# Homework #1

*Stat4DS2+DS*
*https://elearning2.uniroma1.it/course/view.php?id=4951*

### deadline 23/03/2017 (23:55)

## Contents

## My Last+First Name *Umberto_Jr_Mele* My Matricola 1388371

# Question 1

1) 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.

## Part 1.a

a) Assume $Y_1, ..., Y_{100}$ are, conditional on $\theta$, i.i.d. binary random variables with expectation $\theta$. Write down the joint distribution of $Pr(Y_1 = y_1, ..., Y_{100} = y_{100}|\theta)$ in a compact form. Also write down the form of $Pr(\sum_{i=1}^{n} Y_i = y|\theta)$.

**-Answer:**

$Y_i|\theta$ is a random variable with distribution *Bernoulli* and *i.i.d.* so we can write down the joint distribution, like this:

$$Pr(Y_1 = y_1, ......, Y_{100} = y_{100}|\theta) = P(Y_1 = y_1|\theta)P(Y_2 = y_2|\theta)...P(Y_{100} = y_100|\theta) = \prod_{i=1}^{100} \theta^{y_i}(1-\theta)^{1-y_i}$$

And we can say that this distribution is the Likelihood function on the parameter $\theta$.

While, the probability distribution of $Pr(\sum_{i=1}^{100} Y_i = y|\theta)$ is defined by the random variable $Z = Y_1 + .... + Y_{100}$, that is distribuited like a Binomial distribution, so:

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

Note that:

$$Pr(\sum_{i=1}^{100} Y_i = y|\theta) \propto Pr(Y_1 = y_1, ......, Y_{100} = y_{100}|\theta) = L(\theta)$$

## Part 1.b

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

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

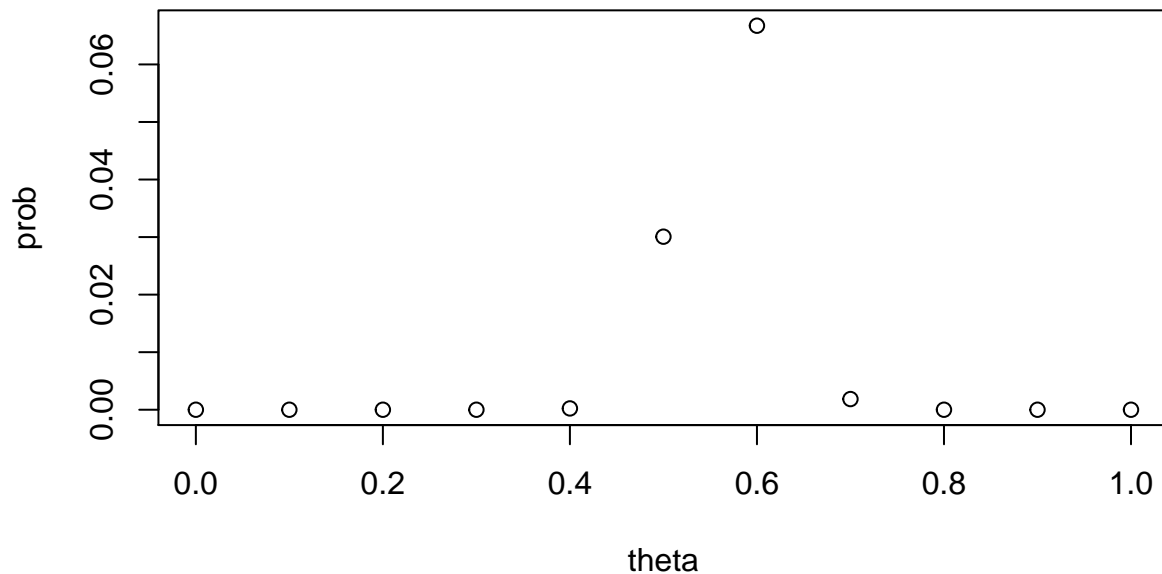for each of these 11 values of $\theta$ and plot these probabilities as a function of $\theta$.

**-Answer:**

The given probability is:

$$Pr(\sum_{i=1}^{100} Y_i = 57|\theta) = Pr(Z = 57) = \binom{100}{57}\theta^{57}(1-\theta)^{100-57}$$

```
theta <- seq(0,1,0.1)
prob_Z.equal.57 <- dbinom(57 ,100, theta)
plot(theta, prob_Z.equal.57, xlab = 'theta' , ylab = 'prob')
```



3

## Part 1.c

c) Now suppose you originally had no prior information to believe one of these $\theta$-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 $\theta$-value. Make a plot of this posterior distribution as a function of $\theta$.
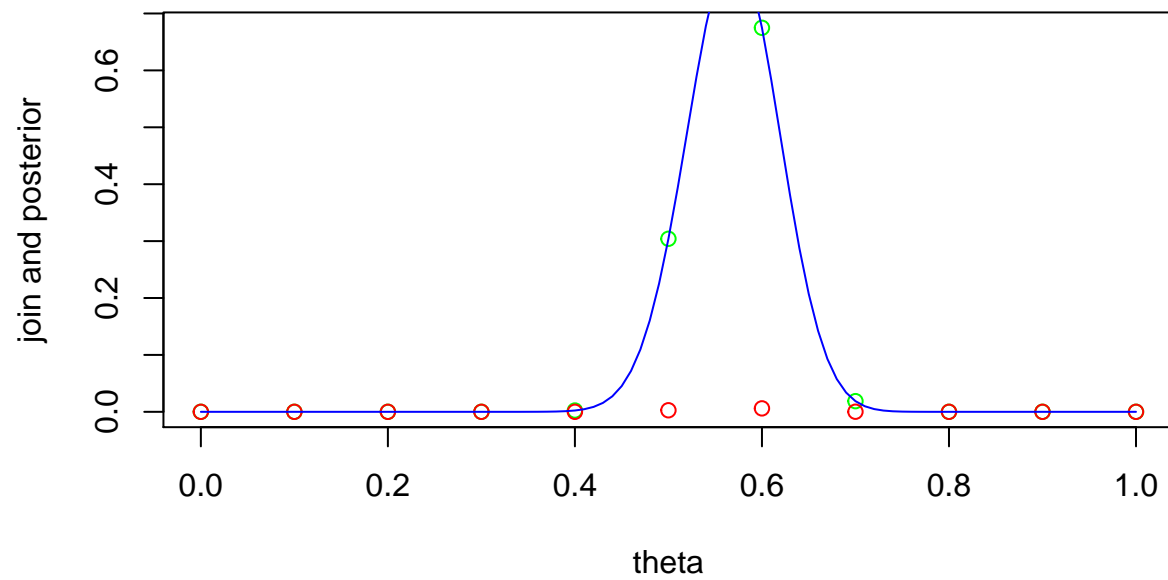
**-Answer:**

