Statistique bayésienne avec R

Julien JACQUES

Préliminaires

Commençons par charger les libraires dont nous aurons besoin

```
library(bayess,quietly = T)
library(rjags,quietly = T)
library(BayesFactor,quietly = T)
```

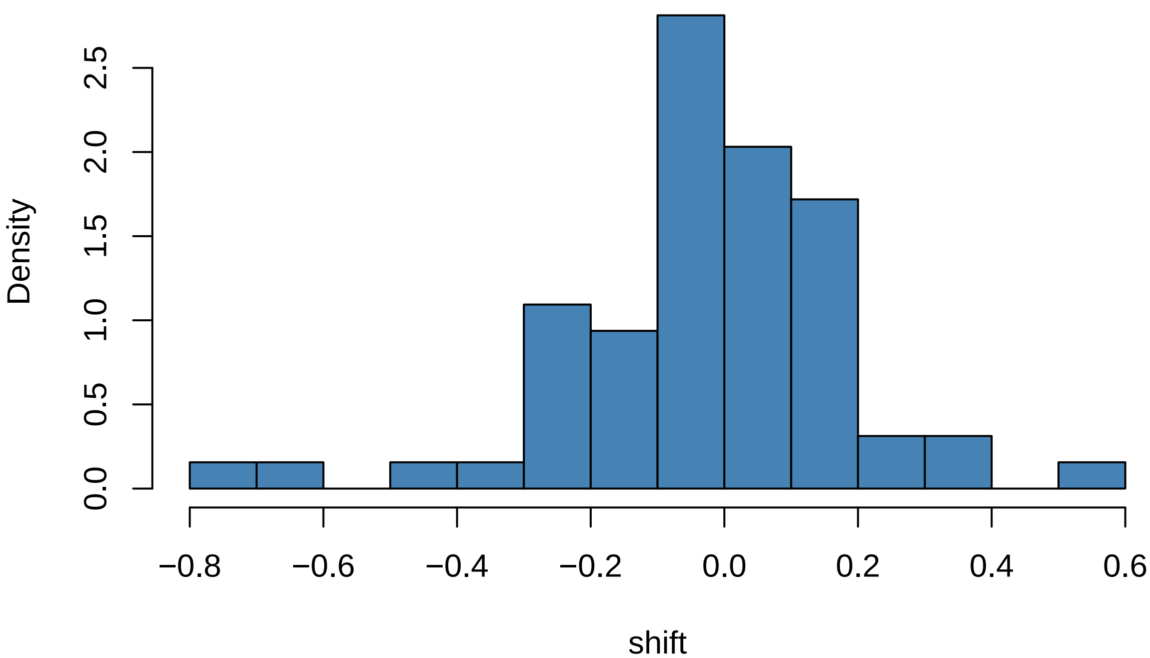
Exemple 1 - normaldata

Chargeons les données normaldata et regardons ce qu'elles contiennent

```
data(normaldata)
?normaldata
```

