Examen

Exercice 1 *Soit $(Y_1, \dots, Y_n)$ un échantillon de v.a. de loi géométrique de paramètre $\theta \in]0,1[$, où θ est un paramètre inconnu. On a donc $P(Y_1 = k) = (1-\theta)\theta^k$ pour $k \geq 0$.*

1. *Exprimer la vraisemblance de ce modèle.*
2. *Donner l'estimateur du maximum de vraisemblance de θ.*
3. *Montrer que cet estimateur est fortement consistant.*
4. *On suppose désormais que l'on observe Y_i seulement si $Y_i < C$ où $C > 0$ est un entier fixé connu. Dans ce modèle, les observations sont les variables aléatoires $L_i := \min(Y_i, C)$.*
   (a) *Que peut-on dire de Y_i si $L_i = C$?*
   (b) *Calculer la probabilité $P(L_i = \ell)$ pour ℓ un entier entre 0 et C. (On discutera selon que $\ell < C$ ou $\ell = C$).*
   (c) *Calculer la vraisemblance du modèle $P(L_i = \ell_i,\, i \in [\![1, n]\!])$. On exprimera cette vraisemblance en fonction de $L := \sum_{i=1}^n \ell_i$ et de $D := \sum_{i=1}^n \mathbf{1}_{\{\ell_i < C\}}$.*
   (d) *En déduire l'estimateur du maximum de vraisemblance de θ dans ce modèle.*

Exercice 2

1. *On teste l'efficacité d'un régime en relevant le poids de n personnes avant et après le régime, notés respectivement $(X_i)_{i=1\dots n}$ et $(Y_i)_{i=1\dots n}$. En supposant que les variables suivent une loi Gaussienne, construire un test statistique pour savoir si le régime est efficace.*
2. *On appelle loi de Bendford la loi sur l'ensemble $\{1, 2, \dots, 9\}$ définie par $p_k = \log_{10}\left(1 + \frac{1}{k}\right)$. La loi de Bendford est connue pour apparaître naturellement dans certaines séries statistiques: en considérant le premier chiffre des données de telles séries, on observe que celui-ci suit à peu près la loi de Bendford. On vous donne une série statistique $X_1, \dots, X_n$. Construire un test statistique permettant de savoir si cette série vérifie bien cette propriété.*

Exercice 3 *Soit la régression linéaire simple* $y_i = ax_i + b + \varepsilon_i$, $i = 1 \dots n$. *On suppose que les* (ε_i) *sont i.i.d. de loi* $\mathcal{N}(0, \sigma^2)$. *(les* x_i *sont supposés non aléatoires). On introduit un paramètre* $\lambda > 0$. *Pour éviter d'avoir une grande variance de l'estimateur de* a, *on choisit des estimateurs* $\tilde{a}$, $\tilde{b}$ *de* a *et* b *qui minimisent la fonction*

$$(a,b) \to \sum_{i=1}^{n} (y_i - (ax_i + b))^2 + \lambda a^2.$$

1. *Trouver l'expression de* $\tilde{a}$ *et* $\tilde{b}$. *On pourra introduire* $S_{xy} := \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ *et* $S_{xx} := \sum_{i=1}^{n}(x_i - \bar{x})^2$.

2. *Calculer le biais de* $\tilde{a}$.

3. *Vérifier que* $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})y_i$. *Calculer la variance de* $\tilde{a}$.

4. *Quelle est la loi de* $\tilde{a}$*?*

5. *Construire un test statistique de* $a = 0$ *contre* $a \neq 0$ *reposant sur l'estimateur* $\tilde{a}$.

Exercice 4 *On veut comparer l'effet de régimes alimentaires sur le poids de poulets. On dispose de deux groupes de poulets: l'un nourri au régime 1 (variable w1) et l'autre au régime 2 (variable w2). On répondra aux questions en s'aidant des sorties R ci-dessous.*

1. *Dire ce que représentent les graphiques sur la dernière page.*

2. *Peut-on supposer que les observations suivent une loi Gaussienne?*

3. *Peut-on supposer que les deux groupes possèdent la même variance? Quelle loi suit la statistique F dans var.test?*

4. *Quelle est le nombre de poulets dans les groupes 1 et 2?*

5. *Donner l'expression de la statistique utilisée dans t.test (on notera* (X_i) *et* (Y_i) *les poids des poulets dans les deux groupes).*

6. *Y a-t-il une différence entre les deux régimes (prendre un risque à 5%)?*

```
> summary(w1)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     91     125     160     174     205     305

> summary(w2)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   74.0   166.2   202.0   210.2   260.5   331.0
```

> ks.test(w1,'pnorm',mean(w1),sd(w1))

```
        One-sample Kolmogorov-Smirnov test

data:  w1
D = 0.11319, p-value = 0.7916
alternative hypothesis: two-sided
```

