Eléments de Modélisation Statistique **2023-2024**
modIA 4A

Feuille de TD de Prérequis

1 Rappels de probabilité

1.1 Variables aléatoires réelles

Exercice 1
Soit X et Y deux variables aléatoires indépendantes. X suit une loi géométrique de paramètre $p \in]0,1[$ et Y satisfait la propriété suivante :

$$\mathbb{P}(Y=1) = 1 - \mathbb{P}(Y=-1) = q \text{ avec } q \in]0,1[.$$

1. Déterminez la loi de $Z = Y(X-1)$.
2. Calculez $\mathrm{Cov}(X,Y)$, $\mathrm{Cov}(Z,Y)$ et $\mathrm{Cov}(Z,X)$.

Exercice 2 - Loi de Pareto
Une variable aléatoire X suit une loi de Pareto de paramètres $a > 0$ et $\alpha > 0$ si elle admet la densité de probabilité suivante

$$f_X(x) = \frac{\alpha a^\alpha}{x^{\alpha+1}} \mathbb{1}_{[a,+\infty[}(x).$$

1. Vérifiez que f_X est bien une densité de probabilité.
2. Déterminez la fonction de répartition F_X de X.
3. Tracez les représentations graphiques de f_X et F_X.
4. Déterminez, si elles existent, $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

1.2 Convergences en loi et en probabilité

Exercice 1
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d de loi $\mathcal{B}(p)$. Montrez que

$$\sqrt{n}\ \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Exercice 2
Soit $X_1, \ldots, X_n$ des v.a. i.i.d de loi uniforme sur $[0,1]$.
On pose $Z_n = n \min(X_1, \ldots, X_n)$.
Déterminez pour tout $n \geq 1$ la fonction de répartition de Z_n et étudiez la convergence en loi de $(Z_n)_{n\geq 1}$.
Indication : Vous pouvez vérifier que la fonction de répartition d'une loi exponentielle de paramètre 1 vaut

$$F(t) = \left(1 - e^{-t}\right) \mathbb{1}_{t>0}.$$

2 Le monde gaussien

2.1 Loi gaussienne unidimensionnelle et lois associées

Exercice 1
Un jardinier récolte des tomates dont le poids est modélisé par une loi normale de moyenne $m = 200$g et d'écart-type $\sigma = 40$g. Il dépose les tomates cueillies au fur et à mesure dans une caisse qui peut supporter un poids maximal de 2kg.

1. Quelle est la probabilité que le jardinier cueille une tomate de plus de 250g.

2. Déterminer la valeur d telle que la probabilité que l'écart entre le poids d'une tomate et m dépasse d soit égale à 0.1.

3. Déterminer le nombre maximum de tomates que l'on peut cueillir pour que la probabilité de la surcharge de la caisse n'excède pas 0.01.

Exercice 2 - Table de $\mathcal{N}(0,1)$

1. Soit X une v.a.r de loi normale centrée réduite. Calculer $\mathbb{P}(-0.3 < X < 0.1)$ et déterminer t pour que $\mathbb{P}(X > t) = 0.7$.

2. Soit Y une v.a.r de loi normale, d'espérance 1 et de variance 4.
Calculer $\mathbb{P}(-1 < Y < 2)$, $\mathbb{P}(Y^2 < 4)$ et déterminer t pour que $\mathbb{P}(0 < Y < t) = 0.5$.

Exercice 3
Soit X, Y suivent une loi $\mathcal{N}(0,1)$, $Z \sim \mathcal{N}(1,2)$, X, Y et Z sont indépendants. Déterminez la loi des variables aléatoires suivantes :

$$A = X - Z \qquad B = X^2 + Y^2$$

$$C = \frac{Z-1}{\sqrt{X^2+Y^2}} \qquad D = \frac{(Z-1)^2}{X^2+Y^2}$$

2.2 Vecteurs gaussiens

Exercice 1
Soient $X_1, X_2, \ldots, X_n$ des v.a. i.i.d de loi $\mathcal{N}(m, \sigma^2)$. Soient

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \text{ et } S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

1. Montrer à l'aide des propriétés des vecteurs gaussiens que

$$\bar{X}_n \sim \mathcal{N}(m, \frac{\sigma^2}{n})$$

2. En appliquant le théorème de Cochran sur un vecteur gaussien Y bien choisi et en considérant la décomposition orthogonale de $\mathbb{R}^n = E_1 \oplus E_1^{\perp}$ avec $E_1 = \text{Vect}\,((1,1,\ldots,1)')$, montrez que

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1) \text{ et } \bar{X}_n \perp\!\!\!\perp S_n^2$$

Exercice 2
Soit $U = (X, Y, Z)'$ un vecteur gaussien d'espérance m et de matrice de variance-covariance Σ où

$$m = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \text{ avec } 0 \leq \rho$$

1. Quelle est la loi de Z ?
2. Pour quelle(s) valeur(s) de ρ les variables aléatoires X, Y et Z sont-elles indépendantes?
3. Soit $V = \begin{pmatrix} X + Y + \alpha \\ X - Y + \sqrt{2\rho} Z \end{pmatrix}$ avec $\alpha \in \mathbb{R}$. Déterminez la loi de V.
4. Que peut-on en conclure sur les coordonnées du vecteur V ?
5. Pour quelles valeurs de α et ρ, le vecteur V est-il centré ?
6. Dans ce cas, on obtient que $V \sim \mathcal{N}_2(0_2, 6I_2)$. Quelle est la loi de $\|V\|_2^2$?

