HW1 - Tardella

Your Last+First Name: Rossini Valerio, Your Matricola: 1613638

1_a)Sample survey: Suppose we are going to sample 100 individuals from a county (of size much larger than 100) and ask each sampled person whether they support policy Z or not. Let $Y_i = 1$ if person i in the sample supports the policy, and $Y_i = 0$ otherwise.

-Answer: Y_i is a r.v. with bernoulli distribution that assumes 1 if person i in the sample supports the policy, 0 otherwise. The joint distribution of $Pr(Y_1 = y_1, ..., Y_{100} = y_{100}|\theta)$ is equal to:

$$Pr(Y_1 = y_1, ..., Y_{100} = y_{100}|\theta) = \prod_{i=1}^{100} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}$$

We also know that if $Y_1, ..., Y_n \sim Ber(\theta)$ then $Y_1 + ... + Y_n \sim Bin(n, \theta)$. So we can write the joint distribution $Pr(\sum_{i=1}^{n} Y_i = y|\theta)$ in this way:

$$Pr(\sum_{i=1}^{n} Y_i = y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y} = \binom{100}{y}\theta^{y}(1-\theta)^{100-y}$$

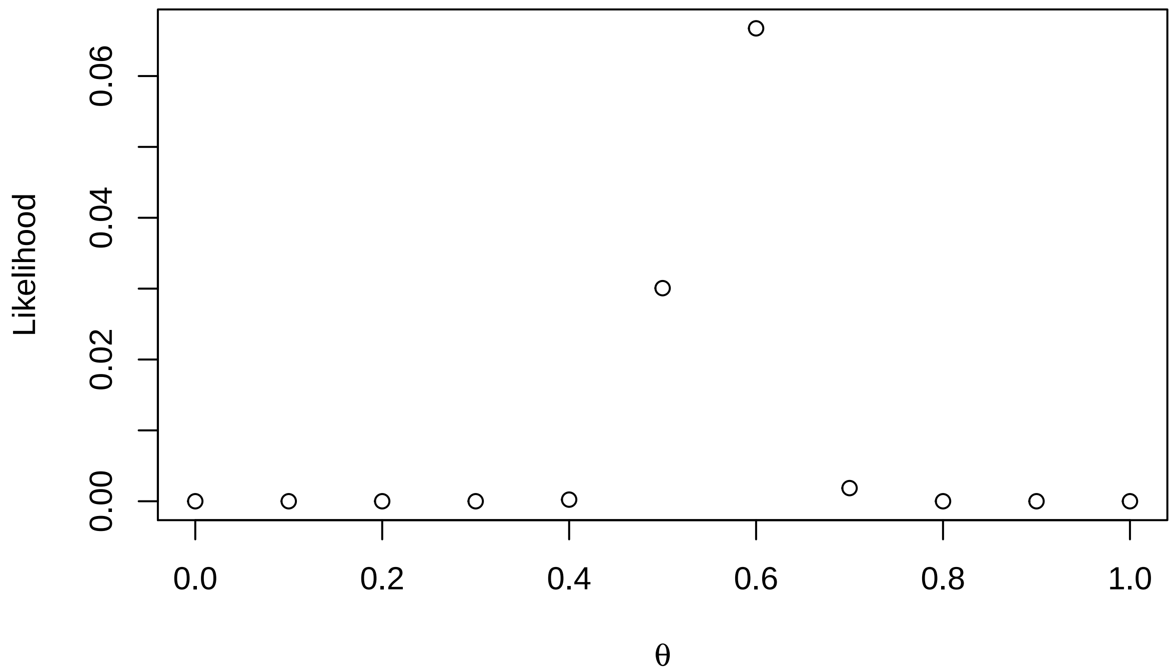
1_b) For the moment, suppose you believed that $\theta \in \{0.0, 0.1, ..., 0.9, 1.0\}$. Given that the results of the survey were $\sum_{i=1}^{n} Y_i = 57$, compute

$$Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$$

for each of these 11 values of θ and plot these probabilities as a function of θ.

-Answer: For each of 11 values of theta (that are a sequence starts from 0.0 to 1.0 with step 0.1) and knowing that $\sum_{i=1}^{n} Y_i = 57$ and $n = 100$, we plot these probabilities, $Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$, as a function of theta

```r
theta=seq(0.0,1,0.1)
likelihood = function(theta) return (dbinom(57,100,prob=theta))
plot(theta,likelihood(theta),xlab=expression(theta),ylab="Likelihood")
```



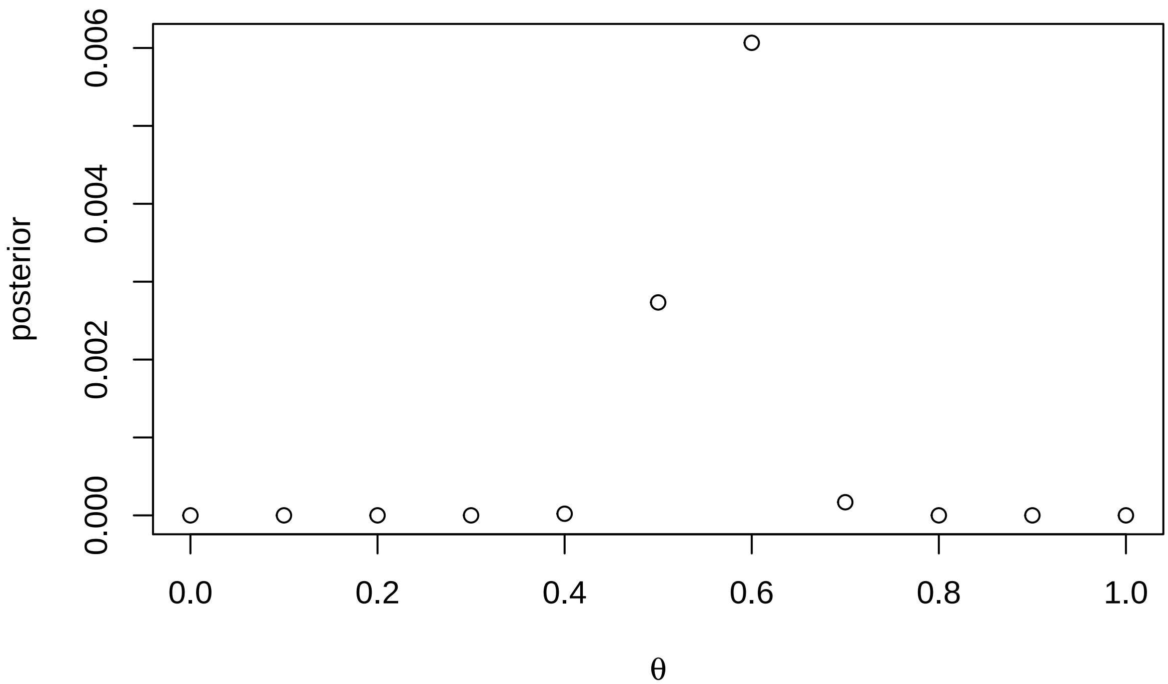

```r
cbind(theta=theta,likelihood=dbinom(57,100,prob=theta))
```

```
##       theta   likelihood
##  [1,]   0.0 0.000000e+00
##  [2,]   0.1 4.107157e-31
##  [3,]   0.2 3.738459e-16
##  [4,]   0.3 1.306895e-08
##  [5,]   0.4 2.285792e-04
##  [6,]   0.5 3.006864e-02
##  [7,]   0.6 6.672895e-02
##  [8,]   0.7 1.853172e-03
##  [9,]   0.8 1.003535e-07
## [10,]   0.9 9.395858e-18
## [11,]   1.0 0.000000e+00
```

1_c) Now suppose you originally had no prior information to believe one of these θ-values over another, and so $Pr(\theta = 0.0) = Pr(\theta = 0.1) = ... = Pr(\theta = 0.9) = Pr(\theta = 1.0)$. Use Bayes rule to compute $\pi(\theta|\sum_{i=1}^{n} Y_i = 57)$ for each θ-value. Make a plot of this posterior distribution as a function of θ.

