Examen: Introduction à la probabilité et aux statistiques

M2:Masco 2025-2026

28-11-2025

Exercice 1:

Dans un jeu de hasard où l'on tire à pile ou face, on gagne lorsque l'on obtient un pile et on perd lorsque l'on obtient un face. On choisit au hasard une pièce dans un sac contenant des pièces truquées de deux types A et B. La probabilité de gagner avec une pièce de type A est de 80%, et celle de gagner avec une pièce de type B est de 10%. On sait également que la probabilité de choisir une pièce de type A est de 30% et celle de choisir une pièce de type B est de 70%.

1. On choisit une pièce au hasard et on la lance une fois. Quelle est la probabilité de perdre ?
2. Sachant que l'on gagne, quelle est la probabilité d'avoir choisi une pièce de type A ?

Exercice 2:

On modélise le nombre de *spikes* au cours d'une expérience par une variable aléatoire X suivant une loi de Poisson de paramètre inconnu $\Theta = \theta > 0$. On écrit :

$$P(X = x \mid \Theta = \theta) = e^{-\theta}\,\frac{\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

À partir de connaissances préalables sur le phénomène, on sait que la variable aléatoire Θ est discrète et prend les valeurs $\{0.1,\ 0.5,\ 1,\ 2\}$ avec les probabilités associées $\{0.25,\ 0.25,\ 0.4,\ 0.1\}$

1. Lors d'une première expérience, on observe 2 spikes. Donner la distribution a posteriori de Θ.
2. Déterminer la moyenne a posteriori de Θ.
3. Calculer la probabilité a posteriori que Θ soit strictement inférieur à 1.
4. On répète l'expérience et on observe cette fois 1 spike. Les observations étant supposées indépendantes, déterminer la distribution a posteriori de Θ sur base des deux expériences.

Exercice 3:

La modification des conditions météorologiques, ou l'ensemencement des nuages, consiste à traiter des nuages individuels ou des systèmes orageux avec divers matériaux inorganiques et organiques dans l'espoir d'obtenir une augmentation des précipitations. Le jeu de données `clouds.csv` contient deux variables donnant le logarithme des précipitations mesurées en acre-foot (environ $1233\,\mathrm{m}^3$) pour 26 nuages non-ensemencés (`unseeded`) et 26 nuages ensemencés (`seeded`).

```r
clouds <- read.csv("clouds.csv")
X <- clouds$unseeded
Y <- clouds$seeded
head(clouds)
```

```
##   unseeded   seeded
## 1 7.092241 7.917755
```

```
## 2 6.721546 7.437089
## 3 5.919969 7.412160
## 4 5.844993 6.885510
## 5 5.772064 6.555926
## 6 5.498397 6.192567
```

1. Expliquez ce que fait chaque ligne du code ci-dessus.

Vous souhaitez examiner si l'ensemencement des nuages a un impact sur les précipitations.

2. Proposez une procédure pour étudier cela (énoncez l'hypothèse nulle, l'hypothèse alternative, les hypothèses nécessaires, etc.).

Vous décidez de réaliser les tests suivants

```r
#a. test 1
shapiro.test(X)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  X
## W = 0.97994, p-value = 0.8727
```

```r
#b. test 2
shapiro.test(Y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Y
## W = 0.96591, p-value = 0.5208
```

```r
#c. test 3
var.test(X,Y)
```

```
##
##  F test to compare two variances
##
## data:  X and Y
## F = 1.0536, num df = 25, denom df = 25, p-value = 0.8971
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4724176 2.3499219
## sample estimates:
## ratio of variances
##           1.053634
```

```r
#d. test 4
t.test(X,Y, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  X and Y
## t = -2.5444, df = 50, p-value = 0.01408
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.046697 -0.240865
```

```
## sample estimates:
## mean of x mean of y
##  3.990406  5.134187
```

3. Décrivez ce que chaque test permet de vérifier et pourquoi on choisit `var.equal = TRUE` pour le test 4.
4. Quelles sont vos conclusions ?

Exercice 4:

1. Voici les valeurs observées de deux échantillons (considérés comme indépendants):

$$A = (176, 391, 314, 248, 373, 393, 330), \qquad B = (307, 310, 374, 343, 360, 280).$$

   On réalise un test de Wilcoxon. Quelle est la valeur de la statistique de test ? Détaillez vos calculs.
2. Si l'on ajoute 200 à chaque élément de A et de B, la valeur de la statistique change-t-elle ?
3. On observe un échantillon $(X_1, \dots, X_n)$ de variables aléatoires indépendantes telles que $X_i \sim \text{Binomial}(10, \theta)$, $\theta \in (0, 1)$. On souhaite estimer θ. On vous propose l'estimateur

$$\hat{\theta} = \frac{n}{10(n-2)}\,\bar{X} - \frac{2}{n}, \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   Calculer le biais et la variance de cet estimateur. Commenter.
4. Comparer et discuter en quelques lignes les approches bayésienne et fréquentiste en inférence statistique.
5. Un expérimentateur affirme : "pour rejeter une hypothèse nulle d'égalité de moyennes, il me suffit de collecter des échantillons suffisamment grands". Commenter cette assertion.
6. Expliquez brièvement ce qu'est l'ANOVA bayésienne et à quoi elle sert.
7. Quel est le facteur bayésien pour deux modèles, M_1 et M_2 ? Comment interpréter $BF_{12} = 4$?
8. Si $BF_{01} = 3$ et $BF_{12} = 2$, quel modèle est le meilleur : M_0 ou M_2 ?