Remembering that the posterior distribution is:

$$\pi(\theta|x) = \frac{L(\theta) \cdot \pi(\theta)}{m(x)}$$

we can compute in a similar way:

$$\pi(\theta| \sum_{i=1}^{n} Y_i = 57) = \frac{f(\sum Y_i = y|\theta) \cdot \pi(\theta)}{\sum_\theta f(\sum Y_i = y|\theta) \cdot \pi(\theta)}$$

```
prior <- function(t)  1/11 *(t < 1 & t >0)
L <-  function(t) dbinom(57,100,t)
m <- function(all.theta){
  su <- 0
  for(t in all.theta){
    su <- su + prior(t)*L(t)
  }
  return(su)
}
jn <- function(t) prior(t)*L(t)
posterior <- function(t) prior(t)*L(t)/m(all.theta = theta)
plot(theta ,posterior(theta),type = 'p', col ='green', ylab='join and posterior')
lines(theta ,jn(theta),type = 'p', col ='red')
curve(posterior(x), add = T, col='blue')
```
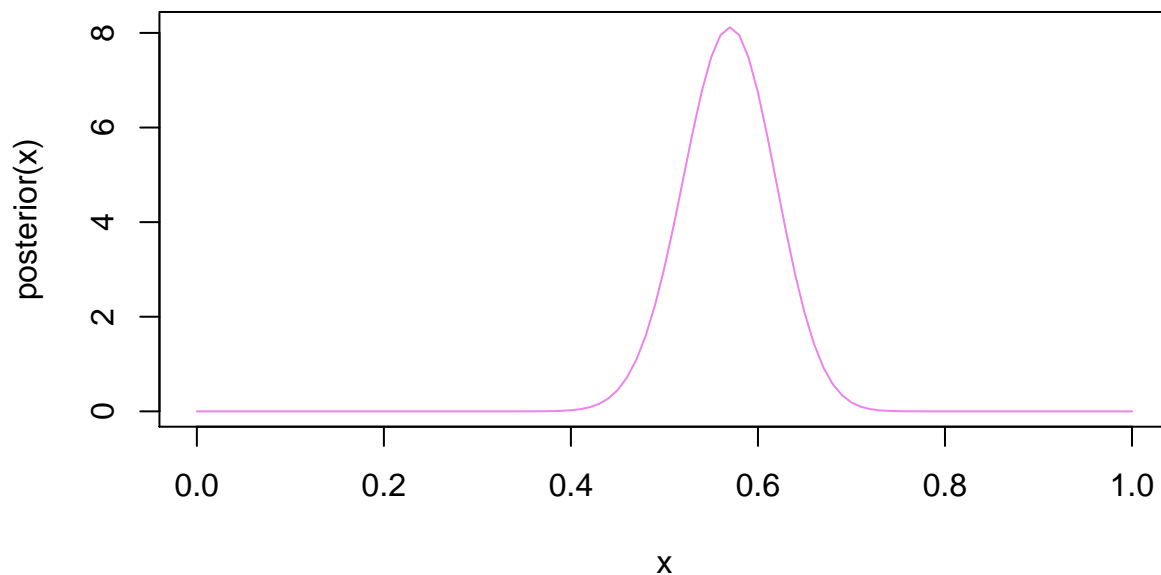
4

## Part 1.d

d) Now suppose you allow $\theta$ 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 $\theta$.

**-Answer:**

```
prior <- function(t) dunif(t)*(t>0 & t<1 )
L <-   function(t) dbinom(57,100,t)
joint <- function(t) L(t)*prior(t)
m <- integrate(joint,0,1)

posterior <- function(t) prior(t)*L(t)/m$value
curve(posterior, col = 'violet')
```



