

Bayes: Homework 3

Leslie Gains-Germain

1. There are several parameters in this paper. First, there are the parameters that describe the run curve. The run curve is not a probability model, it is just a curve that describes the number of fish passing the weir over time. They use both the function for the normal pdf and the skew-normal pdf to describe the run curve. Second, there are the parameters for the process variation models. These models are used to describe the variation of daily passage counts around the counts predicted by the run curve. They use a normal probability model and a negative binomial model to describe the process variation. In the paper, they fit four separate Bayesian models, one for the Normal run curve, normal process variation model; one for the normal run curve, negative binomial process variation model; one for the skew-normal run curve, normal process variation model; and one for the skew normal-normal run curve, negative binomial process variation model.

Below are the priors for the parameters of the normal run curve:

$$\begin{aligned}
 \mu &\sim Unif(150, 300) & f(\mu) &= \frac{1}{150}I(\mu)_{(150,300)} \\
 \sigma &\sim Unif(1, 50) & f(\sigma) &= \frac{1}{49}I(\sigma)_{(1,50)} \\
 \log(S) &\sim N(7.5, 0.25) & f(\log(S)) &= \frac{1}{\sqrt{0.125\pi}}e^{-\frac{(\log(S)-7.5)}{0.125}}
 \end{aligned}$$

Below are the priors for the parameters of the skew-normal run curve:

$$\begin{aligned}
 \xi &\sim Unif(150, 300) & f(\xi) &= \frac{1}{150}I(\xi)_{(150,300)} \\
 \omega &\sim Unif(1, 50) & f(\omega) &= \frac{1}{49}I(\omega)_{(1,50)} \\
 \alpha &\sim Unif(-10, 10) & f(\alpha) &= \frac{1}{20}I(\alpha)_{(-10,10)} \\
 \log(S) &\sim N(7.5, 0.25) & f(\log(S)) &= \frac{1}{\sqrt{0.125\pi}}e^{-\frac{(\log(S)-7.5)}{0.125}}
 \end{aligned}$$

Below are the priors for the parameters of the normal process variation model:

$$\tau \sim Unif(0.1, 1000) \qquad f(\tau) = \frac{1}{999.9} I(\tau)_{(0.1, 1000)}$$

Below are the priors for the parameters of the negative binomial process variation model:

$$\theta \sim Gam(0.1, 0.1) \qquad f(\theta) = \frac{\theta^{-0.9} 0.1^{0.1} e^{-0.1\theta}}{\Gamma(0.1)}$$

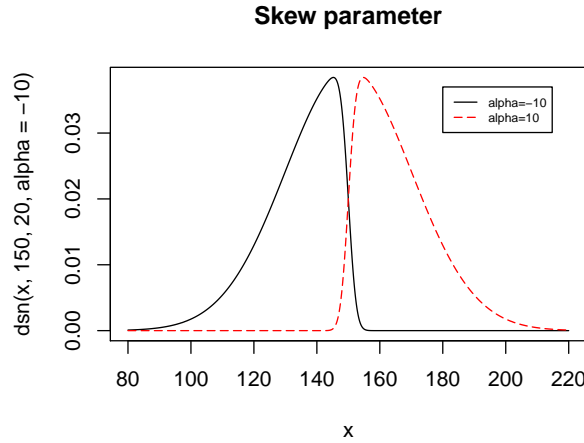
Let me start by saying that they do not discuss their choice of priors anywhere in the paper. In the following paragraphs, I speculate on what they were thinking.

I am going to discuss the priors chosen for the parameters of the model fit with the skew normal run curve and the negative binomial process variation. This is the most confusing model.

ξ represents the peak of the run curve, and they put a weakly informative prior on ξ showing that they expect the peak of the run curve to occur sometime between day 150 and day 300, with equal probability of occurring anywhere in that interval. I'm calling it weakly informative because I'm guessing that have more knowledge about when the peak usually occurs, within this interval, that is not reflected in the prior.

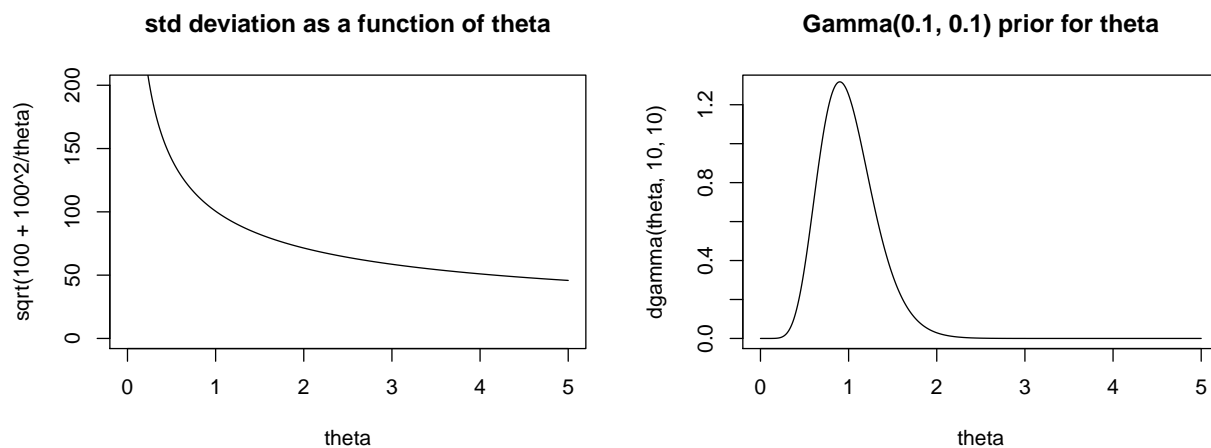
They put a non-informative prior on ω , the spread of the run curve. This parameter represents the length of the run. The uniform prior used reflects the knowledge that the run length is somewhere between 6 and 300 days long. I'm calling this a non-informative prior because I know, even as a lay person, that salmon runs are usually around 60 days long. A interval of 6 to 300 seems really wide for this parameter.

α is the parameter of the skew normal distribution that describes how skewed the distribution is. The sign of α indicates the direction of the skew (negative for left-skewed and positive for right-skewed), and the skewness increases as the magnitude of α increases. They put a uniform prior on the skew parameter between -10 and 10 . The plot below shows the difference in shapes of a skew normal distribution with $\alpha = 10$ and $\alpha = -10$. I would call this a non-informative prior because they aren't even incorporating knowledge about the direction of the skew. But, I think they *are* incorporating prior knowledge by choosing the skew normal distribution to describe the run curve. In this choice, they are reflecting knowledge that the run could either come in fast and then slowly drop off or vice versa.



