

## Master Statistique & Big Data – Université Paris Dauphine

### DM Statistique Bayésienne

#### Importation des données

```
setwd("C:\\Users\\oussa\\Downloads\\Data Science\\Master Statistique Big Data
Dauphine\\Module 2\\Statistique Bayésienne")
list.files()

data=read.csv("mutations2.csv")
require(zoo)#pour la verification de la convergence et du melange

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.3.3

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
```

#### 1) Régression Linéaire

##### 1) Régression Linéaire Bayésienne

Pour faire la régression linéaire bayésienne, on choisit une loi à priori de Zellner :

$$\begin{aligned}\beta|\sigma^2, X &\sim N_{k+1}(0, g\sigma^2(X^T X)^{-1}) \\ \pi(\sigma^2) &\propto \sigma^{-2}\end{aligned}$$

Cette loi est conjuguée la loi à posteriori est :

$$\begin{aligned}\beta|\sigma^2, y, X &\sim N_{k+1}\left(\frac{g}{g+1}\hat{\beta}, \frac{\sigma^2 g}{g+1}(X^T X)^{-1}\right) \\ \sigma^2|y, X &\sim \text{IG}\left(\frac{n}{2}, \frac{s^2}{2} + \frac{1}{2(g+1)}\hat{\beta}^T X^T X \hat{\beta}\right)\end{aligned}$$

et marginalement, on a une loi de Student :

$$\beta|y, X \sim \mathcal{T}_{k+1} \left( n, \frac{g}{g+1} \hat{\beta}, \frac{g(s^2 + \hat{\beta}^T X^T X \hat{\beta} / (g+1))}{n(g+1)} (X^T X)^{-1} \right).$$

le paramètre  $g$  s'interprète comme la quantité d'information disponible dans la loi a priori par rapport à l'échantillon.  $g$  est donc l'inverse de la proportion de l'échantillon équivalente au poids de la loi a priori

```
reg_lin_bayes=function(df){  
  
  y=df[,6]  
  X=as.matrix(df[,7:(ncol(df))])  
  n=length(y)  
  
  #Interpretation de g  
  g=c(1,10,50,nrow(df))  
  
  betahat=lm(y~X)$coefficients  
  #str(betahat)  
  residuals=lm(y~X)$residuals  
  s2=t(residuals)%*%residuals  
  
  #nrow(betahat)  
  for (j in 1:length(betahat)){  
  
    cat(substring(names(betahat)[j],2),":")  
    print(betahat[j]*g/(g+1))  
  }  
  #print(names(betahat))  
  X=cbind(1,X)  
  
  # esperance de sigma^2  
  a=n/2  
  b=s2/2+1/(2*g+2)*((t(betahat)%*%t(X))%*(X%*%betahat))  
  
  sigma_carre=b/(a-1)  
  cat("variance:",sigma_carre)  
  
}
```

Remarque : on peut également faire de la régression linéaire bayésienne avec le package MCMC(régression linéaire avec une loi a priori de Zellner, génération des coefficients linéaires par algorithme de Gibbs):

```
library(MCMCpack)
```

```
## Warning: package 'MCMCpack' was built under R version 3.3.3
```

```
## Loading required package: coda
```

```
## Warning: package 'coda' was built under R version 3.3.3
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 3.3.3
```

```
## ##
```

```
## ## Markov Chain Monte Carlo Package (MCMCpack)
```

```
## ## Copyright (C) 2003-2018 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park
```

```
## ##
```

```
## ## Support provided by the U.S. National Science Foundation
```

```
## ## (Grants SES-0350646 and SES-0350613)
```

```
## ##
```

```
model_bayes=MCMCregress(Barre~.,data=data[,6:ncol(data)])
```

```
summary(model_bayes)
```

```
##
```

```
## Iterations = 1001:11000
```

```
## Thinning interval = 1
```

```
## Number of chains = 1
```

```
## Sample size per chain = 10000
```

```
##
```

```
## 1. Empirical mean and standard deviation for each variable,
```

```
##    plus standard error of the mean:
```

```
##
```

```
##
```

```
## (Intercept) -4.712e+02 5.623e+02 5.623e+00
```

```
## effectif_presents_serie_l 7.479e-01 1.639e+00 1.639e-02
```

```
## effectif_presents_serie_es 2.930e-01 1.222e+00 1.222e-02
```

```
## effectif_presents_serie_s 2.685e-03 1.020e+00 1.020e-02
```

```
## taux_brut_de_reussite_serie_l 3.067e+00 2.546e+00 2.546e-02
```

```
## taux_brut_de_reussite_serie_es 4.719e+00 4.208e+00 4.208e-02
```

```
## taux_brut_de_reussite_serie_s 9.281e+00 6.322e+00 6.322e-02
```

```
## taux_reussite_attendu_serie_l -1.423e+01 6.952e+00 6.952e-02
```

```
## taux_reussite_attendu_serie_es 3.928e+00 8.264e+00 8.264e-02
```

```
## taux_reussite_attendu_serie_s -4.201e+00 9.597e+00 9.597e-02
```

```
## effectif_de_seconde 5.518e-02 6.239e-01 6.239e-03
```

```

## effectif_de_premiere -3.564e-01 7.169e-01 7.169e-03
## taux_acces_brut_seconde_bac 1.071e+01 5.695e+00 5.695e-02
## taux_acces_attendu_seconde_bac -6.940e+00 9.029e+00 9.029e-02
## taux_acces_brut_premiere_bac -2.040e+01 1.065e+01 1.065e-01
## taux_acces_attendu_premiere_bac 3.460e+01 1.916e+01 1.916e-01
## taux_brut_de_reussite_total_series -5.160e+00 1.289e+01 1.289e-01
## taux_reussite_attendu_total_series -4.571e+00 2.217e+01 2.217e-01
## sigma2 1.792e+05 1.147e+04 1.147e+02
## Time-series SE
## (Intercept) 5.571e+00
## effectif_presents_serie_l 1.639e-02
## effectif_presents_serie_es 1.222e-02
## effectif_presents_serie_s 1.020e-02
## taux_brut_de_reussite_serie_l 2.546e-02
## taux_brut_de_reussite_serie_es 4.208e-02
## taux_brut_de_reussite_serie_s 6.070e-02
## taux_reussite_attendu_serie_l 6.952e-02
## taux_reussite_attendu_serie_es 8.264e-02
## taux_reussite_attendu_serie_s 9.597e-02
## effectif_de_seconde 6.239e-03
## effectif_de_premiere 7.169e-03
## taux_acces_brut_seconde_bac 5.695e-02
## taux_acces_attendu_seconde_bac 8.992e-02
## taux_acces_brut_premiere_bac 1.065e-01
## taux_acces_attendu_premiere_bac 1.887e-01
## taux_brut_de_reussite_total_series 1.261e-01
## taux_reussite_attendu_total_series 2.217e-01
## sigma2 1.213e+02
##
## 2. Quantiles for each variable:
##
## 2.5% 25% 50%
## (Intercept) -1.562e+03 -8.509e+02 -4.759e+02
## effectif_presents_serie_l -2.435e+00 -3.562e-01 7.413e-01
## effectif_presents_serie_es -2.143e+00 -5.216e-01 2.956e-01
## effectif_presents_serie_s -1.953e+00 -7.007e-01 -5.803e-03
## taux_brut_de_reussite_serie_l -1.928e+00 1.354e+00 3.084e+00
## taux_brut_de_reussite_serie_es -3.502e+00 1.911e+00 4.690e+00
## taux_brut_de_reussite_serie_s -3.367e+00 5.048e+00 9.328e+00
## taux_reussite_attendu_serie_l -2.782e+01 -1.888e+01 -1.416e+01
## taux_reussite_attendu_serie_es -1.206e+01 -1.708e+00 4.087e+00
## taux_reussite_attendu_serie_s -2.315e+01 -1.077e+01 -4.170e+00
## effectif_de_seconde -1.158e+00 -3.660e-01 5.770e-02
## effectif_de_premiere -1.787e+00 -8.401e-01 -3.521e-01
## taux_acces_brut_seconde_bac -4.918e-01 6.848e+00 1.078e+01
## taux_acces_attendu_seconde_bac -2.501e+01 -1.310e+01 -6.954e+00
## taux_acces_brut_premiere_bac -4.107e+01 -2.765e+01 -2.030e+01
## taux_acces_attendu_premiere_bac -2.656e+00 2.140e+01 3.449e+01
## taux_brut_de_reussite_total_series -3.044e+01 -1.391e+01 -5.068e+00
## taux_reussite_attendu_total_series -4.774e+01 -1.966e+01 -4.622e+00

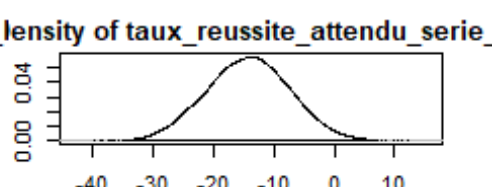
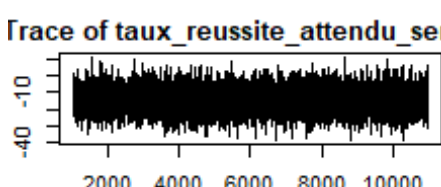
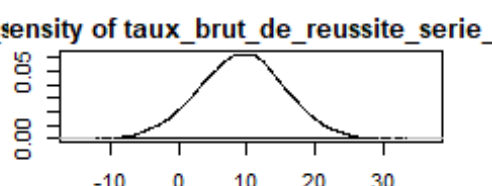
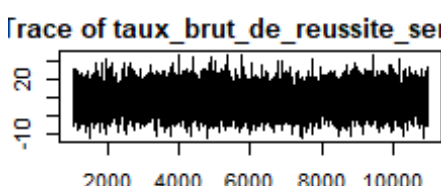
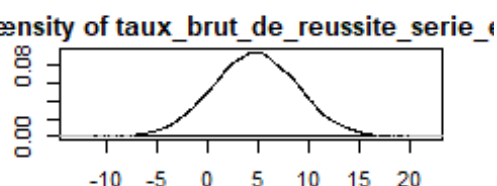
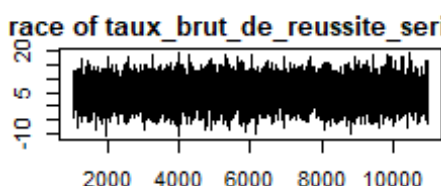
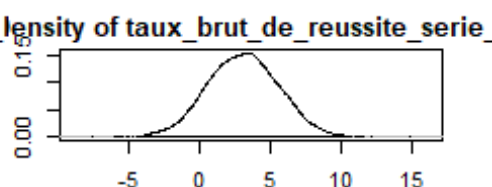
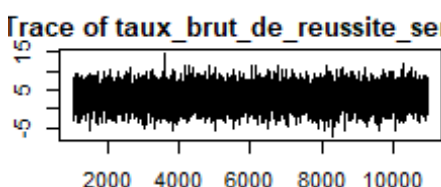
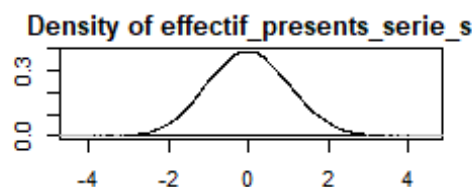
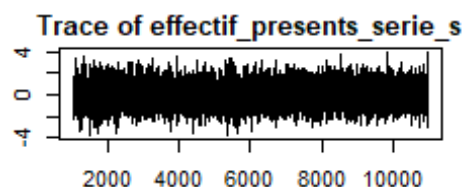
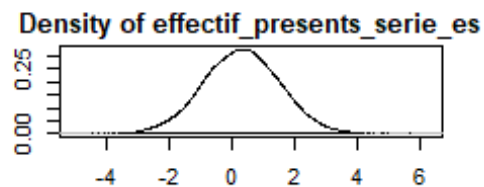
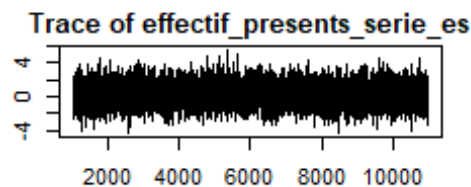
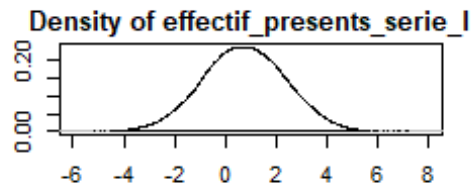
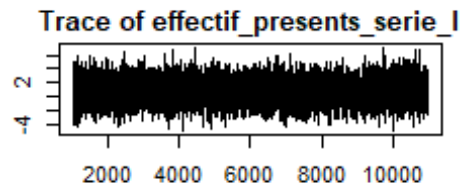
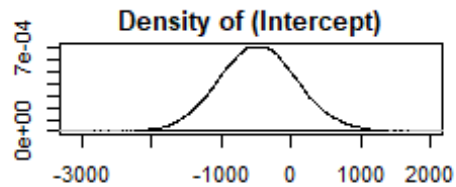
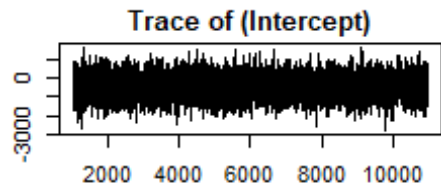
```

## sigma2	1.580e+05	1.713e+05	1.788e+05
##	75%	97.5%	
## (Intercept)	-1.065e+02	6.576e+02	
## effectif_presents_serie_l	1.859e+00	3.948e+00	
## effectif_presents_serie_es	1.109e+00	2.691e+00	
## effectif_presents_serie_s	6.938e-01	2.017e+00	
## taux_brut_de_reussite_serie_l	4.780e+00	8.089e+00	
## taux_brut_de_reussite_serie_es	7.585e+00	1.300e+01	
## taux_brut_de_reussite_serie_s	1.350e+01	2.175e+01	
## taux_reussite_attendu_serie_l	-9.566e+00	-7.551e-01	
## taux_reussite_attendu_serie_es	9.441e+00	2.033e+01	
## taux_reussite_attendu_serie_s	2.229e+00	1.464e+01	
## effectif_de_seconde	4.756e-01	1.266e+00	
## effectif_de_premiere	1.309e-01	1.026e+00	
## taux_acces_brut_seconde_bac	1.454e+01	2.172e+01	
## taux_acces_attendu_seconde_bac	-8.326e-01	1.075e+01	
## taux_acces_brut_premiere_bac	-1.317e+01	2.314e-01	
## taux_acces_attendu_premiere_bac	4.743e+01	7.209e+01	
## taux_brut_de_reussite_total_series	3.474e+00	1.986e+01	
## taux_reussite_attendu_total_series	1.047e+01	3.839e+01	
## sigma2	1.865e+05	2.032e+05	

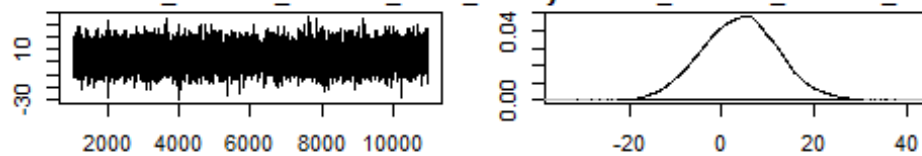
```

par(mfrow=c(4,2))
par(mar = rep(2, 4))
plot(model_bayes)

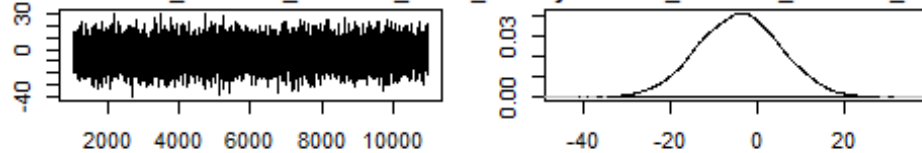
```



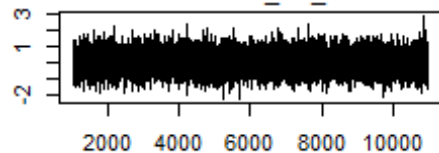
Trace of `taux_reussite_attendu_serie`      Density of `taux_reussite_attendu_serie`



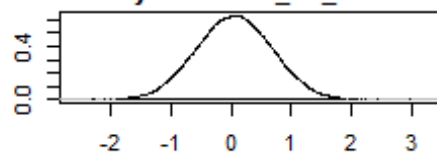
Trace of `taux_reussite_attendu_serie`      Density of `taux_reussite_attendu_serie`



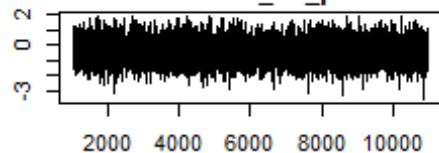
Trace of `effectif_de_seconde`



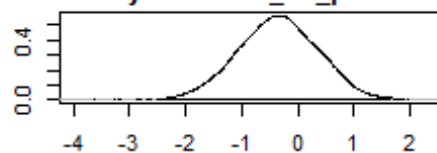
Density of `effectif_de_seconde`



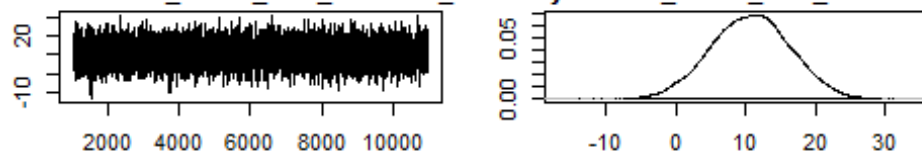
Trace of `effectif_de_premiere`



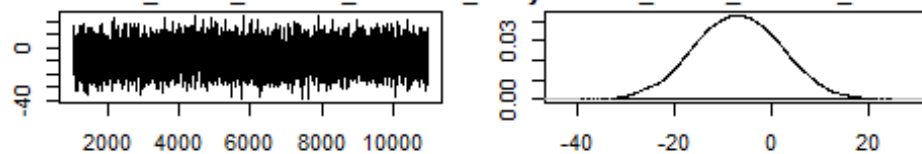
Density of `effectif_de_premiere`



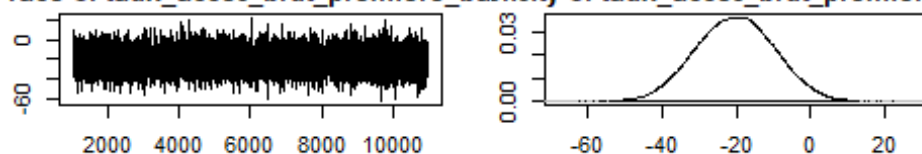
Trace of `taux_acces_brut_seconde`      Density of `taux_acces_brut_seconde`



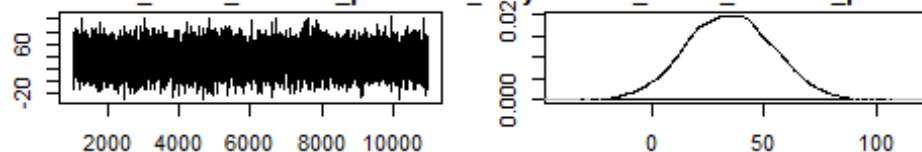
Trace of `taux_acces_attendu_seconde`      Density of `taux_acces_attendu_seconde`