6

## Part 1.e

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

**-Answer:**

We know that the posterior distribution is a $Beta(\alpha_{post} = 1 + 57, \beta_{post} = 1 + 43)$ and if we use the Conjugate Analysis for this problem, we have a $Beta$ distro as Prior:
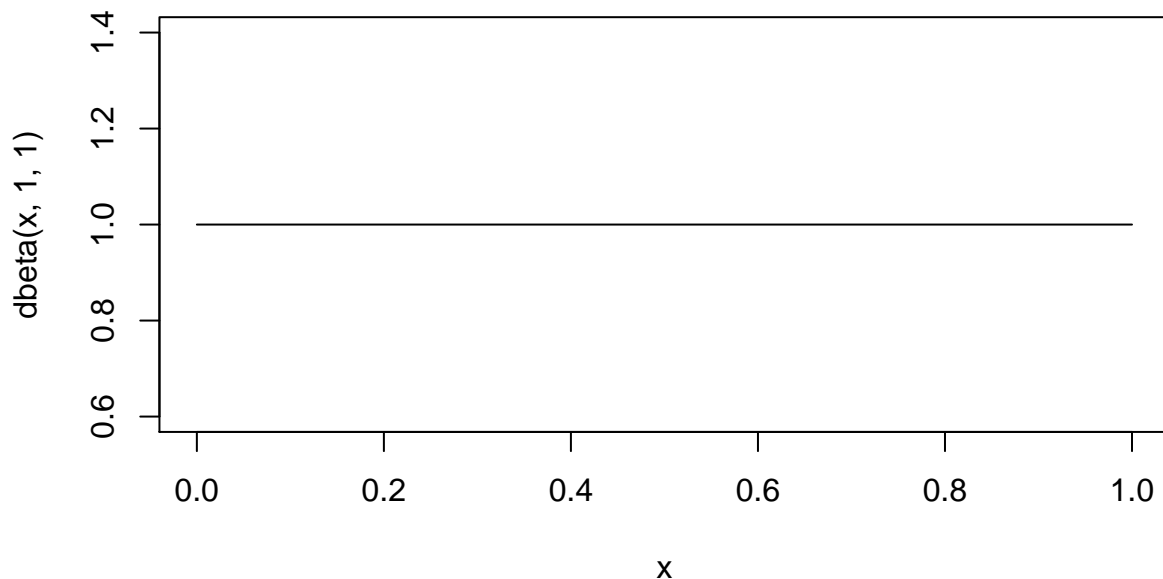
$$\pi(\theta|\bar{Z}) \propto L(\theta) \cdot \pi(\theta) = \binom{100}{57}\theta^{57}(1-\theta)^{100-57} \cdot \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \propto \theta^{57+\alpha-1}(1-\theta)^{43+\beta-1}$$

so the posterior distribution is a $Beta$ distribution with parameters $\dot{\alpha} = 57 + \alpha$ and $\dot{\beta} = 43 + \beta$, so the prior distribution is a $Beta(1,1)$ that is the same distribution of a $Unif(0,1)$.
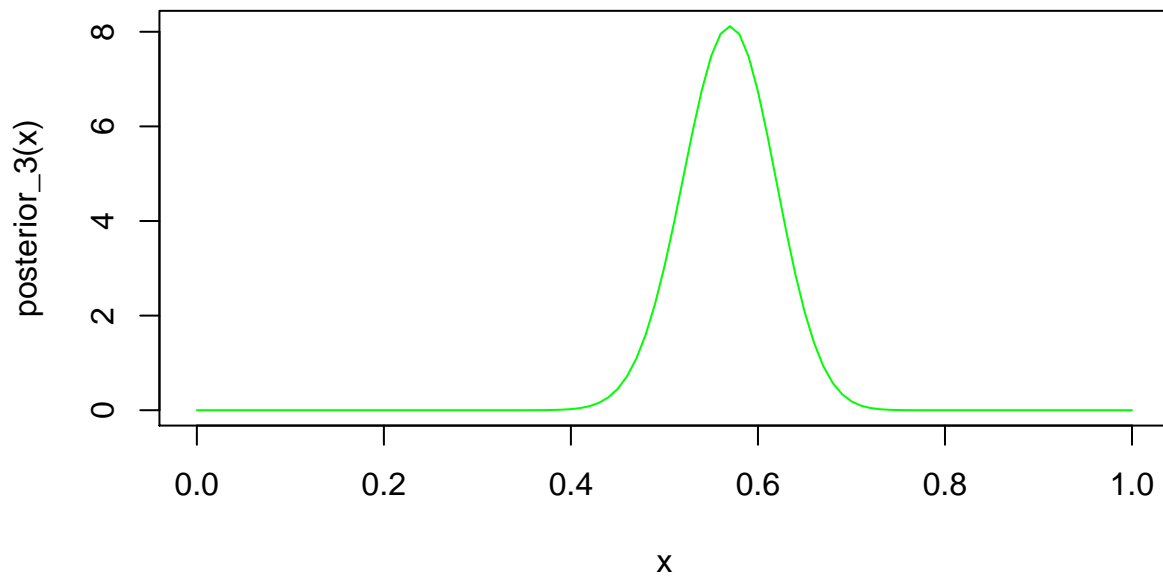
Because:

$$Beta(\theta|\alpha = 1, \beta = 1) \sim \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} = \frac{\theta^{1-1}(1-\theta)^{1-1}}{B(\alpha,\beta)} = constant \quad \forall \theta \in [0,1]$$

```
curve(dbeta(x,1,1))
```



```
prior <- function(t) dbeta(t,1,1)
L <-  function(t) dbinom(57,100,t)
joint <- function(t) L(t)*prior(t)
m <- integrate(joint,0,1)
```

```r
posterior_3 <- function(t) prior(t)*L(t)/m$value
curve(posterior_3, col = 'green')
```



And we can concluding saying that we pose the same supposition for all priors we made (same propbability for all thetas we have), and the posterior distribution is the same of the Likelhood beacause our prior is not changing the distribution of the posterior.

# Question 2

2) Consider a normal statistical model with $X_i \sim N(\theta, \lambda = 1/\sigma^2)$ where the precision parameter is known. Use as a prior distribution on the (conditional) mean $\theta$ a Normal with prior mean $\mu$ and prior precison $\nu$.

## Part 2.a

a) derive the general formula of the prior predictive distribution

**-Answer**

The prior predictive distribution is defined by:

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

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

Now remembering that:

$$N(\mu = \frac{b}{a}; \sigma^2 = \frac{1}{a}) \propto exp\{-\frac{ax^2}{2} + bx\}$$

and..

$$\int \frac{\sqrt{a}}{\sqrt{2\pi}} \cdot e^{-\frac{ax^2}{2}+bx-\frac{b^2}{2a}}dx = 1$$

we can compute:

$$\int e^{-\frac{ax^2}{2}+bx} = \frac{\sqrt{2\pi}}{\sqrt{a}} \cdot e^{\frac{b^2}{2a}}$$

so:

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

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

the last function is a $N(a = (\lambda - \frac{\lambda^2}{\lambda+\nu}); b = \frac{\lambda\nu\mu}{\lambda+\nu})$; so the prior predictive distribution is a normal: $N(\frac{b}{a} = \mu; \frac{1}{a} = \frac{\lambda+\nu}{\lambda\nu})$

Another way to compute this distribution is to think the distribution as a convolution:

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

where $W(w)$ is a Normal $N(0; \frac{1}{\lambda})$.