The parameter S describes the total run size (the total number of fish passing the weir over the course of the run). Run sizes are really large, so they put the prior on $\log(S)$. They used a normal prior, indicating that the log of the run size is somewhere between 7 and 8, which means that the total run size is somewhere between 1097 and 2981 fish. They call this a relatively flat prior for the total run size. I think a flat prior makes sense in this situation because run sizes do vary greatly from year to year. Since they intend to use this model for years and years, they truly do have little knowledge about what the run sizes will be in subsequent years (although they do have a general idea about the smallest and largest run sizes).

Lastly, there is the dispersion parameter for the negative binomial process variation model. They defined the first parameter in the negative binomial model as a function of θ , $\lambda = \frac{\theta}{\theta + c_t}$. They chose this particular function for λ because it allows the passage count variability to increase with the magnitude of the passage while keeping the mean at c_t , the daily passage count on day t . It took me a while to figure out what parameterization of the negative binomial distribution they were using. Thanks to Megan and the openBUGS helpfile, I figured out the mean of the given distribution is in fact c_t and the variance is $c_t + \frac{c_t^2}{\theta}$. They chose a $Gam(0.1, 0.1)$ prior for θ . I picked reasonable arbitrary values of c_t and then explored the probable standard deviations, given the prior on θ . See the plots below. The prior on θ suggests that values of θ between 0.5 and 1.7 are the most probable. If $c_t = 20$, these values correspond to standard deviations between 16 and 77. If $c_t = 100$, these values correspond to standard deviations between 77 and 142. I would say that this is a weakly informative prior; the range of probably standard deviations is pretty large. Although the range makes sense and seems reasonable, it doesn't seem like they are incorporating much prior knowledge.



2. There are difficulties both in deciding whether it is appropriate use a non-informative prior in a Bayesian analysis and then in choosing a prior that is always vague, but non-informative priors can be useful as long as you check that the posterior density is proper and not sensitive to the choice of prior.
3. (a) Gelman's first argument for weakly informative priors is one of convenience; that it often is more convenient, mathematically, to use a weakly informative prior than a strongly informative one. For example, if we have a poisson likelihood, it's more convenient to choose a gamma prior, even if all of the information we have about the parameter can't be expressed as well with the gamma distribution as it could be with another distribution. He does mention here that "arguments for convenience are best justified by the claim that the answer would not have changed much had we been more accurate." I do agree with this argument, but I think it's kind of strange that he doesn't emphasize the importance of doing a sensitivity analysis to check that the posterior is not sensitive to the choice of prior. He emphasizes this in the previous section on non-informative priors, but I would add that it is also important here, especially if he is using convenience as a main argument for using weakly informative priors.

In the next section, he mentions that maybe we don't always want to include all of the relevant information we have about a parameter in the prior. He mentions that you might not want the prior to pull the posterior in any pre-determined direction, especially if the goal of the study is to test a theory. This seems really weird to me, because if your goal is to test some theory, why wouldn't you go about the analysis from a frequentist perspective? It seems like a main strength of the Bayesian analysis is the ability to incorporate prior knowledge; but if you don't really want the prior knowledge to influence results, why would you be doing a Bayesian analysis at all? Even after discussing this in class, I'm not sure I completely agree with or understand the idea of moving the prior distribution in the opposite direction of your existing knowledge about

the parameter in order to strengthen your results.

- (b) One option would be to start with a highly informative prior and then broaden it to incorporate more uncertainty. In this case, we could start with the prior they specified, a $Beta(2.3, 4.28)$ distribution. We could then add more probability in the tails to make it flatter and account for the prior knowledge that 12% of the previous estimates were 1. A second option would be to start with a $Unif(0, 1)$ prior and then change the shape slightly to account for there being more prior estimates between 0 and 0.5.

4. I'm not sure this was the most computationally efficient way to do this, but I simply found the 50th percentile and the 90th percentile for beta distributions with many different combinations of a and b . I then found the a and b that made the 50th percentile closest to 0.3 and the 90th percentile closest to 0.75. The beta distribution that I found was a $Beta(0.79, 1.47)$. The code and plot are shown below.

```
#define a, b and all combinations
a <- seq(0.1,5,0.01)
b <- seq(0.1,5,0.01)
grid.vals <- expand.grid(a, b)

#function to evaluate the median
median.fun <- function(vec) {qbeta(0.5, vec[1], vec[2])}

#find median at all combos of a and b and then subtract desired median of 0.3
median <- matrix(apply(grid.vals,1, median.fun), nrow=length(a), ncol=length(b))
median.min <- abs(median-0.3)

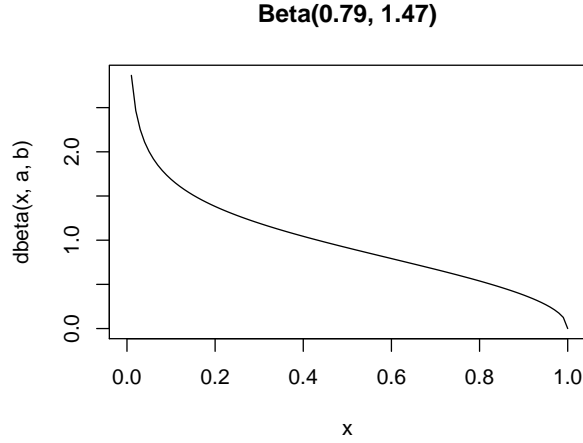
#same process for 90th percentile
fun.90 <- function(vec90) {qbeta(0.9, vec90[1], vec90[2])}
perc.90 <- matrix(apply(grid.vals,1, fun.90), nrow=length(a), ncol=length(b))
min.90 <- abs(perc.90-0.75)

#find a and b that minimize the sum of the differences from 0.3 and 0.75
sum <- median.min+min.90
a <- a[which(sum == min(sum), arr.ind = TRUE)[1]]
b <- b[which(sum == min(sum), arr.ind = TRUE)[2]]

#check that the a,b produce desired result
#qbeta(.5, a, b) = 0.3000459
#qbeta(.9, a, b) = 0.7507122

#plot
```

```
x <- seq(0, 1, by=0.01)
plot(x, dbeta(x, a, b), type="l", main="Beta(0.79, 1.47)")
```



5. (a) Suppose x is fixed and λ is unknown. Then,

$$\begin{aligned} p(\lambda|x) &= \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \frac{e^{-1\lambda} 1^{x+1} \lambda^{(x+1)-1}}{x!} \end{aligned}$$

