Université Lyon 2 — Statistique Inférentielle
Licence 3 MIASHS

TD0 : Variables Aléatoires

Exercice 1 Soient X et Y deux variables aléatoires indépendantes gausiennes, d'espérances respectives μ_X et μ_Y et de variance respective σ_X^2 et σ_Y^2. Calculer

1. $\mathbb{E}[2X+3]$ et $V(2X+3)$

   On utilise la linéarité de l'espérance et on a directement
   $$\mathbb{E}[2X+3] = 2\mathbb{E}[X] + 3 = 2\mu_X + 3.$$
   On se rappelle aussi que $V(a+X) = V(X)$ pour tout réel a et $V(bX) = b^2V(X)$ pour tout réel b. D'où
   $$V(2X+3) = 4V(X) = 4\sigma_X^2.$$

2. $\mathbb{E}[X+Y]$ et $V(X+Y)$

   On utilisera à nouveau la linéarité de l'espérance et le fait que la variance d'une somme de deux variables aléatoires **indépendantes** est égale à la somme des variances.
   $$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \mu_X + \mu_Y.$$
   $$Var(X+Y) = Var(X) + Var(Y) = \sigma_X^2 + \sigma_Y^2.$$

3. $Cov(X,Y)$

   De façon générale, nous avons $Var(X+Y) = Var(X)+Var(Y)+2Cov(X,Y)$. Or les variables aléatoires X et Y sont indépendantes, donc $Cov(X,Y) = 0$.

4. $\mathbb{E}[XY]$

   Le résultat est immédiat en utilisant à nouveau l'indépendance des variables aléatoires X et Y.
   $$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \mu_X\mu_Y.$$

Exercice 2 On considère un dé cubique truqué dont les faces sont numérotées de 1 à 6 et on note X la variable aléatoire donnée par le numéro de la face du dessus. On suppose que le dé est truqué de sorte que la probabilité d'obtenir une face est proportionnelle au numéro inscrit sur cette face.

1. Déterminer la loi de X et calculer son espérance.

   On va construire un tableau donnant les probabilités d'obtenir chaque face. Les probabilités étant proportionnelles au numéro indiqué sur la face supérieure, on notera α ce coefficient de proportionnalité.

X	1	2	3	4	5	6
P(X)	α	2α	3α	4α	5α	6α

   On utilise ensuite le fait que $\sum_{k=1}^{6} P(X = k) = 21\alpha = 1$, ce qui nous donne $\alpha = 1/21$.

   L'espérance de X est alors donné par

   $$\mathbb{E}[X] = \sum_{k=1}^{6} kP(X = k) = \alpha \sum_{k=1}^{6} k^2 = 91/21.$$

2. On pose $Y = 1/X$. Déterminer la loi de Y ainsi que son espérance.

   On va simplement repartir du tableau précédent

X	1	2	3	4	5	6
Y	1	1/2	1/3	1/4	1/5	1/6
P(Y)	1/21	2/21	3/21	4/21	5/21	6/21

   On peut maintenant en déduire l'espérance de Y

   $$\mathbb{E}[Y] = \mathbb{E}[1/X] = \sum_{k=1}^{6} \frac{1}{k} P(X = k) = \sum_{k=1}^{6} \frac{k}{21} \times \frac{1}{k} = \frac{1}{21}.$$

Exercice 3 On lance deux dés parfaitement équilibrés. On note X le plus grand des numéros obtenus. Déterminer la loi de la variable aléatoire X.

On procède comme à l'exercice précédent en décrivant tous les évènements possibles dans un tableau croisé.

Dé 2 / Dé 1	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Il nous reste à compter le nombre d'occurrence de chaque nombre pour en déduire la loi de la variable aléatoire X.

X	1	2	3	4	5	6
P(X)	1/36	1/12	5/36	7/36	1/4	11/36

Exercice 4 On cherche à dépister une maladie détectable à l'aide d'un examen sanguin. On suppose que dans notre population, il y a une proportion p de personnes qui n'ont pas cette maladie. On analyse le sang de r personnes de la population, avec $r \geq 2$. On suppose que l'effectif de la population est suffisamment grand pour que le choix de ces r personnes s'apparente à un tirage avec remise.