-Answer: Since $Pr(\theta = 0.1) = Pr(\theta = 0.2) = ... = Pr(\theta = 1)$ we are talking about a discrete uniform. So knowing that the posterior distribution $\pi(\theta|x)$ is proportional to $\pi(\theta) \cdot L(\theta)$, applying these information in our case study, we have that $\pi(\theta|y) \propto \frac{1}{11} \cdot \theta^{57}(1-\theta)^{100-57}$, where $\frac{1}{11}$ is the probability mass function (pmf) of the discrete uniform of our case study.

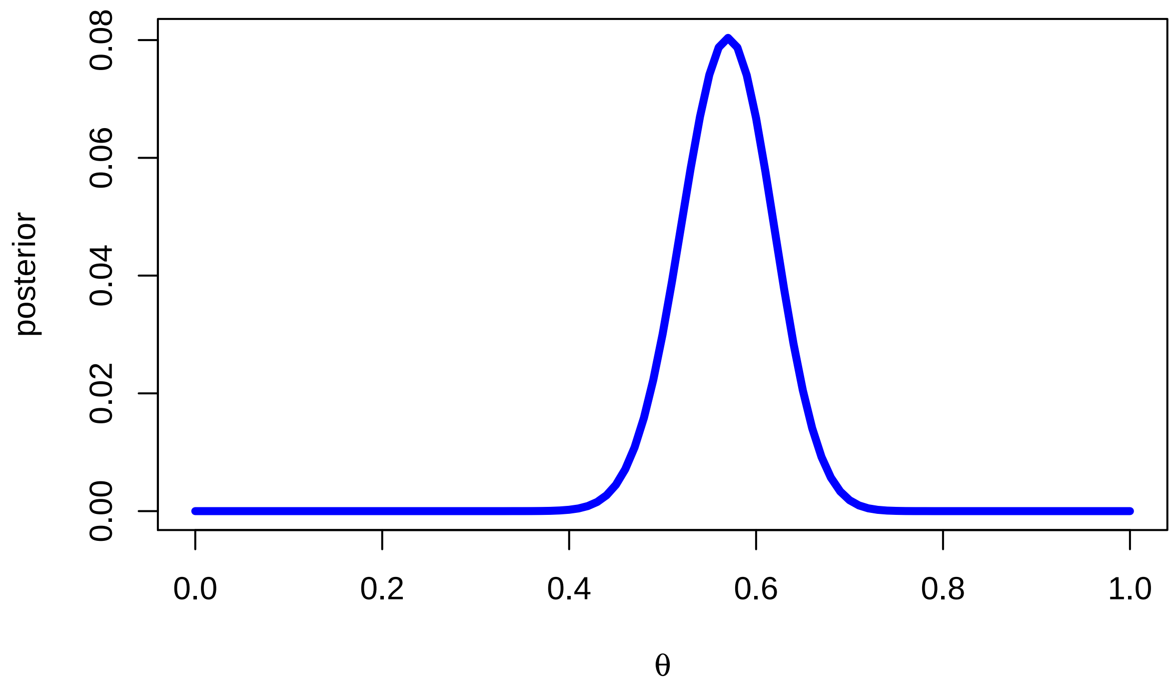
```r
likelihood = function(theta) return (dbinom(57,100,prob=theta))
prior = function(theta) return (1/11*(theta>0 & theta<1))
posterior = function(theta) return (prior(theta)*likelihood(theta))
plot(theta,posterior(theta),xlab=expression(theta), ylab='posterior')
```



1_d) Now suppose you allow θ to be any value in the interval $\Theta = [0, 1]$. Using the uniform prior density for $\theta \in [0, 1]$, so that $\pi(\theta) = I_{[0,1]}(\theta)$, plot $\pi(\theta) \times Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$ as a function of θ.

-Answer: Now θ assumes any value in the interval $\Theta = [0, 1]$, so in this case, we use the command "curve" to rappresent the posterior distribution $\pi(\theta|y)$. Let's note that the uniform prior distribution has $(\pi(\theta)) = 1$ for all $\theta \in [0, 1]$

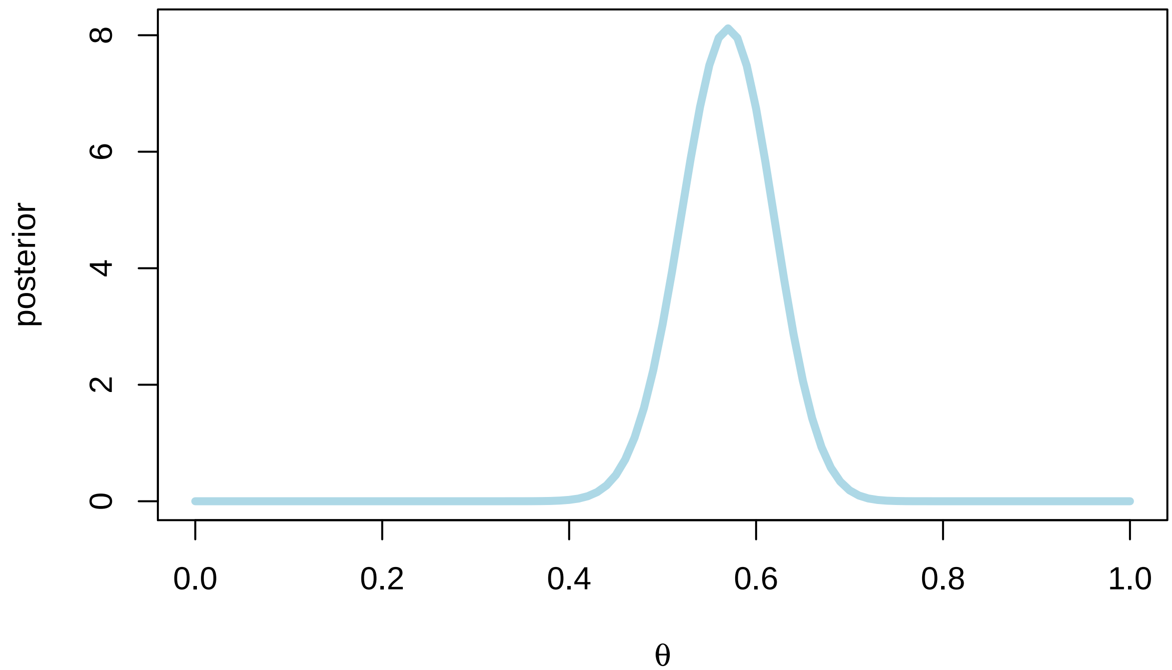
```
likelihood = function(theta) return (dbinom(57,100,prob=theta))
prior = function(theta)  return (dunif(theta)*(theta>0 & theta<1))
posterior = function(theta) return (prior(theta)*likelihood(theta))
curve(posterior(x), from = 0, to = 1, lwd = 4, col = "blue",
xlab = expression(theta), ylab = "posterior")
```



1_e) As discussed in this chapter, the posterior distribution of θ is $Beta(1+57, 1+100-57)$. Plot the posterior density as a function of θ. Discuss the relationships among all of the plots you have made for this exercise.