This is a $Gam(x+1, 1)$ distribution where $x+1 \geq 0$ is fixed and $\lambda > 0$ is unknown. So, if we multiply a Poisson likelihood by a gamma prior, the posterior distribution will be a gamma as well. In general, if $X \sim Poi(\lambda)$ and $\lambda \sim Gam(a, b)$, then posterior distribution of $\lambda|x$ is $Gam(x+a, b+1)$, for one draw from the $Poi(\lambda)$ distribution. See the proof below.

$$\begin{aligned} p(\lambda|x) &= \frac{p(x|\lambda) * p(\lambda)}{m(x)} \\ &= \frac{\frac{e^{-\lambda} \lambda^x}{x!} \frac{\lambda^{a-1} b^a e^{-b\lambda}}{\Gamma(a)}}{m(x)} \\ &= \frac{\frac{b^a}{x! \Gamma(a)} e^{-(b+1)\lambda} \lambda^{x+a-1}}{m(x)} \\ &\propto e^{-(b+1)\lambda} \lambda^{x+a-1} \end{aligned}$$

Suppose X follows a distribution that is a member of the exponential family, and suppose the prior for a parameter θ is also a member of the exponential family. Then, the posterior distribution is also a member of the exponential family. Just as we saw with the poisson and gamma distributions, if you multiply the pdf or pmf of an exponential family distribution by the pdf or pmf of another exponential family distribution, you will get back a member of the exponential family. See the proof below.

$$\begin{aligned}
f(\theta|x) &\propto f(x|\theta)f(\theta) \\
&\propto h(x)c(\theta)e^{\sum_{i=1}^n w_i(\theta)t_i(x)}l(\theta)e^{\sum_{j=1}^{n_j} w_j(\theta)} \\
&\propto h(x)m(\theta)e^{\sum_{k=1}^{n_k} w_k(\theta)t_k(x)}
\end{aligned}$$

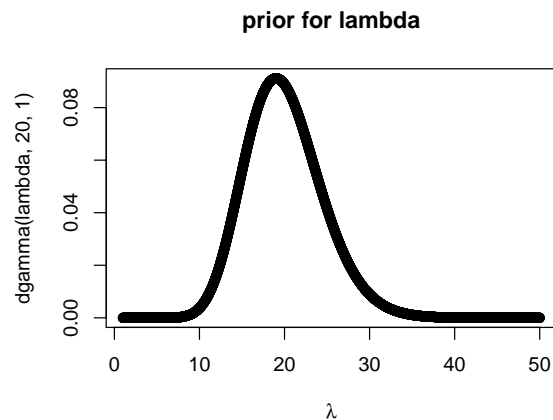
- (b) In part (a) I showed that for one draw from a $Poi(\lambda)$ distribution with a prior of $Gam(a, b)$, the posterior distribution is $Gam(x + a, b + 1)$. Here, I show that for n independent draws from a $Poi(\lambda)$, the posterior distribution is $Gam(\sum_{i=1}^n x_i + a, b + n)$.

$$\begin{aligned}
p(\lambda|x) &= \frac{p(x|\lambda) * p(\lambda)}{m(x)} \\
&= \frac{\frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod_{i=1}^n x_i!} \frac{\lambda^{a-1} b^a e^{-b\lambda}}{\Gamma(a)}}{\int_0^\infty \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod_{i=1}^n x_i!} \frac{\lambda^{a-1} b^a e^{-b\lambda}}{\Gamma(a)} d\lambda} \\
&= \frac{\frac{b^a}{\prod_{i=1}^n x_i! \Gamma(a)} e^{-(n+b)\lambda} \lambda^{\sum x_i + a - 1}}{\frac{b^a \Gamma(\sum x_i + a)}{(1+b)^{\sum x_i + a} \prod_{i=1}^n x_i! \Gamma(a)}} \\
&= \frac{e^{-(n+b)\lambda} \lambda^{\sum x_i + a - 1} (n + b)^{\sum x_i + a}}{\Gamma(\sum x_i + a)}
\end{aligned}$$

- (c) From the likelihood, our estimate of the rate parameter is just x , the number of ‘successes’ observed in one effort. For example, if we are a fishing trawler, and we put out our net once and get 5 fish, we would estimate λ to be 5. The posterior mean for λ given x is $\frac{x+a}{b+1}$. If we observed $x = 5$ and we start with a $Gam(1, 1)$ prior, then the posterior mean for λ is $6/2 = 3$. If we start with a $Gam(3, 3)$ prior, then the posterior mean for λ is

$8/4 = 2$. We can think of this process as adding a successful observations and increasing the total effort by b . For this example, we would be increasing the number of fish caught by a , but then we divide this over $b + 1$ drops of the net.

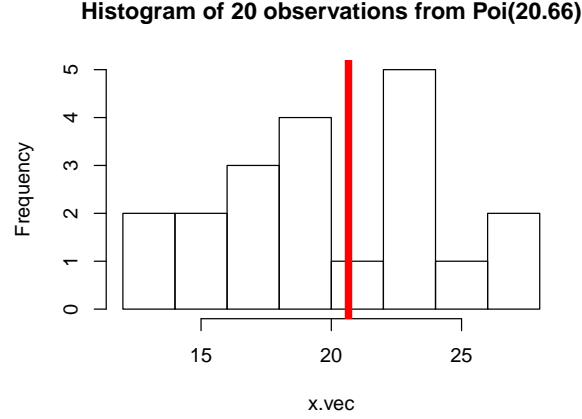
- (d) I started by solving for possible values of α and β that would make the mean of the distribution close to 20 and the standard deviation close to 5. I experimented with different values and found that a $Gam(20, 1)$ looked the best for the given conditions. Based on this distribution, the probability that λ is less than 10 or greater than 35 is $0.00345 + 0.00232 = 0.00577$. The probability that λ is between 10 and 15 or 25 and 35 is $0.1213 + 0.1313 = 0.2526$. The probability that λ is between 15 and 25 is 0.7416. I think this is a reasonable way to satisfy the constraints provided.