On s'intéresse à la seconde variable, shift

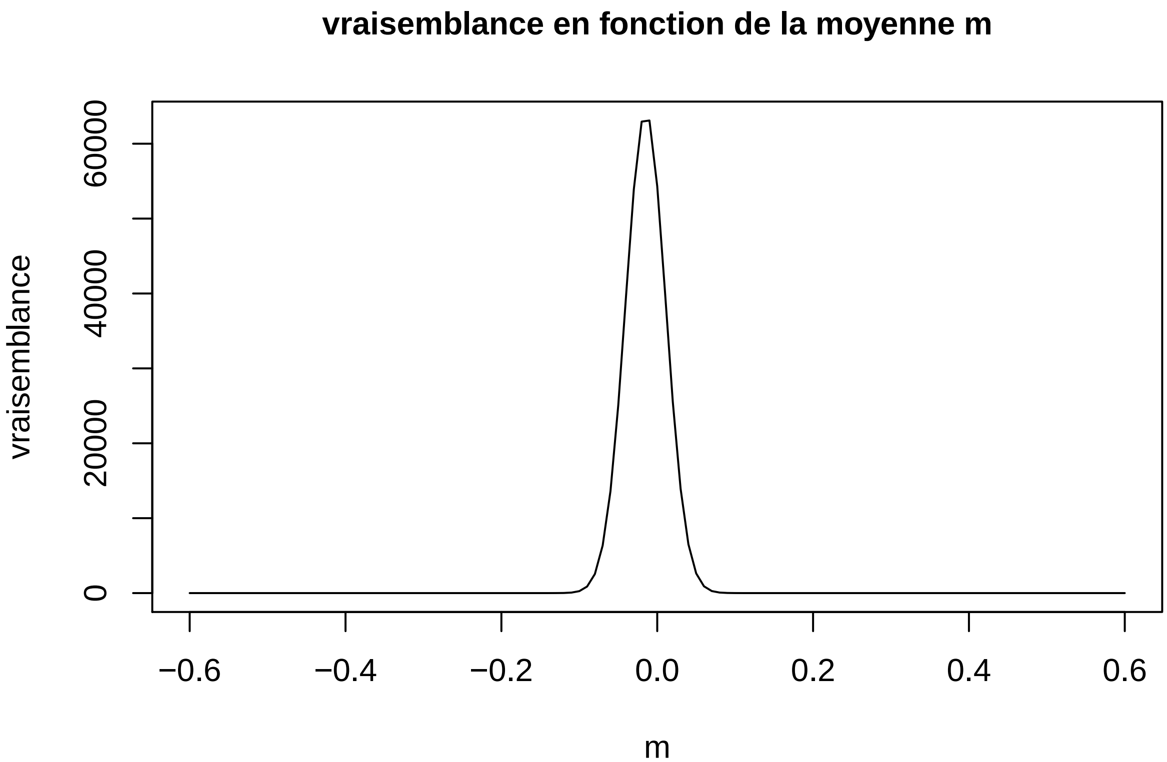
```
shift=normaldata[,2]
hist(shift,nclass=10,col="steelblue",probability=TRUE,main="")
```



Calculons et traçons la vraisemblance des données shift en fonction de la valeur m de l'espérance