So for the properties of convolution we know that:

$$X = W + \theta$$

where $W$ and $\theta$ must be independent and both Gaussians.

$$\theta \sim N(\mu_\pi, \sigma_\pi^2 = \frac{1}{\nu_\pi^2})$$

$$W \sim N(0, \sigma^2 = \frac{1}{\lambda})$$

$$X \sim N(0, \lambda) * N(\mu, \nu) \sim N(\mu_{pred} = \mu_\pi, \sigma^2_{pred} = \frac{\lambda + \nu}{\lambda \nu})$$

where:

$$\mu_{pred} = E(X) = E(W + \theta) = 0 + \mu_\pi$$

and for the independence:

$$\sigma^2_{pred} = Var(X) = Var(W + \theta) = Var(W) + Var(\theta) = \sigma^2_\pi + \sigma^2 = \frac{\lambda + \nu}{\lambda \nu}$$

## Part 2.b

b) derive the general formula of the posterior predictive distribution

**-Anwer**

Remembering that:

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

we know that $\pi(\theta|x)$ is the posterior distribution that is a $N(\mu'_\theta, \lambda'_\theta)$:

$$\mu'_\theta = w \cdot \mu + (1-w) \cdot \bar{x}_n$$

$$\nu'_\theta = \nu + n\lambda$$

and:

$$w = \frac{\nu}{\nu + n\lambda}$$

so it's easy now to compute the posterior predictive distribution:

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

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

because is the same integral we compute before.

So the posterior predictive distribution is a $N(a = (\frac{\lambda\nu'_\theta}{\lambda+\nu'_\theta}); b = \frac{\lambda\nu'_\theta\mu'_\theta}{\lambda+\nu'_\theta}) = N(\mu = \mu'_\theta, \sigma^2 = \frac{\lambda+\nu'_\theta}{\lambda\nu'_\theta})$

We could arrive to the same answer also with convolution method we saw before.

## Part 2.c

c) assume that the known value of $\lambda$ 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 $\theta$ 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 want that $P_\pi(-5 \le \theta \le 5) = 0.96$, and $\mu = 0$ so:

$$P_\pi(-5 \le \theta \le 5) = P_\pi(\frac{-5-\mu}{\sigma} \le \frac{\theta-\mu}{\sigma} \le \frac{5-\mu}{\sigma}) = P_\pi(\frac{-5}{\sigma} \le Z \le \frac{5}{\sigma}) = \Phi(\frac{5}{\sigma}) - \Phi(-\frac{5}{\sigma}) = 1 - 2\Phi(-\frac{5}{\sigma})$$

$$\Phi(-\frac{5}{\sigma}) = 0.02$$

```
qnorm(0.02)
```

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

$$\sigma = \frac{-5}{-2.053749} = 2.4346$$

so the prior is a $N(\mu = 0, \sigma^2 = 5.93)$ or in an other parametrization $N(\mu = 0, \nu = 0.1687)$.

## Part 2.d

d) derive your posterior distribution and represent it graphically

**-Answer**

So:

$$\pi(\theta|\bar{X}) \propto L(\theta) \cdot \pi(\theta) \propto \prod_{i=0}^{10} f(x|\theta) \cdot \pi(\theta)$$

$$\propto \prod_{i=0}^{10} \frac{\sqrt{\lambda}}{\sqrt{2\pi}} \cdot e^{-\frac{\lambda(x_i-\theta)^2}{2}} \cdot \frac{\sqrt{\nu}}{\sqrt{2\pi}} \cdot e^{-\frac{\nu(\theta-\mu)^2}{2}} \propto e^{-\frac{\lambda}{2}\sum_{i=0}^{10}(x_i-\theta)^2} \cdot e^{-\frac{\nu(\theta-\mu)^2}{2}}$$

$$\propto e^{-\frac{\lambda}{2}[\sum(x_i^2-2\theta x_i)+n\theta^2]-\frac{\nu}{2}(\theta^2-2\mu\theta+\mu^2)} \propto e^{-\frac{\lambda}{2}(-2\theta\sum x_i+n\theta^2)-\frac{\nu}{2}(\theta^2-2\mu\theta)} \propto e^{-\frac{1}{2}(n\lambda+\nu)\theta^2+(\nu\mu+\lambda\sum x_i)\theta}$$

So the distribution is a

$$N(a_{post} = (n\lambda + \nu); b_{post} = (\nu\mu + \lambda\sum x_i))$$

where:

$$N(\mu_{post} = \frac{b_{post}}{a_{post}}; \nu_{post} = a_{post})$$

```
par(mfrow = c(2,2))
X.bar <- c(-1.25, 8.77, 1.18, 10.66, 11.81, -6.09, 3.56, 10.85, 4.03, 2.13)

L1 <- function(thet){
  prod=1
  for(x in X.bar){
    prod = prod*dnorm(x, mean=thet, sd=sqrt(3))
  }
  return(prod)
}

#int.1 <- integrate(L1,-Inf,Inf)
#prob1<-function(x) L1(x)/int.1$value
curve(L1, xlab = expression(theta), ylab = expression(L(theta)), from = 0, to = 10)

pr <- function(thet) dnorm(thet, 0, (-5/qnorm(0.02)))
curve(pr, ylab = 'prior', xlab=expression(theta), from = -10, to = 10)

jnt1 <- function(t) L1(t)*pr(t)
m1.x_bar <- integrate(jnt1, -Inf, Inf)
post1 <- function(t) jnt1(t)/m1.x_bar$value

curve(post1, from = 0, to = 10, ylab = expression(posterior),
      xlab = expression(theta), col='green')
```
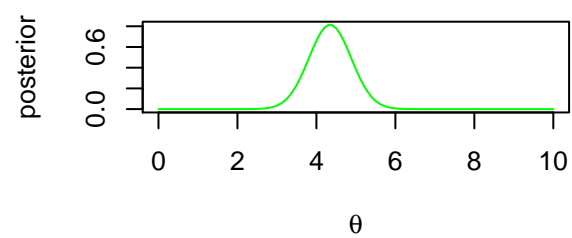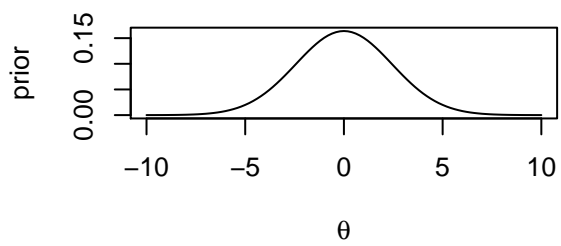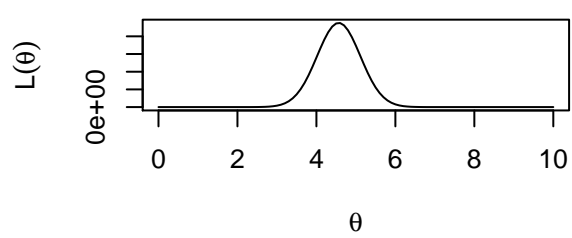
**Part 2.e**

e) derive your favorite point estimate and interval estimate and motivate your choices

**-Answer**

The posterior distribution of $\theta$ is in a Bayesian way to think, the real density function of $\theta$, so we have a good point estimator for our parameter using the expected value of our $\theta$:

$$E(\theta) = \int \theta \cdot \pi(\theta|\bar{x})d\theta$$

and in this case where we have a normal distribution like posterior by computing the mean of our posterior distribution:

$$\mu'_\theta = w \cdot \mu + (1 - w) \cdot \bar{x}_n$$

$$w = \frac{\nu}{\nu + n\lambda} = \frac{0.1687}{0.1687 + 10 \cdot 0.\bar{3}} = 0.048$$

so:

$$\mu'_\theta = 0.044 \cdot 0 + (1 - 0.048) \cdot 4.565 = 4.345$$

While to compute the credible interval we must know the variance of our posterior, so:

$$\nu'_\theta = \nu + n\lambda = 0.1687 + 10 \cdot \frac{1}{3} = 3.50$$

and:

$$\sigma'_\theta = \sqrt{\frac{1}{\nu'_\theta}} = 0.535$$

so:

$$P(\mu'_\theta - z_{\alpha/2} \cdot \sigma'_\theta < \theta|x < \mu'_\theta + z_{\alpha/2} \cdot \sigma'_\theta) = 1 - \alpha$$

for a confidence interval with probabilty of 0.96 we have:

```r
nu = 0.1687154
w = nu/(nu+ 10/3)
mu = (1-w)*mean(X.bar)
nu.tet = nu + 10/3
sig= sqrt(1/nu.tet)
z = qnorm(0.02)
cat('mean =',mu,'\n')
```

```
## mean = 4.345076
```

```r
c(lower = mu + z*sig, upper= mu - z*sig)
```

```
##    lower    upper
## 3.247622 5.442529
```

# Question 3

3) As an alternative model for the previous 10 observations

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

consider the following statistical model where $X_i|\theta$ are i.i.d with

$$X_i|\theta \sim f(x|\theta) = \frac{1}{20}I_{[\theta-10,\theta+10]}(x)$$

Use the same prior elicitation for $\theta$ as in the model of the previous excercise

## Part 3.a

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**

So, remembering that:

$$\pi(\theta|\bar{X}) \propto L(\theta) \cdot \pi(\theta) \propto \prod_{i=0}^{10} f(x|\theta) \cdot \pi(\theta)$$

we have that:

$$L(\theta) = \prod_{i=0}^{10} f(x|\theta) = \prod_{i=0}^{10} \frac{1}{20}I_{[\theta-10,\theta+10]}(x_i)$$

and we can say that:

$$\begin{cases} L(\theta) = 0 & if \quad \theta < \theta_{min} \\ L(\theta) = \frac{1}{20^{10}} & if \quad \theta_{min} \le \theta \le \theta_{max} \\ L(\theta) = 0 & if \quad \theta \ge \theta_{max} \end{cases}$$

where: $\theta_{min} = max(\bar{X}) - 10$ and $\theta_{max} = min(\bar{X}) + 10$

```
theta.min <- max(X.bar) - 10
theta.max = min(X.bar) + 10
c(min = theta.min, max = theta.max)
```

```
##  min  max
## 1.81 3.91
```