- (e) I drew a λ of 20.66 and then drew 20 independent observations from $Poi(20.66)$.

```
set.seed(15)
lambda.draw <- rgamma(1, 20, 1)
x.vec <- rpois(20, lambda.draw)
```

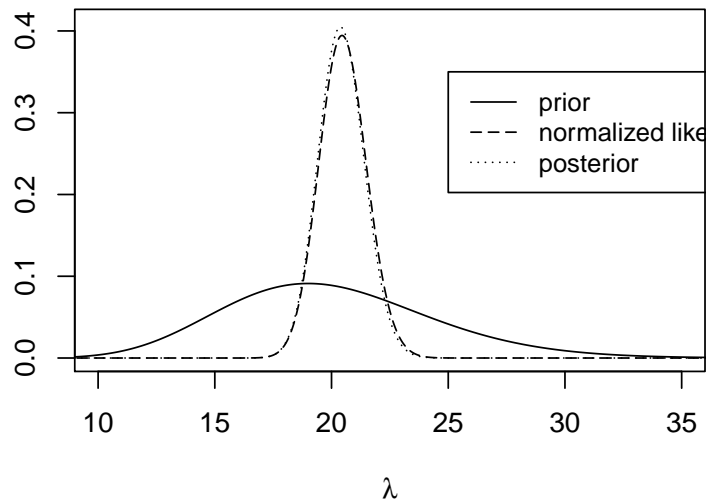
- (f) The plot of the observed data is below.



(g) From (b), the posterior distribution of $\lambda|x$ is $Gam(\sum_{i=1}^n x_i + a, b + n)$ which is $Gam(409 + 20, 1 + 20) = Gam(429, 21)$. My work is below. I used numerical integration to sum under the likelihood curve, and then I used this number to normalize the likelihood.

$$\begin{aligned}
 p(\lambda|x) &= \frac{p(x|\lambda) * p(\lambda)}{m(x)} \\
 &= \frac{\frac{e^{-20\lambda} \lambda^{409}}{\prod_{i=1}^{20} x_i!} \frac{\lambda^{19} 1^9 e^{-1\lambda}}{\Gamma(20)}}{\int_0^\infty \frac{e^{-20\lambda} \lambda^{409}}{\prod_{i=1}^{20} x_i!} \frac{\lambda^{19} 1^{20} e^{-\lambda}}{\Gamma(20)} d\lambda} \\
 &= \frac{\frac{1}{\prod_{i=1}^{20} x_i! \Gamma(20)} e^{-(20+b)\lambda} \lambda^{409+19}}{\frac{\Gamma(409+20)}{(1+20)^{409+20} \prod_{i=1}^{20} x_i! \Gamma(20)}} \\
 &= \frac{e^{-(21)\lambda} \lambda^{429-1} (21)^{429}}{\Gamma(429)}
 \end{aligned}$$

Plot of prior, likelihood, and posterior



- (h) The 90% posterior interval for λ is (19.83, 22.41).

```
c(qgamma(0.275, 429, 21), qgamma(0.975, 429, 21))
```

- (i) The posterior probability that λ is between 10 and 20 is 0.337.

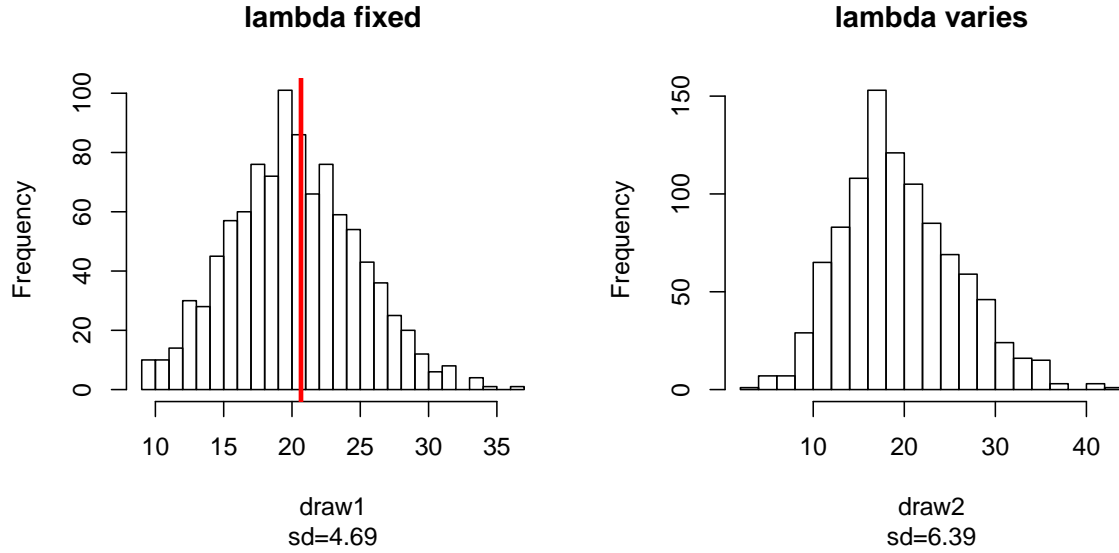
```
pgamma(20, 429, 21)-pgamma(10, 429, 21)
```

- (j) The posterior probability that λ is less than 5 is $7.65 * 10^{-124}$.

```
pgamma(5, 429, 21)
```

- (k) The range of values drawn from the Poisson distribution was narrower when λ was fixed, and the standard deviation was smaller. The range of the observations for the random λ was 3 to 44, compared to a range of 9 to 37 for the fixed λ . The standard deviation of the observations for the random λ was 6.39 compared to 4.69. These results reflect the fact that there are two sources of uncertainty reflected in the second histogram - both uncertainty in the value of λ , and uncertainty in the observation. The first histogram reflects only uncertainty in the value of the observations.

Also, The histograms happened to be centered around the same value, but this occurred because the fixed value of λ that was drawn was close to the mean of the prior distribution (20).



6. (a) Gelman shows that for a binomial likelihood, the Jeffrey's prior on θ is $Beta(.5, .5)$. I do a transformation to find the kernel of the pdf of $\eta = \text{logit}(\theta)$.

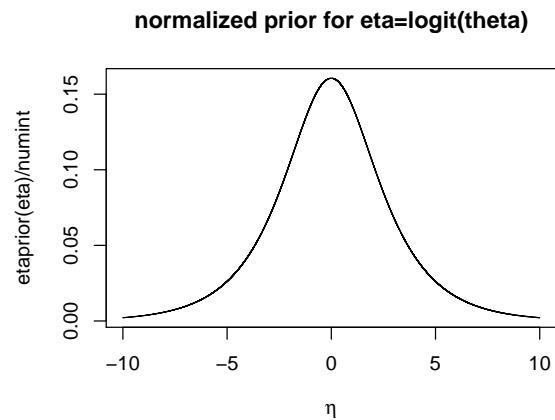
$$\begin{aligned}
 \eta &= \log\left(\frac{\theta}{1-\theta}\right) & \theta &= \frac{e^\eta}{1+e^\eta} \\
 f_\eta(\eta) &= f_\theta(\theta) * d\theta/d\eta \\
 &= f_\theta\left(\frac{e^\eta}{1+e^\eta}\right) * \frac{e^\eta}{(1+e^\eta)^2} \\
 &= \frac{\left(\frac{e^\eta}{1+e^\eta}\right)^{-0.5} \left(\frac{1}{1+e^\eta}\right)^{-0.5} \frac{e^\eta}{(1+e^\eta)^2}}{\beta(0.5, 0.5)} \\
 &\propto \frac{e^{\eta/2}}{1+e^\eta}
 \end{aligned}$$

where $\eta \in (-\infty, \infty)$. I then use R to find the normalizing constant to make the area under this curve sum to 1. In subsequent sections, I use the same code to find the normalizing constant.

```

etaprior <- function(eta) {
  exp(eta/2)/(1+exp(eta))
}
eta <- seq(-10, 10, by=0.0001)
y <- 0.0001*etaprior(eta)
numint <- sum(y)
plot(eta, etaprior(eta)/numint, type="l", main="normalized prior for eta=logit(theta)", xlab=expression(eta))