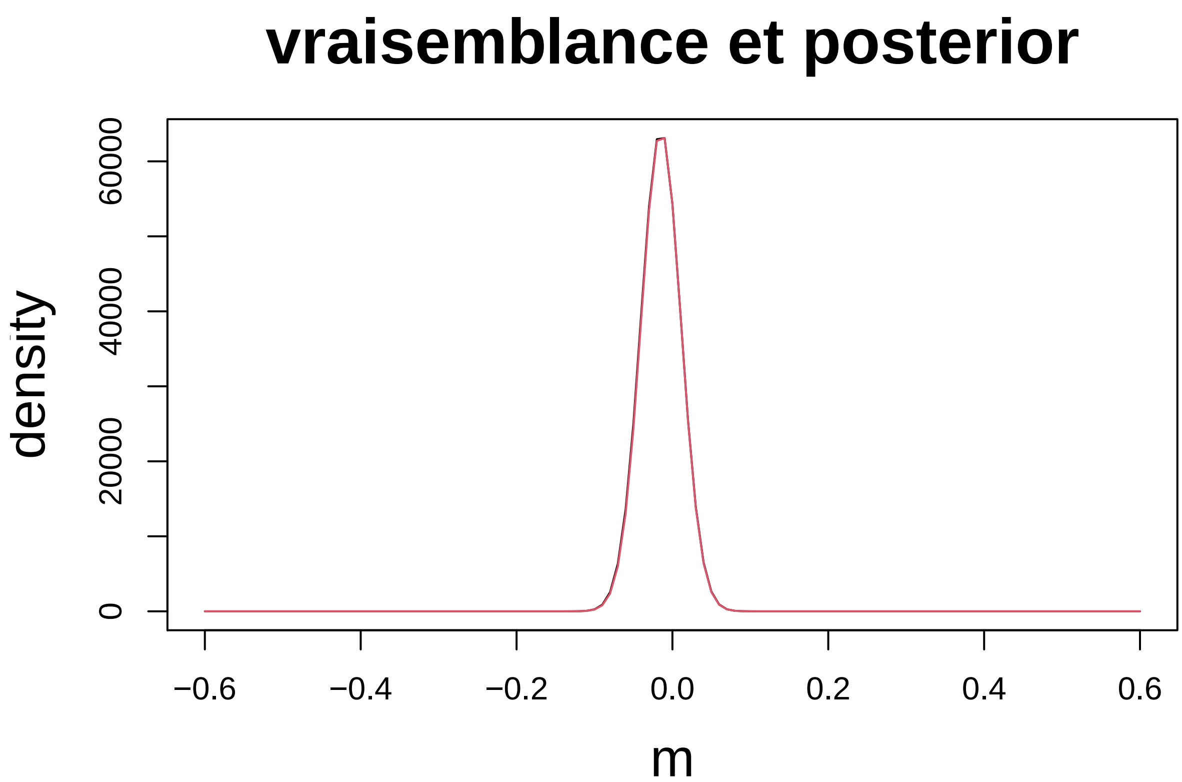
```
m=seq(-0.6,0.6,0.01)
L=NULL
```

```r
for (i in 1:length(m)) L[i]=prod(dnorm(shift,mean=m[i],sd=sd(shift)))
plot(m,L,type="l",ylab="vraisemblance",main="vraisemblance en fonction de la moyenne m",cex.main=1, cex
```



