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

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

Dans notre cas on a $n = 2$ et $p = .95$ et donc

$$\begin{aligned}
\mathbb{P}(Z = 0) &= (1 - .95)^2 = .0025 \\
\mathbb{P}(Z = 1) &= 2\,p(1 - .95) = .095 \\
\mathbb{P}(Z = 2) &= p^2 = .9025.
\end{aligned}$$

(iii) Le week-end, j'envoie cette fois 20 SMS par jour. Proposez une modélisation pour le nombre de SMS arrivés en moins d'une minute.
Même modélisation que pour l'exercice précédant, mais avec cette cette fois $n = 20$ et $Z = Z_1 + Z_2 + \ldots + Z_{20} \sim \mathcal{B}(20, .95)$.
(iv) Quelle est la probabilité pour que le dimanche, au moins la moitié de mes SMS arrive en moins d'une minute ?
On veut calculer $\mathbb{P}(Z > 10)$.

Solution 1 $\mathbb{P}(Z > 10) = P(Z' \leq 9)$ où $Z' \sim \mathcal{B}(20, .05)$ est le nombre de SMS arrivés en plus d'une minute. D'après les tables statistiques disponibles à la fin du manuscrit, $P(Z' \leq 9) = 1$.

Solution 2 On peut faire une approximation Gaussienne en utilisant le fait que $\mathbb{E}[Z] = 20 \times 0.95 = 19.1$ et $\mathrm{Var}(Z) = 20 \times 0.95 \times (1 - 0.95) = 0.95$.

Ainsi,

$$\begin{aligned}
\mathbb{P}(Z \geq 10) &= \mathbb{P}\left(\frac{Z - 19.1}{\sqrt{.95}} > \frac{10 - 19.1}{\sqrt{.95} = -9.34}\right) \\
&\approx \mathbb{P}(W > -9.34).
\end{aligned}$$

où W est une variable Gaussienne centrée réduite $\mathcal{N}(0, 1)$. On procède alors comme dans l'exercice précédent. Dans la table, on trouve que $\mathbb{P}(W > -9.34) \approx 0$ et donc la probabilité d'envoyer plus de 10 messages, chacun partant en moins d'une minute est très proche de 1.

Notons que, normalement, l'approximation par une loi normale gaussienne $\mathcal{N}(np, np(1 - p))$ est valable si :

- $n \geq 30$,
- $np \geq 5$,
- $nq = n(1 - p) \geq 5$.

Certains auteurs modifient un peu ces conditions mais celle présentée est certainement la plus courante.

Solution 3 On aurait pu aussi utiliser une approximation par la loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = \mathbb{E}[Z] = 19$. Dans la Table 1.2, on trouve

$$\mathbb{P}\left(Z \leq 9\right) = 0.0089,$$

ce qui nous donne que

$$\begin{aligned}\mathbb{P}\left(Z \geq 9\right) &= 1 - \mathbb{P}\left(Z \leq 9\right)\\ &= .9911.\end{aligned}$$

Dans le cas présent, en supposant que la loi de Poisson est une approximation plus précise que celle par la loi normale, on prendra le résultat que nous venons de trouver.

Pour effectuer une approximation de la loi Binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(\lambda)$, on suppose que les paramètres sont liés par la relation $\lambda = np$.

L'approximation étant asymptotique, cela suppose que $p \to 0$ lorsque $n \to \infty$.

On retrouve donc le fait que la loi de Poisson sert à modéliser des évènements rares : ici on pourra considérer la probabiltié que le SMS arrrive en plus d'une minute (évènement avec une probabilité de 0.05).

Partant de cet énoncé, on va donc s'intéresser à l'évènement "rare" le SMS arrive en plus d'une minute, dont la variable aléatoire associée Z' peut être approximée par une variable aléatoire, notée $F \sim \mathcal{P}(np)$. On a donc $P(Z > 10) = P(Z' \leq 9) \simeq P(F \leq 9) \simeq 1$.

Exercice 6 En 2011 en France, la durée moyenne des périodes de chômage était de 14 mois. En supposant que la durée d'une période de chômage peut-être modélisée par une loi normale, de moyenne 14 et de variance 36, répondez aux questions suivantes :

(i) quelles sont les limites de cette modélisation ?

Une loi normale $\mathcal{N}(\mu, \sigma^2)$ modélise la distribution d'une variable aléatoire X qui peut *a priori* prendre des valeurs négatives, ce qui n'est pas le cas ici.

(ii) quelle est la probabilité qu'une période de chômage dure plus de 2 ans ?