```

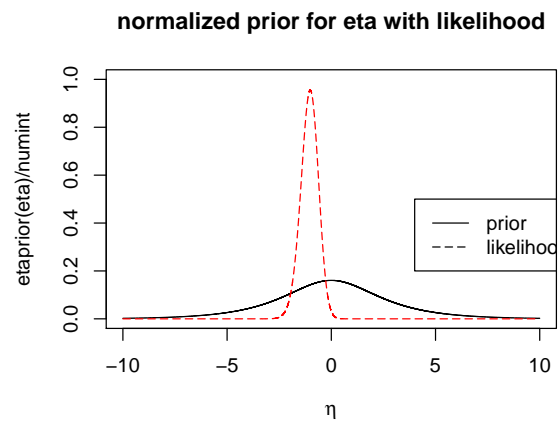


(b) The R function to calculate the likelihood and the plot of the likelihood is shown below.

```

etalike <- function(eta){
  choose(30, 8)*(exp(eta)/(1+exp(eta)))^8*(1/(1+exp(eta)))^22
}

```

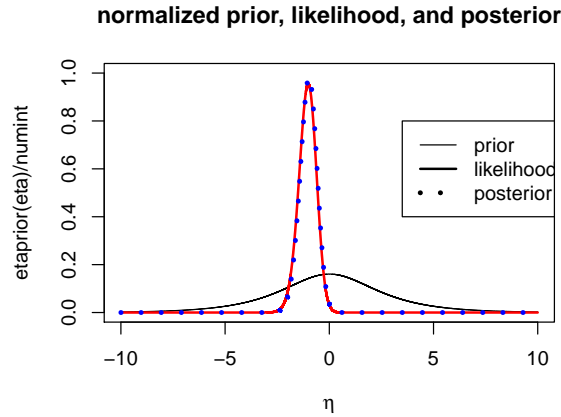


(c) The plot of the normalized prior, likelihood, and posterior is shown below. I am not surprised that the posterior distribution looks almost exactly like the likelihood. The

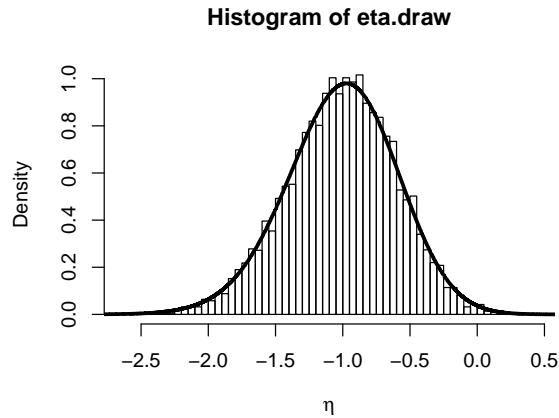
prior is pretty flat, allowing the data to dominate, and the prior is centered near the likelihood.

$$f(\eta|x=8) \propto \binom{30}{8} \left(\frac{e^\eta}{1+e^\eta}\right)^8 \left(\frac{1}{1+e^\eta}\right)^{22} \frac{e^{\eta/2}}{1+e^\eta}$$

$$\propto \frac{e^{17\eta/2}}{(1+e^\eta)^{31}}$$

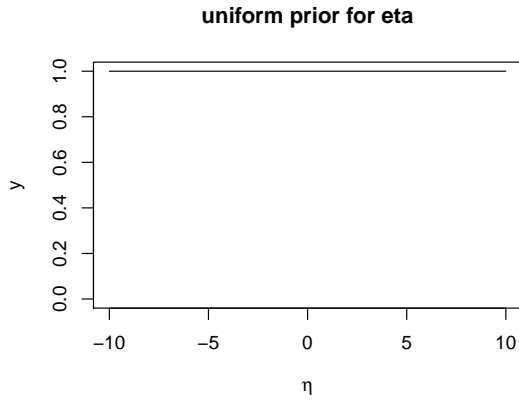


- (d) Using previous results, the posterior distribution for θ is a $Beta(8.5, 22.5)$. The posterior distribution found in (c) approximates the histogram created here. This shows that it is equivalent to find the posterior for η itself or find the posterior for θ and then solve for η .

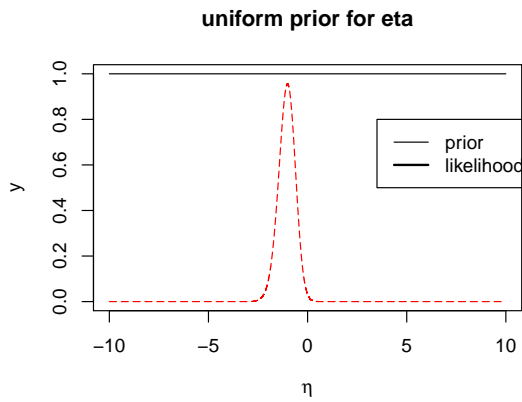


- (e) i. The prior distribution on θ is $Unif(0,1)$, and the prior distribution on η is an

improper uniform distribution over the whole real line so that $f(\eta) = I(\eta)_{(-\infty, \infty)}$.

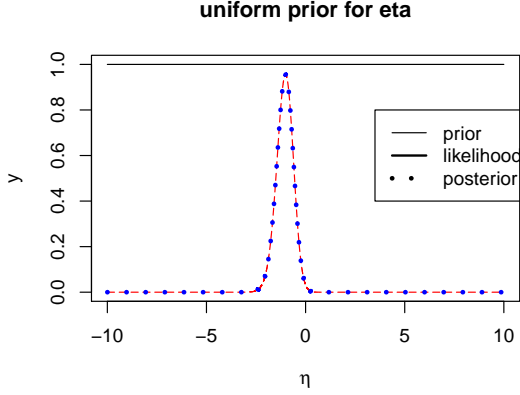


ii. The likelihood function is the same as part (b) and is plotted with the prior below.



iii. Because the prior distribution for η is uniform over the whole real line, the posterior distribution for η is just the normalized likelihood. See work below.