On peut également tracer la loi a posteriori, en la renormalisant pour qu'elle soit sur la même échelle que la vraissemblance qui elle n'est pas une loi de proba

```r
plot(m,L,type="l",ylab="density",main="vraisemblance et posterior",cex.main=2, cex.lab=1.7, cex.sub=1.2)
n=length(shift)
d=dnorm(m,n*mean(shift)/(n+1),sqrt(var(shift)/(n+1)))
lines(m,d/max(d)*max(L),col=2)
```



Le maximum de vraisemblance est le MAP sont très proches :

```
cat('Maximum de vraisemblance = ',mean(shift),'\n')
```

```
## Maximum de vraisemblance =  -0.01484375
```

```
cat('Maximum a posteriori = ',n*mean(shift)/(n+1),'\n')
```

```
## Maximum a posteriori =  -0.01461538
```

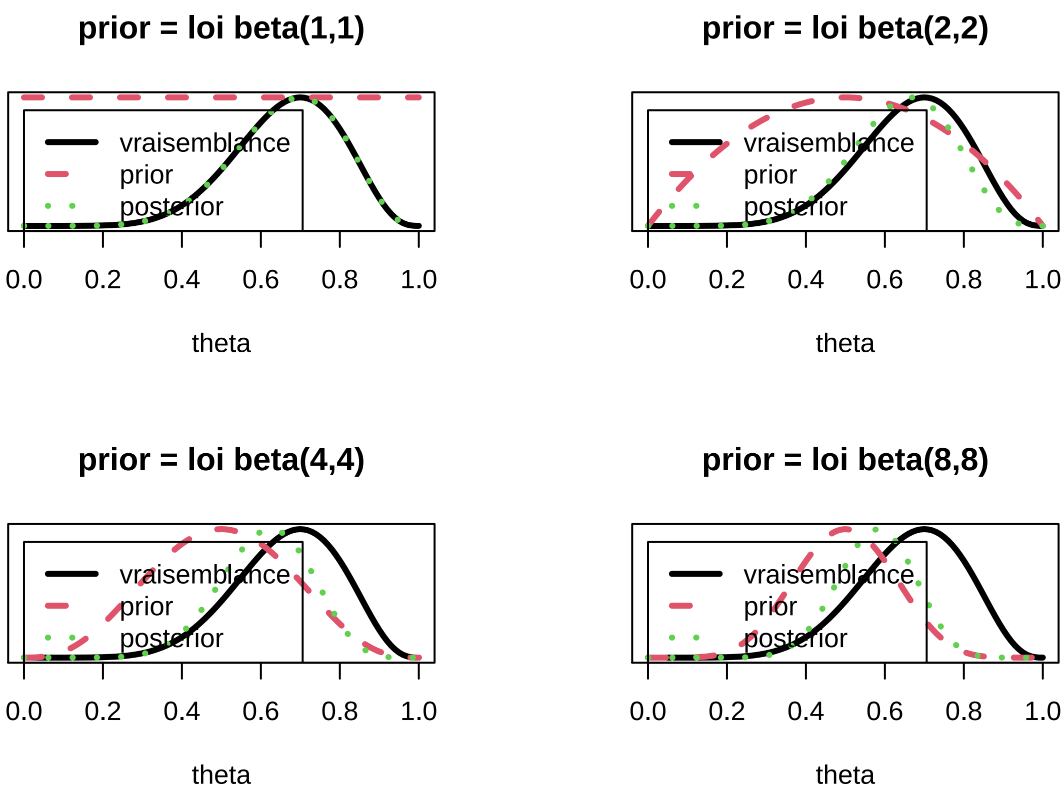
Exemple 2 - pile ou face

Calcul de la vraisemblance

```
theta=seq(0,1,0.01)
vraisemblance=dbinom(7,10,theta)
```

Représentation de la vraisemblance, de la loi a priori et de la loi a posteriori (calculable ici car on a choisi un a priori conjugué), pour différentes valeurs des paramètres de la loi a priori.

```
par(mfrow=c(2,2))
for (i in c(1,2,4,8)){
  a=i;b=i;
  plot(theta,vraisemblance/max(vraisemblance),type='l',ylab="",yaxt="n",lwd=3)
  title(paste("prior = loi beta(",a,",",b,")",sep=""))
  prior=dbeta(theta,a,b)
  lines(theta,prior/max(prior),type='l',col=2,lty=2,lwd=3,ylab="",yaxt="n")
  posterior=dbeta(theta,a+7,b+3)
  lines(theta,posterior/max(posterior),type='l',col=3,lty=3,lwd=3,ylab="",yaxt="n")
  legend(0, .9, legend=c("vraisemblance", "prior","posterior"),col=1:3, lty=1:3, cex=1,lwd=3)
}
```



Interprétation de l'influence de la loi a priori : plus sa variance est faible, plus il “tire à lui” la loi a posteriori.

Sachant que la loi a posteriori est une loi Beta, on peut estimer θ par le mode de la distribution a posteriori (Maximum A Posteriori, MAP). On peut aussi calculer facilement un intervalle de crédibilité à l'aide des quantiles de la loi Beta.

```
for (i in c(1,2,4,8)){
  a=i;b=i;
  cat('Estimation MAP :',(a+6)/(a+b+8),'\n')
  cat('Intervalle de credibilité : [',qbeta(0.025,a+7,b+3),',',qbeta(0.975,a+7,b+3),']','\n')
}
```

```
## Estimation MAP : 0.7
## Intervalle de credibilité : [ 0.3902574 , 0.8907366 ]
## Estimation MAP : 0.6666667
## Intervalle de credibilité : [ 0.3857383 , 0.8614207 ]
## Estimation MAP : 0.625
## Intervalle de credibilité : [ 0.3832837 , 0.815563 ]
## Estimation MAP : 0.5833333
## Intervalle de credibilité : [ 0.3866535 , 0.7559763 ]
```

Monte Carlo en pratique

```
M=matrix(0,100,3)
for (i in 1:100){
  M[i,1]=mean(exp(-runif(100,0,1)^2/2))
  M[i,2]=mean(exp(-runif(10^4,0,1)^2/2))
  M[i,3]=mean(exp(-runif(10^6,0,1)^2/2))
}
boxplot(M)
```

```
abline(h=sqrt(2*pi)*(pnorm(1)-pnorm(0)),col=2)
```