```
#        Cours de Stat Mass 08: Intervalle de confiance pour la fréquence d'un évènement
# RAPPEL: tout ce qui est à droite d'un "dièse"="hash" est ignoré par l'ordinateur
# et n'est écrit là que pour l'humain qui lit ce texte ;-)
 help("runif")
# dunif(x, min = 0, max = 1, log = FALSE)
# punif(q, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
# qunif(p, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
# runif(n, min = 0, max = 1)
# Arguments
# x, q vector of quantiles
# p vector of probabilities.
# n number of observations. If length(n) > 1, the length is taken to be the number required.

# on suppose qu'une proportion p de la population est en faveur d'un référendum
# lorsqu'on demande son avis à une personne au hazard on estime donc q'elle a une
# chance sur p d'être favorable et on estime donc p par la proportion de gens
# qui répondent qu'ils sont en faveur du referendum
# Deux mises en place de simulation de sondages: n=taille de l'échantillon simulé
p=0.54;n=1000
x=floor(p+runif(n));mean(x) # une méthode artisanale
x=rbinom(n,1,p);mean(x)     # la methode R (Bernoulli=binomiale(1,p))
# Voyons ce qui peut se passer, en ne faisant pas un seul, mais S sondages
S=500 # S désigne le nombre de sondages avec échantillons de taille n
Taux.Sondage=numeric(S) # creation de la variable
for (s in 1:S) # une méthode artisanale
{x=floor(p+runif(n));
Taux.Sondage[s]=mean(x)} # fin de la méthode artisanale
# voici l'histogramme des résultats des S sondages
hist(Taux.Sondage,freq=FALSE)
# ajustons deux gaussienne à cet histogramme
# si on connait p :
sigma.th=sqrt(p*(1-p)/n);sigma.th
curve(dnorm(x,p,sigma.th),add=TRUE,col="red")
# si on estime  p  au moyen de S expériences
mu=mean(Taux.Sondage);mu
sigma=sd(Taux.Sondage);sigma
curve(dnorm(x,mu,sigma),add=TRUE,col="green")
# voici la probabilité d'annoncer (à tors, puisque p=0.54) que p<0.5
pnorm(0.5,p,sigma.th)
# voici le nombre de sondages, parmi les S, qui ont donné le mauvais résultat
sum((Taux.Sondage<0.5))
# voici le taux d'erreur correspondant, à rapprocher de la proba théorique
sum((Taux.Sondage<0.5))/S

################## Un autre cas  favorable pour les sondeurs:
p=0.527;n=1000; # ici la bonne réponse pourra être avancée au seuil alpha=5%
sigma.th=sqrt(p*(1-p)/n);sigma.th
pnorm(0.5,p,sigma.th)

########################
# Le cauchemar des sondeurs
p=0.51;n=1000
x=floor(p+runif(n));mean(x) # une méthode artisanale
x=rbinom(n,1,p);mean(x)     # la methode R (Bernoulli=binomiale(1,p))
# notez que la différence n'est due qu'à ce qu'
# on appelle la fluctuation d'échantillon.
# Voyons ce qui peut se passer, en ne faisant pas un seul, mais S sondages
S=500 # S désigne le nombre de sondages avec échantillons de taille n
Taux.Sondage=numeric(S) # creation de la variable
for (s in 1:S) # une méthode artisanale
{x=floor(p+runif(n));
Taux.Sondage[s]=mean(x)} # fin de la méthode artisanale
# voici l'histogramme des résultats des S sondages
hist(Taux.Sondage,freq=FALSE)
# ajustons une gaussienne à cet histogramme
sigma.th=sqrt(p*(1-p)/n);sigma.th
curve(dnorm(x,p,sigma.th),add=TRUE,col="red")
mu=mean(Taux.Sondage);mu
sigma=sd(Taux.Sondage);sigma
curve(dnorm(x,mu,sigma),add=TRUE,col="green")
# voici la probabilité d'annoncer (à tors, puisque p=0.51) que p<0.5
pnorm(0.5,p,sigma.th)
# voici le nombre de sondages, parmi les S, qui ont donné le mauvais résultat
sum((Taux.Sondage<0.5))
```

```
# voici le taux d'erreur correspondant, à rapprocher de la proba théorique
sum((Taux.Sondage<0.5))/S

# influence de la taille n de l'échantillon:
# à partir de quel p>0.5 peut-on assurer au seuil alpha qu'il y a une majorité favorable
p.min=0.4;p.max=0.6
pp=seq(p.min,p.max,0.01) # des valeurs de p de p.min a p.max par pas de 0.001
mon.sigma.th=function(p,n){return(sqrt(p*(1-p)/n))}
majorité=function(x){return(0.5+0*x)} # la majorité à 50% NB l'introduction de x
curve(majorité,0.4,0.6) # comment tracer le niveau 0.5
alpha=0.05 # le seuil choisi
points(pp,qnorm(alpha,pp,mon.sigma.th(pp,n)),pch=1) # des ronds pour n=1000
points(pp,qnorm(alpha,pp,mon.sigma.th(pp,4*n)),pch=2) # des triangles (n=4000)
points(pp,qnorm(alpha,pp,mon.sigma.th(pp,16*n)),pch=3) # des + (n=16000)

# Compléments
# la méthode binomiale pour effectuer la même expérience
Sondages=rbinom(S,n,p);Sondages # notez que le pluriel "Sondages"<>"Sondage"
hist(Sondages)
Taux.Sondages=Sondages/n
hist(Taux.Sondages,freq=FALSE)
sum(Sondages<n/2)  # nombre de sondages annonçant une issue erronée
sum(Sondages<n/2)/S  # taux d'erreur
mu=mean(Taux.Sondages);mu
sigma=sd(Taux.Sondages);sigma
curve(dnorm(x,mu,sigma),add=TRUE,col="green")

pnorm(0.5,mu,sigma)# proba d'annoncer un rejet
rbinom(100,1,0.5) # un échantillon de taille 100 de Bernoulli pour p=0.5
sum(dbinom(46:54, 100, 0.5))# Computes P(45 < X < 55) for X Binomial(100,0.5)
```