwd = 1, angle = 45, density = 5)
rect(-1, 0, -0.75, heights[5], lwd = 2, angle = 45, density = 10)
points(x = -1, heights[5], pch = 19)
axis(1, at = c(-1, -0.75), line = 0.5, tcl = 0.3, hadj = -0.7, padj = -1,
     labels = c(expression(paste(Delta,x)), ''))
axis(2, at = c(0, heights[5]), line = 0.5, tcl = 0.3, hadj = -2.4, padj = 1,
     labels = c(expression(f(x[i])), ''))
text(x = -1.2, y = 0.26,
     labels = expression(paste('(', x[i], ', ', f('(', x[i], ')')'))))

```

## Problem 8(c)

```

posterior <- function(x){
  return(((1-x)^39)/x)
}

curve(posterior(x), from = 0, to = 1,
      main = 'Unnormalized Posterior Density', xlab = expression(lambda),
      ylab = expression(paste('p(', lambda, '|y)'))))

resultstab <- cbind('Right Endpoint' = endpoints[-length(endpoints)],
                    'Width' = deltas,
                    'f(Right Endpoint)' = heights,
                    'Area' = areas)
print(xtable(resultstab, digits = c(0, 5, 5, 2, 5)),
      include.rownames = FALSE, floating = FALSE)

```