> ks.test(w2,'pnorm',mean(w2),sd(w2))

```
        One-sample Kolmogorov-Smirnov test

data:  w2
D = 0.10327, p-value = 0.9684
alternative hypothesis: two-sided
```

> var.test(w1,w2)

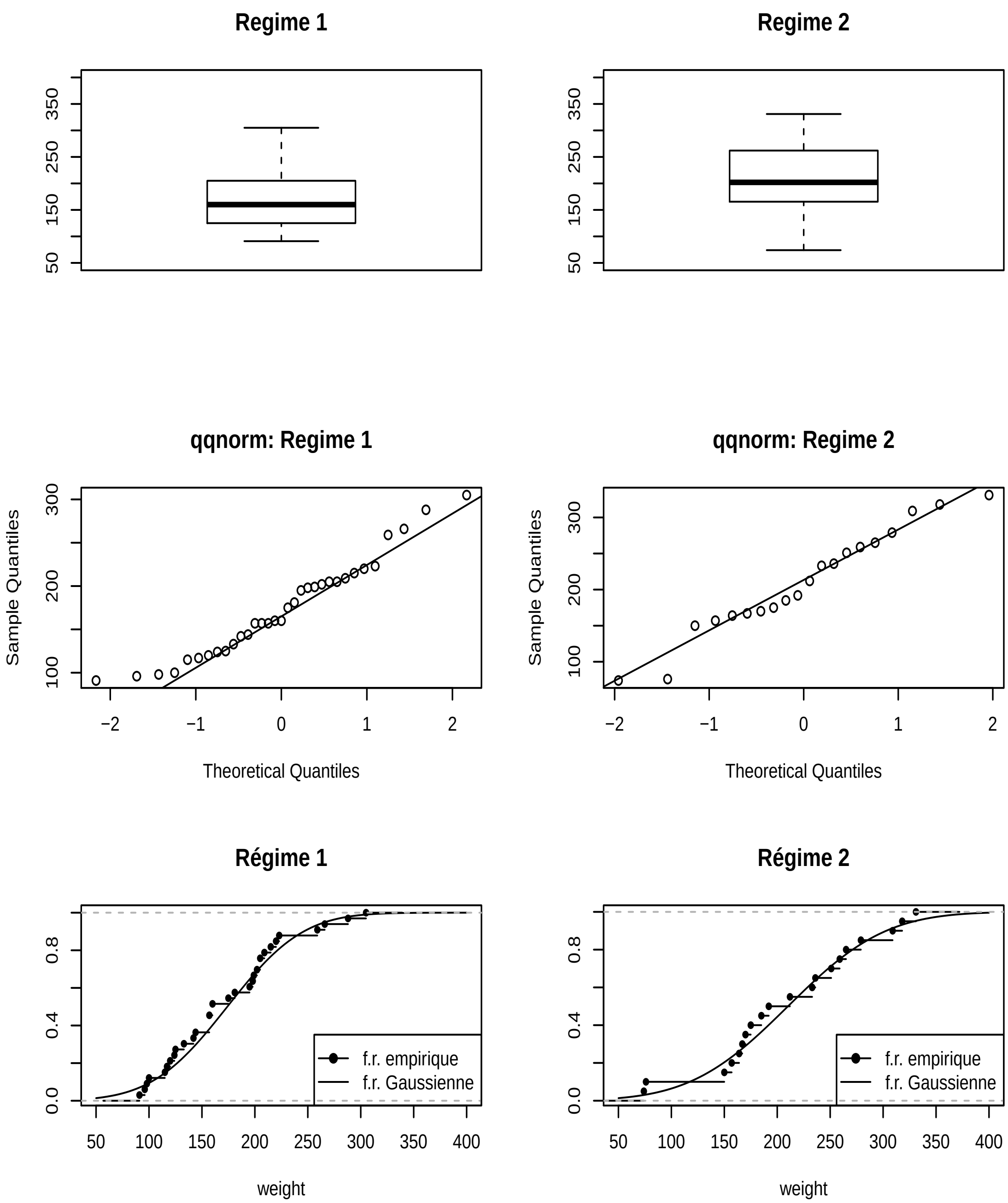
```
        F test to compare two variances

data:  w1 and w2
F = 0.6028, num df = 32, denom df = 19, p-value = 0.201
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2534088 1.3148254
sample estimates:
ratio of variances
         0.6027972
```

> t.test(w1,w2,var.equal=T)

```
        Two Sample t-test

data:  w1 and w2
t = -2.0333, df = 51, p-value = 0.04724
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -71.9030588  -0.4575473
sample estimates:
mean of x mean of y
 173.9697  210.1500
```


Regime 1
Regime 2
50
150
250
350
qqnorm: Regime 1
qqnorm: Regime 2
Sample Quantiles
Theoretical Quantiles
100
200
300
−2
−1
0
1
2
Régime 1
Régime 2
f.r. empirique
f.r. Gaussienne
0.0
0.4
0.8
50
100
150
200
250
300
350
400
weight