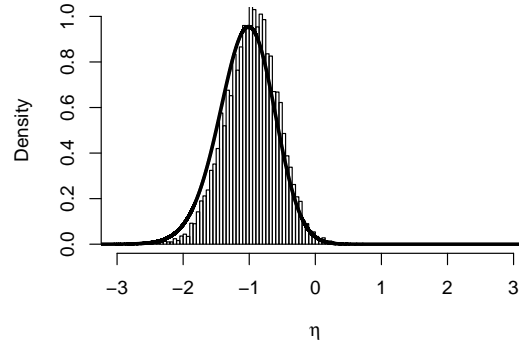
$$\eta|x \propto \binom{30}{8} \left(\frac{e^\eta}{1+e^\eta}\right)^8 \left(\frac{1}{1+e^\eta}\right)^{22} I(\eta)_{(-\infty, \infty)}$$



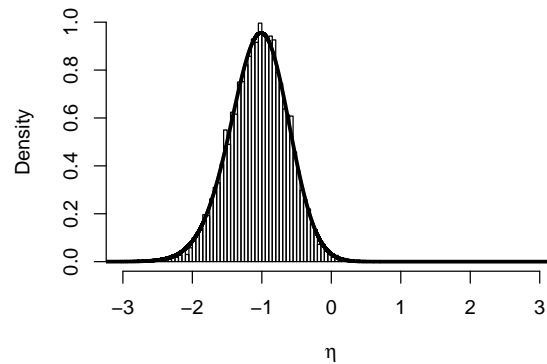
iv. I show below that the posterior distribution for θ is $Beta(9, 23)$.

$$\begin{aligned}\theta|x &\propto \theta^8(1-\theta)^{22}I(\theta)_{(0,1)} \\ &\propto \theta^{9-1}(1-\theta)^{23-1}I(\theta)_{(0,1)}\end{aligned}$$

I obtain 10000 independent draws from a $Beta(9, 23)$, and then I calculate η for each draw. It turns out that the histogram of draws does not match up exactly with the posterior distribution for η found by starting with a uniform prior on η . I think this shows that you can't just slap a flat prior on θ and a flat prior η and expect to get the same results. A flat uniform prior will give us different results depending on what parameterization of θ we apply the prior to. If we really want to express the same beliefs by applying a prior to θ or a prior to a transformed version of θ , we need to start with a Jeffrey's prior for θ then do the transformation to find the prior for the transformed version of θ .



- (f) If we start with a $Beta(0,0)$ prior on θ , the posterior distribution for θ is $Beta(8,22)$. I obtained 10000 draws from a $Beta(8,22)$, and it matched up pretty closely with the posterior of η obtained from the uniform prior on η .



Seeing this, I went back and wondered what would happen if we started with a $Beta(0,0)$ prior on θ and then transformed to find the prior for η . It turns out that, after transformation, the prior for η is the improper uniform distribution over the whole real line! See my work below. Even though the $Beta(0,0)$ distribution is improper, it does follow Jeffrey's principle. We get the same result whether we use the $Beta(0,0)$ prior for θ or whether we use the improper uniform prior on η .

$$\begin{aligned}
\eta &= \log\left(\frac{\theta}{1-\theta}\right) & \theta &= \frac{e^\eta}{1+e^\eta} \\
f_\eta(\eta) &= f_\theta(\theta) * d\theta/d\eta \\
&= f_\theta\left(\frac{e^\eta}{1+e^\eta}\right) * \frac{e^\eta}{(1+e^\eta)^2} \\
&= \frac{\left(\frac{e^\eta}{1+e^\eta}\right)^{-1} \left(\frac{1}{1+e^\eta}\right)^{-1} \frac{e^\eta}{(1+e^\eta)^2}}{\beta(0.5, 0.5)} \\
&\propto 1I(\eta)_{(-\infty, \infty)}
\end{aligned}$$

(g) I constructed a 92% Wald confidence interval for θ by hand, $0.267 \pm 1.75\sqrt{0.267 * 0.733/30} = (0.126, 0.408)$. The posterior intervals under the given priors are shown in the following table.

Prior	92% posterior	posterior interval	width of interval
Wald interval		(0.126, 0.408)	0.282
Beta(0, 0)	Beta(8, 22)	(0.139, 0.416)	0.277
Beta(0.5, 0.5)	Beta(8.5, 22.5)	(0.147, 0.422)	0.275
Beta(1, 1)	Beta(9, 23)	(0.154, 0.427)	0.273

The $Beta(0, 0)$ prior gives an interval that is closest to the Wald interval. I am actually surprised that the interval from the $Beta(0, 0)$ prior isn't closer to the Wald interval, because I think that if you use a non-informative prior, inference should be similar to what a frequentist would find. The Wald interval isn't completely trustworthy because the normal approximation may not be very good with a relatively small sample size of 30, so I experimented to see if the Wald interval is closer to the $Beta(0, 0)$ interval for larger sample sizes. If we increase the sample size to 80, the Wald interval for θ is actually (0.222, 0.311), which is very close to the $Beta(0, 0)$ interval of (0.223, 0.312).

Also, the Wald interval is wider than the posterior intervals, and among the posterior intervals, the widths decrease as the posterior parameters increase. Also, the intervals

are shifted farther right as the posterior parameters increase, reflecting the contribution of the prior. For a symmetric *Beta* distribution, it will always be true that the posterior distribution gets pulled towards 0.5 because the $Beta(\alpha, \alpha)$ prior adds α successes and α failures to what was observed.

7. Next page.

...