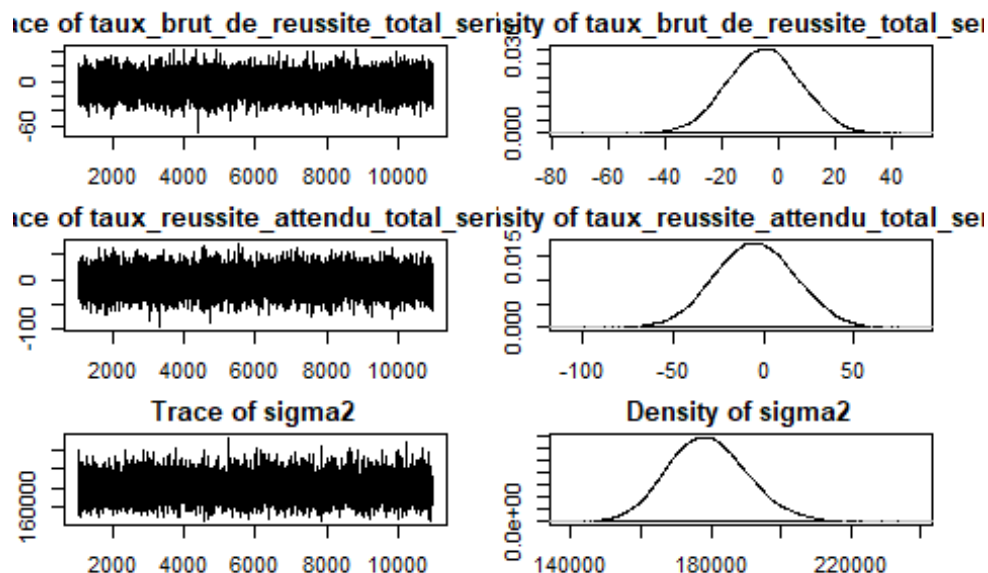


Trace of `taux_acces_brut_premiere`      Density of `taux_acces_brut_premiere`



Trace of `taux_acces_attendu_premiere`      Density of `taux_acces_attendu_premiere`





```
raftery.diag(model_bayes)
```

```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
##
```

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
## (Intercept)	2	3897	3746	1.040
## effectif_presents_serie_l	2	3620	3746	0.966
## effectif_presents_serie_es	2	3802	3746	1.010
## effectif_presents_serie_s	2	3710	3746	0.990
## taux_brut_de_reussite_serie_l	2	3834	3746	1.020
## taux_brut_de_reussite_serie_es	2	3771	3746	1.010
## taux_brut_de_reussite_serie_s	2	3834	3746	1.020
## taux_reussite_attendu_serie_l	2	3680	3746	0.982
## taux_reussite_attendu_serie_es	2	3620	3746	0.966
## taux_reussite_attendu_serie_s	2	3771	3746	1.010
## effectif_de_seconde	2	3710	3746	0.990
## effectif_de_premiere	2	3834	3746	1.020
## taux_acces_brut_seconde_bac	2	3710	3746	0.990
## taux_acces_attendu_seconde_bac	2	3590	3746	0.958
## taux_acces_brut_premiere_bac	2	3710	3746	0.990
## taux_acces_attendu_premiere_bac	2	3834	3746	1.020
## taux_brut_de_reussite_total_serie	2	3834	3746	1.020



##	taux_reussite_attendu_total_series	2	3929	3746	1.050
##	sigma2	2	3680	3746	0.982

## 2) Choix de modele

*##Choix de modele\**

*#option 1 : on utilise l'algorithme de Gibbs vu en TD*

*#fonction pour la log marginal likelihood*

```
marglkd=function(df,gamma){
  y=df[,6]
  X=as.matrix(df[,7:(ncol(df))])
  X=cbind(1,X)
  n=length(y)
  g=nrow(df)
  q=sum(gamma)
  X1=X[,c(T,gamma)]
  if(q==0){return(q/2*log(g+1) -n/2*log(t(y)**y))}
  m = -q/2*log(g+1) -n/2*log(t(y)**y - g/(g+1)* t(y)** X1 **
    solve(t(X1)**X1) **t(X1)**y)
  return(m)
}
```

```
choix_modele_bayes=function(df){
```

*#algorithme de Gibbs: on veut construire une chaine de Markov stationnaire dont la loi stationnaire suit la loi voulue*

```
niter=1e4 # nombre d'iterations
```

```
k=ncol(df[,7:ncol(df)])
```

```
print(k)
```

```
gamma=matrix(F,nrow=niter,ncol=k)
```

```
gamma0=sample(c(T,F),size=k, replace=TRUE) #valeur initiale aleatoire
```

```
lkd=rep(0,niter)
```

```
modelnumber=rep(0,niter)
```

```
oldgamma=gamma0
```

```
for(i in 1:niter){
```

```
  newgamma=oldgamma
```

```
  for(j in 1:k){
```

```
    g1=newgamma; g1[j]=TRUE
```

```
    g2=newgamma; g2[j]=FALSE
```

```
    ml1=marglkd(df,g1)
```

```
    ml2=marglkd(df,g2)
```

```

    p=c(m11,m12)-min(m11,m12)
    # On souhaite tirer depuis une Bernoulli, avec probabilite de tirer TRUE
    egale a exp(p[1])/(exp(p[1])+exp(p[2])).
    # C'est ce que fait la ligne suivante. Notons que la fonction sample()
    calcule la constante de normalisation.
    newgamma[j]=sample(c(T,F), size=1, prob=exp(p))
  }
  gamma[i,]=newgamma
  lkd[i]=marglkd(df,newgamma)
  modelnumber[i]=sum(newgamma*2^(0:(k-1)))
  oldgamma=newgamma
}

apply(gamma, 2, "mean")

# Verifications Le melange de la chaine de Markov a l'aide des
autocorrelations.
par(mfrow=c(4,2))
par(mar = rep(2, 4))
for(i in 1:17) acf(as.numeric(gamma[,i]))# a adapter

# Verifications la convergence + Le melange a l'aide de la trace, ie valeur
prise a chaque iteration (on utilise une moyenne glissante puisque les
valeurs sont binaires).

for(i in 2:k) plot(rollapply(gamma[,i], width=100, FUN=mean), type="l")

burnin=500 # 500 iterations de burn-in
gammab=modelnumber[(burnin+1):niter]
res=as.data.frame(table(gammab))
odo=order(res$Freq, decreasing=T)[1:50]
modcho=res$gammab[odo]
probttop50=res$Freq[odo]/(niter-burnin)

indices=match(modcho,modelnumber)
resultat=cbind(probttop50,gamma[indices,])
print(resultat)
cat("frequence apparition covariable :",colMeans(resultat[, -c(1)]))

}

reg_lin_bayes(data)

## Intercept) :[1] -236.2641 -429.5710 -463.2629 -471.6141
## effectif_presents_serie_l :[1] 0.3890274 0.7073225 0.7627988 0.7765499
## effectif_presents_serie_es :[1] 0.1462204 0.2658552 0.2867066 0.2918751
## effectif_presents_serie_s :[1] 0.004847157 0.008813013 0.009504230
0.009675564
## taux_brut_de_reussite_serie_l :[1] 1.560912 2.838022 3.060612 3.115785

```

```
## taux_brut_de_reussite_serie_es :[1] 2.405544 4.373716 4.716752 4.801782
## taux_brut_de_reussite_serie_s :[1] 4.692469 8.531761 9.200919 9.366785
## taux_reussite_attendu_serie_l :[1] -7.142413 -12.986206 -14.004732 -
14.257196
## taux_reussite_attendu_serie_es :[1] 1.907175 3.467591 3.739559 3.806972
## taux_reussite_attendu_serie_s :[1] -2.149611 -3.908384 -4.214924 -4.290907
## effectif_de_seconde :[1] 0.02153071 0.03914674 0.04221707 0.04297812
## effectif_de_premiere :[1] -0.1760354 -0.3200643 -0.3451674 -0.3513898
## taux_acces_brut_seconde_bac :[1] 5.372220 9.767672 10.533764 10.723657
## taux_acces_attendu_seconde_bac :[1] -3.538743 -6.434078 -6.938712 -
7.063796
## taux_acces_brut_premiere_bac :[1] -10.19337 -18.53340 -19.98700 -20.34731
## taux_acces_attendu_premiere_bac :[1] 17.22171 31.31220 33.76806 34.37680
## taux_brut_de_reussite_total_series :[1] -2.696179 -4.902143 -5.286625 -
5.381927
## taux_reussite_attendu_total_series :[1] -2.035931 -3.701692 -3.992021 -
4.063985
## variance: 228523.2 182966 175025.7 173057.5
```

Comparaison au cas frequentiste :

```
modele_complet=lm(data$Barre~.,data=data[,6:ncol(data)])
ncol(data)

## [1] 23

summary(modele_complet)

##
## Call:
## lm(formula = data$Barre ~ ., data = data[, 6:ncol(data)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -429.72 -205.90 -122.25  -8.55 1645.96
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.725e+02   5.586e+02  -0.846   0.3980
## effectif_presents_serie_l    7.781e-01   1.638e+00   0.475   0.6351
## effectif_presents_serie_es    2.924e-01   1.234e+00   0.237   0.8128
## effectif_presents_serie_s    9.694e-03   1.019e+00   0.010   0.9924
## taux_brut_de_reussite_serie_l    3.122e+00   2.559e+00   1.220   0.2232
## taux_brut_de_reussite_serie_es    4.811e+00   4.205e+00   1.144   0.2531
## taux_brut_de_reussite_serie_s    9.385e+00   6.383e+00   1.470   0.1421
## taux_reussite_attendu_serie_l   -1.428e+01   6.879e+00  -2.077   0.0383
## taux_reussite_attendu_serie_es    3.814e+00   8.261e+00   0.462   0.6445
## taux_reussite_attendu_serie_s   -4.299e+00   9.586e+00  -0.448   0.6540
## effectif_de_seconde    4.306e-02   6.229e-01   0.069   0.9449
## effectif_de_premiere   -3.521e-01   7.182e-01  -0.490   0.6242
## taux_acces_brut_seconde_bac    1.074e+01   5.655e+00   1.900   0.0580
## taux_acces_attendu_seconde_bac  -7.077e+00   9.038e+00  -0.783   0.4340
```

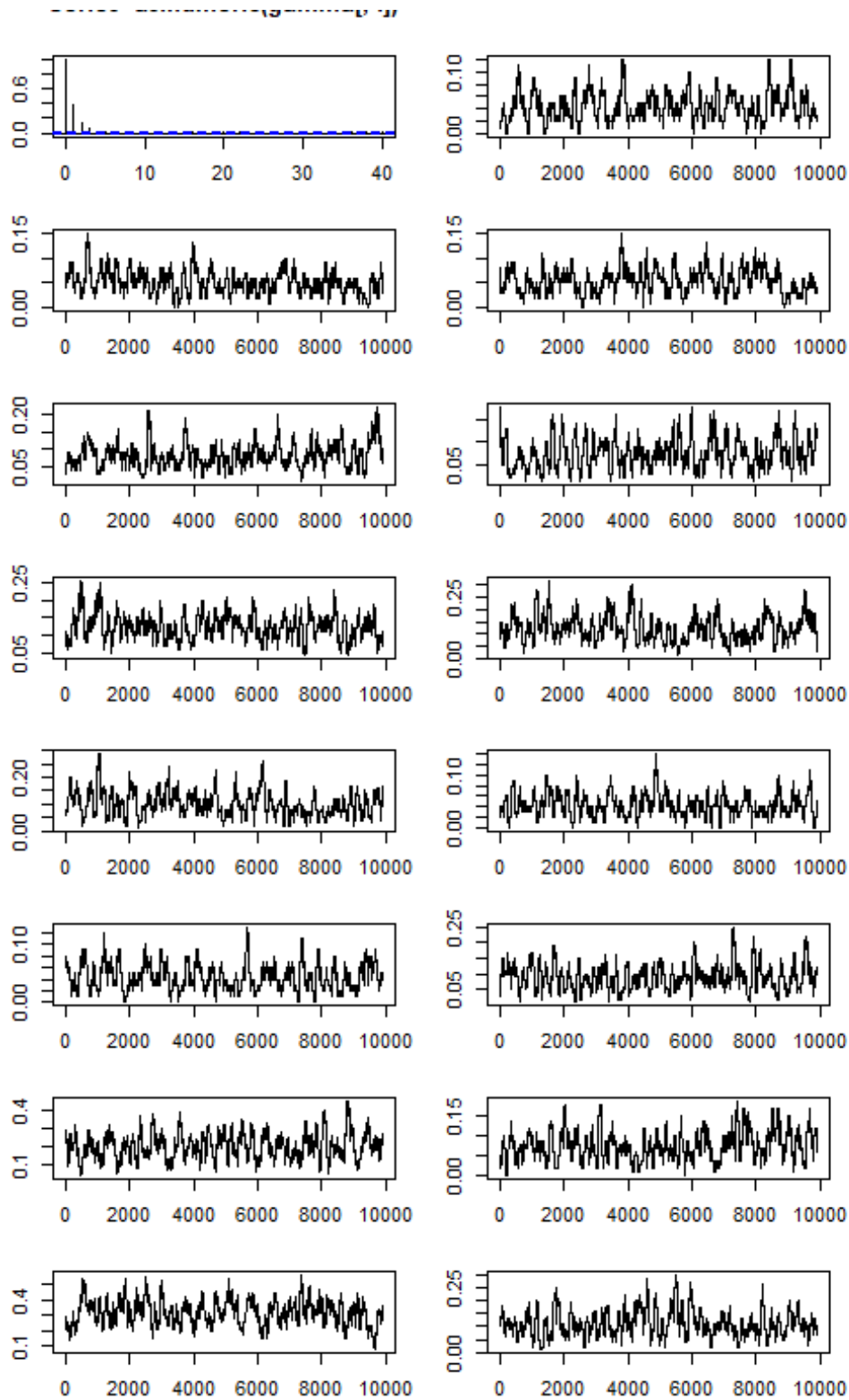
```
## taux_acces_brut_premiere_bac      -2.039e+01  1.071e+01  -1.904  0.0575
## taux_acces_attendu_premiere_bac    3.444e+01  1.916e+01   1.797  0.0729
## taux_brut_de_reussite_total_series -5.392e+00  1.288e+01  -0.419  0.6757
## taux_reussite_attendu_total_series -4.072e+00  2.202e+01  -0.185  0.8534
##
## (Intercept)
## effectif_presents_serie_l
## effectif_presents_serie_es
## effectif_presents_serie_s
## taux_brut_de_reussite_serie_l
## taux_brut_de_reussite_serie_es
## taux_brut_de_reussite_serie_s
## taux_reussite_attendu_serie_l      *
## taux_reussite_attendu_serie_es
## taux_reussite_attendu_serie_s
## effectif_de_seconde
## effectif_de_premiere
## taux_acces_brut_seconde_bac        .
## taux_acces_attendu_seconde_bac
## taux_acces_brut_premiere_bac       .
## taux_acces_attendu_premiere_bac    .
## taux_brut_de_reussite_total_series
## taux_reussite_attendu_total_series
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 422.4 on 498 degrees of freedom
## Multiple R-squared:  0.04068,    Adjusted R-squared:  0.007931
## F-statistic: 1.242 on 17 and 498 DF,  p-value: 0.2267
```

On remarque que quand  $g$  tend vers le nombre d'observations, les espérances des coefficients de régressions bayesiens convergent vers les coefficients de régression de la méthode fréquentiste (obtenus par optimisation du maximum de vraisemblance, ou de manière équivalente, minimisation des moindres carrés) et la variance de l'estimateur diminue

```
choix_modele_bayes(data)
```

```
## [1] 17
```



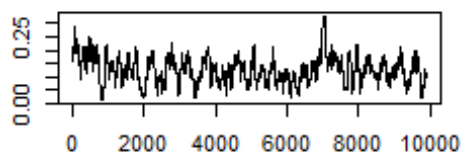


```
##          probtop50
## [1,] 0.126105263 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [2,] 0.092842105 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
## [3,] 0.048315789 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
```

```

## [4,] 0.044105263 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
## [5,] 0.037368421 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
## [6,] 0.036000000 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
## [7,] 0.027684211 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0
## [8,] 0.024736842 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
## [9,] 0.022631579 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
## [10,] 0.020105263 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [11,] 0.018947368 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [12,] 0.012210526 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1
## [13,] 0.009578947 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0
## [14,] 0.007894737 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
## [15,] 0.007894737 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1
## [16,] 0.006736842 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [17,] 0.006631579 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
## [18,] 0.006526316 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [19,] 0.006421053 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [20,] 0.006315789 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0
## [21,] 0.006315789 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0
## [22,] 0.006105263 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0
## [23,] 0.005894737 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0
## [24,] 0.005894737 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0
## [25,] 0.005684211 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0
## [26,] 0.005157895 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0
## [27,] 0.005157895 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0
## [28,] 0.005052632 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0
## [29,] 0.005052632 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0
## [30,] 0.004842105 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
## [31,] 0.004736842 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [32,] 0.004526316 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1
## [33,] 0.004315789 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0
## [34,] 0.004210526 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0
## [35,] 0.004105263 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0
## [36,] 0.004000000 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
## [37,] 0.003894737 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0
## [38,] 0.003789474 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
## [39,] 0.003789474 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0
## [40,] 0.003684211 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0
## [41,] 0.003684211 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0
## [42,] 0.003578947 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
## [43,] 0.003578947 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0
## [44,] 0.003473684 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
## [45,] 0.003473684 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0
## [46,] 0.003263158 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [47,] 0.003157895 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0
## [48,] 0.003052632 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0
## [49,] 0.003052632 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0
## [50,] 0.003052632 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1
## frequence apparition covariable : 0.04 0.04 0.08 0.06 0.08 0.06 0.08 0.12
0.08 0.04 0.04 0.08 0.34 0.08 0.34 0.1 0.1

```



On peut voir que l'autocorrélation décroît rapidement et que la chaîne de Markov explore bien toute la loi : l'algorithme de Gibbs semble avoir donc convergé

Résultat du choix de modèle bayésien toutes matières confondues : c'est un modèle avec uniquement la covariable 15 ("taux\_acces\_attendu\_premiere\_bac") qui arrive en premier avec un probabilité d'environ 12,6%

Comparaison au choix de modele frequentiste:

```
modele=step(lm(data$Barre~.,data=data[,6:ncol(data)]))

## Start: AIC=6257.03
## data$Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##   taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##   taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##   taux_reussite_attendu_serie_s + effectif_de_seconde +
effectif_de_premiere +
##   taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##   taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##   taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
##
```

	Df	Sum of Sq	RSS	AIC
## - effectif_presents_serie_s	1	16	88840876	6255.0
## - effectif_de_seconde	1	852	88841713	6255.0
## - taux_reussite_attendu_total_series	1	6098	88846958	6255.1
## - effectif_presents_serie_es	1	10012	88850872	6255.1
## - taux_brut_de_reussite_total_series	1	31248	88872109	6255.2
## - taux_reussite_attendu_serie_s	1	35882	88876742	6255.2
## - taux_reussite_attendu_serie_es	1	38030	88878890	6255.2
## - effectif_presents_serie_l	1	40236	88881096	6255.3
## - effectif_de_premiere	1	42865	88883725	6255.3
## - taux_acces_attendu_seconde_bac	1	109391	88950251	6255.7
## - taux_brut_de_reussite_serie_es	1	233522	89074382	6256.4
## - taux_brut_de_reussite_serie_l	1	265394	89106254	6256.6
## <none>			88840860	6257.0
## - taux_brut_de_reussite_serie_s	1	385706	89226567	6257.3
## - taux_acces_attendu_premiere_bac	1	576250	89417110	6258.4
## - taux_acces_brut_seconde_bac	1	644017	89484878	6258.8
## - taux_acces_brut_premiere_bac	1	646888	89487748	6258.8
## - taux_reussite_attendu_serie_l	1	769345	89610205	6259.5
##				



```
## Step: AIC=6255.03
## data$Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##   taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##   taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##   effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##   taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##   taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series +
##   taux_reussite_attendu_total_series
##
```