So:

$$L(\theta) \propto U(1.81, 3.91)$$

Our prior is: $\pi(\theta) = N(\mu = 0, \nu = 0.1687)$

So, finally we can derive the posterior distribution:

$$\pi(\theta|\bar{X}) \propto L(\theta) \cdot \pi(\theta) \propto \frac{1}{20^{10}} \cdot I_{(\theta \in [1.81;3.91])} \cdot \frac{\sqrt{\nu}}{\sqrt{2\pi}}e^{-\frac{\nu}{2}\theta^2} \propto \frac{\sqrt{\nu}}{\sqrt{2\pi}}e^{-\frac{\nu}{2}\theta^2} \cdot I_{(\theta \in [1.81;3.91])}$$
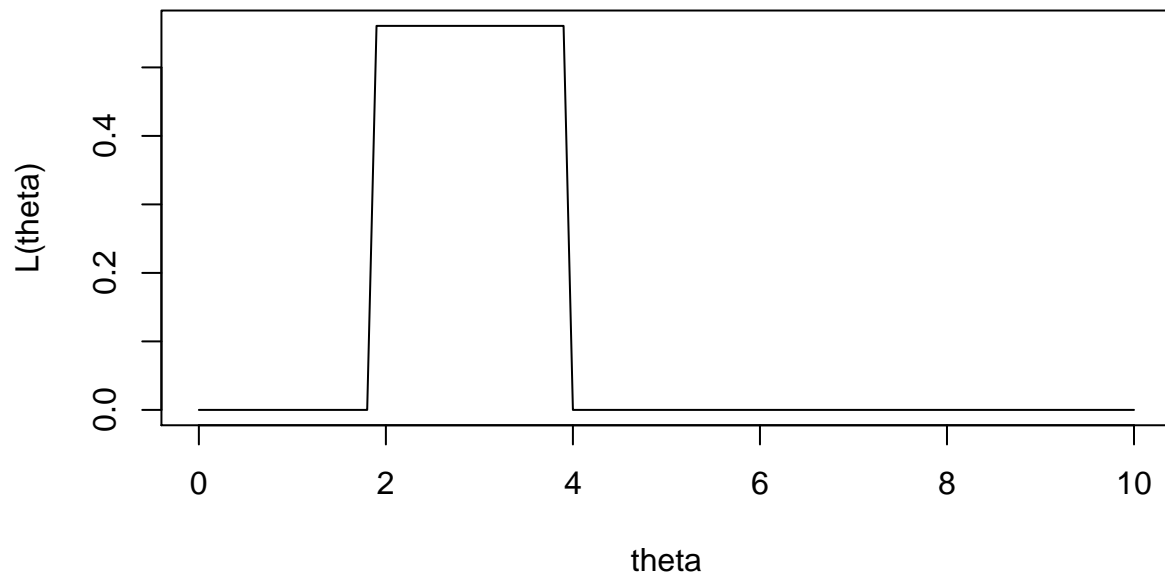
That is a truncated Normal.

16

```r
X.bar <- c(-1.25, 8.77, 1.18, 10.66, 11.81, -6.09, 3.56, 10.85, 4.03, 2.13)

L2 <- function(thet){
  prod=1
  for(x in X.bar){
    prod = 1/12*prod*(x>=thet-10 & x<=thet+10)
  }
  return(prod)
}

int2 <- integrate(L2,-Inf,Inf)
prob2 <- function(x) L2(x)/int2$value
curve(prob2, xlab = 'theta', ylab = 'L(theta)', from = 0, to = 10)
```
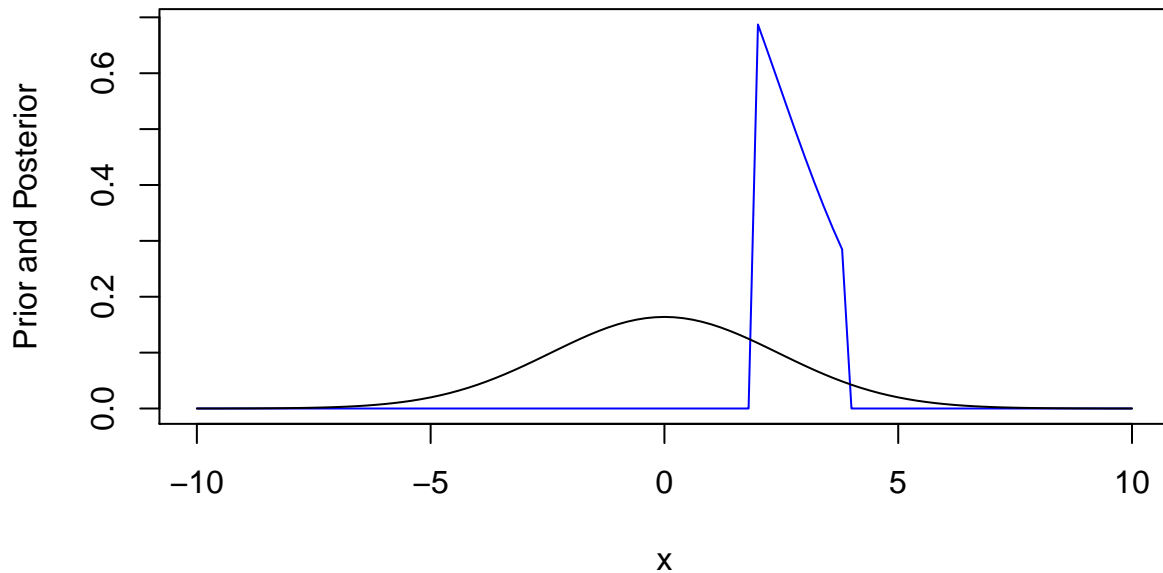
```
pr <- function(thet) dnorm(thet, 0, (-5/qnorm(0.02)))
```

```
jnt2 <- function(t) L2(t)*pr(t)
m2.x_bar <- integrate(jnt2, -Inf, Inf)
post2 <- function(t) jnt2(t)/m2.x_bar$value
curve(post2, from = -10, to = 10, col='blue',ylab = 'Prior and Posterior')
curve(pr, ylab = 'prior', xlab='theta', add = T)
```



At last, comparing the posterior distribution of previous model and this last model, we can infer that the first model allows all possible thetas but mainly in the interval $[3.25; 5.44]$ with expected value equal 4.35, instead this last model recommends a finite interval of thetas that has to be in the interval $[1.81; 3.91]$ with probability 1, and with expected value 2.76

```
mean.int <- function(x) x*post2(x)
E <- integrate(mean.int, -Inf, Inf)
E
```

```
## 2.758036 with absolute error < 0.00031
```

