Méthodes à retenir :

- Il est utile de savoir représenter les marginales dans un tableau pour un couple de variables aléatoires qui ne prennent qu'un nombre fini de valeurs.

I. Applications directes du cours

Exercice 1 ✰

On considère le couple (X,Y) de variables aléatoires à valeurs dans $\{0,1\}$, dont la loi est donnée par :
$\mathbf{P}((0,0)) = 1/8$, $\mathbf{P}((0,1)) = 1/4$,
$\mathbf{P}((1,0)) = 3/8$, $\mathbf{P}((1,1)) = 1/4$.

1. Déterminer les lois marginales $\mathbf{P}_X$ et $\mathbf{P}_Y$, à l'aide d'un tableau.
2. X et Y sont-elles indépendantes ?

Exercice 2 ✰✰

Soient X et Y deux variables aléatoires indépendantes suivant des lois de Bernoulli $b(1/2)$. Les variables aléatoires $X+Y$ et $X-Y$ sont-elles indépendantes ?

Exercice 3 ✰✰ *Sondage*

Une population de personnes présente une propriété donnée avec une proportion inconnue $p \in]0,1[$.
On choisit un échantillon de n personnes et l'on pose $X_i = 1$ si le i-ème individu présente la propriété étudiée, 0 sinon. On considère que les variables aléatoires X_i ainsi définies sont indépendantes et suivent toute une loi de Bernoulli de paramètre p.
a) Quelle est la loi suivie par

$$S_n = X_1 + \cdots + X_n\,?$$

b) Déterminer espérance et variance de S_n/n.
c) Soit $\varepsilon > 0$. Etablir

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant \frac{1}{4n\varepsilon^2}$$

d) Pour $\varepsilon = 0,05$, quelle valeur de n choisir pour que S_n/n soit voisin de p à ε près avec une probabilité supérieure à 95 % ?

Exercice 4 ✰✰ *Approximation Binomiale Poisson*

Soit $p \in]0,1[$ fixé.
Pour tout $k \in \mathbb{N}$ fixé, vérifier que

$$\binom{n}{k} p^k (1-p)^{n-k} \underset{n\to+\infty}{\sim} \frac{e^{-np}(np)^k}{k!}$$

En déduire, pour $n \in \mathbb{N}$ et $\lambda = np$ que

$$\lim_{n\to+\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Interprétation : La relation $\lambda_n = np$ montre que pour n grand, la probabilités $\binom{n}{k} p^k(1-p)^{n-k}$ qu'une variable de loi $\mathcal{B}(n,p)$ soit égale à k est très proche de la probabilités $e^{-\lambda}\dfrac{\lambda^k}{k!}$ qu'une variable de loi $\mathcal{P}(\lambda)$ soit égale à k : on dit que la loi de Poisson est la loi des évènements "rares".

Exercice 5 *approximation binomiale-Poisson* ✰✰

Des machines fabriquent des plaques de tôle destinées à être empilées ; on estime à 0.1% la proportion de plaques inutilisables. L'utilisation de ces plaques consiste à en empiler n, numérotées de 1 à n en les prenant au hasard.

1. Pour $n = 2000$, quelle est la loi suivie par la variable aléatoire N "nombre de plaques inutilisables parmi les 2000" ? (on utilisera une loi de probabilité adaptée) ;
2. Quelle est la probabilité pour que N soit strictement inférieure à 3 ?
3. quelle est la probabilité pour que N soit inférieure ou égale à 3 ?

II. A savoir rédiger

Exercice 6 ✰✰

Soit X et Y deux variables aléatoires réelles sur l'espace probabilisé $(\Omega, \mathbf{P})$.
Montrer que :

$$|\operatorname{Cov}(X,Y)| \leqslant \sqrt{\mathbb{V}(X)\mathbb{V}(Y)}$$

Exercice 7 ✰

On note S la somme de deux dés équilibrés.

1. Déterminer les lois des variables aléatoires D_1 et D_2 représentant les résultats respectifs des deux dés.
2. Déterminer la loi de S.
3. Déterminer l'espérance de S.
4. Déterminer la variance de S.

III. Exercices

Exercice 8 ✰✰✰ *Mines-Telecom*

X_1, X_2 et Y sont trois variables aléatoires indépendantes. X_1 et X_2 suivent une même loi géométrique de paramètre p. Y est définie telle que : $Y(\Omega) = \{-1, 1\}$ et $P(Y) = 1$. On considère $M = \begin{pmatrix} X_1 & YX_2 \\ X_2 & X_1 \end{pmatrix}$.

