TD 1
Bases de probabilités et statistique

Exercice 1

Un sondage a receuilli des informations sur le prix (au kilogramme) de certains fruits et legumes. Le tableau suivant donne les effectifs pour chaque paire type / classe de prix.

	pomme	poire	courgette	aubergine
2 €	12	24	54	23
3 €	45	26	72	16
4 €	34	63	34	33

Par exemple, dans le sondage il y a 12 pommes qui coutent 2 € (au kilogramme).

1. Calculer la probabilité que le fruit/légume du sondage soit une pomme sachant que son prix est 2 €.
2. Calculer la probabilité que le fruit/légume du sondage coute 3 € ou plus sachant que c'est une aubergine.
3. Calculer la probabilité que le fruit/légume du sondage soit un fruit sachant qu'il coute 3 € ou plus.
4. Calculer l'espérance conditionelle d'un fruit/légume du sondage sachant que c'est une poire (le prix moyen d'une poire dans le sondage).
5. Calculer l'espérance conditionelle d'un fruit/légume du sondage sachant que c'est un légume (le prix moyen d'un légume dans le sondage).
6. Calculer la variance conditionelle d'un fruit/légume du sondage sachant que c'est une courgette.

Exercice 2

On considère un couple de variables aléatoires (X, Y) sur $[0,1]^2$ dont la densité de probabilité est la fonction $f_{X,Y} : [0,1]^2 \to \mathbb{R}^+$ définie par, pour $x, y \in [0,1]^2$,

$$f_{X,Y}(x,y) = \frac{1}{c}\exp(-|x-y|),$$

où $c > 0$ est une constante (ne dépendant pas de x, y).

1. Calculer la constante c.
2. Pour $x \in [0,1]$, calculer la densité de X en x, c'est à dire la fonction $f_X : [0,1] \to \mathbb{R}^+$.
3. Calculer la fonction de densité conditionelle de Y sachant X, c'est à dire la fonction $f_{Y|X} : [0,1]^2 \to \mathbb{R}^+$.
4. Calculer l'espérance de Y sachant que X vaut 1.
5. Calculer la variance de Y sachant que X vaut 1.
6. Calculer la probabilité que $Y \geq 1/2$ sachant que X vaut 1.

Elément de cours : la loi de Poisson La loi de Poisson de paramètre $\lambda > 0$, notée $\mathcal{P}(\lambda)$, est une loi de probabilité sur $\mathbb{N} = \{0, 1, 2, \ldots\}$. Pour une variable aléatoire X telle que $X \sim \mathcal{P}(\lambda)$ on a

$$\mathbb{P}[X = k] = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}$$

et

$$\mathbb{E}[X] = \lambda.$$

Exercice 3

Soient $X_1, \ldots, X_n$ iid selon une loi de Poisson $\mathcal{P}(\lambda)$, où λ est inconnu.

1. Matériel de base créé par François Bachoc et Adrien Mazoyer.

1. Pour $x_1, \dots, x_n \in \mathbb{N}^n$, calculer la probabilité que $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ en fonction de λ.
2. Le maximum de vraissemblance consiste à maximiser la probabilité précédente en fonction de λ, lorsque $(x_1, \dots, x_n)$ est fixé et égal à $(X_1, \dots, X_n)$ (ce dernier vecteur est appelé vecteur des observations). Calculer l'estimateur $\widehat{\lambda}_{\mathrm{ML}}$ du maximum de vraisemblance de λ.

Exercice 4

Soient $X_1, \dots, X_n$ iid selon une loi Uniforme $\mathcal{U}(0, t)$, où $t \geq 0$ est inconnu.

1. Pour $(x_1, \dots, x_n) \in [0, +\infty[^n$, calculer la valeur de la fonction densité de probabilité de $(X_1, \dots, X_n)$ évaluée en $(x_1, \dots, x_n)$, en fonction de t.
2. Le maximum de vraissemblance consiste à maximiser la densité de probabilité précédente en fonction de t, lorsque $(x_1, \dots, x_n)$ est fixé et égal à $(X_1, \dots, X_n)$ (ce dernier vecteur est appelé vecteur des observations).Calculer l'estimateur $\widehat{t}_{\mathrm{ML}}$ du maximum de vraisemblance de t.
3. On considère maintenant que les variables sont iid selon une loi Uniforme $\mathcal{U}(t, t+1)$. Montrer alors que, quel que soit $t \in \mathbb{R}$, presque surement (avec probabilité 1), $\max(X_1, \dots, X_n) \leq \min(X_1, \dots, X_n) + 1$.
4. Calculer la valeur de la fonction densité de probabilité de $(X_1, \dots, X_n)$ évaluée en $(x_1, \dots, x_n)$, en fonction de t, lorsque $\max(x_1, \dots, x_n) \leq \min(x_1, \dots, x_n) + 1$.
5. Trouver l'ensemble des t qui maximisent la vraissemblance (il peut y en avoir plusieurs).

Exercice 5

Une maladie se propage dans une population, avec un taux de contamination de 1 personne pour 1000. Un nouveau test de dépistage de cette maladie est proposé avec les taux de détection suivants. Une personne malade obtiendra bien un test positif avec probabilité 99%. Une personne saine en revanche pourra obtenir un résultat positif avec probabilité 0.2%.

Calculez la probabilité qu'une personne soit effectivement malade si son test est positif.