	Df	Sum of Sq	RSS	AIC
## - effectif_de_seconde	1	865	88841741	6253.0
## - taux_reussite_attendu_total_series	1	6193	88847069	6253.1
## - effectif_presents_serie_es	1	13489	88854365	6253.1
## - taux_brut_de_reussite_total_series	1	31796	88872672	6253.2
## - taux_reussite_attendu_serie_es	1	38045	88878922	6253.2
## - taux_reussite_attendu_serie_s	1	39332	88880208	6253.3
## - effectif_presents_serie_l	1	41800	88882677	6253.3
## - effectif_de_premiere	1	47941	88888817	6253.3
## - taux_acces_attendu_seconde_bac	1	109396	88950272	6253.7
## - taux_brut_de_reussite_serie_es	1	239195	89080071	6254.4
## - taux_brut_de_reussite_serie_l	1	284746	89125623	6254.7
## <none>			88840876	6255.0
## - taux_brut_de_reussite_serie_s	1	388299	89229176	6255.3
## - taux_acces_attendu_premiere_bac	1	594557	89435433	6256.5
## - taux_acces_brut_premiere_bac	1	648535	89489412	6256.8
## - taux_acces_brut_seconde_bac	1	652246	89493123	6256.8
## - taux_reussite_attendu_serie_l	1	785988	89626865	6257.6

```
## Step: AIC=6253.03
## data$Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##   taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##   taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##   effectif_de_premiere + taux_acces_brut_seconde_bac +
taux_acces_attendu_seconde_bac +
##   taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##   taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
```

	Df	Sum of Sq	RSS	AIC
## - taux_reussite_attendu_total_series	1	6227	88847968	6251.1
## - effectif_presents_serie_es	1	15416	88857157	6251.1
## - taux_reussite_attendu_serie_es	1	37271	88879012	6251.2
## - taux_brut_de_reussite_total_series	1	37979	88879721	6251.3
## - taux_reussite_attendu_serie_s	1	40401	88882142	6251.3
## - effectif_presents_serie_l	1	48080	88889821	6251.3
## - taux_acces_attendu_seconde_bac	1	118553	88960294	6251.7
## - effectif_de_premiere	1	134075	88975816	6251.8

```

## - taux_brut_de_reussite_serie_es      1      244004 89085745 6252.4
## - taux_brut_de_reussite_serie_l      1      294118 89135859 6252.7
## <none>                                88841741 6253.0
## - taux_brut_de_reussite_serie_s      1      403157 89244898 6253.4
## - taux_acces_attendu_premiere_bac    1      635847 89477589 6254.7
## - taux_acces_brut_seconde_bac       1      711945 89553686 6255.1
## - taux_acces_brut_premiere_bac      1      759053 89600794 6255.4
## - taux_reussite_attendu_serie_l      1      786303 89628045 6255.6
##
## Step: AIC=6251.07
## data$Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##      taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##      taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##      taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##      effectif_de_premiere + taux_acces_brut_seconde_bac +
taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##      taux_brut_de_reussite_total_series
##
##                               Df Sum of Sq      RSS      AIC
## - effectif_presents_serie_es      1      14162 88862130 6249.1
## - taux_reussite_attendu_serie_es    1      31749 88879717 6249.3
## - effectif_presents_serie_l        1      57217 88905185 6249.4
## - taux_brut_de_reussite_total_series 1      69607 88917574 6249.5
## - taux_reussite_attendu_serie_s     1     103349 88951317 6249.7
## - taux_acces_attendu_seconde_bac    1     116796 88964764 6249.7
## - effectif_de_premiere              1     137175 88985143 6249.9
## <none>                             88847968 6251.1
## - taux_brut_de_reussite_serie_l     1     360537 89208505 6251.2
## - taux_brut_de_reussite_serie_es     1     369900 89217868 6251.2
## - taux_brut_de_reussite_serie_s     1     452319 89300287 6251.7
## - taux_acces_brut_seconde_bac       1     727783 89575751 6253.3
## - taux_acces_brut_premiere_bac      1     754145 89602113 6253.4
## - taux_acces_attendu_premiere_bac    1     769511 89617479 6253.5
## - taux_reussite_attendu_serie_l      1    1369102 90217070 6257.0
##
## Step: AIC=6249.15
## data$Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##      taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##      taux_reussite_attendu_serie_s + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series
##
##                               Df Sum of Sq      RSS      AIC
## - taux_reussite_attendu_serie_es     1      26069 88888198 6247.3
## - effectif_presents_serie_l          1      54065 88916195 6247.5
## - taux_brut_de_reussite_total_series  1      66785 88928915 6247.5
## - taux_reussite_attendu_serie_s      1     106010 88968140 6247.8

```

```

## - taux_acces_attendu_seconde_bac      1    115309 88977439 6247.8
## - effectif_de_premiere                 1    148299 89010428 6248.0
## <none>                                88862130 6249.1
## - taux_brut_de_reussite_serie_l       1    346484 89208614 6249.2
## - taux_brut_de_reussite_serie_es      1    361321 89223451 6249.2
## - taux_brut_de_reussite_serie_s       1    442312 89304441 6249.7
## - taux_acces_brut_seconde_bac         1    716539 89578669 6251.3
## - taux_acces_brut_premiere_bac        1    747339 89609469 6251.5
## - taux_acces_attendu_premiere_bac     1    846319 89708448 6252.0
## - taux_reussite_attendu_serie_l       1   1357447 90219577 6255.0
##
## Step:  AIC=6247.3
## data$Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##      taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_s +
##      effectif_de_premiere + taux_acces_brut_seconde_bac +
taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##      taux_brut_de_reussite_total_series
##
##              Df Sum of Sq      RSS      AIC
## - effectif_presents_serie_l      1      55457 88943656 6245.6
## - taux_brut_de_reussite_total_series 1      80406 88968604 6245.8
## - taux_reussite_attendu_serie_s    1      81899 88970097 6245.8
## - taux_acces_attendu_seconde_bac   1     105994 88994192 6245.9
## - effectif_de_premiere             1     148972 89037170 6246.2
## - taux_brut_de_reussite_serie_l    1     341460 89229658 6247.3
## <none>                             88888198 6247.3
## - taux_brut_de_reussite_serie_s    1     436786 89324985 6247.8
## - taux_brut_de_reussite_serie_es   1     469446 89357645 6248.0
## - taux_acces_brut_seconde_bac      1     698500 89586698 6249.3
## - taux_acces_brut_premiere_bac     1     721678 89609876 6249.5
## - taux_acces_attendu_premiere_bac  1     914637 89802835 6250.6
## - taux_reussite_attendu_serie_l    1    1348477 90236675 6253.1
##
## Step:  AIC=6245.62
## data$Barre ~ taux_brut_de_reussite_serie_l +
taux_brut_de_reussite_serie_es +
##      taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##      taux_reussite_attendu_serie_s + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series
##
##              Df Sum of Sq      RSS      AIC
## - taux_brut_de_reussite_total_series 1      80103 89023759 6244.1
## - effectif_de_premiere             1     102654 89046310 6244.2
## - taux_acces_attendu_seconde_bac    1     105438 89049094 6244.2
## - taux_reussite_attendu_serie_s     1     110788 89054443 6244.3
## <none>                             88943656 6245.6

```

```

## - taux_brut_de_reussite_serie_l      1      354498 89298153 6245.7
## - taux_brut_de_reussite_serie_s      1      411465 89355121 6246.0
## - taux_brut_de_reussite_serie_es      1      444436 89388092 6246.2
## - taux_acces_brut_seconde_bac         1      657684 89601340 6247.4
## - taux_acces_brut_premiere_bac        1      698047 89641702 6247.7
## - taux_acces_attendu_premiere_bac     1     1015616 89959272 6249.5
## - taux_reussite_attendu_serie_l       1     1328732 90272388 6251.3
##
## Step: AIC=6244.09
## data$Barre ~ taux_brut_de_reussite_serie_l +
taux_brut_de_reussite_serie_es +
##      taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##      taux_reussite_attendu_serie_s + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac
##
##
##              Df Sum of Sq      RSS      AIC
## - taux_reussite_attendu_serie_s      1      73327 89097086 6242.5
## - effectif_de_premiere                1      89797 89113556 6242.6
## - taux_acces_attendu_seconde_bac      1     102883 89126642 6242.7
## - taux_brut_de_reussite_serie_l      1     276792 89300551 6243.7
## <none>                                89023759 6244.1
## - taux_brut_de_reussite_serie_s      1     365827 89389586 6244.2
## - taux_brut_de_reussite_serie_es      1     374522 89398281 6244.3
## - taux_acces_brut_seconde_bac         1     680453 89704212 6246.0
## - taux_acces_brut_premiere_bac        1     928129 89951888 6247.4
## - taux_acces_attendu_premiere_bac     1     935917 89959676 6247.5
## - taux_reussite_attendu_serie_l      1    1288056 90311815 6249.5
##
## Step: AIC=6242.51
## data$Barre ~ taux_brut_de_reussite_serie_l +
taux_brut_de_reussite_serie_es +
##      taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##      effectif_de_premiere + taux_acces_brut_seconde_bac +
taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##
##              Df Sum of Sq      RSS      AIC
## - effectif_de_premiere                1      76781 89173867 6241.0
## - taux_acces_attendu_seconde_bac      1      82820 89179906 6241.0
## - taux_brut_de_reussite_serie_l      1     218004 89315090 6241.8
## - taux_brut_de_reussite_serie_s      1     294818 89391904 6242.2
## - taux_brut_de_reussite_serie_es      1     317631 89414717 6242.3
## <none>                                89097086 6242.5
## - taux_acces_brut_seconde_bac         1     640439 89737525 6244.2
## - taux_acces_brut_premiere_bac        1     859668 89956754 6245.5
## - taux_acces_attendu_premiere_bac     1     910795 90007881 6245.8
## - taux_reussite_attendu_serie_l      1    1410562 90507648 6248.6
##

```

```

## Step: AIC=6240.96
## data$Barre ~ taux_brut_de_reussite_serie_l +
taux_brut_de_reussite_serie_es +
##      taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##      taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## - taux_acces_attendu_seconde_bac    1      76775 89250642 6239.4
## - taux_brut_de_reussite_serie_l      1     234055 89407922 6240.3
## - taux_brut_de_reussite_serie_s      1     259466 89433332 6240.5
## - taux_brut_de_reussite_serie_es     1     291185 89465052 6240.6
## <none>                                89173867 6241.0
## - taux_acces_brut_seconde_bac        1     563976 89737842 6242.2
## - taux_acces_brut_premiere_bac        1     788549 89962415 6243.5
## - taux_acces_attendu_premiere_bac     1     880326 90054193 6244.0
## - taux_reussite_attendu_serie_l       1    1407485 90581352 6247.0
##
## Step: AIC=6239.4
## data$Barre ~ taux_brut_de_reussite_serie_l +
taux_brut_de_reussite_serie_es +
##      taux_brut_de_reussite_serie_s + taux_reussite_attendu_serie_l +
##      taux_acces_brut_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## - taux_brut_de_reussite_serie_l      1     221048 89471690 6238.7
## - taux_brut_de_reussite_serie_s      1     221840 89472482 6238.7
## - taux_brut_de_reussite_serie_es     1     307546 89558188 6239.2
## <none>                                89250642 6239.4
## - taux_acces_brut_seconde_bac        1     534092 89784734 6240.5
## - taux_acces_brut_premiere_bac        1     713231 89963873 6241.5
## - taux_acces_attendu_premiere_bac     1    1306414 90557056 6244.9
## - taux_reussite_attendu_serie_l       1    1333142 90583784 6245.1
##
## Step: AIC=6238.68
## data$Barre ~ taux_brut_de_reussite_serie_es +
taux_brut_de_reussite_serie_s +
##      taux_reussite_attendu_serie_l + taux_acces_brut_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## - taux_brut_de_reussite_serie_s      1     267456 89739145 6238.2
## - taux_brut_de_reussite_serie_es     1     287265 89758955 6238.3
## <none>                                89471690 6238.7
## - taux_acces_brut_seconde_bac        1     437670 89909360 6239.2
## - taux_acces_brut_premiere_bac        1     528829 90000519 6239.7
## - taux_acces_attendu_premiere_bac     1    1125625 90597315 6243.1
## - taux_reussite_attendu_serie_l       1    1162829 90634519 6243.3
##

```

```
## Step: AIC=6238.22
## data$Barre ~ taux_brut_de_reussite_serie_es +
taux_reussite_attendu_serie_l +
##      taux_acces_brut_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## - taux_brut_de_reussite_serie_es    1    178959 89918104 6237.2
## - taux_acces_brut_premiere_bac      1    286408 90025554 6237.9
## - taux_acces_brut_seconde_bac      1    301845 90040990 6237.9
## <none>                                89739145 6238.2
## - taux_reussite_attendu_serie_l      1    944393 90683538 6241.6
## - taux_acces_attendu_premiere_bac    1   1274207 91013352 6243.5
##
## Step: AIC=6237.24
## data$Barre ~ taux_reussite_attendu_serie_l + taux_acces_brut_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## - taux_acces_brut_premiere_bac      1    184758 90102863 6236.3
## - taux_acces_brut_seconde_bac      1    318211 90236315 6237.1
## <none>                                89918104 6237.2
## - taux_reussite_attendu_serie_l      1    840123 90758227 6240.0
## - taux_acces_attendu_premiere_bac    1   1283071 91201175 6242.6
##
## Step: AIC=6236.3
## data$Barre ~ taux_reussite_attendu_serie_l + taux_acces_brut_seconde_bac +
##      taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## - taux_acces_brut_seconde_bac      1    154234 90257096 6235.2
## <none>                                90102863 6236.3
## - taux_reussite_attendu_serie_l      1    693957 90796819 6238.3
## - taux_acces_attendu_premiere_bac    1   1254445 91357307 6241.4
##
## Step: AIC=6235.19
## data$Barre ~ taux_reussite_attendu_serie_l +
taux_acces_attendu_premiere_bac
##
##              Df Sum of Sq      RSS      AIC
## <none>                                90257096 6235.2
## - taux_reussite_attendu_serie_l      1    574942 90832038 6236.5
## - taux_acces_attendu_premiere_bac    1    1913707 92170804 6244.0

summary(modele)

##
## Call:
## lm(formula = data$Barre ~ taux_reussite_attendu_serie_l +
taux_acces_attendu_premiere_bac,
```

```
##      data = data[, 6:ncol(data)])
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -387.32 -196.56 -130.83  -14.95 1696.20
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -494.324     260.593   -1.897   0.05840 .
## taux_reussite_attendu_serie_l    -7.882       4.360   -1.808   0.07124 .
## taux_acces_attendu_premiere_bac  17.833       5.407    3.298   0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 419.5 on 513 degrees of freedom
## Multiple R-squared:  0.02539,    Adjusted R-squared:  0.02159
## F-statistic: 6.681 on 2 and 513 DF,  p-value: 0.001366
```

résultat choix de modèle fréquentiste : 2 covariables sont retenues : “taux\_reussite\_attendu\_serie\_l” et “taux\_acces\_attendu\_premiere\_bac”

Le choix de modèle fréquentiste diffère donc du choix de modèle bayésien, même si on retrouve la covariable “taux\_acces\_attendu\_premiere\_bac” dans les 2 cas.

### 3) Etude des matières Mathématiques et Anglais

```
data_maths=data[data$Matiere=="MATHS",]
data_anglais=data[data$Matiere=="ANGLAIS",]
data_maths_anglais=data[data$Matiere=="ANGLAIS" | data$Matiere=="MATHS" ,]

reg_lin_bayes(data_maths)

## Intercept) :[1] 32.89372 59.80676 64.49749 64.69098
## effectif_presents_serie_l :[1] 0.2030316 0.3691484 0.3981012 0.3992955
## effectif_presents_serie_es :[1] -0.1929207 -0.3507650 -0.3782760 -
0.3794108
## effectif_presents_serie_s :[1] 0.5860546 1.0655539 1.1491267 1.1525741
## taux_brut_de_reussite_serie_l :[1] -1.834131 -3.334783 -3.596334 -3.607124
## taux_brut_de_reussite_serie_es :[1] 3.623369 6.587944 7.104646 7.125960
## taux_brut_de_reussite_serie_s :[1] 0.7477495 1.3595446 1.4661756 1.4705741
## taux_reussite_attendu_serie_l :[1] -2.063615 -3.752027 -4.046304 -4.058443
## taux_reussite_attendu_serie_es :[1] -3.100392 -5.637077 -6.079201 -
6.097439
## taux_reussite_attendu_serie_s :[1] -1.027492 -1.868167 -2.014690 -2.020734
## effectif_de_seconde :[1] 0.4017633 0.7304787 0.7877711 0.7901345
## effectif_de_premiere :[1] -0.7020843 -1.2765169 -1.3766358 -1.3807657
## taux_acces_brut_seconde_bac :[1] 6.548822 11.906949 12.840827 12.879350
## taux_acces_attendu_seconde_bac :[1] -4.615612 -8.392022 -9.050220 -
9.077371
## taux_acces_brut_premiere_bac :[1] -6.298805 -11.452373 -12.350599 -
12.387650
```

```
## taux_acces_attendu_premiere_bac :[1] 2.820641 5.128439 5.530669 5.547261
## taux_brut_de_reussite_total_series :[1] 0.7275528 1.3228233 1.4265741
1.4308539
## taux_reussite_attendu_total_series :[1] 5.660323 10.291496 11.098672
11.131968
## variance: 31048.99 13625.3 10588.49 10463.23

modele_complet_maths=lm(Barre~.,data=data_maths[,6:ncol(data_maths)])
#ncol(data_maths)
summary(modele_complet_maths)

##
## Call:
## lm(formula = Barre ~ ., data = data_maths[, 6:ncol(data_maths)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -173.08  -69.00  -21.11   75.56  273.03
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      65.7874    489.3234   0.134   0.8937
## effectif_presents_serie_l      0.4061     1.2463   0.326   0.7462
## effectif_presents_serie_es     -0.3858     0.9616  -0.401   0.6903
## effectif_presents_serie_s      1.1721     0.8454   1.386   0.1731
## taux_brut_de_reussite_serie_l    -3.6683     2.1240  -1.727   0.0917 .
## taux_brut_de_reussite_serie_es     7.2467     4.1641   1.740   0.0893 .
## taux_brut_de_reussite_serie_s      1.4955     5.5932   0.267   0.7905
## taux_reussite_attendu_serie_l     -4.1272     5.6497  -0.731   0.4692
## taux_reussite_attendu_serie_es    -6.2008     7.1700  -0.865   0.3922
## taux_reussite_attendu_serie_s     -2.0550     8.3611  -0.246   0.8071
## effectif_de_seconde      0.8035     0.5236   1.535   0.1325
## effectif_de_premiere     -1.4042     0.6320  -2.222   0.0319 *
## taux_acces_brut_seconde_bac     13.0976     6.2043   2.111   0.0409 *
## taux_acces_attendu_seconde_bac    -9.2312     7.2998  -1.265   0.2132
## taux_acces_brut_premiere_bac    -12.5976    10.4840  -1.202   0.2364
## taux_acces_attendu_premiere_bac     5.6413    14.9814   0.377   0.7084
## taux_brut_de_reussite_total_series  1.4551    11.0152   0.132   0.8956
## taux_reussite_attendu_total_series 11.3206    17.7198   0.639   0.5265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 116.4 on 41 degrees of freedom
## Multiple R-squared:  0.3826, Adjusted R-squared:  0.1267
## F-statistic: 1.495 on 17 and 41 DF,  p-value: 0.1451
```