8. (a) Yes. I show below that the resulting posterior distribution is a $Beta(10, 40)$ which is a proper distribution.

$$\begin{aligned}\theta|x &\propto \theta^{10}(1-\theta)^{30}\theta^{-1}(1-\theta)^{-1}I(\theta)_{(0,1)} \\ &\propto \theta^{10-1}(1-\theta)^{30-1}I(\theta)_{(0,1)}\end{aligned}$$

- (b) No. I show below that if we observe 0 or 40 successes, the resulting Beta distribution would have a parameter that is 0. Since both parameters in the Beta distribution must be greater than 0, this is not a proper distribution.

$$\begin{aligned}\theta|x &\propto \theta^0(1-\theta)^{40}\theta^{-1}(1-\theta)^{-1}I(\theta)_{(0,1)} \\ &\propto \theta^{0-1}(1-\theta)^{40-1}I(\theta)_{(0,1)} \\ \theta|x &\propto \theta^{40}(1-\theta)^0\theta^{-1}(1-\theta)^{-1}I(\theta)_{(0,1)} \\ &\propto \theta^{40-1}(1-\theta)^{0-1}I(\theta)_{(0,1)}\end{aligned}$$

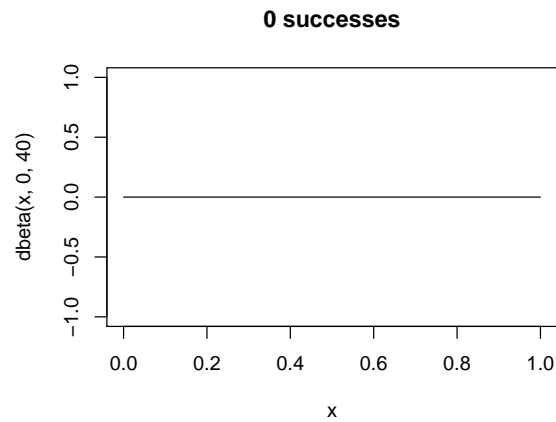
In the code below, I find the area under the $Beta(10, 40)$ distribution by numerical integration. It sums to 1 and is a proper distribution.

```
#define the support of the beta distribtuion
x <- seq(0, 0.9999, by=0.0001)
#left hand riemann sum
is.pdf <- 0.0001*dbeta(x, 10, 40)
#sum(is.pdf)=1
```

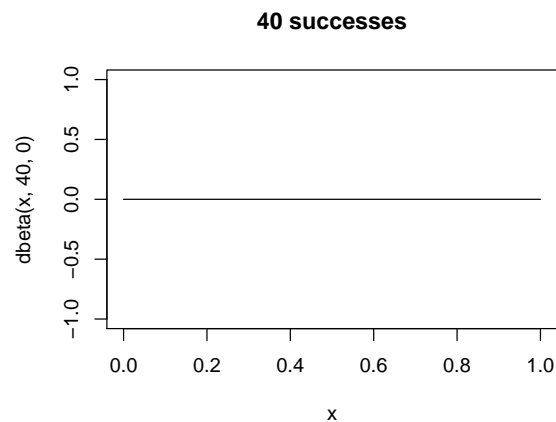
In the code below, I find the area under the $Beta(0, 40)$ and $Beta(40, 0)$ curves by numerical integration. We can see that both curves are just horizontal lines at 0. I find the area under the 0 successes curve to be infinite, and I find the area under the 40 successes curve to be 0.

```
#define the support of the beta distribtuion
x <- seq(0, 0.9999, by=0.0001)
#left hand riemann sum
is.pdf0 <- 0.0001*dbeta(x, 0, 40)
```

```
#sum(is.pdf0)=Inf
plot(x, dbeta(x, 0, 40), type="l", main="0 successes")
```



```
#define the support of the beta distribtuion
x <- seq(0, 0.9999, by=0.0001)
#left hand riemann sum
is.pdf40 <- 0.0001*dbeta(x, 40, 0)
#sum(is.pdf40)=0
plot(x, dbeta(x, 40, 0), type="l", main="40 successes")
```



I think this is really interesting because in part (a) we started with an improper prior and ended up with a proper posterior. In part (b), however, we started with an improper prior and ended up with an improper posterior. I think the lesson here is that sometimes you can use an improper prior distribution, and sometimes you can't. You just have to put in the work to check that the resulting posterior distribution is proper.