## Part 3.b

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**

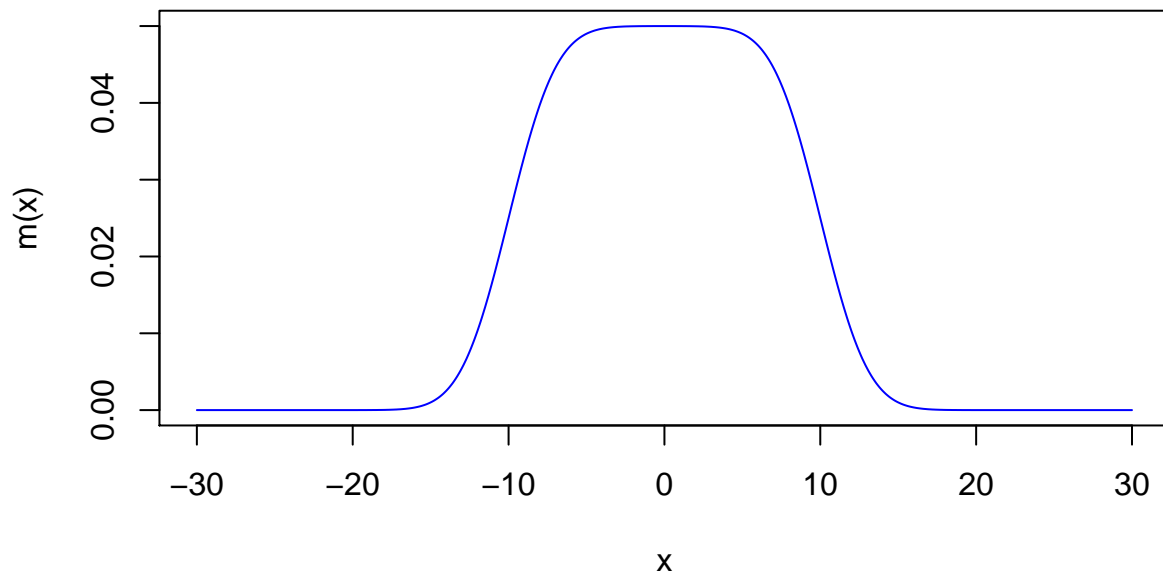Like always the prior predictive distribution is:

$$m(\cdot) = \int f(\cdot|\theta)\pi(\theta)d\theta$$

so:

$$m(x) = \int \frac{1}{20} \cdot I_{(\theta \in [x-10;x+10])} \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu}{2}\theta^2} d\theta = \int_{x-10}^{x+10} \frac{1}{20} \cdot \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu}{2}\theta^2} d\theta$$

$$= \frac{1}{20} \cdot P(x-10 < \theta_{prior} < x+10) = \frac{1}{20}(\Phi(\frac{x+10}{\sigma}) - \Phi(\frac{x-10}{\sigma})) = \frac{1}{20}(\Phi(\frac{x+10}{2.43}) - \Phi(\frac{x-10}{2.43}))$$

```
x <- seq(-30,30,0.1)
marginal <- function(x) 1/20*(pnorm((x+10)*sqrt(nu)) - pnorm((x-10)*sqrt(nu)))

x.prob <- marginal(x)
plot(x,x.prob, type= 'l', col='blue',ylab='m(x)')
```



And we can see this distribution looks like an uniform distribution but is not.