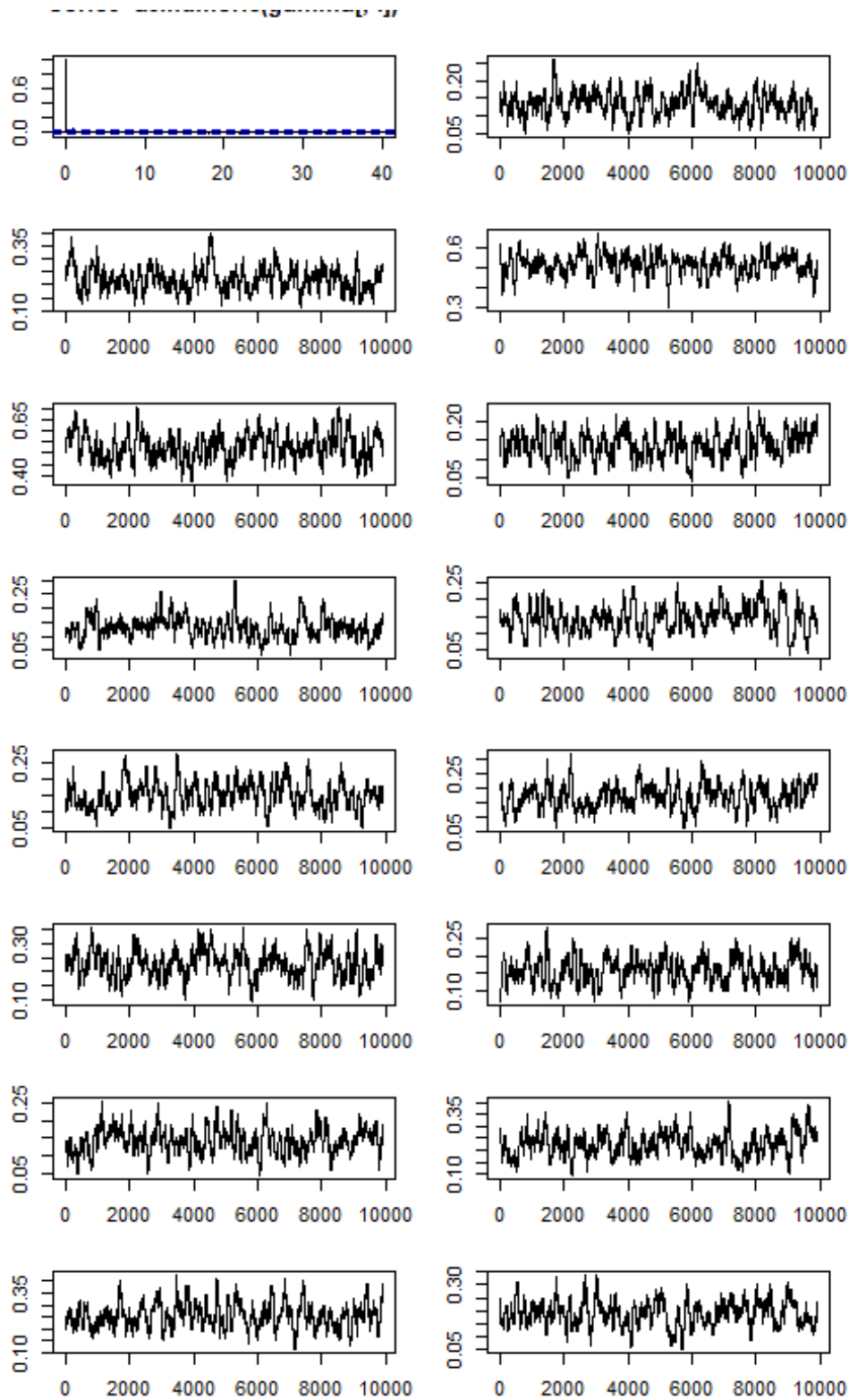
Interpretation des coefficients : Verification algo converge

```
choix_modele_bayes(data_maths)
```

```
## [1] 17
```





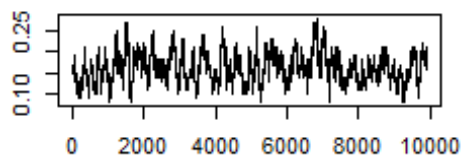


```
##          probtop50
## [1,] 0.023368421 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [2,] 0.021789474 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
## [3,] 0.012842105 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0
```

```

## [4,] 0.008736842 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0
## [5,] 0.008526316 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0
## [6,] 0.008315789 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [7,] 0.007684211 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0
## [8,] 0.006526316 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
## [9,] 0.006526316 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0
## [10,] 0.005894737 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [11,] 0.005263158 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0
## [12,] 0.004947368 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0
## [13,] 0.004736842 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0
## [14,] 0.004526316 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1
## [15,] 0.004421053 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0
## [16,] 0.004315789 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
## [17,] 0.004105263 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0
## [18,] 0.004000000 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [19,] 0.004000000 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
## [20,] 0.004000000 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
## [21,] 0.004000000 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0
## [22,] 0.003894737 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
## [23,] 0.003894737 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0
## [24,] 0.003894737 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
## [25,] 0.003789474 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0
## [26,] 0.003578947 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
## [27,] 0.003578947 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
## [28,] 0.003473684 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
## [29,] 0.003368421 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [30,] 0.003368421 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0
## [31,] 0.003368421 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0
## [32,] 0.003368421 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1
## [33,] 0.003263158 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0
## [34,] 0.003157895 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0
## [35,] 0.003157895 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0
## [36,] 0.003052632 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
## [37,] 0.002947368 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
## [38,] 0.002947368 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0
## [39,] 0.002842105 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 0
## [40,] 0.002736842 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [41,] 0.002526316 0 0 1 1 1 0 0 0 0 0 1 0 0 0 0 0 0
## [42,] 0.002526316 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0
## [43,] 0.002526316 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 1
## [44,] 0.002421053 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0
## [45,] 0.002421053 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0
## [46,] 0.002315789 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
## [47,] 0.002315789 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
## [48,] 0.002210526 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0
## [49,] 0.002210526 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1
## [50,] 0.002105263 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
## frequence apparition covariable : 0.04 0.04 0.1 0.48 0.68 0.06 0.04 0.06
0.06 0.06 0.1 0.06 0.04 0.12 0.1 0.12 0.1

```



L'autocorrélogramme décroît vers 0 et la chaîne explore toute la loi.

Résultat du choix de modèle bayésien pour les maths : un modèle avec uniquement la covariable 5 ("taux\_brut\_de\_reussite\_series\_es") et une probabilité de 2.8%

```
modele_maths=step(lm(Barre~.,data=data_maths[,6:ncol(data_maths)]))

## Start:  AIC=575.9
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##   taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##   taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##   taux_reussite_attendu_serie_s + effectif_de_seconde +
effectif_de_premiere +
##   taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##   taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##   taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
##               Df Sum of Sq   RSS   AIC
## - taux_brut_de_reussite_total_series  1      237 556179 573.93
## - taux_reussite_attendu_serie_s      1      819 556761 573.99
## - taux_brut_de_reussite_serie_s      1      969 556912 574.00
## - effectif_presents_serie_l          1     1439 557381 574.05
## - taux_acces_attendu_premiere_bac     1     1923 557865 574.11
## - effectif_presents_serie_es          1     2183 558125 574.13
## - taux_reussite_attendu_total_series  1     5534 561477 574.49
## - taux_reussite_attendu_serie_l       1     7236 563179 574.67
## - taux_reussite_attendu_serie_es      1    10142 566084 574.97
## <none>                                555942 575.90
## - taux_acces_brut_premiere_bac        1     19578 575520 575.94
## - taux_acces_attendu_seconde_bac      1     21684 577626 576.16
## - effectif_presents_serie_s           1     26066 582009 576.61
## - effectif_de_seconde                 1     31939 587881 577.20
## - taux_brut_de_reussite_serie_l       1     40444 596386 578.05
## - taux_brut_de_reussite_serie_es      1     41066 597008 578.11
## - taux_acces_brut_seconde_bac         1     60430 616372 579.99
## - effectif_de_premiere                 1     66945 622887 580.61
##
## Step:  AIC=573.93
```

```

## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##   taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##   taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##   taux_reussite_attendu_serie_s + effectif_de_seconde +
effectif_de_premiere +
##   taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##   taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##   taux_reussite_attendu_total_series
##
##
##           Df Sum of Sq    RSS    AIC
## - taux_reussite_attendu_serie_s      1      1108 557287 572.04
## - effectif_presents_serie_l          1      1407 557586 572.08
## - taux_acces_attendu_premiere_bac     1      1942 558121 572.13
## - effectif_presents_serie_es          1      2092 558271 572.15
## - taux_brut_de_reussite_serie_s       1      2837 559016 572.23
## - taux_reussite_attendu_total_series   1       7953 564131 572.76
## - taux_reussite_attendu_serie_l        1      8352 564530 572.81
## - taux_reussite_attendu_serie_es       1     14733 570912 573.47
## <none>                                556179 573.93
## - taux_acces_attendu_seconde_bac      1      21977 578156 574.21
## - taux_acces_brut_premiere_bac        1      24179 580358 574.44
## - effectif_presents_serie_s           1      26048 582227 574.63
## - effectif_de_seconde                 1      33240 589419 575.35
## - taux_brut_de_reussite_serie_l       1      42255 598433 576.25
## - taux_acces_brut_seconde_bac         1      60204 616383 577.99
## - taux_brut_de_reussite_serie_es      1      60525 616703 578.02
## - effectif_de_premiere                1      68211 624390 578.75
##
## Step:  AIC=572.04
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##   taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##   taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##   effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##   taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##   taux_acces_attendu_premiere_bac + taux_reussite_attendu_total_series
##
##
##           Df Sum of Sq    RSS    AIC
## - taux_acces_attendu_premiere_bac     1       1529 558816 570.21
## - taux_brut_de_reussite_serie_s       1       1814 559100 570.24
## - effectif_presents_serie_l           1       1837 559124 570.24
## - effectif_presents_serie_es          1       2767 560054 570.34
## - taux_reussite_attendu_total_series   1       7022 564309 570.78
## - taux_reussite_attendu_serie_l        1       7271 564558 570.81
## - taux_reussite_attendu_serie_es       1      15536 572823 571.67
## <none>                                557287 572.04
## - taux_acces_attendu_seconde_bac      1      21900 579187 572.32
## - taux_acces_brut_premiere_bac        1      23351 580638 572.47

```

```

## - effectif_de_seconde          1      32228 589515 573.36
## - effectif_presents_serie_s    1      39755 597042 574.11
## - taux_brut_de_reussite_serie_l 1      59333 616620 576.01
## - taux_acces_brut_seconde_bac  1      61330 618617 576.20
## - taux_brut_de_reussite_serie_es 1      61952 619239 576.26
## - effectif_de_premiere         1      69157 626443 576.95
##
## Step: AIC=570.21
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##          effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##          taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##          taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##          effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##          taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##          taux_reussite_attendu_total_series
##
##                               Df Sum of Sq    RSS    AIC
## - taux_brut_de_reussite_serie_s    1      961 559776 568.31
## - effectif_presents_serie_l        1     2103 560919 568.43
## - effectif_presents_serie_es       1     3157 561973 568.54
## - taux_reussite_attendu_serie_l    1     9449 568264 569.20
## - taux_reussite_attendu_serie_es   1    17098 575914 569.98
## <none>                                558816 570.21
## - taux_reussite_attendu_total_series 1     21363 580179 570.42
## - taux_acces_brut_premiere_bac     1     21823 580639 570.47
## - taux_acces_attendu_seconde_bac   1     22795 581610 570.57
## - effectif_de_seconde              1     40576 599392 572.34
## - effectif_presents_serie_s        1     45803 604619 572.85
## - taux_acces_brut_seconde_bac      1     60943 619759 574.31
## - taux_brut_de_reussite_serie_es   1     61171 619987 574.34
## - taux_brut_de_reussite_serie_l    1     61515 620331 574.37
## - effectif_de_premiere              1     84418 643234 576.51
##
## Step: AIC=568.31
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##          effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##          taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_l +
##          taux_reussite_attendu_serie_es + effectif_de_seconde +
effectif_de_premiere +
##          taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##          taux_acces_brut_premiere_bac + taux_reussite_attendu_total_series
##
##                               Df Sum of Sq    RSS    AIC
## - effectif_presents_serie_l        1     1809 561586 566.50
## - effectif_presents_serie_es       1     4165 563941 566.74
## - taux_reussite_attendu_serie_l    1    10769 570546 567.43
## - taux_reussite_attendu_serie_es   1    18138 577914 568.19
## <none>                                559776 568.31
## - taux_acces_brut_premiere_bac     1    20950 580727 568.48

```

```

## - taux_acces_attendu_seconde_bac      1      22172 581949 568.60
## - taux_reussite_attendu_total_series  1      30849 590626 569.47
## - effectif_de_seconde                  1      39616 599393 570.34
## - effectif_presents_serie_s            1      45566 605342 570.92
## - taux_acces_brut_seconde_bac          1      60050 619826 572.32
## - taux_brut_de_reussite_serie_l        1      60567 620343 572.37
## - taux_brut_de_reussite_serie_es       1      62481 622257 572.55
## - effectif_de_premiere                  1      85953 645729 574.74
##
## Step:  AIC=566.5
## Barre ~ effectif_presents_serie_es + effectif_presents_serie_s +
##      taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##      taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##      effectif_de_seconde + effectif_de_premiere +
##      taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_reussite_attendu_total_series
##
##              Df Sum of Sq    RSS    AIC
## - effectif_presents_serie_es      1      4043 565629 564.92
## - taux_reussite_attendu_serie_l    1     10170 571755 565.56
## - taux_reussite_attendu_serie_es   1     17296 578882 566.29
## <none>                                561586 566.50
## - taux_acces_brut_premiere_bac     1     19827 581413 566.55
## - taux_acces_attendu_seconde_bac   1     20674 582260 566.63
## - taux_reussite_attendu_total_series 1     29797 591383 567.55
## - effectif_presents_serie_s        1     43757 605343 568.92
## - effectif_de_seconde              1     44174 605760 568.97
## - taux_acces_brut_seconde_bac      1     58495 620080 570.34
## - taux_brut_de_reussite_serie_l    1     60965 622551 570.58
## - taux_brut_de_reussite_serie_es   1     61523 623108 570.63
## - effectif_de_premiere              1     85258 646844 572.84
##
## Step:  AIC=564.92
## Barre ~ effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_l +
##      taux_reussite_attendu_serie_es + effectif_de_seconde +
##      effectif_de_premiere +
##      taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_reussite_attendu_total_series
##
##              Df Sum of Sq    RSS    AIC
## - taux_reussite_attendu_serie_l    1      9224 574854 563.88
## - taux_reussite_attendu_serie_es   1     16284 581913 564.60
## - taux_acces_brut_premiere_bac     1     18155 583784 564.79
## <none>                                565629 564.92
## - taux_acces_attendu_seconde_bac   1     21462 587091 565.12
## - taux_reussite_attendu_total_series 1     27485 593114 565.72
## - effectif_presents_serie_s        1     41639 607268 567.11
## - effectif_de_seconde              1     43637 609266 567.31

```

```

## - taux_acces_brut_seconde_bac      1      57562 623191 568.64
## - taux_brut_de_reussite_serie_l    1      60161 625790 568.88
## - taux_brut_de_reussite_serie_es   1      67512 633141 569.57
## - effectif_de_premiere              1      90217 655846 571.65
##
## Step:  AIC=563.88
## Barre ~ effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_es +
##      effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_reussite_attendu_total_series
##
##                               Df Sum of Sq    RSS    AIC
## - taux_acces_brut_premiere_bac      1      11320 586174 563.03
## - taux_reussite_attendu_serie_es     1      15247 590100 563.42
## - taux_acces_attendu_seconde_bac     1      17365 592218 563.63
## - taux_reussite_attendu_total_series  1      18297 593151 563.72
## <none>                                574854 563.88
## - effectif_presents_serie_s          1      40343 615196 565.88
## - effectif_de_seconde                 1      40414 615267 565.88
## - taux_acces_brut_seconde_bac        1      48608 623462 566.66
## - taux_brut_de_reussite_serie_es     1      61109 635963 567.84
## - taux_brut_de_reussite_serie_l      1      70381 645235 568.69
## - effectif_de_premiere                1      85121 659975 570.02
##
## Step:  AIC=563.03
## Barre ~ effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_es +
##      effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_reussite_attendu_total_series
##
##                               Df Sum of Sq    RSS    AIC
## - taux_reussite_attendu_total_series  1       9542 595716 561.98
## - taux_reussite_attendu_serie_es     1      10590 596763 562.08
## - taux_acces_attendu_seconde_bac     1      10631 596804 562.09
## <none>                                586174 563.03
## - effectif_de_seconde                 1      30816 616990 564.05
## - effectif_presents_serie_s          1      32617 618790 564.22
## - taux_acces_brut_seconde_bac        1      43613 629787 565.26
## - taux_brut_de_reussite_serie_es     1      49810 635983 565.84
## - effectif_de_premiere                1      74972 661146 568.13
## - taux_brut_de_reussite_serie_l      1      81238 667412 568.68
##
## Step:  AIC=561.98
## Barre ~ effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_es +
##      effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +

```



```

##      taux_acces_attendu_seconde_bac
##
##              Df Sum of Sq    RSS    AIC
## - taux_reussite_attendu_serie_es  1      2322 598038 560.21
## - taux_acces_attendu_seconde_bac  1      3867 599583 560.36
## <none>                                595716 561.98
## - effectif_de_seconde                1     33600 629317 563.22
## - taux_acces_brut_seconde_bac        1     40070 635786 563.82
## - effectif_presents_serie_s          1     40832 636548 563.89
## - taux_brut_de_reussite_serie_es     1     41551 637268 563.96
## - taux_brut_de_reussite_serie_l      1     72530 668246 566.76
## - effectif_de_premiere                1     85868 681584 567.92
##
## Step:  AIC=560.21
## Barre ~ effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + effectif_de_seconde +
effectif_de_premiere +
##      taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac
##
##              Df Sum of Sq    RSS    AIC
## - taux_acces_attendu_seconde_bac  1     10570 608608 559.24
## <none>                                598038 560.21
## - effectif_de_seconde                1     32156 630194 561.30
## - effectif_presents_serie_s          1     38998 637036 561.94
## - taux_acces_brut_seconde_bac        1     39260 637298 561.96
## - taux_brut_de_reussite_serie_es     1     41177 639215 562.14
## - taux_brut_de_reussite_serie_l      1     79602 677640 565.58
## - effectif_de_premiere                1     83570 681608 565.93
##
## Step:  AIC=559.24
## Barre ~ effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + effectif_de_seconde +
effectif_de_premiere +
##      taux_acces_brut_seconde_bac
##
##              Df Sum of Sq    RSS    AIC
## <none>                                608608 559.24
## - effectif_presents_serie_s          1     29049 637656 559.99
## - taux_acces_brut_seconde_bac        1     30328 638935 560.11
## - effectif_de_seconde                1     33166 641773 560.37
## - taux_brut_de_reussite_serie_es     1     35732 644340 560.61
## - taux_brut_de_reussite_serie_l      1     75541 684148 564.15
## - effectif_de_premiere                1     77694 686302 564.33

summary(modele_maths)

##
## Call:
## lm(formula = Barre ~ effectif_presents_serie_s +
taux_brut_de_reussite_serie_l +