R code appendix

```
require(sn)
x <- seq(80, 220, by=0.01)
plot(x, dsn(x, 150, 20, alpha=-10), type="l", main="Skew parameter")
lines(x, dsn(x, 150, 20, alpha=10), col="red", lty=5)
legend(180, 0.037, c("alpha=-10", "alpha=10"), lty=c(1,5), col=c("black", "red"), cex=0.6)

par(mfrow=c(1,2))
theta <- seq(0, 5, by=0.01)
plot(theta, sqrt(100+100^2/theta), main="std deviation as a function of theta", type="l", ylim=c(0, 200))
plot(theta, dgamma(theta, 10, 10), type="l", ylim=c(0, 1.3), main="Gamma(0.1, 0.1) prior for theta")

#define a, b and all combinations
a <- seq(0.1,5,0.01)
b <- seq(0.1,5,0.01)
grid.vals <- expand.grid(a, b)

#function to evaluate the median
median.fun <- function(vec) {qbeta(0.5, vec[1], vec[2])}

#find median at all combos of a and b and then subtract desired median of 0.3
median <- matrix(apply(grid.vals,1, median.fun), nrow=length(a), ncol=length(b))
median.min <- abs(median-0.3)

#same process for 90th percentile
fun.90 <- function(vec90) {qbeta(0.9, vec90[1], vec90[2])}
perc.90 <- matrix(apply(grid.vals,1, fun.90), nrow=length(a), ncol=length(b))
min.90 <- abs(perc.90-0.75)

#find a and b that minimize the sum of the differences from 0.3 and 0.75
sum <- median.min+min.90
a <- a[which(sum == min(sum), arr.ind = TRUE)[1]]
b <- b[which(sum == min(sum), arr.ind = TRUE)[2]]

#check that the a,b produce desired result
#qbeta(.5, a, b) = 0.3000459
#qbeta(.9, a, b) = 0.7507122

#plot
x <- seq(0, 1, by=0.01)
plot(x, dbeta(x, a, b), type="l", main="Beta(0.79, 1.47)")

lambda <- seq(1,50, by=0.01)
#plot(x, dgamma(x, 40,2))
#plot(x, dgamma(x, 400/15,20/15))
plot(lambda, dgamma(lambda, 20, 1), xlab=expression(lambda), main="prior for lambda")
#pgamma(10, 20, 1)
#1-pgamma(35,20,1)
#pgamma(15, 20, 1) - pgamma(10, 20, 1)
#pgamma(35, 20, 1) - pgamma(25, 20, 1)
#pgamma(25, 20, 1) - pgamma(15, 20, 1)
```

```
set.seed(15)
lambda.draw <- rgamma(1, 20, 1)
x.vec <- rpois(20, lambda.draw)
```

```
hist(x.vec, main="Histogram of 20 observations from Poi(20.66)", nclass=10)
abline(v=lambda.draw, col="red", lwd=6)
```

```
loglike.fun <- function(lambda){
  -length(x.vec)*lambda+sum(x.vec)*log(lambda)-sum(log(factorial(x.vec)))
}
like.fun <- function(lambda){exp(loglike.fun(lambda))}
plot(lambda, dgamma(lambda, 20, 1), type="l", xlim=c(10,35), ylim=c(0, 0.41), xlab=expression(lambda), ylab="")
lines(lambda, dgamma(lambda, 429, 21), lty=3)
lines(lambda, like.fun(lambda)/(2.08304*10^{-25}), lty=5)
legend(25,0.35, c("prior", "normalized likelihood", "posterior"), lty=c(1,5,3),
      cex=0.9)

#numerical integration so that I can plot normalized likelihood function
int <- seq(1,50, .00001)
y <- 0.00001*like.fun(int)
lhsum <- sum(y)
```

```
c(qgamma(0.275, 429, 21), qgamma(0.975, 429, 21))
```

```
pgamma(20, 429, 21)-pgamma(10, 429, 21)
```

```
pgamma(5, 429, 21)
```

```
set.seed(26)
par(mfrow=c(1,2))
draw1 <- rpois(1000, lambda.draw)
hist(draw1, main="lambda fixed", sub="sd=4.69", nclass=20)
abline(v=lambda.draw, col="red", lwd=3)

lambda.first <- matrix(rgamma(1000, 20, 1), nrow=1000, ncol=1)
draw.fun <- function(x){rpois(1, x)}
draw2 <- apply(lambda.first, 1, draw.fun)
hist(draw2, nclass=20, main="lambda varies", sub="sd=6.39")
```

```
etaprior <- function(eta) {
  exp(eta/2)/(1+exp(eta))
}
eta <- seq(-10, 10, by=0.0001)
y <- 0.0001*etaprior(eta)
numint <- sum(y)
plot(eta, etaprior(eta)/numint, type="l", main="normalized prior for eta=logit(theta)", xlab=expression(eta))
```



```
etalike <- function(eta){
  choose(30, 8)*(exp(eta)/(1+exp(eta)))^8*(1/(1+exp(eta)))^22
}
```

```
y2 <- 0.0001*etalike(eta)
numint2 <- sum(y2)
plot(eta, etaprior(eta)/numint, type="l", main="normalized prior for eta with likelihood", ylim=c(0,1), xlab=expression(eta))
lines(eta, etalike(eta)/numint2, lty=5, col="red")
legend(4, 0.5, c("prior", "likelihood"), lty=c(1,5))
```

```
etapost <- function(eta){
  exp(17*eta/2)/((1+exp(eta))^31)
}
y3 <- 0.0001*etapost(eta)
numint3 <- sum(y3)
plot(eta, etaprior(eta)/numint, type="l", lwd=1, main="normalized prior, likelihood, and posterior", ylim=c(0,1), xlab=expression(eta))
lines(eta, etalike(eta)/numint2, lwd=2, col="red")
lines(eta, etapost(eta)/numint3, lwd=4, lty=3, col="blue")
legend(3.5, 0.8, c("prior", "likelihood", "posterior"), lwd=c(1,2,4), lty=c(1,1,3))
```

```
theta.draw <- rbeta(10000, 8.5, 22.5)
eta.draw <- log(theta.draw/(1-theta.draw))
hist(eta.draw, nclass=100, xlab=expression(eta), freq=FALSE)
lines(eta, etapost(eta)/numint3, lwd=3)
```

```
etaunifprior <- function(eta) {
  1
}
eta <- matrix(seq(-10, 10, by=0.0001))
y <- apply(eta, 1, etaunifprior)
etalike <- function(eta){
  choose(30, 8)*(exp(eta)/(1+exp(eta)))^8*(1/(1+exp(eta)))^22
}
plot(eta, y, type="l", main="uniform prior for eta", xlab=expression(eta), ylim=c(0,1))
```

```
etaunifprior <- function(eta) {
  1
}
eta <- matrix(seq(-10, 10, by=0.0001))
y <- apply(eta, 1, etaunifprior)
etalike <- function(eta){
  choose(30, 8)*(exp(eta)/(1+exp(eta)))^8*(1/(1+exp(eta)))^22
}
plot(eta, y, type="l", main="uniform prior for eta", xlab=expression(eta), ylim=c(0,1))
lines(eta, etalike(eta)/numint2, lty=5, col="red")
legend(3.5, 0.8, c("prior", "likelihood"), lwd=c(1,2), lty=c(1,1))
```

```

etaunifprior <- function(eta) {
  1
}
eta <- matrix(seq(-10, 10, by=0.0001))
y <- apply(eta, 1, etaunifprior)
etalike <- function(eta){
  choose(30, 8)*(exp(eta)/(1+exp(eta)))^8*(1/(1+exp(eta)))^22
}
plot(eta, y, type="l", main="uniform prior for eta", xlab=expression(eta), ylim=c(0,1))
lines(eta, etalike(eta)/numint2, lty=5, col="red")
lines(eta, etalike(eta)/numint2, lwd=4, lty=3, col="blue")
legend(3.5, 0.8, c("prior", "likelihood", "posterior"), lwd=c(1,2,4), lty=c(1,1,3))

```

```

set.seed(11)
theta.draw <- rbeta(10000, 9, 23)
eta.draw <- log(theta.draw/(1-theta.draw))
hist(eta.draw, nclass=100, xlab=expression(eta), freq=FALSE, ylim=c(0,1), xlim=c(-3,3), main="")
lines(eta, etalike(eta)/numint2, lwd=3)

```

```

set.seed(11)
theta.draw <- rbeta(10000, 8, 22)
eta.draw <- log(theta.draw/(1-theta.draw))
hist(eta.draw, nclass=100, xlab=expression(eta), freq=FALSE, ylim=c(0,1), xlim=c(-3,3), main="")
lines(eta, etalike(eta)/numint2, lwd=3)

```

```

#define the support of the beta distribtuion
x <- seq(0, 0.9999, by=0.0001)
#left hand riemann sum
is.pdf <- 0.0001*dbeta(x, 10, 40)
#sum(is.pdf)=1

```

```

#define the support of the beta distribtuion
x <- seq(0, 0.9999, by=0.0001)
#left hand riemann sum
is.pdf0 <- 0.0001*dbeta(x, 0, 40)
#sum(is.pdf0)=Inf
plot(x, dbeta(x, 0, 40), type="l", main="0 successes")

```

```

#define the support of the beta distribtuion
x <- seq(0, 0.9999, by=0.0001)
#left hand riemann sum
is.pdf40 <- 0.0001*dbeta(x, 40, 0)
#sum(is.pdf40)=0
plot(x, dbeta(x, 40, 0), type="l", main="40 successes")

```