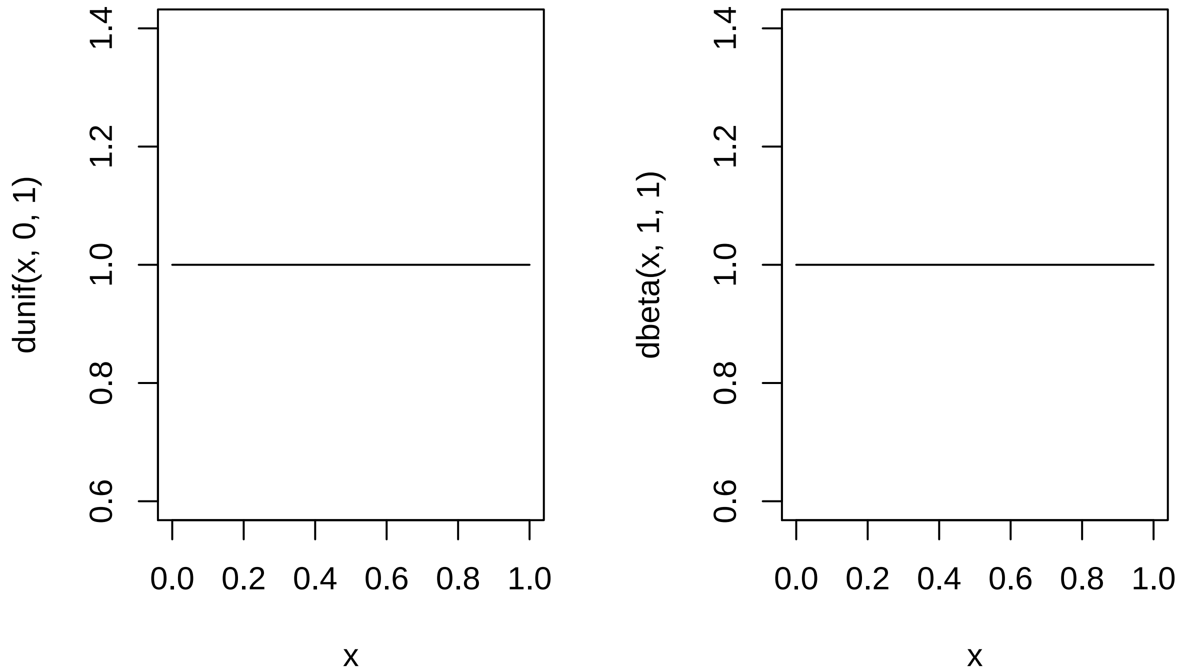
-Answer: As said in the previous point, the uniform prior distribution has $(\pi(\theta)) = 1$ for all $\theta \in [0, 1]$ This distribution can be thought of as a beta prior distribution with parameters a = 1, b = 1. In fact we can write: $\pi(\theta) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\theta^{1-1}(1-\theta)^{1-1} = 1$. So if $\theta \sim beta(1, 1)$ and $Y_1 + ... + Y_n \sim Bin(n, \theta)$ then $\theta|y \sim beta(1 + y, 1 + n - y)$

```
posterior = function(theta) return (dbeta(theta, 1+57, 1+100-57))
curve(posterior(x), from = 0, to = 1, lwd = 4, col = "lightblue",
xlab = expression(theta), ylab = "posterior")
```



We can say that if we use as prior distribution a $\beta(1,1)$ is the same thing of using as prior distribution a $unif(0,1)$

```
par(mfrow = c(1,2))
curve(dunif(x,0,1))
curve(dbeta(x,1,1))
```



2_a) derive the general formula of the prior predictive distribution

-Answer:

$$m(\cdot) = \int f(\cdot|\theta)\pi(\theta)d\theta = \int \frac{\sqrt{\lambda}}{\sqrt{2\pi}}e^{-\frac{\lambda(x-\theta)^2}{2}} \cdot \frac{\sqrt{\nu}}{\sqrt{2\pi}}e^{-\frac{\nu(\theta-\mu)^2}{2}}d\theta$$

$$\propto \int e^{-\frac{\lambda(x^2-2x\theta+\theta)^2}{2}} \cdot e^{-\frac{\nu(\theta^2-2\theta\mu+\mu^2)}{2}}d\theta$$

$$\propto \int e^{-\frac{\lambda x^2}{2}+x\theta\lambda-\frac{\theta^2\lambda}{2}-\frac{\nu\theta^2}{2}+\theta\mu\nu-\frac{\nu\mu^2}{2}}d\theta$$

$$\propto \int e^{-\frac{\lambda+\nu}{2}\theta^2+(x\lambda+\mu\nu)\theta} \cdot e^{-\frac{\lambda x^2}{2}-\frac{\nu\mu^2}{2}}d\theta$$

Knowing that if $\theta \sim N(\frac{b}{a}, \frac{1}{a}) \Rightarrow p(\theta) \propto e^{-\frac{a}{2}\theta^2+b\theta}$, and that $1 = \int \frac{\sqrt{a}}{\sqrt{2\pi}}e^{-\frac{a}{2}(\theta-\frac{b}{a}^2)}d\theta = \frac{\sqrt{a}}{\sqrt{2\pi}}\int e^{-\frac{a}{2}\theta^2+b\theta-\frac{b^2}{2a}}d\theta = \frac{\sqrt{a}}{\sqrt{2\pi}}e^{-\frac{b^2}{2a}}\int e^{-\frac{a}{2}\theta^2+b\theta}d\theta$, we have also that $\frac{\sqrt{2\pi}}{\sqrt{a}} = e^{\frac{b^2}{2a}}\int e^{-\frac{a}{2}\theta^2+b\theta}d\theta$. Applying these information on your case study, we have that:

$$m(\cdot) = \int f(\cdot|\theta)\pi(\theta)d\theta \propto e^{-\frac{\lambda x^2}{2}-\frac{\nu\mu^2}{2}} \cdot \int e^{-\frac{\lambda+\nu}{2}\theta^2+(x\lambda+\mu\nu)\theta}d\theta \propto e^{-\frac{\lambda x^2}{2}-\frac{\nu\mu^2}{2}} \cdot e^{\frac{(x\lambda+\mu\nu)^2}{2(\lambda+\nu)}}$$

$$\propto e^{\frac{x^2\lambda^2+2x\lambda\mu\nu}{2(\lambda+\nu)}} \cdot e^{-\frac{\lambda x^2}{2}} \propto e^{-\frac{\lambda\nu}{2(\lambda+\nu)}x^2+\frac{2\lambda\nu\mu}{2(\lambda+\nu)}x} \propto e^{-\frac{1}{2}\frac{\lambda\nu}{\lambda+\nu}x^2+\frac{\lambda\nu\mu}{\lambda+\nu}x}$$

If we consider $a = \frac{\lambda\nu}{\lambda+\nu}$ and $b = \frac{\lambda\nu\mu}{\lambda+\nu}$ the prior predictive distribution is a normal of parameters: $\frac{b}{a} = \mu$ and $\frac{1}{a} = \frac{\lambda+\nu}{\lambda\nu}$