On introduit la variable centrée-réduite associée

$$Z = \frac{X - \mu}{\sqrt{\sigma^2}} = \frac{X - 14}{\sqrt{36}}.$$

Cette variable aléatoire suit une loi normale $\mathcal{N}(0, 1)$. On cherche[1] $\mathbb{P}(X > 24)$. On peut

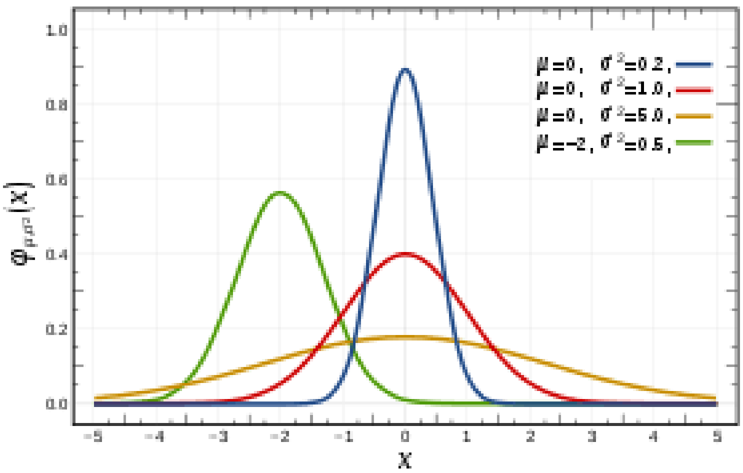

FIGURE 1 – Loi normale

réexprimer cette quantité en fonction de Z, dont on connaît la loi, de la manière suivante :

$$\begin{aligned}\mathbb{P}(X > 24) &= \mathbb{P}\left(\frac{(X-\mu)}{\sqrt{\sigma^2}} > \frac{(24-\mu)}{\sqrt{\sigma^2}}\right),\\ &= \mathbb{P}\left(Z > \frac{24-14}{6}\right),\\ &= \mathbb{P}\left(Z > 1.66666\right).\end{aligned}$$

La table de la loi normale centrée réduite consigne les valeurs de la fonction de répartition de Z, c'est à dire les valeurs de $\mathbb{P}(Z \geq x)$ pour certaines valeur de x. Comme

$$\mathbb{P}\left(Z > 1.66666\right) = 1 - \mathbb{P}\left(Z \leq 1.66666\right)$$

on peut trouver dans la Table 1.3 du polycopié que $\mathbb{P}\left(Z \leq 1.66666\right) = 0.9515$ et donc $1 - \mathbb{P}\left(Z \leq 1.66666\right) = 0.0485$, ce qui donne la réponse à la question $\mathbb{P}(X > 24) = 0.0485$.

(iii) quelle est la probabilité qu'une période de chômage dure moins de 6 mois ?

La méthode est la même. On cherche $\mathbb{P}(X \leq 6)$. On peut réexprimer cette quantité en fonction de Z, dont on connaît la loi, de la manière suivante[2] :

$$\begin{aligned}\mathbb{P}(X < 6) &= \mathbb{P}\left(\frac{(X-\mu)}{\sqrt{\sigma^2}} < \frac{(6-\mu)}{\sqrt{\sigma^2}}\right),\\ &= \mathbb{P}\left(Z < \frac{6-14}{6}\right),\\ &= \mathbb{P}\left(Z < -1.33333\right).\end{aligned}$$

1. sachant que 2 ans font 24 mois
2. sachant que 1 an fait 12 mois

Comme les valeurs négatives ne se trouvent pas dans la table, il faut faire une transformation simple, utilisant la symétrie de la Gaussienne centrée-réduite :

$$\begin{aligned}\mathbb{P}\,(Z < -1.33333) &= \mathbb{P}\,(Z > 1.33333)\\ &= 1 - \mathbb{P}\,(Z \leq 1.33333)\end{aligned}$$

on peut trouver dans la Table 1.3 du polycopié que $\mathbb{P}\,(Z \leq 1.33333) = 0.9082$ et donc $1 - \mathbb{P}\,(Z \leq 1.333333) = 0.0918$, ce qui donne la réponse à la question $\mathbb{P}(X > 6) = 0.0918$.

(iv) quelle est la probabilité qu'une période de chômage dure entre 6 mois et 1 an ?