We can compute also an approssimation for this distribution function using a simulation with the Acceptance-rejection method;

```
marginal <- function(x) 1/20*(pnorm((x+10)*sqrt(nu)) - pnorm((x-10)*sqrt(nu)))
obs <- c()
for(i in 1:100000){
  y <- runif(1,-20,20)
  z <- rbinom(1,1,marginal(y)/0.06)
  if(z==1){
    obs=c(obs,y)
  }
}
hist(obs, probability = T, breaks = 20)
curve(marginal(x), type= 'l',add = T, col='blue')
d <- density(obs)
lines(d,type= 'l', col='green')
legend(8.3,0.05,c('prior predictive','Kernel of obs'), lty = c(1,1),
       col=c('blue','green'), lwd = c(2.5,2.5),cex = 0.75)
```
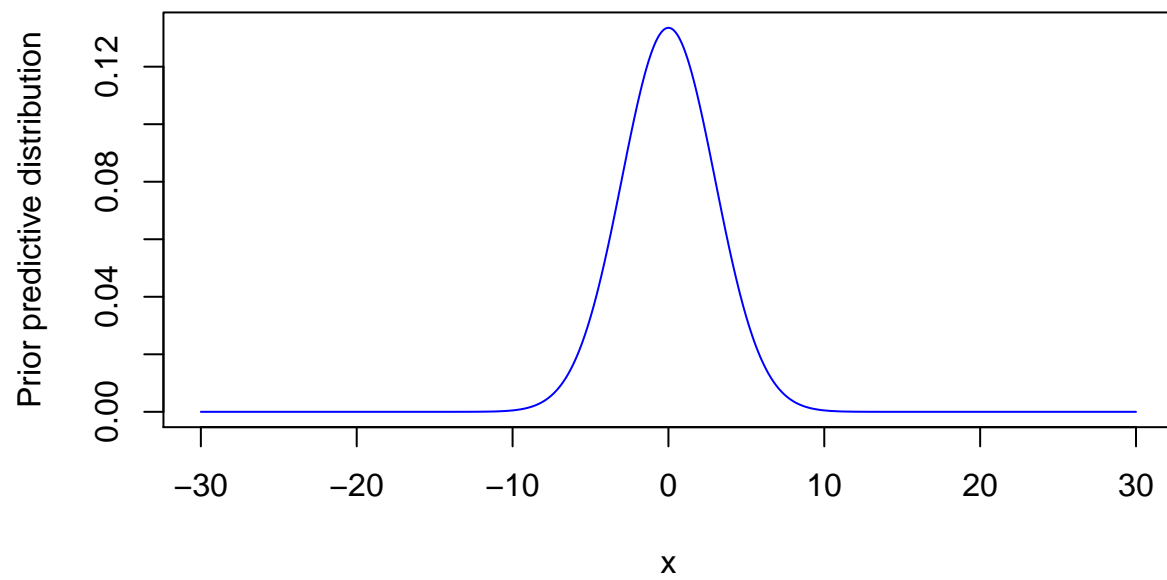
## Histogram of obs
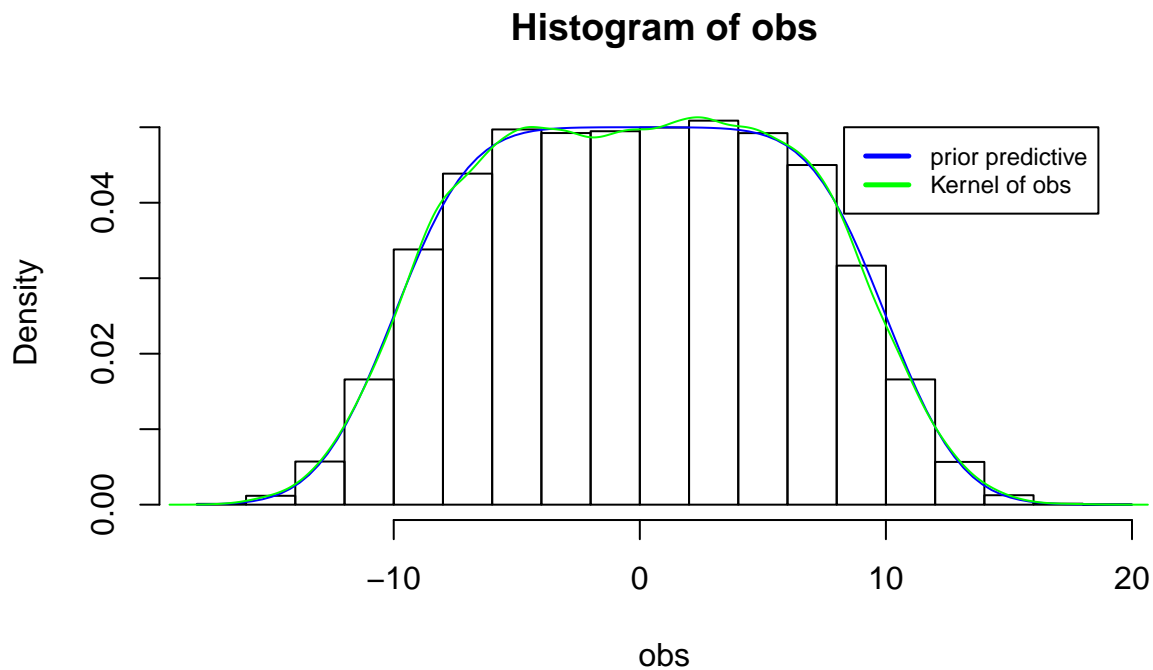


Where I took 0.06 like the value of $c \cdot g(y)$ that is for shure greater than $f(x)$, so is a good bounder for it.

While the pior predictive distribution for the previous model is: $N(\mu = 0, \sigma = 8.927)$

```
plot(x, dnorm(x,sd=sqrt(8.927)),type = 'l',col='blue',ylab = 'Prior predictive distribution')
```

This prior predictive distribution has higher probability for x close to 0 and fewer to values further, and don't have a behaviour similar to a Uniform distribution of course.

## Part 3.c

c) Consider the same discrete (finite) grid of values as parameter space $\Theta$ for the conditional mean $\theta$ 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**

Using the properties of Bayesian multimodel inference, we have that:

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

and:

$$J(data, m) = \int J(m, \theta, data)d\theta = prior(m) \cdot b(m|data)$$

where $b(m|data) = J(data|m) = \int \pi(\theta|m) \cdot f(data|\theta, m)d\theta$.

So to compute the posterior distribution, we should:

$$post(m|data) = \frac{prior(m) \cdot b(m|data)}{\sum_{m_i} prior(m_i) \cdot b(m_i|data)}$$

Finally we cann compare both models using:

$$\frac{post(m_1|data)}{post(m_2|data)} = \frac{prior(m_1)}{prior(m_2)} \cdot \frac{b(m_1|data)}{b(m_2|data)}$$

and using the same prior probabilty for both models, we have that:

$$\frac{post(m_1|data)}{post(m_2|data)} = \frac{b(m_1|data)}{b(m_2|data)}$$

that is the Bayes Factor!!!

$$BF = \frac{\int \prod_{i=0}^{10} \frac{\sqrt{\lambda}}{\sqrt{2\pi}} \cdot e^{-\frac{\lambda(x_i - \theta)^2}{2}} \cdot \frac{\sqrt{\nu}}{\sqrt{2\pi}} \cdot e^{-\frac{\nu(\theta - \mu)^2}{2}} d\theta}{\int \frac{1}{20^{10}} \cdot I_{(\theta \in [1.81;3.91])} \cdot \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{\nu}{2}\theta^2} d\theta}$$

```
bayes.factor <- m1.x_bar$value/m2.x_bar$value
bayes.factor
```

```
## [1] 1.307783e-19
```

At the end, we can see that computing the BF, we infer that the **second model** is a better choice.

---

© 2016-2017 - Stat4DS2+CS - Luca Tardella

```
## This homework will be graded and it will be part of your final evaluation
##
##

## Last update by LT: Thu Mar 23 17:16:06 2017
```