2_b) derive the general formula of the posterior predictive distribution

-Answer: Knowing that

$$m(x_{new}|x) = \int f(x_{new}|\theta)\pi(\theta|x)d\theta$$

and that $\pi(\theta|x) = \pi(\theta|x_1,...,x_n) \sim N(\mu^* = w\mu + (1-w)\bar{x}_n, \nu^* = \nu + \lambda)$ where $w = \frac{\nu}{\nu+n\lambda}$, we have that the posterior predictive distribution is egual to:

$$m(x_{new}|x) = \int f(x_{new}|\theta)\pi(\theta|x)d\theta = \int \frac{\sqrt{\lambda}}{\sqrt{2\pi}}e^{-\frac{\lambda(x-\theta)^2}{2}} \cdot \frac{\sqrt{\nu^*}}{\sqrt{2\pi}}e^{-\frac{\nu^*(\theta-\mu^*)^2}{2}}d\theta \propto \int e^{-\frac{\lambda(x^2-2x\theta+\theta)^2}{2}} \cdot e^{-\frac{\nu^*(\theta^2-2\theta\mu^*+(\mu^*)^2)}{2}}d\theta$$

$$\propto \int e^{-\frac{\lambda x^2}{2}+x\theta\lambda-\frac{\theta^2\lambda}{2}-\frac{\nu^*\theta^2}{2}+\theta\mu^*\nu^*-\frac{\nu^*(\mu^*)^2}{2}}d\theta \propto \int e^{-\frac{\lambda+\nu^*}{2}\theta^2+(x\lambda+\mu^*\nu^*)\theta} \cdot e^{-\frac{\lambda x^2}{2}-\frac{\nu^*(\mu^*)^2}{2}}d\theta$$

$$\propto e^{-\frac{\lambda x^2}{2}-\frac{\nu^*(\mu^*)^2}{2}} \cdot \int e^{-\frac{\lambda+\nu^*}{2}\theta^2+(x\lambda+\mu^*\nu^*)\theta}d\theta \propto e^{-\frac{\lambda x^2}{2}-\frac{\nu^*(\mu^*)^2}{2}} \cdot e^{\frac{(x\lambda+\mu^*\nu^*)^2}{2(\lambda+\nu^*)}}$$

$$\propto e^{\frac{x^2\lambda^2+2x\lambda\mu^*\nu^*}{2(\lambda+\nu^*)}} \cdot e^{-\frac{\lambda x^2}{2}} \propto e^{-\frac{\lambda\nu^*}{2(\lambda+\nu^*)}x^2+\frac{2\lambda\nu^*\mu^*}{2(\lambda+\nu^*)}x} \propto e^{-\frac{1}{2}\frac{\lambda\nu^*}{\lambda+\nu^*}x^2+\frac{\lambda\nu^*\mu^*}{\lambda+\nu^*}x}$$

If we consider $a = \frac{\lambda\nu^*}{\lambda+\nu^*}$ and $b = \frac{\lambda\nu^*\mu^*}{\lambda+\nu^*}$ the posterior predictive distribution is a normal of parameters: $\frac{b}{a} = \mu^*$ and $\frac{1}{a} = \frac{\lambda+\nu^*}{\lambda\nu^*}$

2_c) assume that the known value of λ is 1/3 and suppose you have observed the following data

$$-1.25 \;\; 8.77 \;\; 1.18 \;\; 10.66 \;\; 11.81 \;\; -6.09 \;\; 3.56 \;\; 10.85 \;\; 4.03 \;\; 2.13$$

Elicit your prior distribution on the unknown θ in such a way that your prior mean is 0 and you believe that the unknown theta is in the interval $[-5, 5]$ with prior probability 0.96

-Answer: We have that $P(-5 \leq \theta \leq 5) = 0.96$. So applying some simply steps:

$$Pr(-5 \leq \theta \leq 5) = Pr(\frac{-5-\mu}{\sigma} \leq \frac{\theta-\mu}{\sigma} \leq \frac{5-\mu}{\sigma}) = Pr(-\frac{5}{\sigma} \leq \frac{\theta}{\sigma} \leq \frac{5}{\sigma}) = Pr(-\frac{5}{\sigma} \leq Z \leq \frac{5}{\sigma}) = \Phi(\frac{5}{\sigma}) - \Phi(-\frac{5}{\sigma})$$

Now, exploiting the properties of the normal distribution (particularly, the property of the symmetry), we have that $\Phi(-z) + \Phi(z) = 1$ and so that $\Phi(-z) = 1 - \Phi(z)$. Applying this property: $\Phi(-\frac{5}{\sigma}) = 1 - \Phi(\frac{5}{\sigma})$, and we can write:

$$\Phi(\frac{5}{\sigma}) - \Phi(-\frac{5}{\sigma}) = \Phi(\frac{5}{\sigma}) - (1 - \Phi(\frac{5}{\sigma})) = \Phi(\frac{5}{\sigma}) - 1 + \Phi(\frac{5}{\sigma}) = -1 + 2\Phi(\frac{5}{\sigma})$$

Putting $-1 + 2\Phi(\frac{5}{\sigma}) = 0.96$ and resolving the equation, we have that:

$$-1 + 2\Phi(\frac{5}{\sigma}) = 0.96 \Leftrightarrow 2\Phi(\frac{5}{\sigma}) = 1 + 0.96 \Leftrightarrow \frac{2}{2}\Phi(\frac{5}{\sigma}) = \frac{1.96}{2} \Leftrightarrow \Phi(\frac{5}{\sigma}) = 0.98$$

Applying the inverse function Φ^{-1} to $\Phi(\frac{5}{\sigma}) = 0.98$:

$$\Phi(\frac{5}{\sigma}) = 0.98 \Leftrightarrow \Phi^{-1}(\Phi(\frac{5}{\sigma})) = \Phi^{-1}(0.98) \Leftrightarrow \frac{5}{\sigma} = \Phi^{-1}(0.98)$$

We find the value of the quantile $\Phi^{-1}(0.98)$ with the following code:

```
qnorm(0.98)
```

```
## [1] 2.053749
```

And finally we can also find the value of sigma:

$$\frac{5}{\sigma} = \Phi^{-1}(0.98) \Leftrightarrow \frac{5}{\sigma} = 2.053749 \Leftrightarrow \sigma = \frac{5}{2.053749} \Leftrightarrow \sigma = 2.43457209$$

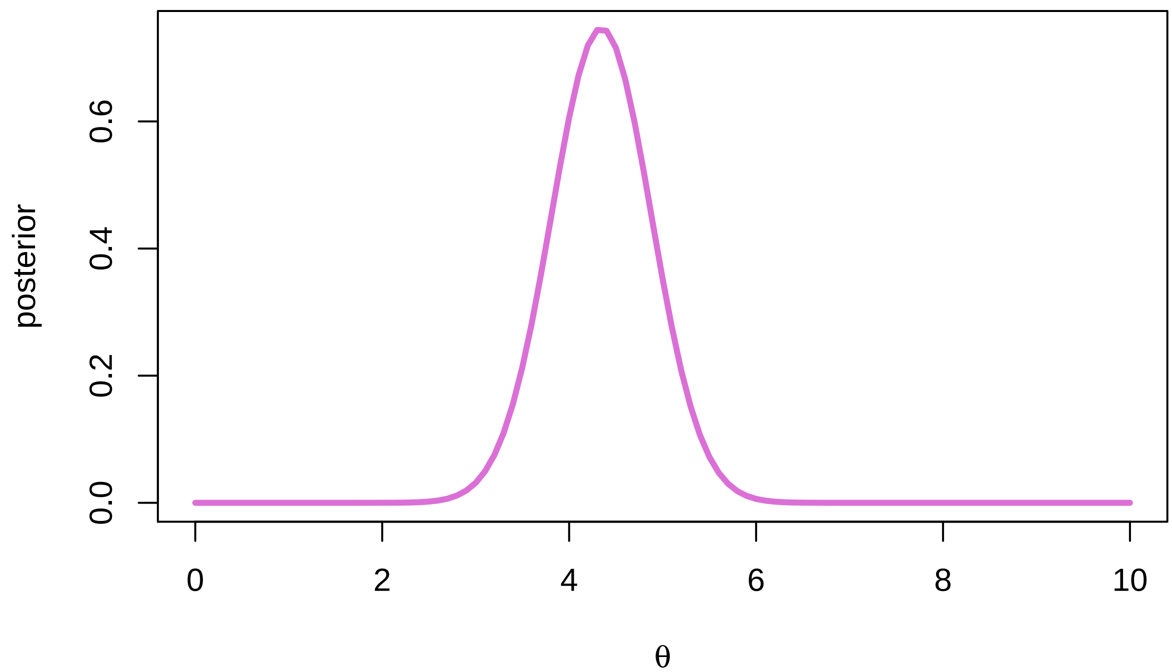
Hence the prior is a normal of parameters $\mu = 0$ and $\sigma^2 = 5.927141$

