

Feuille-réponse 2 Approximation gaussienne d'un échantillon

Remarque générale importante : Il est impératif de donner pour chaque question **une explication** telle *j'utilise la commande ...*, ou *je lis sur la figure* et éventuellement les commandes ayant produit le résultat ou la figure, si elles n'ont pas été données.

1 Exploration des densités gaussiennes

1. Que produisent les deux lignes suivantes ?

```
mu=0;sigma=1;gaussienne1=function(x) {return (dnorm(x,mu,sigma))}  
plot(gaussienne1,-3,3)
```

2. Calculer la valeur de cette fonction en $x = 0$ et $x = -3$.

3. A l'aide des deux commandes suivantes, ajouter une autre gaussienne ayant un μ différent et comparer les deux graphes.

```
mu=2;gaussienne2=function(x) {return (dnorm(x,mu,sigma))}  
plot(gaussienne2,-3,3,col="red",add=TRUE)
```

4. Ajouter aux deux gaussiennes précédentes, en bleu, une troisième pour laquelle $\mu = -1$ cette fois et expliquer ce que produit le changement de μ sur la forme de la gaussienne.

5. Tracer à nouveau la première gaussienne (qu'on appelle "centrée-réduite") et lui ajouter cette fois, en rouge et en bleu, deux autres gaussiennes ayant le même μ mais des sigmas valant respectivement $2/3$ et $5/3$. Expliquer comment évoluent les gaussiennes en fonction de σ .

2 Histogrammes des rendements observés

On reprend le fichier de données appelé “rubber.csv” correspondant aux prix et rendements journaliers du caoutchouc que vous pourrez charger à nouveau au moyen de la commande `mydata<-read.csv2(file.choose())` puis désigner par x et y les prix et rendements observés.

1. A l'aide de la commande `h=hist(y)` tracer l'histogramme des rendements (en effectifs) puis expliquer ce que produisent les commandes `names(h)` et `h$counts`.
2. Donner la liste des prix appartenant à la classe la plus petite.
3. A l'aide de la commande `d=hist(x,freq=F)` tracer cette fois l'histogramme en densité puis indiquer ce que produit la commande `d$density` et `sum(d$density)`. Expliquer comment sont calculées les densités de chaque classe.

3 Ajustement d'une gaussienne aux rendements observés

On considère à présent le code suivant, où N est la longueur du vecteur y :

```
yy=y[2:N]
hist(yy,freq=F)
mu=mean(yy);sigma=sd(yy)
gaussienne<-function(x)return(dnorm(x,mu,sigma))
plot(gaussienne,min(yy),max(yy),col="red",add=T)
```

1. Quelle est la différence entre y et yy ; pourquoi avoir remplacé y par yy ?
2. Décrire la figure obtenue et expliquez pourquoi avoir choisi ces valeurs pour μ et σ . Essayer avec d'autres valeurs de ces 2 paramètres. Qu'observez-vous ?
3. Que se passe-t-il si l'histogramme est tracé sans l'option `freq=F` ? Expliquer.
4. En revenant cette fois aux prix (et non aux rendements), tracer l'histogramme des prix en densité et ajuster une gaussienne. Quelles valeurs de μ et σ choisissez-vous ?
5. A votre avis, l'ajustement d'un modèle gaussien est-il plus adapté pour les prix ou pour les rendements ? Expliquer.