```

```
##      taux_brut_de_reussite_serie_es + effectif_de_seconde +
effectif_de_premiere +
##      taux_acces_brut_seconde_bac, data = data_maths[, 6:ncol(data_maths)])
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -222.62   -73.59   -20.98    54.23   292.99
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -77.4081    182.6176  -0.424   0.6734
## effectif_presents_serie_s      0.7892     0.5009   1.575   0.1212
## taux_brut_de_reussite_serie_l  -3.9693     1.5624  -2.541   0.0141 *
## taux_brut_de_reussite_serie_es  3.8648     2.2119   1.747   0.0865 .
## effectif_de_seconde           0.6536     0.3883   1.683   0.0983 .
## effectif_de_premiere          -1.1044     0.4286  -2.576   0.0129 *
## taux_acces_brut_seconde_bac    4.2086     2.6145   1.610   0.1135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 108.2 on 52 degrees of freedom
## Multiple R-squared:  0.3242, Adjusted R-squared:  0.2462
## F-statistic: 4.157 on 6 and 52 DF,  p-value: 0.001744
```

resultat choix de modele frequentiste pour les maths : bien plus de covariables retenues

effectif\_presents\_serie\_s taux\_brut\_de\_reussite\_serie\_l taux\_reussite\_attendu\_serie\_es  
effectif\_de\_seconde effectif\_de\_premiere taux\_acces\_brut\_seconde\_bac

**reg\_lin\_bayes**(data\_anglais)

```
## Intercept) :[1] -643.2191 -1169.4893 -1261.2139 -1262.1658
## effectif_presents_serie_l :[1] -2.621737 -4.766795 -5.140661 -5.144541
## effectif_presents_serie_es :[1] -0.8269082 -1.5034694 -1.6213886 -
1.6226123
## effectif_presents_serie_s :[1] 0.3955658 0.7192106 0.7756193 0.7762046
## taux_brut_de_reussite_serie_l :[1] -2.557733 -4.650424 -5.015164 -5.018949
## taux_brut_de_reussite_serie_es :[1] 2.997441 5.449892 5.877335 5.881770
## taux_brut_de_reussite_serie_s :[1] 1.501459 2.729925 2.944037 2.946259
## taux_reussite_attendu_serie_l :[1] -1.894983 -3.445423 -3.715653 -3.718457
## taux_reussite_attendu_serie_es :[1] 15.46847 28.12450 30.33034 30.35323
## taux_reussite_attendu_serie_s :[1] -12.31814 -22.39662 -24.15321 -24.17144
## effectif_de_seconde :[1] 0.623326 1.133320 1.222208 1.223130
## effectif_de_premiere :[1] -0.5085573 -0.9246496 -0.9971711 -0.9979237
## taux_acces_brut_seconde_bac :[1] 4.385296 7.973265 8.598619 8.605109
## taux_acces_attendu_seconde_bac :[1] -15.21503 -27.66369 -29.83339 -
29.85591
## taux_acces_brut_premiere_bac :[1] -21.01288 -38.20524 -41.20173 -41.23283
## taux_acces_attendu_premiere_bac :[1] 51.51593 93.66533 101.01163
101.08787
## taux_brut_de_reussite_total_series :[1] 16.41253 29.84096 32.18142
```

```

32.20571
## taux_reussite_attendu_total_series :[1] -31.16928 -56.67142 -61.11623 -
61.16236
## variance: 99308.41 64176.66 58053.48 57989.94

modele_complet_anglais=lm(Barre~.,data=data_anglais[,6:ncol(data_anglais)])
ncol(data_anglais)

## [1] 23

summary(modele_complet_anglais)

##
## Call:
## lm(formula = Barre ~ ., data = data_anglais[, 6:ncol(data_anglais)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -363.03 -133.21  -24.75   97.93 1044.32
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1286.4382   1860.5131  -0.691   0.4940
## effectif_presents_serie_l      -5.2435     5.6318  -0.931   0.3584
## effectif_presents_serie_es     -1.6538     2.7501  -0.601   0.5516
## effectif_presents_serie_s       0.7911     2.3383   0.338   0.7372
## taux_brut_de_reussite_serie_l    -5.1155     5.8121  -0.880   0.3850
## taux_brut_de_reussite_serie_es     5.9949     8.9053   0.673   0.5054
## taux_brut_de_reussite_serie_s     3.0029    14.7626   0.203   0.8400
## taux_reussite_attendu_serie_l     -3.7900    19.4963  -0.194   0.8470
## taux_reussite_attendu_serie_es    30.9369    17.5094   1.767   0.0862
## taux_reussite_attendu_serie_s    -24.6363    20.1117  -1.225   0.2290
## effectif_de_seconde       1.2467     1.5277   0.816   0.4202
## effectif_de_premiere      -1.0171     1.9638  -0.518   0.6079
## taux_acces_brut_seconde_bac       8.7706    13.6909   0.641   0.5261
## taux_acces_attendu_seconde_bac   -30.4301    23.0086  -1.323   0.1948
## taux_acces_brut_premiere_bac    -42.0258    25.2706  -1.663   0.1055
## taux_acces_attendu_premiere_bac  103.0319    52.3436   1.968   0.0572
## taux_brut_de_reussite_total_series  32.8251    31.4412   1.044   0.3038
## taux_reussite_attendu_total_series -62.3386    54.2693  -1.149   0.2587
##
## (Intercept)
## effectif_presents_serie_l
## effectif_presents_serie_es
## effectif_presents_serie_s
## taux_brut_de_reussite_serie_l
## taux_brut_de_reussite_serie_es
## taux_brut_de_reussite_serie_s
## taux_reussite_attendu_serie_l
## taux_reussite_attendu_serie_es
## taux_reussite_attendu_serie_s

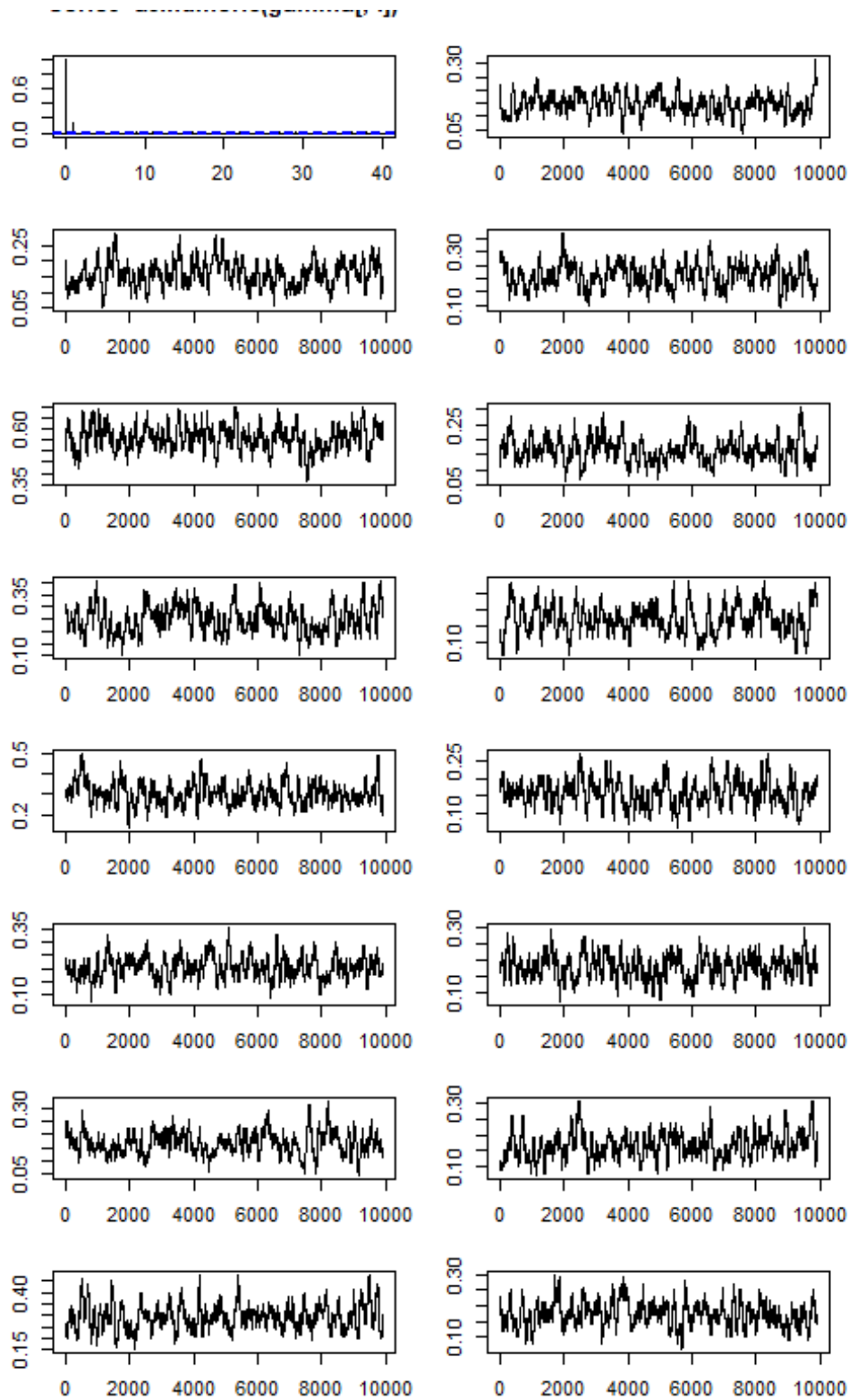
```

```
## effectif_de_seconde
## effectif_de_premiere
## taux_acces_brut_seconde_bac
## taux_acces_attendu_seconde_bac
## taux_acces_brut_premiere_bac
## taux_acces_attendu_premiere_bac .
## taux_brut_de_reussite_total_series
## taux_reussite_attendu_total_series
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 287.9 on 34 degrees of freedom
## Multiple R-squared:  0.3659, Adjusted R-squared:  0.0489
## F-statistic: 1.154 on 17 and 34 DF,  p-value: 0.3492

choix_modele_bayes(data_anglais)

## [1] 17
```



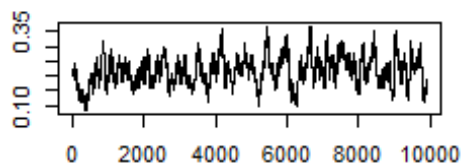


```
##          probtop50
## [1,] 0.012315789 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
## [2,] 0.009368421 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [3,] 0.008947368 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0
```

```

## [4,] 0.006736842 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
## [5,] 0.006631579 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
## [6,] 0.006105263 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0
## [7,] 0.005894737 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [8,] 0.005684211 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [9,] 0.005368421 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
## [10,] 0.005368421 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
## [11,] 0.004736842 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [12,] 0.004736842 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0
## [13,] 0.004736842 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1
## [14,] 0.004526316 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0
## [15,] 0.004421053 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
## [16,] 0.004315789 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
## [17,] 0.004210526 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
## [18,] 0.004210526 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [19,] 0.004210526 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
## [20,] 0.004000000 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
## [21,] 0.003894737 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0
## [22,] 0.003789474 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
## [23,] 0.003684211 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [24,] 0.003368421 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0
## [25,] 0.003368421 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
## [26,] 0.003263158 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
## [27,] 0.003263158 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0
## [28,] 0.003263158 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
## [29,] 0.003157895 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0
## [30,] 0.003052632 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
## [31,] 0.003052632 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0
## [32,] 0.002947368 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
## [33,] 0.002736842 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
## [34,] 0.002736842 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
## [35,] 0.002631579 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
## [36,] 0.002631579 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0
## [37,] 0.002631579 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1
## [38,] 0.002421053 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0
## [39,] 0.002421053 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1
## [40,] 0.002315789 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [41,] 0.002315789 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0
## [42,] 0.002315789 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0
## [43,] 0.002315789 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0
## [44,] 0.002315789 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1
## [45,] 0.002210526 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0
## [46,] 0.002210526 1 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0
## [47,] 0.002210526 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 0
## [48,] 0.002210526 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0
## [49,] 0.002000000 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0
## [50,] 0.002000000 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
## frequence apparition covariable : 0.08 0.04 0.02 0.08 0.6 0.04 0.26 0.06
0.18 0.06 0.06 0.08 0.04 0.06 0.18 0.06 0.1

```



L'autocorrélogramme décroît vers 0 et la chaîne explore toute la loi.

Résultat du choix de modèle bayésien pour l'anglais : covariables 5 ("taux\_brut\_de\_reussite\_series\_es") et 7 ("taux\_reussite\_attendu\_serie\_l") avec une probabilité de 1%

```
modele_anglais=step(lm(Barre~.,data=data_anglais[,6:ncol(data_anglais)]))

## Start:  AIC=602.82
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##   effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
##   taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
##   taux_reussite_attendu_serie_l + taux_reussite_attendu_serie_es +
##   taux_reussite_attendu_serie_s + effectif_de_seconde +
effectif_de_premiere +
##   taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
##   taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##   taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
##              Df Sum of Sq    RSS    AIC
## - taux_reussite_attendu_serie_l      1      3133 2821613 600.88
## - taux_brut_de_reussite_serie_s      1      3430 2821910 600.89
## - effectif_presents_serie_s          1      9490 2827970 601.00
## - effectif_de_premiere                1     22237 2840717 601.23
## - effectif_presents_serie_es          1     29979 2848460 601.37
## - taux_acces_brut_seconde_bac         1     34020 2852500 601.45
## - taux_brut_de_reussite_serie_es      1     37567 2856047 601.51
## - effectif_de_seconde                 1     55203 2873683 601.83
## - taux_brut_de_reussite_serie_l       1     64216 2882697 602.00
## - effectif_presents_serie_l           1     71859 2890339 602.13
## - taux_brut_de_reussite_total_series  1     90355 2908835 602.47
## - taux_reussite_attendu_total_series  1    109381 2927861 602.80
## <none>                                2818480 602.82
## - taux_reussite_attendu_serie_s       1    124391 2942872 603.07
## - taux_acces_attendu_seconde_bac      1    144997 2963478 603.43
## - taux_acces_brut_premiere_bac        1    229265 3047745 604.89
## - taux_reussite_attendu_serie_es      1    258792 3077272 605.39
## - taux_acces_attendu_premiere_bac     1    321183 3139663 606.44
##
## Step:  AIC=600.88
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
```



```
## effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
## taux_brut_de_reussite_serie_es + taux_brut_de_reussite_serie_s +
## taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
## effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
## taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
## taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series +
## taux_reussite_attendu_total_series
```

	Df	Sum of Sq	RSS	AIC
## - taux_brut_de_reussite_serie_s	1	1584	2823197	598.91
## - effectif_presents_serie_s	1	7629	2829242	599.02
## - effectif_de_premiere	1	19704	2841317	599.24
## - effectif_presents_serie_es	1	28139	2849752	599.40
## - taux_acces_brut_seconde_bac	1	32718	2854331	599.48
## - taux_brut_de_reussite_serie_es	1	37387	2859000	599.57
## - effectif_de_seconde	1	55865	2877478	599.90
## - taux_brut_de_reussite_serie_l	1	105620	2927233	600.79
## - effectif_presents_serie_l	1	108663	2930276	600.85
## <none>			2821613	600.88
## - taux_reussite_attendu_serie_s	1	121395	2943008	601.07
## - taux_brut_de_reussite_total_series	1	146797	2968410	601.52
## - taux_acces_attendu_seconde_bac	1	148001	2969614	601.54
## - taux_acces_brut_premiere_bac	1	261543	3083156	603.49
## - taux_reussite_attendu_serie_es	1	266530	3088143	603.58
## - taux_reussite_attendu_total_series	1	298943	3120556	604.12
## - taux_acces_attendu_premiere_bac	1	417952	3239565	606.06

```
##
## Step: AIC=598.91
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
## effectif_presents_serie_s + taux_brut_de_reussite_serie_l +
## taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_es +
## taux_reussite_attendu_serie_s + effectif_de_seconde +
effectif_de_premiere +
## taux_acces_brut_seconde_bac + taux_acces_attendu_seconde_bac +
## taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
## taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
```

	Df	Sum of Sq	RSS	AIC
## - effectif_presents_serie_s	1	8008	2831205	597.06
## - effectif_de_premiere	1	19057	2842255	597.26
## - taux_acces_brut_seconde_bac	1	31695	2854893	597.49
## - effectif_presents_serie_es	1	31773	2854971	597.49
## - taux_brut_de_reussite_serie_es	1	44442	2867639	597.72
## - effectif_de_seconde	1	58215	2881412	597.97
## <none>			2823197	598.91
## - effectif_presents_serie_l	1	123442	2946640	599.14
## - taux_reussite_attendu_serie_s	1	123635	2946832	599.14
## - taux_brut_de_reussite_serie_l	1	125040	2948237	599.16

```

## - taux_acces_attendu_seconde_bac      1    146929 2970126 599.55
## - taux_acces_brut_premiere_bac        1    260235 3083433 601.50
## - taux_reussite_attendu_serie_es      1    265745 3088942 601.59
## - taux_brut_de_reussite_total_series  1    274427 3097624 601.73
## - taux_reussite_attendu_total_series  1    301322 3124520 602.18
## - taux_acces_attendu_premiere_bac     1    420547 3243745 604.13
##
## Step:   AIC=597.06
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##      taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##      taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##      effectif_de_seconde + effectif_de_premiere +
taux_acces_brut_seconde_bac +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series +
##      taux_reussite_attendu_total_series
##
##                               Df Sum of Sq      RSS      AIC
## - effectif_de_premiere      1      14262 2845467 595.32
## - effectif_presents_serie_es 1      23818 2855024 595.49
## - taux_acces_brut_seconde_bac 1      28601 2859807 595.58
## - taux_brut_de_reussite_serie_es 1      51525 2882731 596.00
## - effectif_de_seconde      1      62931 2894136 596.20
## <none>                        2831205 597.06
## - taux_brut_de_reussite_serie_l 1     120087 2951292 597.22
## - taux_acces_attendu_seconde_bac 1     146753 2977958 597.69
## - effectif_presents_serie_l    1     150667 2981872 597.75
## - taux_reussite_attendu_serie_s 1     185466 3016672 598.36
## - taux_acces_brut_premiere_bac 1     252565 3083770 599.50
## - taux_reussite_attendu_serie_es 1     268123 3099329 599.76
## - taux_brut_de_reussite_total_series 1     284412 3115617 600.04
## - taux_reussite_attendu_total_series 1     293698 3124903 600.19
## - taux_acces_attendu_premiere_bac 1     461014 3292219 602.90
##
## Step:   AIC=595.32
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##      taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##      taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##      effectif_de_seconde + taux_acces_brut_seconde_bac +
taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##      taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
##                               Df Sum of Sq      RSS      AIC
## - taux_acces_brut_seconde_bac      1      17731 2863198 593.64
## - effectif_presents_serie_es        1      37677 2883145 594.00
## - taux_brut_de_reussite_serie_es    1      53214 2898681 594.28
## - effectif_de_seconde              1      82312 2927779 594.80
## <none>                              2845467 595.32