2_d) derive your posterior distribution and represent it graphically

-Answer: Knowing that $\pi(\theta|x) = \pi(\theta|x_1,...,x_n) \sim N(\mu^* = w\mu + (1-w)\bar{x}_n, \nu^* = \nu + \lambda)$

```r
# Data
x=c(-1.25, 8.77, 1.18, 10.66, 11.81, -6.09, 3.56, 10.85, 4.03, 2.13)
n=10
mean_x=sum(x)/n
mu=0
lambda=1/3
sigma=2.43457209
nu=1/5.927141
w=nu/(nu+n*lambda)
mu_star=w*mu+(1-w)*mean_x
nu_star=nu+n*lambda
sigma_star=1/sqrt(nu_star)

# Plot
posterior_d=function(theta) return (dnorm(theta,mu_star, sigma_star))
curve(posterior_d(x), from=0, to=10,lwd = 3, col = "orchid",
xlab = expression(theta), ylab = "posterior")
```

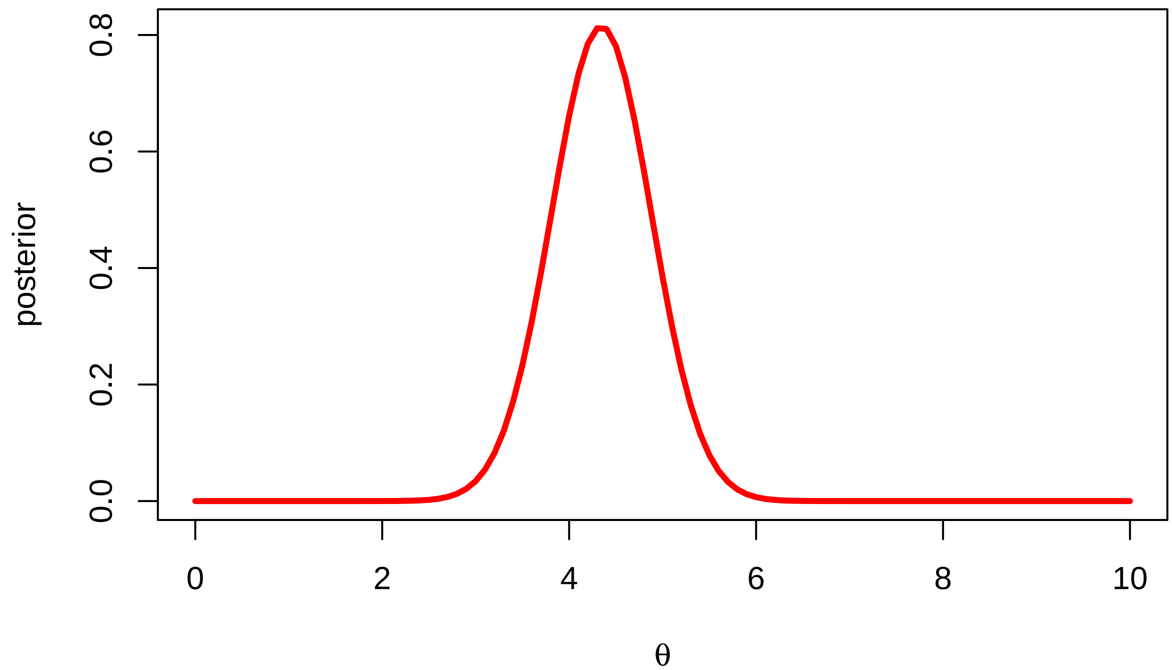


Another way to compute the posterior distribution is finding all the "ingredients": likelihood, prior distribution and marginal likelihood. We can find the posterior distribution as product between the likelihood and the prior, and divide for the marginal likelihood. In formulas: $\pi(\theta|x) = \frac{\pi(\theta)\cdot L(\theta)}{m(x)}$

```r
x=c(-1.25, 8.77, 1.18, 10.66, 11.81, -6.09, 3.56, 10.85, 4.03, 2.13)
mu=0
sigma=2.43457209
likelihood=function(theta){
  prod=1
```

```
  for(data in x)
    prod = prod*dnorm(data,theta,sqrt(3))
  return(prod)
}
prior=function(theta) return (dnorm(theta, mu, sigma))

numerator=function(theta) return (likelihood(theta)*prior(theta))
denominator=integrate(numerator, -Inf, Inf)
posterior_s=function(theta) return (numerator(theta)/denominator$value)
curve(posterior_s(x), from=0, to=10,lwd = 3, col = "red",
xlab = expression(theta), ylab = "posterior")
```



2_e) derive your favorite point estimate and interval estimate and motivate your choices

-Answer: For a normal statistical model a point estimate is given by:

$$\hat{\theta}_B = E[\theta|x_1, ..., x_n] = \int_{\Theta} \theta \cdot \pi(\theta|x_1, ..., x_n) d\theta = \mu^* = w\mu + (1-w) \cdot \bar{X}_n$$

Computing the mean of our posterior distribution: $\mu^* = w\mu + (1-w)\bar{x}_n$ where $w = \frac{\nu}{\nu+n\lambda}$ and substituting the corresponding values we have:

$$w = \frac{\nu}{\nu + n\lambda} = \frac{0.1687154}{0.1687154 + 10\frac{1}{3}} = 0.0481762$$

and

$$\mu^* = 0.0481762 \cdot 0 + (1 - 0.0481762) \cdot 4.565 = 4.345076$$

Before of writing our confidence interval, we need to compute the variance that is egual to:

$$\nu^* = \nu + n\lambda = 0.1687154 + 10 \cdot \frac{1}{3} = 3.502049 \Longrightarrow \sigma^* = \sqrt{\frac{1}{\nu^*}} = 0.5343661$$

Now we have all the "ingredients" for writing our confidence interval:

$$1 - \alpha = Pr(\mu^* - z_{1-\frac{\alpha}{2}} \cdot \sigma^* < \theta | x < \mu^* + z_{1-\frac{\alpha}{2}} \cdot \sigma^*)$$

```
x=c(-1.25, 8.77, 1.18, 10.66, 11.81, -6.09, 3.56, 10.85, 4.03, 2.13)
n=10
mean_x=sum(x)/n
mu=0
lambda=1/3
sigma=2.43457209
nu=1/5.927141
w=nu/(nu+n*lambda)
mu_star=w*mu+(1-w)*mean_x
nu_star=nu+n*lambda
sigma_star=1/sqrt(nu_star)
c(L=mu_star-qnorm(0.98)*sigma_star, U=mu_star+qnorm(0.98)*sigma_star)
```

```
##        L        U
## 3.247622 5.442529
```

3_a) Provide a fully Bayesian analysis for these data explaining all the basic ingredients and steps for carrying it out. In particular, compare your final inference on the uknown $\theta = E[X|\theta]$ with the one you have derived in the previous point 2)

-Answer:

$$f(x) = \begin{cases} \frac{1}{20} & if \quad \theta - 10 \leq x \leq \theta + 10 \\ 0 & if \quad otherwise \end{cases} \qquad = \frac{1}{20} I_{[\theta-10,\theta+10](x)}$$

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{1}{20} I_{[\theta-10,\theta+10]}(x_i) = \frac{1}{20^n} \prod_{i=1}^{n} I_{[\theta-10,\theta+10]}(x_i)$$