1. Quelle est la probabilité qu'aucune de ces personnes ne soit atteinte de la maladie ?

   La probabilité qu'aucune de ces personnes ne soit atteinte de la maladie est p^r. En effet, si on note X la variable aléatoire traduisant le fait d'avoir ou non la maladie (*e.g.* $X = 0$ pour une personne saine et $X = 1$ pour une personne malade), alors

   $$P(X_1 = 0, X_2 = 0, \cdots, X_r = 0) = P(X_1 = 0)P(X_2 = 0) \cdots P(X_r = 0) = \prod_{i=1}^{r} P(X = 0) = p^r,$$

   où la première égalité est due à l'hypothèse d'indépendance entre les variables X_i qui suivent la même distribution que X.
2. On regroupe le sang de ces r personnes, puis on procède à l'analyse de sang. Si l'analyse est positive, on toutes les analyses individuelles (on avait pris soin de conserver une partie du sang recueilli avant l'analyse groupée). On note Y la variable aléatoire qui donne le nombre d'analyses de sang effectuées. Donner la loi de probabilité de Y et calculer son espérance en fonction de r et p.

   Il faut décortiquer un petit peu l'énoncé. On remarque que l'on a deux possibilités :*on effectue une seule analyse* ce qui arrive uniquement si tous les individus sont sains, *i.e.* l'évènement une seule analyse est effectuée se produit avec une probabilité p^r (d'après la question précédente). En revanche on va effectuer $r + 1$ *analyse* (nombre de personnes + analyse groupée) si au moins un des individus est malade, *i.e.* on va effectuer $(r + 1)$ analyses avec une probabilité de $1 - p^r$. On en déduit le tableau de la loi de Y suivant :

Y	1	r+1
P(Y)	p^r	$1 - p^r$

On en déduit directement l'espérance de cette variable aléatoire

$$\mathbb{E}[Y] = p^r + (r+1)(1-p^r) = 1 + r(1-p^r)$$

3. On s'intéresse à une population de n personnes, et on effectue des analyses collectives après avoir mélangé les prélèvements par groupe de r personnes, où r est un diviseur de n. Montrer que le nombre d'analyses que l'on peut espérer économiser, par rapport à la démarche consistant à tester immédiatement toutes les personnes, est égal à $np^r - n/r$.
   On nous indique que r divise n, il existe donc un entier k tel que $n = kr$ et cet entier k représente le nombre de groupes d'analyses. Ensuite, il faut voir qu'à chacun de ces groupes on peut associer une variable aléatoire Y_k qui aura la même distribution que Y.

   Ensuite, on parle du nombre d'analyse que l'on peut économiser : sachant que le nombre d'analsyesà effectuer est de n et que le nombre moyen d'analyses à effectuer sur un groupe avec le processus décrit est $\mathbb{E}[Y_k] = \mathbb{E}[Y] = 1 + r(1-p^r)$.

   Alors le nombre de d'analyses que l'on peut espérer économiser est égal

$$n - k(1 + r(1-p^r)) = n - \frac{n}{r}(1 + r(1-p^r)) = n - n/r - r + np^r = np^r - n/r.$$

Exercice 5 Mon opérateur de téléphonie mobile m'assure que 95% des SMS que j'envoie seront transmis en moins d'une minute.

(i) Quelle est la probabilité qu'un SMS envoyé arrive en moins d'une minute ?
On définit une variable aléatoire X qui modélise le temps mis par un SMS pour être transmis. Dans ce modèle, on peut interpréter l'information donnée comme

$$\mathbb{P}(X \leq 1) = .95.$$

(ii) J'envoie chaque jour 2 SMS. Quelle est la probabilité que le nombre de SMS arrivés en moins d'une minute soit : 0, 1, 2 ?
Pour chaque SMS envoyé, on définit la variable de Benoulli $\mathcal{B}(.95)$ Z_i, $i = 1, 2$ qui prend la valeur 0 si le SMS numéro i a été envoyé après 1 minute et la valeur 1 sinon. Le nombre de SMS envoyés en moins d'une minute est donc $Z = Z_1 + Z_2$. Il s'agit d'une variable Binomiale $\mathcal{B}(2, .95)$. Pour une variable binomiale générale $\mathcal{B}(n, p)$, Z, on a

$$\begin{aligned}
\mathbb{P}(Z = 0) &= C_n^0 \; p^0(1-p)^{n-0} \\
\mathbb{P}(Z = 1) &= C_n^1 \; p^1(1-p)^{n-1} \\
\vdots \quad & \quad \vdots \\
\mathbb{P}(Z = n) &= C_n^n \; p^n(1-p)^{n-n}
\end{aligned}$$