```

```

## - taux_brut_de_reussite_serie_l      1    111821 2957288 595.32
## - effectif_presents_serie_l          1    190244 3035711 596.69
## - taux_acces_attendu_seconde_bac     1    196708 3042175 596.80
## - taux_reussite_attendu_serie_s      1    231462 3076929 597.39
## - taux_acces_brut_premiere_bac       1    245961 3091428 597.63
## - taux_reussite_attendu_serie_es     1    256182 3101649 597.80
## - taux_brut_de_reussite_total_series 1    280121 3125589 598.20
## - taux_reussite_attendu_total_series 1    306123 3151590 598.63
## - taux_acces_attendu_premiere_bac    1    640416 3485884 603.88
##
## Step: AIC=593.64
## Barre ~ effectif_presents_serie_l + effectif_presents_serie_es +
##      taux_brut_de_reussite_serie_l + taux_brut_de_reussite_serie_es +
##      taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##      effectif_de_seconde + taux_acces_attendu_seconde_bac +
taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series +
##      taux_reussite_attendu_total_series
##
##              Df Sum of Sq      RSS      AIC
## - effectif_presents_serie_es      1      28893 2892091 592.16
## - effectif_de_seconde              1      82912 2946110 593.13
## - taux_brut_de_reussite_serie_es   1      88596 2951794 593.23
## <none>                             2863198 593.64
## - taux_brut_de_reussite_serie_l    1     117920 2981118 593.74
## - effectif_presents_serie_l        1     179961 3043160 594.81
## - taux_acces_attendu_seconde_bac   1     221349 3084547 595.51
## - taux_reussite_attendu_serie_s    1     227807 3091005 595.62
## - taux_reussite_attendu_serie_es   1     240986 3104184 595.84
## - taux_acces_brut_premiere_bac     1     249140 3112338 595.98
## - taux_brut_de_reussite_total_series 1     262470 3125668 596.20
## - taux_reussite_attendu_total_series 1     288418 3151616 596.63
## - taux_acces_attendu_premiere_bac  1     690292 3553490 602.87
##
## Step: AIC=592.16
## Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_es +
##      taux_reussite_attendu_serie_s + effectif_de_seconde +
taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##      taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
##              Df Sum of Sq      RSS      AIC
## - effectif_de_seconde              1      54161 2946252 591.13
## - taux_brut_de_reussite_serie_es   1      94058 2986149 591.83
## - taux_brut_de_reussite_serie_l    1      97540 2989630 591.89
## <none>                             2892091 592.16
## - effectif_presents_serie_l        1     169374 3061464 593.12
## - taux_reussite_attendu_serie_s    1     213406 3105497 593.87

```

```

## - taux_acces_brut_premiere_bac      1    235858 3127949 594.24
## - taux_reussite_attendu_serie_es     1    240226 3132316 594.31
## - taux_acces_attendu_seconde_bac     1    244666 3136756 594.39
## - taux_brut_de_reussite_total_series 1    262965 3155055 594.69
## - taux_reussite_attendu_total_series 1    303474 3195565 595.35
## - taux_acces_attendu_premiere_bac    1    677085 3569176 601.10
##
## Step:  AIC=591.13
## Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##      taux_brut_de_reussite_serie_es + taux_reussite_attendu_serie_es +
##      taux_reussite_attendu_serie_s + taux_acces_attendu_seconde_bac +
##      taux_acces_brut_premiere_bac + taux_acces_attendu_premiere_bac +
##      taux_brut_de_reussite_total_series +
taux_reussite_attendu_total_series
##
##              Df Sum of Sq      RSS      AIC
## - taux_brut_de_reussite_serie_es      1    100460 3046712 590.87
## - taux_brut_de_reussite_serie_l       1    101804 3048056 590.90
## <none>                                2946252 591.13
## - effectif_presents_serie_l           1    162158 3108410 591.92
## - taux_reussite_attendu_serie_s       1    187981 3134233 592.35
## - taux_acces_attendu_seconde_bac      1    221767 3168019 592.90
## - taux_acces_brut_premiere_bac        1    223082 3169334 592.92
## - taux_reussite_attendu_serie_es      1    229817 3176068 593.04
## - taux_brut_de_reussite_total_series  1    269681 3215933 593.68
## - taux_reussite_attendu_total_series   1    289485 3235737 594.00
## - taux_acces_attendu_premiere_bac     1    627708 3573960 599.17
##
## Step:  AIC=590.87
## Barre ~ effectif_presents_serie_l + taux_brut_de_reussite_serie_l +
##      taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##      taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##      taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series +
##      taux_reussite_attendu_total_series
##
##              Df Sum of Sq      RSS      AIC
## <none>                                3046712 590.87
## - effectif_presents_serie_l           1    141143 3187854 591.23
## - taux_brut_de_reussite_serie_l       1    147794 3194505 591.34
## - taux_reussite_attendu_serie_s       1    209765 3256476 592.34
## - taux_acces_attendu_seconde_bac      1    227532 3274244 592.62
## - taux_acces_brut_premiere_bac        1    242989 3289701 592.86
## - taux_reussite_attendu_total_series   1    392672 3439384 595.18
## - taux_brut_de_reussite_total_series   1    472968 3519679 596.38
## - taux_reussite_attendu_serie_es      1    526469 3573181 597.16
## - taux_acces_attendu_premiere_bac     1    632733 3679444 598.69

```

`summary(modele_anglais)`

```
##
## Call:
## lm(formula = Barre ~ effectif_presents_serie_l +
taux_brut_de_reussite_serie_l +
##     taux_reussite_attendu_serie_es + taux_reussite_attendu_serie_s +
##     taux_acces_attendu_seconde_bac + taux_acces_brut_premiere_bac +
##     taux_acces_attendu_premiere_bac + taux_brut_de_reussite_total_series +
##     taux_reussite_attendu_total_series, data = data_anglais[,
##     6:ncol(data_anglais)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -420.1  -116.3   -33.2    88.1  1089.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1696.950     822.088   -2.064  0.04521
## effectif_presents_serie_l      -3.095       2.219   -1.395  0.17038
## taux_brut_de_reussite_serie_l     -6.030       4.225   -1.427  0.16087
## taux_reussite_attendu_serie_es     37.414     13.888    2.694  0.01010
## taux_reussite_attendu_serie_s    -26.011     15.296   -1.700  0.09643
## taux_acces_attendu_seconde_bac    -26.632     15.037   -1.771  0.08381
## taux_acces_brut_premiere_bac     -29.358     16.041   -1.830  0.07432
## taux_acces_attendu_premiere_bac   101.217     34.272    2.953  0.00513
## taux_brut_de_reussite_total_series  44.714     17.511    2.553  0.01439
## taux_reussite_attendu_total_series -74.814     32.156   -2.327  0.02488
##
## (Intercept) *
## effectif_presents_serie_l
## taux_brut_de_reussite_serie_l
## taux_reussite_attendu_serie_es *
## taux_reussite_attendu_serie_s .
## taux_acces_attendu_seconde_bac .
## taux_acces_brut_premiere_bac .
## taux_acces_attendu_premiere_bac **
## taux_brut_de_reussite_total_series *
## taux_reussite_attendu_total_series *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 269.3 on 42 degrees of freedom
## Multiple R-squared:  0.3146, Adjusted R-squared:  0.1677
## F-statistic: 2.142 on 9 and 42 DF, p-value: 0.04689
```

Résultat choix de modèle frequentiste : bien plus de covariables retenues :

```
effectif_presents_serie_l
taux_brut_de_reussite_serie_l taux_reussite_attendu_serie_es taux_reussite_attendu_serie_s
taux_acces_attendu_seconde_bac taux_acces_brut_premiere_bac
```

taux\_acces\_attendu\_premiere\_bac taux\_brut\_de\_reussite\_total\_series  
taux\_reussite\_attendu\_total\_series

En conclusion dans tous ces cas le choix de modèle entre bayésien et fréquentistes sont différents. L'avantage du choix de modèle bayésien est que nous disposons de la probabilité du modèle, et on peut donc faire un "mélange de modèles", ie une régression pondérée par la probabilité du choix de modèle.

Influence des covariables sur les 2 matières :

Intuitivement on peut penser que les covariables n'ont pas la même influence et que par exemple les professeurs de maths seront plus attentifs aux filières S et ES et moins à la filière L, alors que les profs d'anglais seront plus sensibles à la filière littéraire. D'autre part le résultat du choix de modèle Bayésien n'est pas le même pour ces 2 matières.

Pour le vérifier, il suffit de regarder la fréquence d'apparition de la covariable `taux_reussite_attendu_serie_L` :

fréquences apparitions pour les maths: 0.06 0.06 0.1 0.46 0.62 0.06 0.04 0.06 0.06 0.04 0.08  
0.06 0.04 0.18 0.14 0.12 0.08

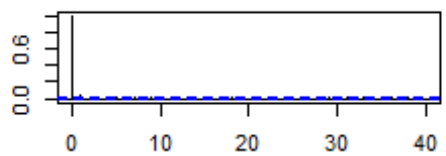
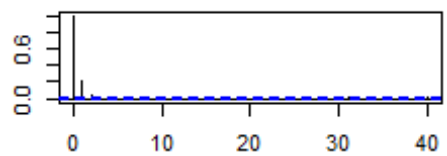
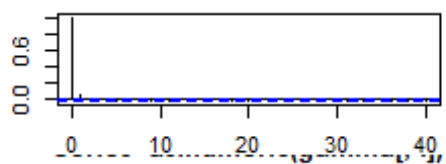
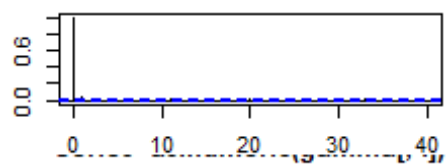
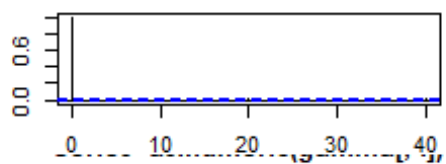
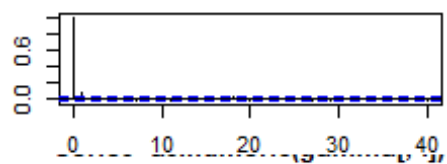
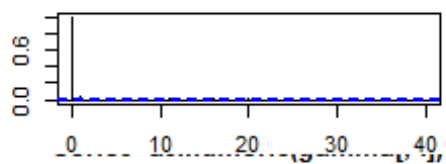
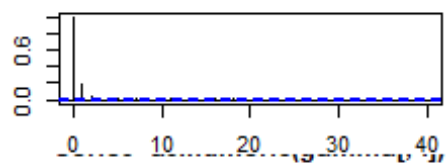
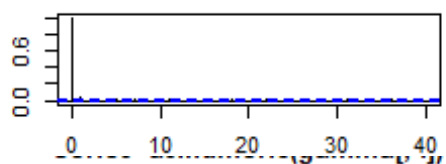
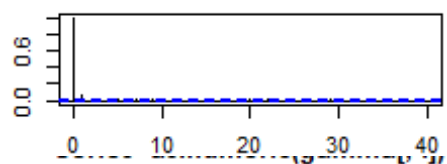
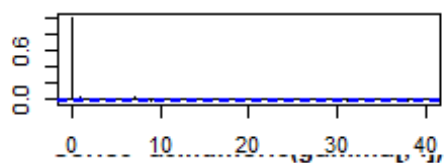
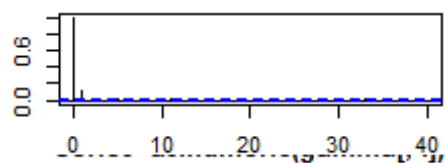
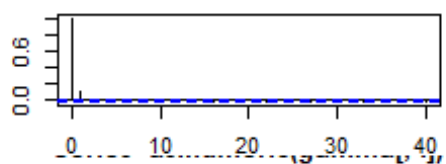
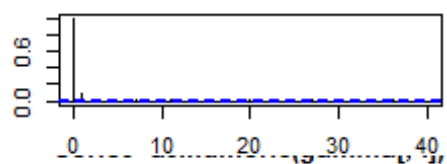
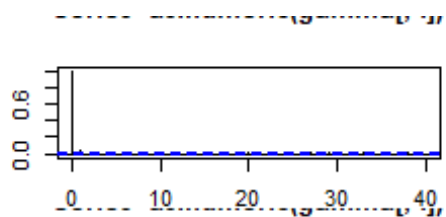
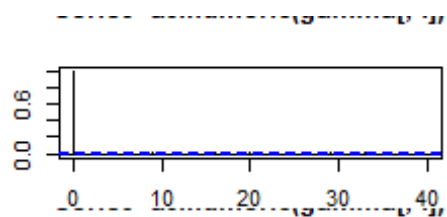
fréquences apparitions pour l'anglais: 0.06 0.04 0.04 0.08 0.62 0.04 0.22 0.06 0.18 0.04 0.08  
0.1 0.04 0.06 0.18 0.06 0.08

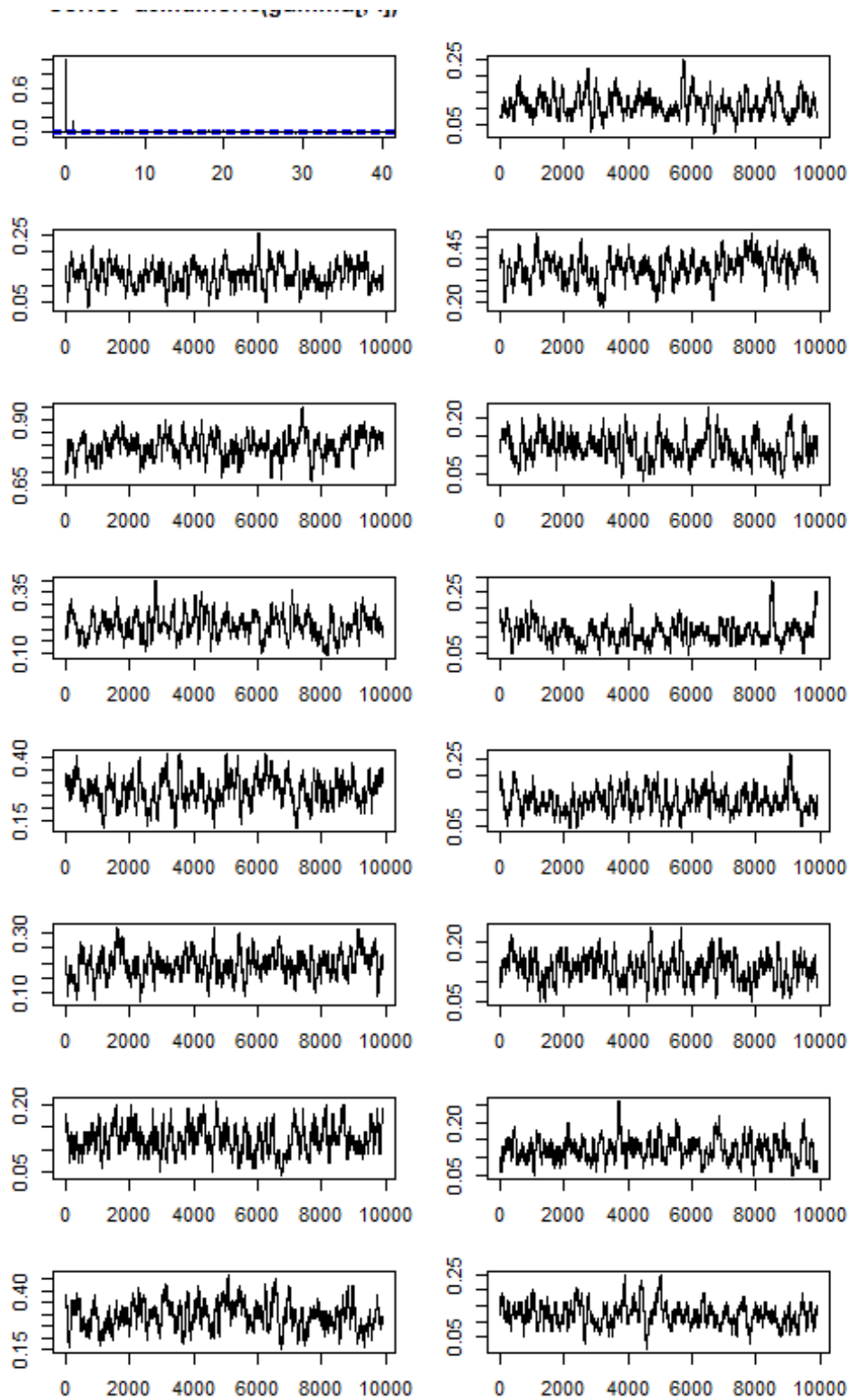
On voit que les 2 fréquences d'apparitions sont différentes pour certaines covariables

D'autre part on dispose déjà de la fonction log vraisemblance marginale, on peut donc comparer les 2 modèles suivants par calcul du facteur de Bayes : Hypothèse 1) : les covariables agissent de la même manière sur anglais et maths ( $Y = \beta_0 + \beta_1 X_1 + \dots$ ) Hypothèse 2) : on a deux modèles différents ( $Y = \beta'_0 + \beta'_1 X_1 + \dots$  pour l'anglais et  $Y = \beta''_0 + \beta''_1 X_1 + \dots$  pour les maths)

```
choix_modele_bayes(data_maths_anglais)
```

```
## [1] 17
```





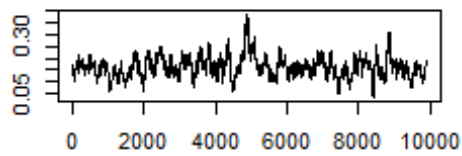
```
##          probtop50
## [1,] 0.037578947 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
## [2,] 0.033789474 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [3,] 0.019684211 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
```



```

## [4,] 0.017157895 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0
## [5,] 0.010526316 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0
## [6,] 0.010421053 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
## [7,] 0.009578947 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0
## [8,] 0.008631579 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0
## [9,] 0.008000000 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
## [10,] 0.007157895 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0
## [11,] 0.007052632 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
## [12,] 0.007052632 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1
## [13,] 0.006526316 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
## [14,] 0.006000000 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0
## [15,] 0.005894737 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0
## [16,] 0.005894737 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
## [17,] 0.005684211 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0
## [18,] 0.005368421 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [19,] 0.005263158 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0
## [20,] 0.005157895 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [21,] 0.005157895 0 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0
## [22,] 0.004947368 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0
## [23,] 0.004842105 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0
## [24,] 0.004736842 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0
## [25,] 0.004736842 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1
## [26,] 0.004631579 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0
## [27,] 0.004631579 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0
## [28,] 0.004631579 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1
## [29,] 0.004526316 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0
## [30,] 0.004421053 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0
## [31,] 0.004315789 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0
## [32,] 0.004315789 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0
## [33,] 0.004000000 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0
## [34,] 0.004000000 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0
## [35,] 0.003894737 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 1
## [36,] 0.003473684 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0
## [37,] 0.003263158 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [38,] 0.003263158 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0
## [39,] 0.003263158 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0
## [40,] 0.003052632 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
## [41,] 0.002947368 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0
## [42,] 0.002947368 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0
## [43,] 0.002842105 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
## [44,] 0.002736842 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
## [45,] 0.002631579 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0
## [46,] 0.002631579 0 0 0 0 1 0 0 0 1 0 0 0 0 1 1 0
## [47,] 0.002631579 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1
## [48,] 0.002526316 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
## [49,] 0.002526316 1 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0
## [50,] 0.002526316 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1 0
## frequence apparition covariable : 0.06 0.04 0.02 0.36 0.86 0.04 0.14 0.04
0.22 0.08 0.12 0.06 0.04 0.06 0.26 0.06 0.08