Since the likelihood is a function of θ (not of x_i), to further simplify this formula we need to express the indicators as functions of θ:

$$x_i \in [\theta - 10, \theta + 10] \Leftrightarrow \theta \in [x - 10, x + 10]$$

hence, for each $i \in 1, ..., n$

$$I_{[\theta-10,\theta+10]}(x_i) = 1 \Leftrightarrow I_{[x_i-10,x_i+10]}(\theta) = 1$$

making the product, we get:

$$\prod_{i=1}^{n} I_{[x_i-10,x_i+10]}(\theta) = 1 \Leftrightarrow \theta \in \cap_{i=1}^{n} [x_i - 10, x_i + 10] \Leftrightarrow I_{\cap_{i=1}^{n}[x_i-10,x_i+10]}(\theta) = 1$$

$$I_{\cap_{i=1}^{n}[x_i-10,x_i+10]}(\theta) = I_{[x_{max}-10,x_{min}+10]}(\theta)$$

So finally we have:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{1}{20} I_{[\theta-10,\theta+10]}(x_i) = \frac{1}{20^n} \prod_{i=1}^{n} I_{[\theta-10,\theta+10]}(x_i) = \frac{1}{20^n} I_{[x_{max}-10,x_{min}+10]}(\theta)$$

Let's compute the likelihood:

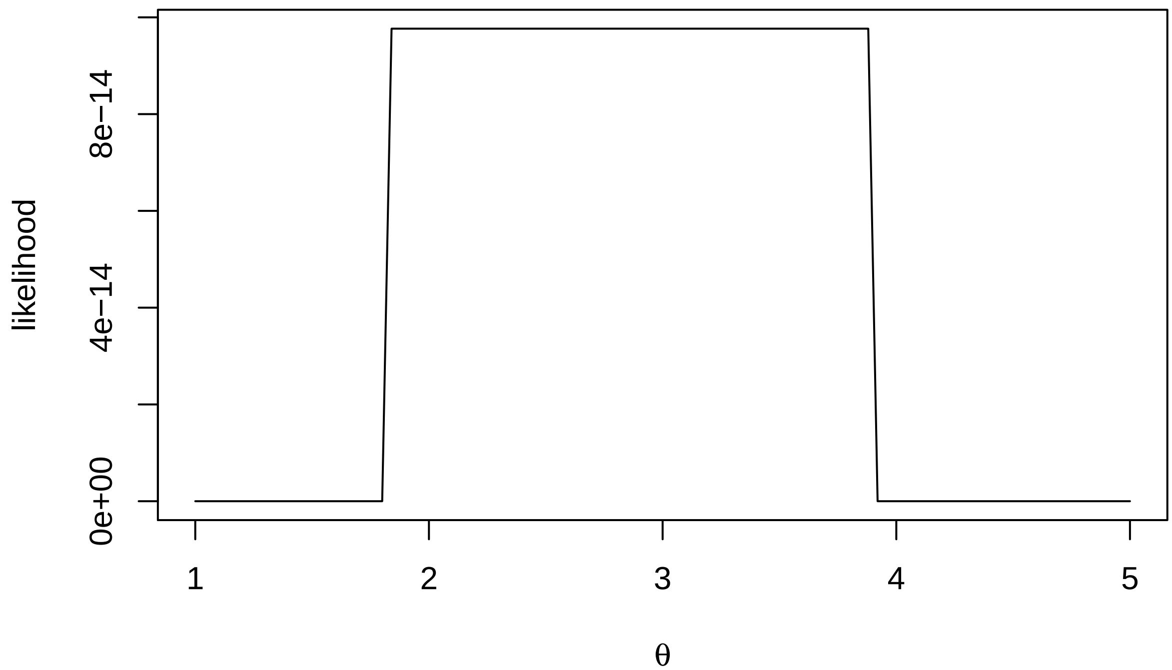
```r
x=c(-1.25, 8.77, 1.18, 10.66, 11.81, -6.09, 3.56, 10.85, 4.03, 2.13)
a = max(x)-10
a
```

```
## [1] 1.81
```

```r
b = min(x)+10
b
```

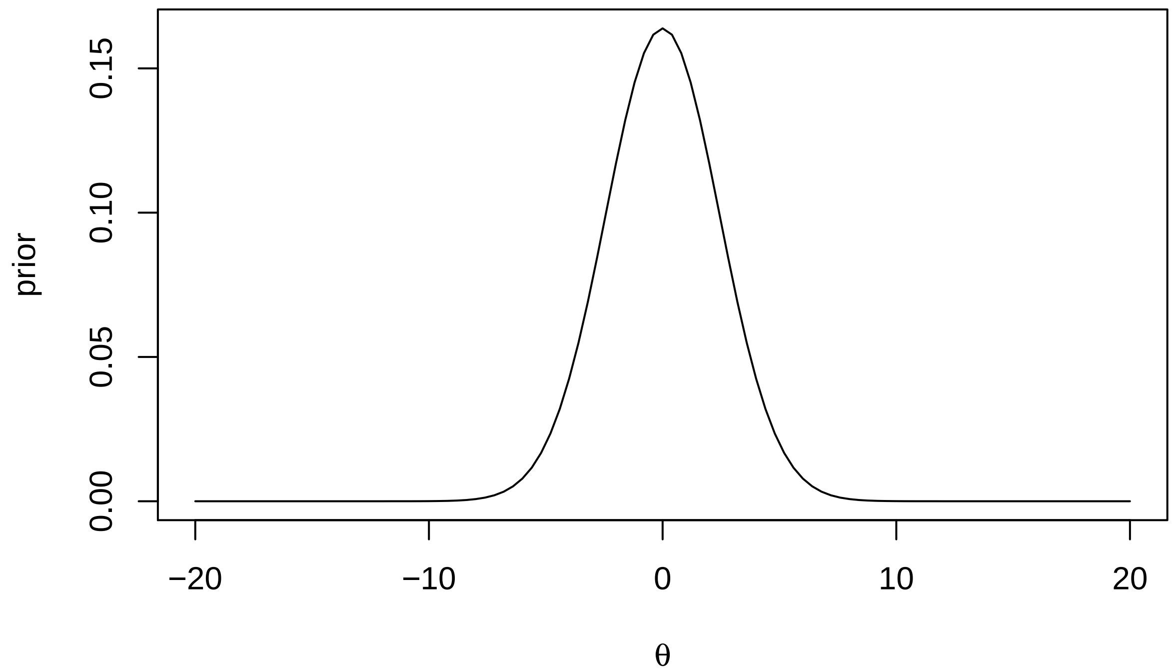
```
## [1] 3.91
```

```r
L=function(theta) return (1/20^10*(theta>=1.81 & theta<=3.91))
curve(L(x),from=1,to=5,xlab=expression(theta),ylab="likelihood")
```



Let's compute the prior distribution: $\pi(\theta) = \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(\theta-\mu)^2}{2}}$

```r
prior = function(theta) return(dnorm(theta,mu,sigma))
curve(prior(x),from=-20, to=20,xlab=expression(theta),ylab="prior")
```



Finally we can compute the posterior distribution:

$$\pi(\theta|x) = \frac{L(\theta)\cdot\pi(\theta)}{m(x)} = \frac{L(\theta)\cdot\pi(\theta)}{\int_\Theta L(t)\cdot\pi(t)dt} = \frac{\frac{1}{20^n} I_{[x_{max}-10, x_{min}+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(\theta-\mu)^2}{2}}}{\int_{-\infty}^{+\infty} \frac{1}{20^n} I_{[x_{max}-10, x_{min}+10]}(t)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(t-\mu)^2}{2}} dt}$$

$$= \frac{\frac{1}{20^n} I_{[x_{max}-10, x_{min}+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(\theta-\mu)^2}{2}}}{\int_{-\infty}^{+\infty} \frac{1}{20^n} I_{[x_{max}-10, x_{min}+10]}(t)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(t-\mu)^2}{2}} dt}$$

$$= \frac{\frac{1}{20^n} I_{[x_{max}-10, x_{min}+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(\theta-\mu)^2}{2}}}{\frac{1}{20^n}\int_{max(x)-10}^{min(x)+10} \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(t-\mu)^2}{2}} dt}$$

$$= \frac{I_{[x_{max}-10, x_{min}+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(\theta-\mu)^2}{2}}}{\int_{max(x)-10}^{min(x)+10} \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(t-\mu)^2}{2}} dt}$$

