Statistique mathématique: TD2

Exercice 1. *Soient X et Y deux variables aléatoires admettant un moment d'ordre deux.*

1. *Montrer que*
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\mathrm{Cov}(X,Y) + \mathrm{Var}(Y).$$
2. *Si $X \perp\!\!\!\perp Y$, montrer que $\mathrm{Cov}(X,Y) = 0$. Soient $X \sim \mathcal{N}(0,1)$, Z une variable de Rademacher indépendante de X et $Y = ZX$. Montrer que les variables X et Y ne sont pas indépendantes.*
3. *Montrer que $\mathrm{Cov}(X,Y) = 0$.*
4. *En tirez-vous un enseignement sur la relation entre covariance et indépendance ?*

Exercice 2. *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in]0,1[$.*

1. *Calculez l'espérance et la variance de X_1.*
2. *Quelle est la loi de $\sum_{i=1}^n X_i$?*
3. *Soit Y_n une variable aléatoire binomiale $\mathcal{B}(n,p)$. En s'aidant des questions précédentes, calculer l'espérance de Y ? Quelle est la variance de Y ?*
4. *Que peut-on dire de la suite de variables $(Y_n/n)_{n\geq 1}$?*

Exercice 3. *Le nombre de pièeces sortant d'une usine en une journée est une variable aléatoire d'espérance 50. On veut estimer la probabilité que la production de demain dépasse 75 pièces.*

1. *En utilisant l'inégalité de Markov, quelle estimation obtient-on sur cette probabilité ?*
2. *Que peut-on dire de plus sur cette probabilité si on sait que l'écart-type de la production quotidienne est 5 ?*

Exercice 4. *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires de loi $\mathcal{E}(n)$. Montrez que la suite $(X_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers 0.*

Exercice 5. *On lance n fois un dé à 6 faces et on regarde le nombre d'obtentions de la face 6. Si on appelle X_n ce nombre, quelle est la loi de X_n ? La fréquence d'obtention de la face 6 est définie par $Y_n = X_n/n$. Quelle est l'espérance de Y_n ? Peut-on majorer la variance de Y_n ?*

Démontrer, à l'aide de l'inégalité de Tchebychev, que la suite de variables aléatoires $(Y_n)_{n\in\mathbb{N}}$ converge en probabilité vers $1/6$.

Exercice 6. *On sait qu'à chaque naissance, la probabilité p que l'enfant soit une fille est très légèrement inférieure à 50%. On veut estimer cette probabilité en observant un grand nombre, noté n, de naissances et en comptant le nombre X de filles parmi ces bébés.*

1. *En utilisant l'inégalité de Tchebychev, construire un intervalle de confiance pour p au niveau de confiance 99%.*

2. *On observe les naissances en France pendant une année, on constate que* 349105 *bébés sont des filles et* 364924 *des garcons (données INSEE de* 2019*). Quelle est la valeur observée de l'intervalle de confiance obtenu ?*
3. *Combien faudrait-il observer de naissances pour obtenir une estimation de la valeur de p à plus ou moins 10^{-3} près, toujours au niveau de confiance* 99% *?*

Exercice 7. *Soit $(X_i)_{i\leq 50}$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0, 20]$. Les X_i modélisent une note aléatoire obtenue à un examen de probabilité. Soit V_{50} le nombre d'étudiants et étudiantes ayant obtenu plus de* 10/20 *et T_{50} le nombre d'étudiants et étudiantes ayant obtenu plus de* 16/20*.*

1. *Quelle est loi loi de V_{50} ? de T_{50} ?*
2. *Quelles sont les espérances et variances de V_{50} et T_{50} ?*
3. *Donner une majoration de la probabilité que plus de* 30 *étudiants aient plus de* 16/20 *?*
4. *Donner un intervalle de confiance à* 95% *sur le nombre d'étudiants et étudiantes qui valident le cours.*
5. *Si on note n le nombre total d'étudiants, quelles seraient les limites lorsque n tend vers l'infini de la moyenne de la promotion ? Du taux de validation du cours ?*

Bonus : en quoi cette modélisation est critiquable ? Que proposeriez-vous pour l'améliorer ?

Exercice 8 (Un contre exemple important)**.** *On dit qu'une variable aléatoire suit la loi de Cauchy de paramètre $a > 0$, notée $\mathcal{C}(a)$ si sa densité est*

$$f(x) = \frac{1}{\pi}\frac{a^2}{a^2 + x^2}.$$

On admet la propriété suivante (qui peut être démontrée à l'aide des fonctions caractéristiques) :

Si $X \sim \mathcal{C}(a)$, $Y \sim \mathcal{C}(b)$ et $X \perp\!\!\!\perp Y$ alors $X + Y \sim \mathcal{C}(a + b)$.

Soit $(X_n)_{n\in NN}$ une suite de variables aléatoires indépendantes, de loi de Cauchy de paramètre 1*. On s'intéresse à la convergence de la suite de variables aléatoires*

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. *Montrer que si X est une variable aléatoire de loi $\mathcal{C}(1)$, alors $a \cdot X$ est une variable aléatoire de loi $\mathcal{C}(a)$.*
2. *En déduire la loi de $\overline{X_n}$.*
3. *La suite $(\overline{X_n})_{n\in\mathbb{N}}$ converge-t-elle en probabilité ? en loi ?*
4. *En quoi cela ne contredit-il pas la loi des grand nombres ?*
5. *BONUS : Simulez les variables $\overline{X_n}$ et tracez la fonction $n \mapsto \overline{X_n}$ sur ordinateur.*