```



```
#meilleur modele maths_anglais: 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
gamma_maths_anglais=c(F,F,F,T,T,F,F,F,F,F,F,F,F,F,F,F)

#meilleur modele laths: 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
gamma_maths=c(F,F,F,F,T,F,F,F,F,F,F,F,F,F,F,F)

#meilleur modele anglais: 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0

gamma_anglais=c(F,F,F,F,T,F,T,F,F,F,F,F,F,F,F,F)

#log marginal likelihood hypothese un modele unique pour les 2 matieres
m1=marglkd(data_maths_anglais,gamma_maths_anglais)

#log marginal likelihood hypothese 2 modeles differents
m21=marglkd(data_maths,gamma_maths)
m22=marglkd(data_anglais,gamma_anglais)

#cat(m1,m21,m22)
#facteur de Bayes
#cat(exp(m1))

BF=exp(m1-m21-m22)
print(BF)

##           [,1]
## [1,] 6.30343e-26

print(log10(BF))#-24 <-2 : L'hypothese 2 est decisive
```

```
##           [,1]
## [1,] -25.20042
```

En conclusion il est fort plausible que les covariables n'agissent pas de la même manière sur les 2 disciplines, ce qui est conforme a notre intuition initiale

## *II) Loi de Pareto*

### *4) Generation de Pareto*

```
library(EnvStats)

## Warning: package 'EnvStats' was built under R version 3.3.3

##
## Attaching package: 'EnvStats'

## The following object is masked from 'package:MASS':
##
##      boxcox

## The following objects are masked from 'package:stats':
##
##      predict, predict.lm

## The following object is masked from 'package:base':
##
##      print.default

?rpareto

## starting httpd help server ...

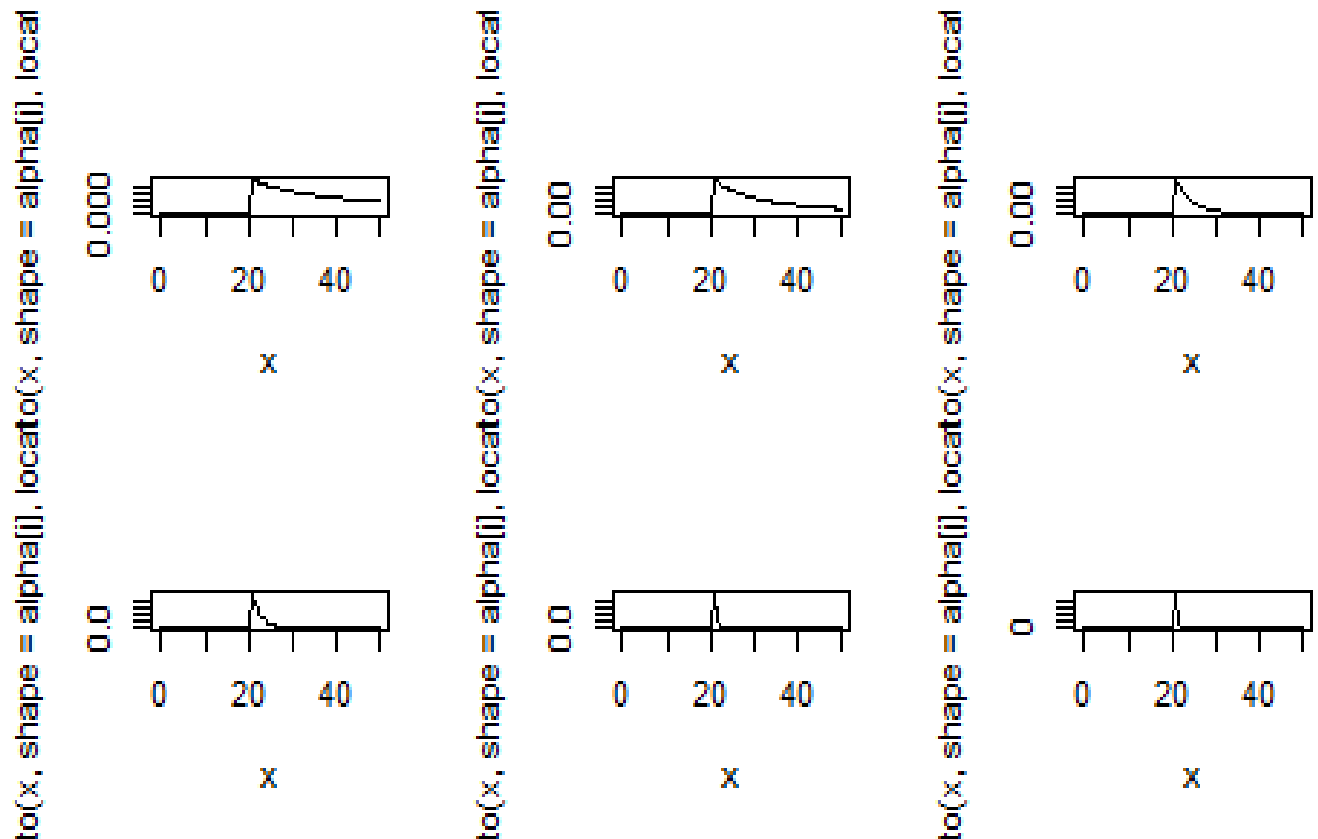
## done

n=1e4
m=21
alpha=c(0.1,1,5,10,50,100)

par(mfrow=c(3,3))

for (j in 1:length(alpha)){
  curve(dporeto(x,shape = alpha[j],location=m),xlim=c(0,50))
}

#quand alpha tends vers L'infini, La Loi de Pareto converge vers une Dirac en
m
```



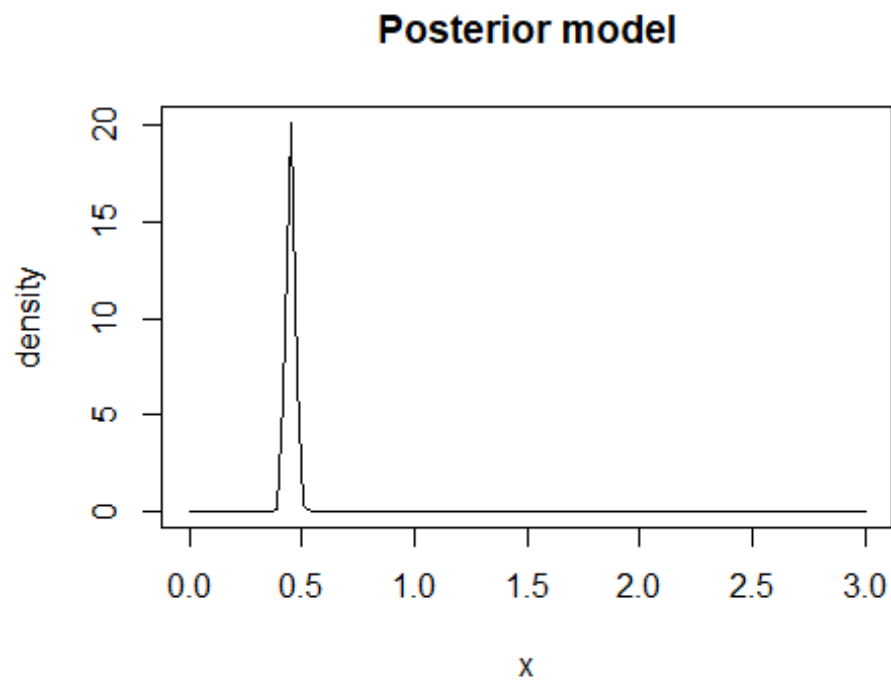
### 5) Loi a priori pour alpha

On va choisir une loi a priori  $\text{gamma}(a,b)$  pour  $\alpha$ . En effet on aura alors une loi a posteriori conjuguée (pour le voir, l'astuce consiste à reecrire la loi de Pareto :  $(m/z)^\alpha = \exp(\alpha * (\ln(m/z)))$ )

6) Loi a posteriori pour alpha Soit  $n$  les nombre d'observations, et  $z_i$  la valeur de Barre pour l'observation  $i$  La loi a posteriori est alors  $\Gamma(a + n, b + \sum \ln(z_i) - n * \ln(m))$

```
a=2
b=2
n=nrow(data)

curve(dgamma(x,a+n,b+sum(log(data$Barre))-n*log(m)), xlim=c(0,3),
main="Posterior model", ylab="density")
```



```
#esperance
print((a+n)/(b+sum(log(data$Barre))-n*log(m)))

## [1] 0.451164

#interval de credibilite a 95 %
qgamma(c(.025, .975), a+n, b+sum(log(data$Barre))-n*log(m))

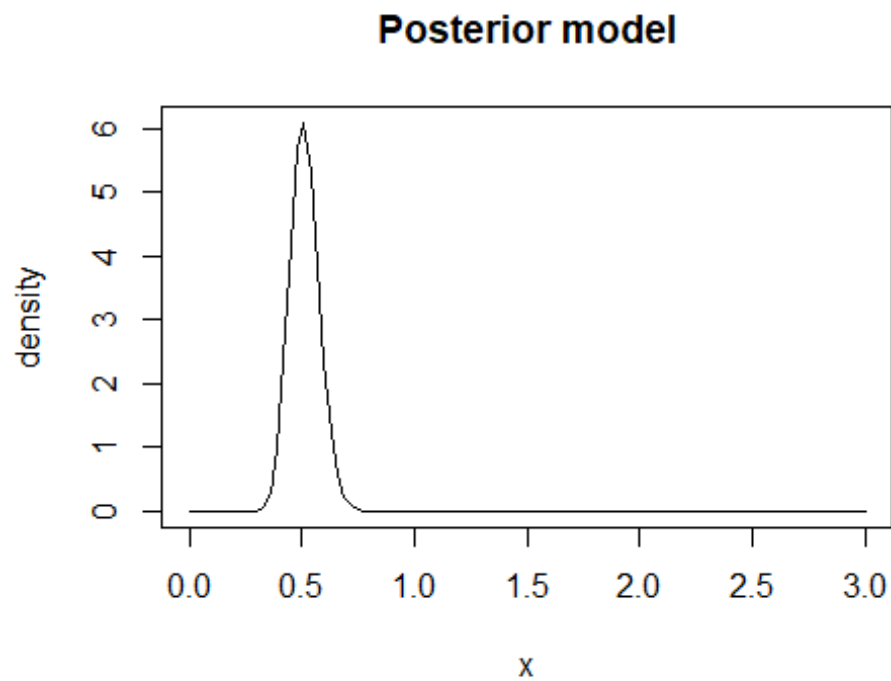
## [1] 0.4131430 0.4908346
```

7)Loi a posteriori pour alpha - mathématiques et anglais

mathématiques

```
a=2
b=2
n=nrow(data_maths)

curve(dgamma(x,a+n,b+sum(log(data_maths$Barre))-n*log(m)), xlim=c(0,3),
main="Posterior model", ylab="density")
```



```
#esperance
print((a+n)/(b+sum(log(data_maths$Barre))-n*log(m)))

## [1] 0.512894

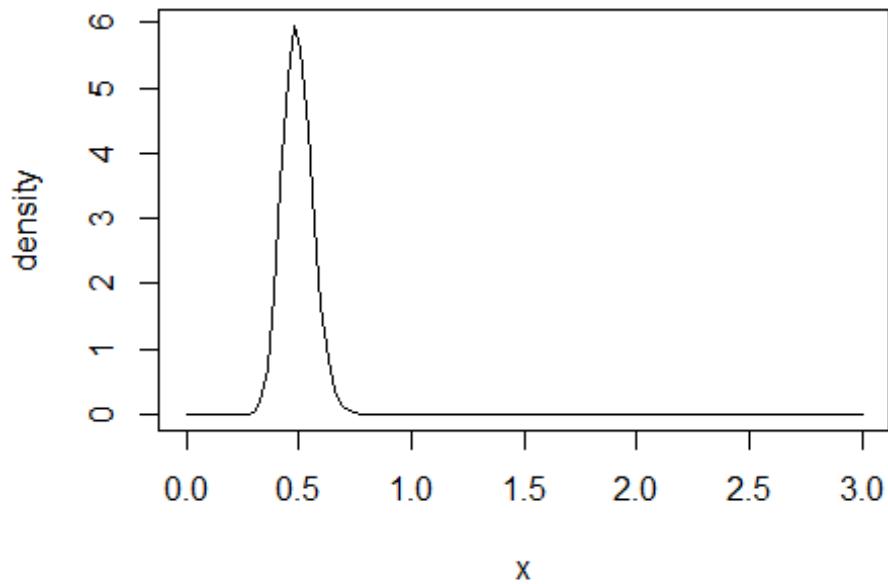
#interval de credibilite a 95 %
qgamma(c(.025, .975), a+n, b+sum(log(data_maths$Barre))-n*log(m))

## [1] 0.3923232 0.6493733

anglais
a=2
b=2
n=nrow(data_anglais)

curve(dgamma(x,a+n,b+sum(log(data_anglais$Barre))-n*log(m)), xlim=c(0,3),
main="Posterior model", ylab="density")
```

## Posterior model



```
#esperance
print((a+n)/(b+sum(log(data_anglais$Barre))-n*log(m)))

## [1] 0.4946058

#interval de credibilite a 95 %
qgamma(c(.025, .975), a+n, b+sum(log(data_anglais$Barre))-n*log(m))

## [1] 0.3715631 0.6349758
```

### 8) Test bayesien de l'hypothese sur $\alpha$

On va estimer un facteur de Bayes par la méthode de Monte-Carlo pour tester l'hypothese

$$\alpha_{\text{maths}} = \alpha_{\text{anglais}}$$

```
#Q5: Monte Carlo standard
# Fonctions pour la vraisemblance
lkd.model1=function(log_z,n,alpha,m){
  #return(exp(-n*Lambda+y*Log(Lambda)))
  return(exp(n*(log(alpha)+alpha*log(m))-(alpha)*log_z))
}

lkd.model2=function(log_z1,n1,log_z2,n2,alpha1,alpha2,m){
  return(exp(n1*(log(alpha1)+alpha1*log(m))-(alpha1)*log_z1
    +n2*(log(alpha2)+alpha2*log(m))-(alpha2)*log_z2))
  #return(Lambda1^y1*exp(-n1*Lambda1)*Lambda2^y2*exp(-n2*Lambda2))
}
```

**Remarque : normalement dans la vraisemblance, c'est  $(\alpha+1)\log_z$  au lieu de  $(\alpha+1)\log_z$  cependant le "+1" se neutralise dans le facteur de Bayes**

```
BF_MC=function(a,b,log_z1,n1,log_z2,n2,M,m,log_z){
  alpha1=rgamma(M,a,b)

  m1=cumsum(lkd.model1((log_z),(n1+n2),alpha1,m))/(1:M)
  alpha2.1=rgamma(M,a,b)
  alpha2.2=rgamma(M,a,b)
  m2=cumsum(lkd.model2(log_z1,n1,log_z2,n2,alpha2.1,alpha2.2,m))/(1:M)
  return(m2/m1)
}

M=1e7
m=21
a=2
b=2

log_z=sum(log(data_maths_anglais$Barre))
log_z1=sum(log(data_maths$Barre))
n1=nrow(data_maths)
log_z2=sum(log(data_anglais$Barre))
n2=nrow(data_anglais)

resMC=BF_MC(a,b,log_z1,n1,log_z2,n2,M,m,log_z)
resMC[M]

## [1] 0.1750169

resMC[1000:1100]

## [1] 0.1159330 0.1159347 0.1160420 0.1160420 0.1150490 0.1150490
0.1150491
## [8] 0.1150491 0.1147564 0.1145575 0.1145575 0.1145575 0.1145575
0.1145575
## [15] 0.1145575 0.1140789 0.1140652 0.1140652 0.1140652 0.1140646
0.1140640
## [22] 0.1140640 0.1140640 0.1140640 0.1140640 0.1141202 0.1141202
0.1141202
## [29] 0.1141202 0.1141202 0.1132530 0.1122656 0.1122656 0.1122656
0.1122655
## [36] 0.1122648 0.1122648 0.1122648 0.1120102 0.1120102 0.1120102
0.1120102
## [43] 0.1120102 0.1110179 0.1109891 0.1109891 0.1109891 0.1109891
0.1109890
## [50] 0.1109890 0.1109890 0.1099044 0.1099044 0.1099044 0.1099044
0.1088603
## [57] 0.1088178 0.1088179 0.1151163 0.1147742 0.1147742 0.1147742
0.1147742
```



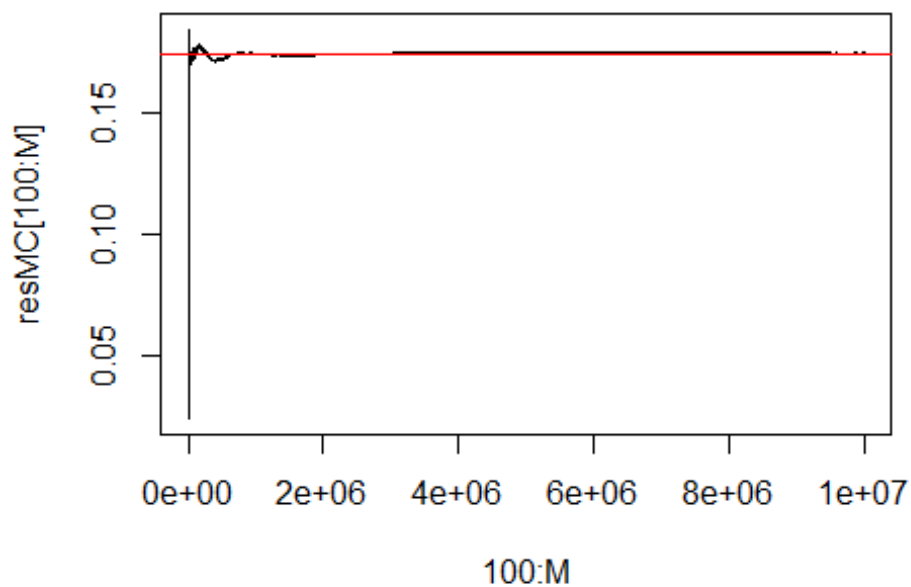
```
## [64] 0.1147742 0.1147763 0.1147763 0.1147763 0.1137036 0.1137036
0.1137036
## [71] 0.1137034 0.1137040 0.1146884 0.1147028 0.1137514 0.1137514
0.1137514
## [78] 0.1137831 0.1137746 0.1139904 0.1139934 0.1139934 0.1139934
0.1139934
## [85] 0.1139877 0.1129657 0.1129574 0.1129571 0.1119685 0.1119686
0.1119685
## [92] 0.1119692 0.1119675 0.1119675 0.1110656 0.1110656 0.1116583
0.1116583
## [99] 0.1110979 0.1110159 0.1106933

plot(100:M,resMC[100:M], type="l")
#abline(h=trueBF, col=2)

log10(resMC[M])#-0.75

## [1] -0.7569201

abline(h=resMC[M],col="red")
```



Le log base 10 du facteur de Bayes obtenu est -0.75: ceci est sur l'échelle de Jeffreys substantiel en faveur du dénominateur "m1" ie  $\alpha_{maths} = \alpha_{anglais}$ . Ceci est coherent avec les intervalles de credibilite trouvés à la question precedente.