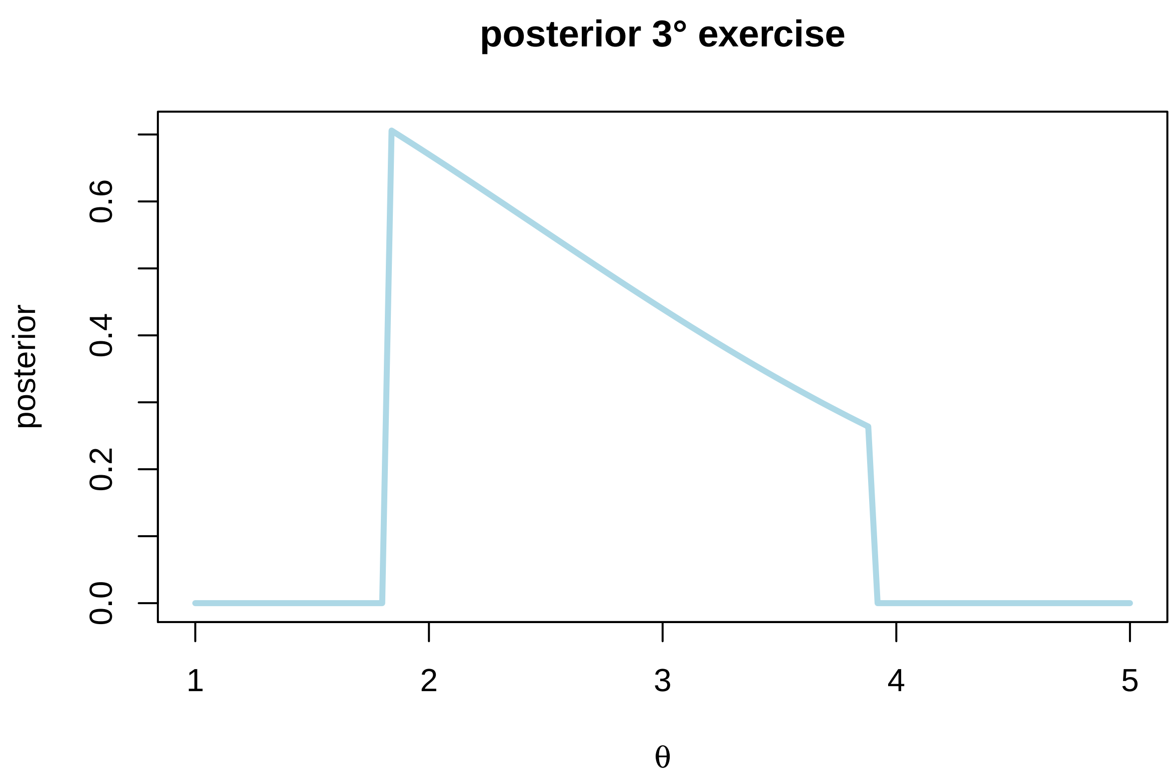
$$= \frac{I_{[x_{max}-10, x_{min}+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(\theta-\mu)^2}{2}}}{\int_{1.81}^{3.91} \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu(t-\mu)^2}{2}} dt}$$

```
m = pnorm(b,mu,sigma)-pnorm(a,mu,sigma)
m
```

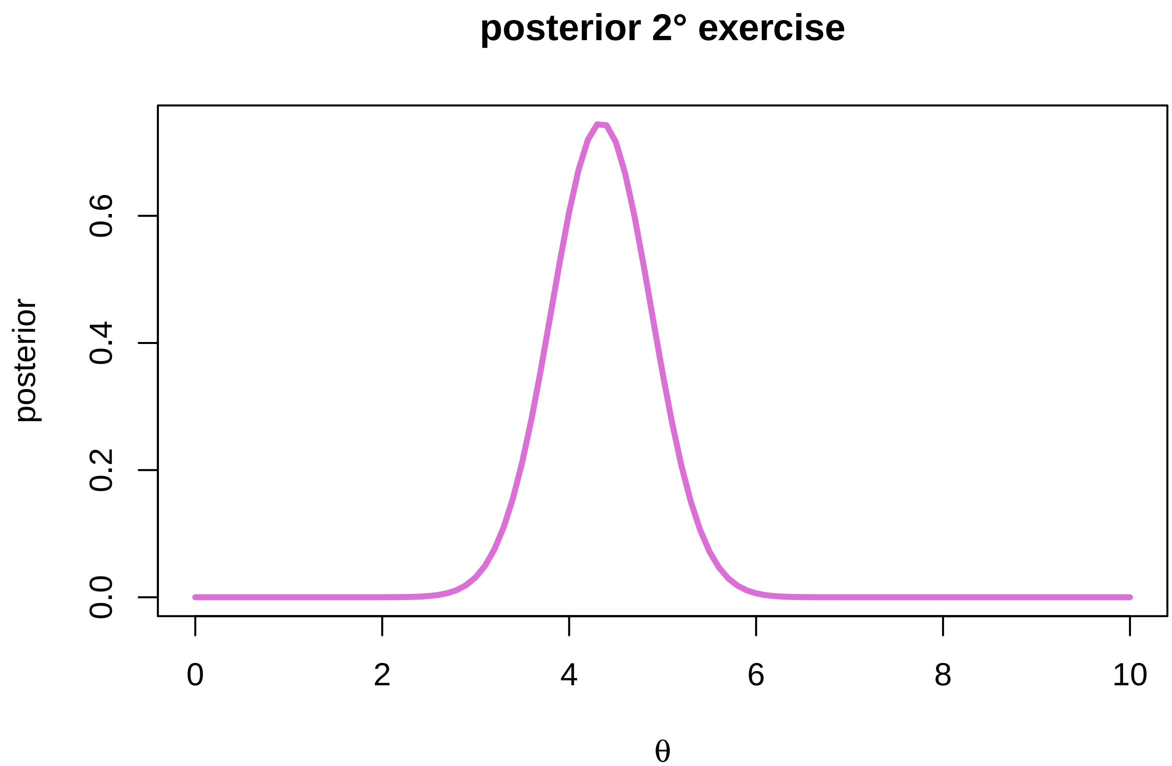
```
## [1] 0.174469
```

```r
I = function(theta)  return (1*(theta>=1.81 & theta<=3.91))
prior=function(theta) return (dnorm(theta, mu, sigma))
post=function(theta) return (I(theta)*prior(theta)/m)

# posterior 3° exercise
curve(post(x), from=1, to=5, xlab=expression(theta),
      ylab="posterior", col='lightblue', lwd=3, main = 'posterior 3° exercise')
```




```r
#posterior 2° exercise
posterior_d=function(x) return (dnorm(x,mu_star, sigma_star))
curve(posterior_d(x), from=0, to=10,lwd = 3, col = "orchid",
xlab = expression(theta), ylab = "posterior",main = 'posterior 2° exercise')
```



Now we can compare our final inference on the uknown $\theta = E[X|\theta]$ with the one of the previous exercise

```
theta_second_exercise=mu_star
theta_third_exercise=integrate(function(x) post(x)*x,-Inf,Inf)
cbind(theta_sec=theta_second_exercise,theta_th=theta_third_exercise$value)
```

```
##      theta_sec theta_th
## [1,]  4.345076 2.689806
```

3_b) Write the formula of the prior predictive distribution of a single observation and explain how you can simulate i.i.d random drws from it. Use the simulated values to represent approximately the predictive density in a plot and compare it with the prior predictive density of a single observation of the previous model

