Université de Nice — L1MASS, année 2014-2015
Département de Mathématiques — Statistique
NOM : Date : .
PRENOM : Groupe : .

Feuille-réponse 2
Approximation gaussienne d'un échantillon

Remarque générale importante : Lorsqu'on vous demande une valeur, il est impératif de donner une explication telle **j'utilise la commande ...**, ou **je lis sur la figure** et éventuellement donner la commande ayant produit cette figure, si elle n'a pas été donnée avant, etc.

1 Le caractère `Wr.Hnd` (empan de la main qui écrit) : effet de la largeur des classes

Toutes les données de cette séance sont dans le fichier `survey` de la bibliothèque `MASS` qui figure dans toutes les distributions du logiciel `R`. Nous n'utiliserons que les individus pour lesquels on dispose de tous les caractères collectés (`complete.case`). Rappelons que pour former l'échantillon `x` du caractère `Wr.Hnd` il suffit d'utiliser les commandes suivantes :

```
library(MASS) # MASS = Modern Applied Statistics with S
data(survey) #lecture des données de survey
survey.cc <- survey[complete.cases(survey),]
x <- survey.cc$Wr.Hnd
```

1. Que produit la commande `hist(x)` ? Précisez le nombre et la largeur des classes choisies par `R`. Que représente la hauteur des barres représentées ?

2. Qu'est-ce qui change lorsqu'on utilise la commande `hist(x,freq=F)` ? Comment traduit-on en français le mot `freq`(=frequency). Que signifie le `F` ? Expliquez ce choix de syntaxe.

3. Donnez la hauteur approximative de la plus grande barre. Expliquez cette hauteur (comparez avec `hist(x)` et l'effectif total de l'échantillon).

4. Que produisent les trois commandes suivantes ?

```
mu=mean(x);mu
sigma=sd(x);sigma
magaussienne=function(x) {return (dnorm(x,mu,sigma))}
```

5. Exécutez la commande suivante (après être revenu à un histogramme des densités)

```
plot(magaussienne,min(x),max(x),col="red",add=T)
```

Pourquoi a-t-on mis l'option `add=T` ? Commentez le résultat obtenu (nature et position de la courbe rouge ajoutée par rapport à l'histogramme).

6. On choisit à présent la largeur des barres, au moyen de l'option `breaks=`. Exécutez les instructions suivantes :

```
largeur=2
bornes=seq(xmin,xmax,largeur)
hist(x,breaks=bornes,freq=F)
plot(magaussienne,min(x),max(x),col="red",add=T)
```

Que vaut `bornes` ? Commentez le graphique obtenu, notamment la qualité de l'approximation de l'histogramme par une gaussienne.

7. Recommencez avec `largeur` égale à 3 puis 4. Comparez avec la figure précédente.

8. Recommencez avec `largeur` égale à 5. Expliquez.

9. Recommencez avec `largeur` égale à 0.5 puis 0.05. Comparez avec les résultats précédents.

2 Le caractère `Height` (taille de l'individu)

Former le caractère `u` constitué des valeurs de `Height` (la taille) et représenter son histogramme.

– combien y a-t-il de classes ? Quelle est la largeur $b_k - b_{k-1}$ des classes de cet histogramme-ci ? Quelles sont les valeurs extrêmes du caractère `u` ?

– Calculer et donner la taille moyenne `mu` de cet échantillon. Appartient-elle à la classe d'effectif maximal ?

– Calculer et donner l'écart-type `sigma` de cet échantillon. Y a-t-il des individus en dehors de l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$? Si oui, quelle est leur taille ?

3 Le caractère “Age”

1. Former le caractère `v` constitué des valeurs de `Age` et représenter l'histogramme des densités. Qu'obervez vous ?

2. Définir la fonction `magaussienneAge`. Donnez ici votre code.

3. Ajoutez le graphe de cette fonction à votre histogramme. Commentez la figure obtenue.