## Importance sampling

Soit  $n$  les nombre d'observations, et  $z_i$  la valeur de Barre pour l'observation  $i$  La loi a posteriori est alors  $\Gamma(a + n, b + \sum \ln(z_i) - n * \ln(m))$

```
BF_IS=function(a,b,log_z1,n1,log_z2,n2,M,m,log_z){
#BF_IS=function(a,b,y1,n1,y2,n2,M)

#mean1=(a+y1+y2)/(b+n1+n2)
#mean2.1=(a+y1)/(b+n1)
#mean2.2=(a+y2)/(b+n2)

mean1=(a+n1+n2)/(b+log_z-(n1+n2)*log(m))

mean2.1=(a+n1)/(b+log_z1-n1*log(m))
mean2.2=(a+n2)/(b+log_z2-n2*log(m))

#sigma1=sqrt((a+y1+y2)/(b+n1+n2)^2)
#sigma2.1=sqrt((a+y1)/(b+n1)^2)
#sigma2.2=sqrt((a+y2)/(b+n2)^2)

sigma1=sqrt((a+n1+n2)/(b+log_z-(n1+n2)*log(m))^2)
sigma2.1=sqrt((a+n1)/(b+log_z1-n1*log(m))^2)
sigma2.2=sqrt((a+n2)/(b+log_z2-n2*log(m))^2)

#lambda1=rnorm(M,mean1,sigma1)
alpha1=rnorm(M,mean1,sigma1)

#lkd.model1((log_z),(n1+n2),alpha1,m)
#lkd.model1(y1+y2,n1+n2,lambda1)

m1=cumsum( lkd.model1((log_z),(n1+n2),alpha1,m) *
           dgamma(alpha1,a, b) / dnorm(alpha1,mean1,sigma1))/(1:M)

alpha2.1=rnorm(M,mean2.1, sigma2.1)
alpha2.2=rnorm(M,mean2.2, sigma2.2)

m2=cumsum( lkd.model2(log_z1,n1,log_z2,n2,alpha2.1,alpha2.2,m) *
           dgamma(alpha2.1,a, b) * dgamma(alpha2.2,a,b) /
           (dnorm(alpha2.1,mean2.1,sigma2.1) *
            dnorm(alpha2.2,mean2.2,sigma2.2))) / (1:M)
```

```

    return(m2/m1)
}

#resIS=BF_IS(2,2,qh,nh,qf,nf,M)

M=1e7
m=21
a=2
b=2

log_z=sum(log(data_maths_anglais$Barre))
log_z1=sum(log(data_maths$Barre))
n1=nrow(data_maths)
log_z2=sum(log(data_anglais$Barre))
n2=nrow(data_anglais)

resIS=BF_IS(a,b,log_z1,n1,log_z2,n2,M,m,log_z)

resIS[M]

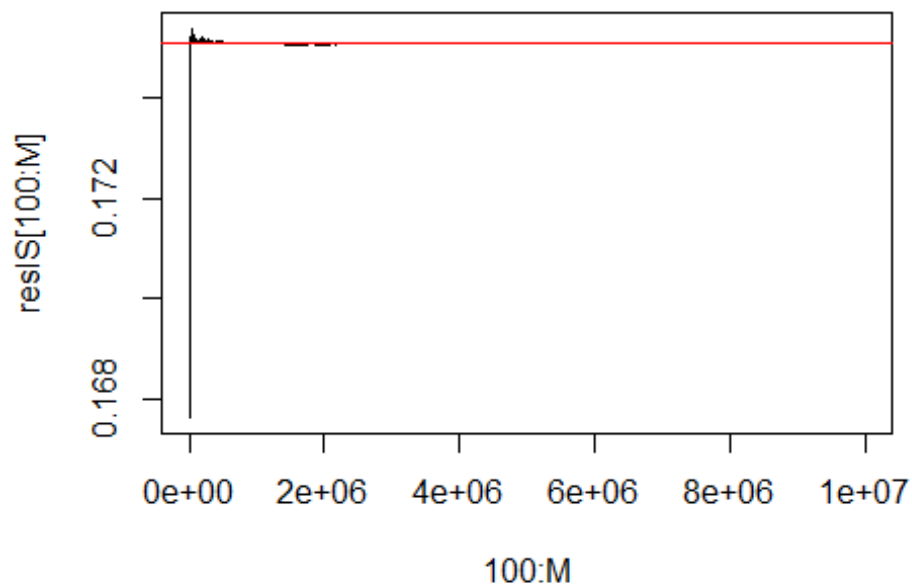
## [1] 0.1750963

log10(resMC[M])#on retrouve Le meme facteur de Bayes

## [1] -0.7569201

plot(100:M,resIS[100:M], type="l")
abline(h=resIS[M], col=2)

```



On retrouve le même résultat que précédemment avec une meilleure précision.

### III) Loi de Pareto Generalisee

#### 9) Choix de la loi a priori pour $\alpha$

Apparemment difficile d'obtenir une loi conjuguée ici,  $\tau$  étant un réel positif, on peut choisir une loi gamma pour  $\tau$

```
a=1
b=1
c=10#3
d=10#3

?dgamma
prior=function(beta){
  return(dgamma(beta[1], a, scale = b)*dgamma(beta[2], c, d))
}#beta correspond au vecteur (alpha, tau)

logprior=function(beta){
  return(dgamma(beta[1], a, b, log=T) + dgamma(beta[2], c, d, log = T))
}
```

#### 10) Loi a posteriori

$\tau$  et  $\alpha$  étant indépendantes, la loi a posteriori devient ne ressemble pas à une loi connue : on va avoir du mal à travailler avec directement

### 11) Echantillon de la loi a posteriori

On utilise un algorithme de Metropolis-Hastings

```
# loi a posteriori
posterior=function(beta, t){

  #p=pnorm(beta[1]+x*beta[2])
  #lkd=prod(p^y)*prod((1-p)^(1-y))

  lkd=((beta[1]/(m*beta[2]))^nrow(data))*prod((1+(t-m)/(beta[2]*m))^( -(beta[1]+1)))
  return(lkd*prior(beta))
}

logposterior=function(beta,t){

  #p=pnorm(beta[1]+x*beta[2])

  #loglkd=sum(y*log(p)) + sum((1-y)*log(1-p))

  loglkd=nrow(data)*log(beta[1]/(m*beta[2]))+(-(beta[1]+1))*sum(log(1+(t-m)/(beta[2]*m)))
  #print(nrow(data)*log(beta[1]/(m*beta[2])))
  if(!is.finite(loglkd)) return(-Inf)

  return(loglkd+logprior(beta))
}

t0=data$Barre

beta_0=c(0.5,1)

logposterior(beta_0,t0)

## [1] -3648.125

#print(rbind(c(1,0),c(0,1)))
#rm(beta0)

# algorithme de Metropolis-Hastings

require(mvtnorm)

## Loading required package: mvtnorm
```

```

MH=function(beta0, niter, sigma,t){
  beta=matrix(NA, nrow=niter, ncol=2)
  beta[1,]=beta0
  acc=0 # nombre d'acceptations

  Id=rbind(c(1,0),c(0,1))

  for(i in 2:niter){
    proposal=rmvnorm(1, beta[i-1,], sigma^2*Id)
    #print(proposal)
    #print(logposterior(proposal, t))
    #print(logposterior(beta[i-1,], t))
    logalpha=logposterior(proposal, t)-logposterior(beta[i-1,], t)
    #print(logalpha)
    if(log(runif(1))<logalpha){
      beta[i,]=proposal
      acc=acc+1
    }
    else{
      beta[i,]=beta[i-1,]
    }
  }
  print(acc/niter) #proportion d'acceptations
  return(beta)
}

niter=2e4
b1=MH(c(1,5), niter, 0.24,data$Barre)

## [1] 0.2401

b2=MH(c(1,5), niter, 0.23,data$Barre)

## [1] 0.2415

#b3=MH(c(1,1), niter, 0.38,data$Barre)
b3=MH(c(1,5), niter, 0.22,data$Barre)

## [1] 0.262

#Etudions la sortie de l'algorithme

#pour alpha :
par(mfcol=c(3,3))
i=1
# trace
plot(b1[,i], type="l")

plot(b2[,i], type="l")
plot(b3[,i], type="l")

```

```

# autocorrelations
acf(b1[100:niter,i],lag.max = 300)
acf(b2[100:niter,i],lag.max = 300)
acf(b3[100:niter,i],lag.max = 300)

# histogrammes
hist(b1[100:niter,i], breaks=50,lag.max = 300)

## Warning in plot.window(xlim, ylim, "", ...): "lag.max" n'est pas un
## paramètre graphique

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "lag.max" n'est pas un paramètre graphique

## Warning in axis(1, ...): "lag.max" n'est pas un paramètre graphique
## Warning in axis(2, ...): "lag.max" n'est pas un paramètre graphique

hist(b2[100:niter,i], breaks=50,lag.max = 300)

## Warning in plot.window(xlim, ylim, "", ...): "lag.max" n'est pas un
## paramètre graphique

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "lag.max" n'est pas un paramètre graphique

## Warning in axis(1, ...): "lag.max" n'est pas un paramètre graphique
## Warning in axis(2, ...): "lag.max" n'est pas un paramètre graphique

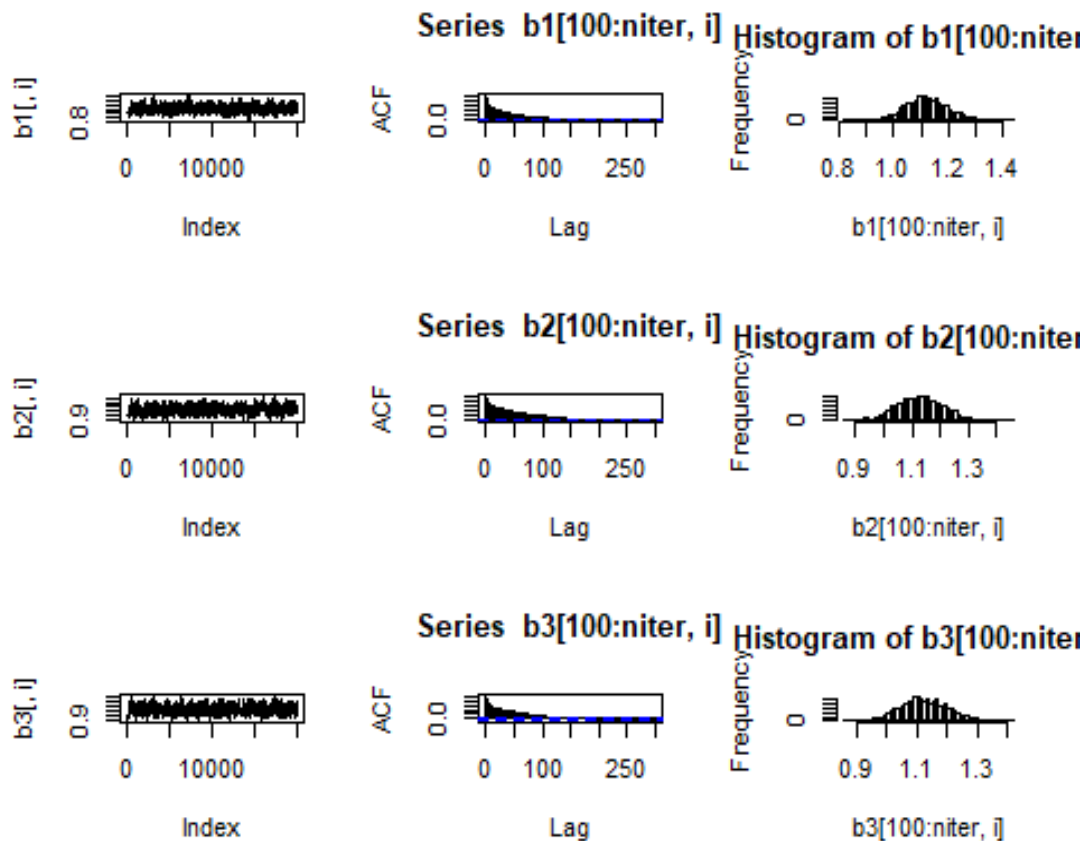
hist(b3[100:niter,i], breaks=50,lag.max = 300)

## Warning in plot.window(xlim, ylim, "", ...): "lag.max" n'est pas un
## paramètre graphique

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "lag.max" n'est pas un paramètre graphique

## Warning in axis(1, ...): "lag.max" n'est pas un paramètre graphique
## Warning in axis(2, ...): "lag.max" n'est pas un paramètre graphique

```



*#Etudions la sortie de L'algorithme*

*#pour tau :*

`par(mfcol=c(3,3))`

`i=2`

*# trace*

`plot(b1[,i], type="l")`

`plot(b2[,i], type="l")`

`plot(b3[,i], type="l")`

*# autocorrÃ©lations*

`acf(b1[100:niter,i],lag.max = 300)`

`acf(b2[100:niter,i],lag.max = 300)`

`acf(b3[100:niter,i],lag.max = 300)`

*# histogrammes*

`hist(b1[100:niter,i], breaks=50,lag.max = 300)`

## Warning in plot.window(xlim, ylim, "", ...): "lag.max" n'est pas un  
## paramètre graphique

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):  
## "lag.max" n'est pas un paramètre graphique



```
## Warning in axis(1, ...): "lag.max" n'est pas un paramètre graphique
## Warning in axis(2, ...): "lag.max" n'est pas un paramètre graphique
hist(b2[100:niter,i], breaks=50,lag.max = 300)

## Warning in plot.window(xlim, ylim, "", ...): "lag.max" n'est pas un
## paramètre graphique

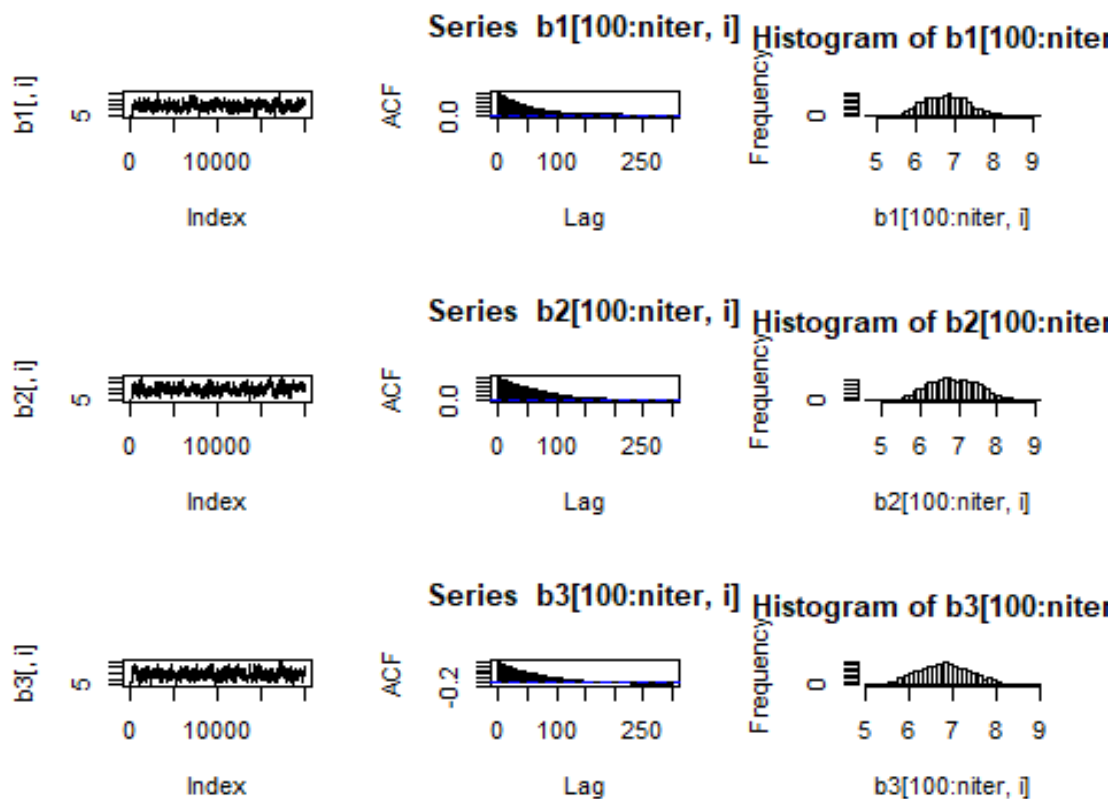
## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "lag.max" n'est pas un paramètre graphique

## Warning in axis(1, ...): "lag.max" n'est pas un paramètre graphique
## Warning in axis(2, ...): "lag.max" n'est pas un paramètre graphique
hist(b3[100:niter,i], breaks=50,lag.max = 300)

## Warning in plot.window(xlim, ylim, "", ...): "lag.max" n'est pas un
## paramètre graphique

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "lag.max" n'est pas un paramètre graphique

## Warning in axis(1, ...): "lag.max" n'est pas un paramètre graphique
## Warning in axis(2, ...): "lag.max" n'est pas un paramètre graphique
```



Interprétation : nous avons dans ces 3 simulations des chaines de Markov qui explorent bien toutes les lois, des autocorrélations qui décroissent vite vers 0 et des taux d'acceptation proches de 0.234 (conformément à la loi de Robert, Gelman et Gilks) : l'algorithme de Metropolis-Hastings a bien fonctionné.

Remarque : il semblerait que les paramètres  $a$ ,  $b$ ,  $c$  et  $d$  des lois à priori aient une influence sur les valeurs prises par  $\alpha$  et  $\tau$ , ainsi que la variance a posteriori..

```
# Effective Sample Size
niter/(2*sum(acf(b1[100:niter,1], plot=F)$acf)-1)
## [1] 532.0766

niter/(2*sum(acf(b2[100:niter,1], plot=F)$acf)-1)
## [1] 448.0705

niter/(2*sum(acf(b3[100:niter,1], plot=F)$acf)-1)
## [1] 496.7695

#Esperance et Variance de alpha et tau

colMeans(b1[100:niter,])
## [1] 1.120897 6.786428

var(b1[100:niter,])
##           [,1]      [,2]
## [1,] 0.00618944 0.03845107
## [2,] 0.03845107 0.39730526
```

Les variances sont plutôt faibles.

## 12) Hypothèse sur $\tau$

Au vu de l'espérance a posteriori de  $\tau$ , et de la variance, il est peu probable que  $\tau$  prenne la valeur 1.

Question : ce modèle ( $\tau$ ,  $\alpha$ ) est-il meilleur que celui de la question précédente ( $\tau=1$ ,  $\alpha$  environ égal à 0.5) ? On peut penser que comme dans le modèle précédent on avait fixé le paramètre  $\tau$  à 1, et qu'ici nous avons 2 degrés de liberté, il est fort plausible que le modèle de Pareto généralisé généré avec l'algorithme de Metropolis-Hastings avec  $\alpha$ ,  $\tau$  ayant des espérances de respectivement environ 1.1 et 6.9 soit meilleur que le modèle avec  $\tau=1$ .

Pour le vérifier on pourrait calculer un nouveau facteur de Bayes (que je n'arrive pas à calculer sur R) entre le modèle de la partie 2 et le modèle Pareto généralisé.