-Answer:

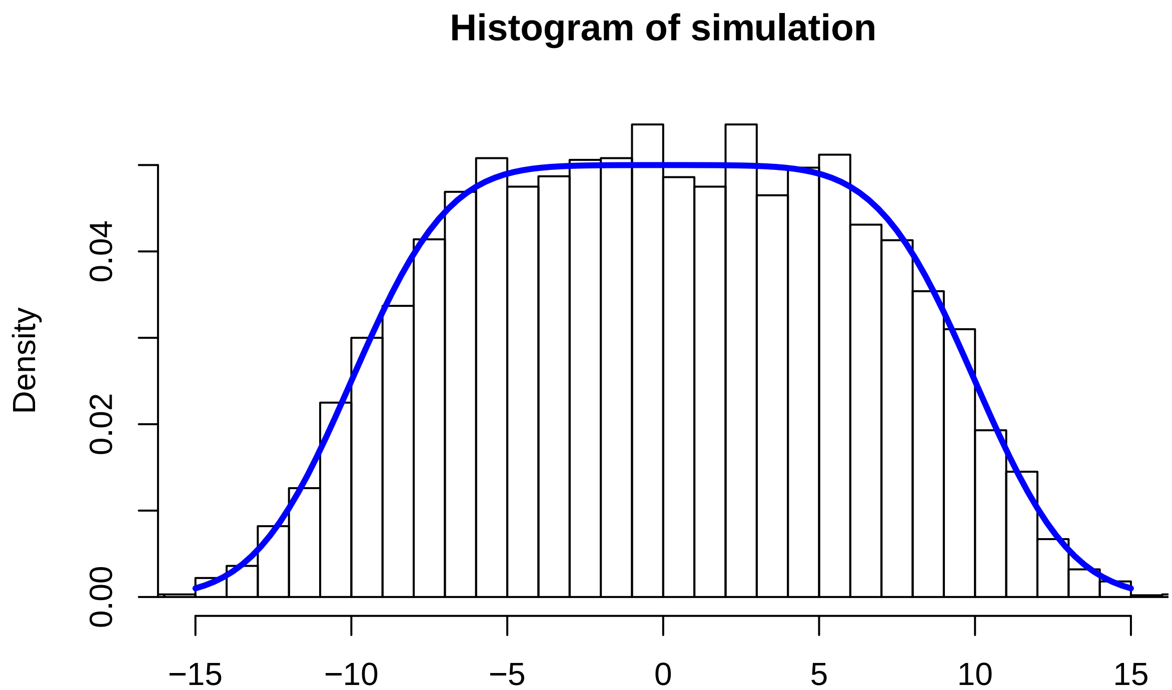
$$m(\cdot) = \int f(\cdot|\theta)\pi(\theta)d\theta = \int \frac{1}{20} I_{[x-10,x+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}}e^{-\frac{\nu(\theta-\mu)^2}{2}}d\theta = \frac{1}{20}\int I_{[x-10,x+10]}(\theta)\frac{\sqrt{\nu}}{\sqrt{2\pi}}e^{-\frac{\nu(\theta-\mu)^2}{2}}d\theta$$
$$= \frac{1}{20}\int_{x-10}^{x+10}\frac{\sqrt{\nu}}{\sqrt{2\pi}}e^{-\frac{\nu(\theta-\mu)^2}{2}}d\theta = \frac{1}{20}\cdot(\Phi(\sqrt{\nu}\cdot(x+10-\mu)) - \Phi(\sqrt{\nu}\cdot(x-10-\mu)))$$
$$= \frac{1}{20}\cdot(\Phi(\frac{x+10-\mu}{\sigma}) - \Phi(\frac{x-10-\mu}{\sigma})) = \frac{1}{20}\cdot(\Phi(\frac{x+10}{\sigma}) - \Phi(\frac{x-10}{\sigma}))$$

After writing the prior predictive distribution of a single observation, we can simulate i.i.d random drws from it with the following command:

```
set.seed(123)
m = function(x,mu,sigma) return(1/20*(pnorm((x+10-mu)/sigma)-pnorm((x-10-mu)/sigma)))
n=10000
simulation=rep(NA,n)
```

```r
for (i in 1:n)
{
  th.hat=rnorm(1,mu,sigma)
  simulation[i]=runif(1,th.hat-10,th.hat+10)
}

hist(simulation,prob=T,breaks=50,xlim=c(-15,15),xlab="")
curve(m(x,mu,sigma),lwd = 3,add=T,col="blue")
```



3_c) Consider the same discrete (finite) grid of values as parameter space Θ for the conditional mean θ in both models. Use this simplified parametric setting to decide whether one should use the Normal model rather than the Uniform model in light of the observed data.

-Answer: About the bayesian inference in the presence of alternative models we have that:

$$J(m, \theta, data) = pri(m)\pi(\theta|m)f(data|\theta, m)$$

the joint distribution of data and model m is equal to:

$$J(data|m) = \int_{\Theta_m} pri(m)f(data|\theta, m)\pi(\theta|m)d\theta = pri(m)\int_{\Theta_m} f(data|\theta, m)\pi(\theta|m)d\theta = pri(m)J(data|m) = pri(m)b(m|data)$$

the posterior model probability for model m given the data is egual to:

$$post(m|data) = \frac{J(data, m)}{\sum_{m'} J(data, m')} = \frac{pri(m)b(m|data)}{\sum_{m'} pri(m')b(data|m')}$$

and finally, the posterior odds between two alternative models (m_i and m_j) is equal to:

$$\frac{post(m_i|data)}{post(m_j|data)} = \frac{\frac{q(data|m_i)}{\sum_{m'} q(data|m')}}{\frac{q(data|m_j)}{\sum_{m'} q(data|m')}} = \frac{q(data|m_i)}{q(data|m_j)} = \frac{pri(m_i)}{pri(m_j)} \cdot \frac{b(m_i|data)}{b(m_j|data)}$$

where the ratio $BF_{ij} = \frac{b(m_i|data)}{b(m_j|data)}$ is called Bayes Factor and it is that we want to find to decide whether one should use the model m_i rather than the model m_j in light of the observed data If we suppose that $\pi(\theta|m_i)$ and $\pi(\theta|m_j)$ are the same, we have that:

$$BF_{ij} = \frac{b(m_i|data)}{b(m_j|data)} = \frac{\int_{\Theta_{m_i}} f(data|\theta, m_i)\pi(\theta|m_i)d\theta}{\int_{\Theta_{m_j}} f(data|\theta, m_j)\pi(\theta|m_j)d\theta} = \frac{\int_{\Theta_{m_i}} f(data|\theta, m_i)d\theta}{\int_{\Theta_{m_j}} f(data|\theta, m_j)d\theta}$$

in our case study, since we consider the same discrete (finite) grid of values as parameter space Θ for the conditional mean θ in both models, we have that:

$$BF_{ij} = \frac{b(m_i|data)}{b(m_j|data)} = \frac{\sum_{\Theta_{m_i}} f(data|\theta, m_i)}{\sum_{\Theta_{m_j}} f(data|\theta, m_j)}$$

```
theta=seq(-100,100,0.1)

model_1=rep(NA,length(theta))
model_2=rep(NA,length(theta))

for(i in 1:length(theta)){
  model_1[i]=prod(dnorm(x,theta[i],sqrt(3)))
  model_2[i]=prod(dunif(x,theta[i]-10,theta[i]+10))
}

BF=sum(model_1)/sum(model_2)
BF
```

```
## [1] 6.374436e-17
```

Since the ratio $\frac{\sum_{\Theta_{m_i}} f(data|\theta,m_i)}{\sum_{\Theta_{m_j}} f(data|\theta,m_j)}$ is smaller than 1, we can conclude that we chose the second model(i.e. the Uniform model).