3 Estimation ponctuelle et IC

3.1 Estimation ponctuelle

Exercice 1
Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$.

1. Déterminez l'EMV $\hat{\mu}$ pour le paramètre μ.
2. L'estimateur $\hat{\mu}$ est-il sans biais ?
3. L'estimateur $\hat{\mu}$ est-il consistant ?
4. L'estimateur $\hat{\mu}$ est-il efficace ?

Exercice 2
Soit $(Y_1, \ldots, Y_n)$ un n-échantillon de densité de probabilité

$$f_Y(y) = \frac{|y|}{\mu} \exp\left(-\frac{y^2}{\mu}\right)$$

avec $\mu > 0$ paramètre inconnu.
Reprenez les questions de l'exercice précédent
Indication : vérifiez que $X = Y^2$ *suit une loi exponentielle* $\mathcal{E}(\frac{1}{\mu})$.

3.2 Estimation par intervalle de confiance

Exercice 1
Le responsable qualité d'une usine de boissons gazeuses s'intéresse à la stabilité du système de remplissage des bouteilles. Il décide de prélever $n = 16$ bouteilles remplies par la machine et mesure la hauteur du liquide dans chacune des bouteilles. On pose X_i la hauteur de liquide dans la $i$ème bouteille et on suppose que $(X_1, \ldots, X_{16})$ forment un 16-échantillon de loi $\mathcal{N}(m, \sigma^2)$. Les résultats numériques obtenus sont $\sum_{i=1}^{16} x_i = 376$ et $\sum_{i=1}^{16} x_i^2 = 8837$.

1. Proposez un estimateur pour m, précisez sa loi et donnez une estimation pour m.

2. On suppose dans cette question que $\sigma^2 = 0.05$. Construisez un intervalle de confiance pour m au niveau de confiance de 0,95.

3. Le responsable qualité met en doute la valeur de σ^2.

   a) Proposez un estimateur et une estimation pour σ^2.

   b) Construisez un intervalle de confiance pour m au niveau de confiance de 0,95 en supposant σ^2 inconnu.

Exercice 2
La durée de vie d'un composant électronique suit une loi exponentielle de paramètre $\frac{1}{\theta}$ avec $\theta > 0$. On observe les durées de vie de n composants notées $X_1, \ldots, X_n$.

1. Proposez un estimateur pour θ. Calculez l'espérance et la variance de cet estimateur.

2. Construisez un intervalle bilatéral asymptotique de niveau de confiance 0.95 pour θ.

4 Tests paramétriques

Exercice 1
Un fournisseur commercialise des canettes de soda de 33cl. Ce fournisseur livre une commande à un de ses clients. Lors de la livraison, le fournisseur et le client décident de contrôler la qualité des produits. Ils prélèvent pour cela un même échantillon $X_1, \ldots, X_n$ de taille n et mesurent la quantité de soda en cl dans les canettes. On suppose que les X_i suivent une loi $\mathcal{N}(\mu, \sigma^2)$, où l'écart-type σ est supposé connu égal à 0.7.

1. Le fournisseur veut savoir si la moyenne μ est bien égale à 33cl. Il souhaite contrôler le risque de se voir rejeter un lot conforme. écrivez les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ associées à cette problématique et construisez un test statistique.

2. Le client a lui besoin d'une quantité minimale de 32cl pour commercialiser les canettes de soda. De son point de vue, il souhaite pouvoir contrôler la probabilité d'accepter un lot non conforme. Par contre, il a des doutes sur la valeur annoncée de σ^2, il suppose donc cette quantité inconnue. Construisez un test statistique adéquate pour le client.

3. Sur un échantillon de taille 10, on trouve $\bar{x}_{10} = 32.5$ et $s_{10}^2 = 0.64$. Quelles sont les conclusions du client et du fournisseur?

Exercice 2
On s'intéresse à l'influence sur la consommation électrique de l'utilisation d'un adoucisseur d'eau pour la machine à laver. On dispose de $n = 25$ mesures de consommation d'électricité avec adoucisseur et $p = 17$ mesures sans adoucisseur. On suppose que les consommations avec adoucisseur $(X_1, \ldots, X_n)$ et sans adoucisseur $(Y_1, \ldots, Y_p)$ sont gaussiennes de même variance σ^2 inconnue et indépendantes.
Testez l'influence de l'utilisation d'un adoucisseur d'eau sur la consommation électrique d'une machine à laver.

```
Two Sample t-test
data:  X and Y
t = -2.8141, df = 40, p-value = 0.00755
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.16086264 -0.02638442
sample estimates:
mean of x mean of y
0.8052000 0.8988235
```

Exercice 3
On décide de tester l'hypothèse qu'une pièce de monnaie est parfaite, en adoptant la règle de décision suivante : on accepte l'hypothèse de perfection si et seulement si le nombre de faces obtenues dans un échantillon de 100 jets est compris entre 40 et 60.

1. Quelle est la probabilité de rejeter l'hypothèse de perfection alors qu'elle est vraie ?

2. Quelle est la probabilité de ne pas rejeter l'hypothèse de perfection alors que la probabilité p d'avoir un face est de 0.7 ?