On cherche $\mathbb{P}(6 < X \leq 12)$. On peut réexprimer cette quantité en fonction de Z, dont on connaît la loi, de la manière suivante [3] :

$$\begin{aligned}\mathbb{P}(6 < X \leq 12) &= \mathbb{P}\left(\frac{(6-\mu)}{\sqrt{\sigma^2}} < \frac{(X-\mu)}{\sqrt{\sigma^2}} \leq \frac{(12-\mu)}{\sqrt{\sigma^2}}\right),\\ &= \mathbb{P}\left(\frac{6-14}{6} < Z \leq \frac{12-14}{6}\right),\\ &= \mathbb{P}\,(-1.33333 < Z \leq -.33333)\,.\end{aligned}$$

Notons que l'événement $\{-1.33333 < Z \leq -.33333\}$ est égal à l'événement $\{Z \leq -.33333\}$ auquel on enlève l'événement $\{Z < -1.33333\}$. On obtient donc

$$\mathbb{P}(6 < X \leq 12) = \mathbb{P}\,(Z \leq -.33333) - \mathbb{P}\,(Z < -1.33333)\,.$$

Comme les valeurs négatives ne se trouvent pas dans la table, il faut faire une transformation simple, utilisant la symétrie de la Gaussienne centrée-réduite :

$$\begin{aligned}\mathbb{P}\,(Z < -1.33333) &= \mathbb{P}\,(Z > 1.33333)\\ &= 1 - \mathbb{P}\,(Z \leq 1.33333)\end{aligned}$$

On a déjà calculé que $\mathbb{P}(Z < -1.33333) = 0.0918$. De la même manière, on calcule que $\mathbb{P}(Z < -0.33333) = 1 - \mathbb{P}(Z \leq 0.33333) = 1 - 0.6293 = 0.3707$. Ainsi,

$$\mathbb{P}(6 < X \leq 12) = 0.3707 - 0.0918 = 0.2789.$$

Exercice 7 On suppose que la distance en mètre parcouru par un javelot lancé par un athlète A suit une loi normale. Au cours d'un entraînement, on constate que exactement 10% des javelots atteignent plus de 75 mètres et exactement 25% des javelots atteignent moins de 50 mètres. Calculer la longueur moyenne parcourue par un javelot ainsi que l'écart-type de cette longueur.

On pourrait reformuler cet exerce par : déterminer la moyenne et l'écart-type de la loi normale, ce qui laisse présager la résolution d'un système d'équations à deux inconnues ... ça tombe, on a deux données à utiliser !

On va noter X la variable aléatoire traduisant la longueur parcourue par un javelot.

3. sachant que 2 ans font 24 mois

- la phrase 10% *des javelots atteignent plus de* 75 *mètres* peut se traduire par

$$P(X \geq 75) = 0.1,$$
$$P(X \leq 75) = 0.9.$$

- la phrase 25% *des javelots atteignent moins de* 50 *mètres* peut se traduire par

$$P(X \leq 50) = 0.25.$$

Il nous reste à faire apparaître nos paramètre de moyennes et écart-type en se ramenant à une variable aléatoire Z centrée-réduite. Ce qui nous donne

$$\begin{aligned} P(X \leq 75) &= 0.9, \\ P\left(\frac{X-\mu}{\sigma} \leq \frac{75-\mu}{\sigma}\right) &= 0.9 \\ P\left(Z \leq \frac{75-\mu}{\sigma}\right) &= 0.9 \end{aligned}$$

Il nous reste à trouver pour quelle valeur de z, $P(Z < z) = 0.9$. On trouve $z = 1.28$. On procède de la même façon avec la deuxième équation et on trouve :

$$\begin{aligned} P(X \leq 50) &= 0.25, \\ P\left(\frac{X-\mu}{\sigma} \leq \frac{50-\mu}{\sigma}\right) &= 0.25 \\ P\left(Z \leq \frac{50-\mu}{\sigma}\right) &= 0.25 \end{aligned}$$

Il nous reste à trouver pour quelle valeur de z, $P(Z < z) = 0.25$, ce qui est pareil que de trouver la valeur de z pour laquelle $P(Z < -z) = 0.75$ (par symétrie de la loi normale. Cela nous donne $z = -0.68$. on aboutit au système suivant :

$$\begin{cases} \dfrac{75-\mu}{\sigma} & = & 1.28 \\ \dfrac{50-\mu}{\sigma} & = & -0.68 \end{cases} \rightarrow \begin{cases} \mu & = & 75 - 1.28\sigma \\ 50 - (75 - 1.28\sigma) & = & -0.68\sigma \end{cases}$$

La deuxième équation nous donne $1.96\sigma = 25$ soit $\sigma = 12.76$ et la première équation nous donne alors $\mu = 75 - 1.28\dfrac{25}{1.96} = 58.67$.