1. Calculez la probabilité pour que la matrice M soit inversible.

2. Montrer que : $\operatorname{Sp}(M) \subset \mathbb{R}$.

Exercice 9 ✰✰✰ *CCINP MP*

Soient X et Y deux variables aléatoires indépendantes qui suivent des lois géométriques de paramètres respectifs λ et μ avec $\lambda, \mu \in]0,1[$.

1) Déterminer la loi de $X + Y$.

2) Si X et Y sont deux variables aléatoires quelconques indépendantes, $X + Y$ et $X - Y$ sont-elles indépendantes ?

Exercice 10 ✰✰✰ *CCINP PSI*

Soit $n \in \mathbb{N}$, avec $n \geqslant 2$, $p, q \in]0,1[$, avec $p + q = 1$. Soit X, Y deux variables aléatoires discrètes à valeurs dans $[\![0, n]\!]$ telles que $\forall (j,k) \in [\![0,n]\!] \times [\![1,n]\!]$, $P(X = j, Y = k) = \begin{cases} \dbinom{n}{k} p^k q^{n-k} & \text{si } k = j \neq 0 \\ \dfrac{q^n}{n} & \text{si } j = 0 \\ 0 & \text{si } k \neq j \text{ et } j \neq 0 \end{cases}$

1. Lois de X et Y ? Calculer l'espérance de Y.
2. Les variables X et Y sont-elles indépendantes ?
3. Déterminer la loi conditionnelle de Y sachant $(X = j)$.
4. Calculer la covariance de X, Y. Existe-t-il des valeurs de q telles que X et Y soient décorrélées ?

Exercice 11 ✰✰

On dispose de n boîtes numérotées de 1 à n . La boîte k contient k boules numérotées de 1 à k.

On choisit au hasard une boîte puis une boule dans cette boîte.

Soit X le numéro de la boîte et Y le numéro de la boule.

1. Déterminer la loi du couple (X, Y).
2. Calculer $\mathbf{P}(X = Y)$.
3. X et Y sont-elles indépendantes ?

Exercice 12 ✰✰✰

Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$.
On suppose que la loi conjointe de X et Y vérifie

$$\mathbf{P}(X = j, Y = k) = \frac{a}{2^{j+1}\, k!} \text{ avec } a \in \mathbb{R}$$

a) Déterminer la valeur de a.
b) Reconnaître les lois marginales de X et Y.
c) Les variables X et Y sont elles indépendantes ?

IV. Pour aller plus loin

Exercice 13 ☆☆ *Bluff et théorie des jeux*

On considère le jeu suivant :

- le joueur A lance une pièce en espérant obtenir Pile. Il garde pour lui le résultat du lancé. Il peut alors soit abandonner en déclarant avoir obtenu Face et payer un euro, soit déclarer avoir obtenu Pile (mais cela peut être du bluff).

- Si A déclare "Pile"le joueur B a alors deux options :

- ou bien il ne conteste pas l'affirmation, et perd 1 euro.
- ou bien il conteste le résultat affirmé. S'il a raison il gagne 2 euros. S'il a tort, il verse 2 euros.

1. *Etude du jeu sans bluff :*
   Le joueur A s'interdit ici le bluff.
   On note X la variable aléatoire représentant le gain du joueur A. Déterminer la loi de X. Le joueur B ne contestera jamais l'affirmation de A.
   En déduire l'espérance de X.
2. *Etude du jeu avec bluff :* En cas d'échec lors du lancer, le joueur A s'autorise à déclarer "Pile" au lieu du résultat obtenu avec une probabilité $p \in [0, 1]$.
   En cas de déclaration "Pile" par le joueur A à l'issue du lancer, le joueur B s'autorise à contester l'affimation de A avec une probabilité $q \in [0, 1]$.
   On note Y la variable aléatoire représentant le gain du joueur A. Déterminer la loi de Y.
   En déduire que l'espérance de Y est $\mathbb{E}[Y] = p + q\left(\frac{1}{2} - \frac{3}{2}p\right)$
3. Quelle est alors la valeur de p optimale pour maximiser le gain du joueur A ?
   Commentaire ?
4. En réécrivant $\mathbb{E}[Y] = \frac{1}{2}q + p\left(1 - \frac{3}{2}q\right)$, montrer que B a une stratégie optimale pour limiter ses pertes.

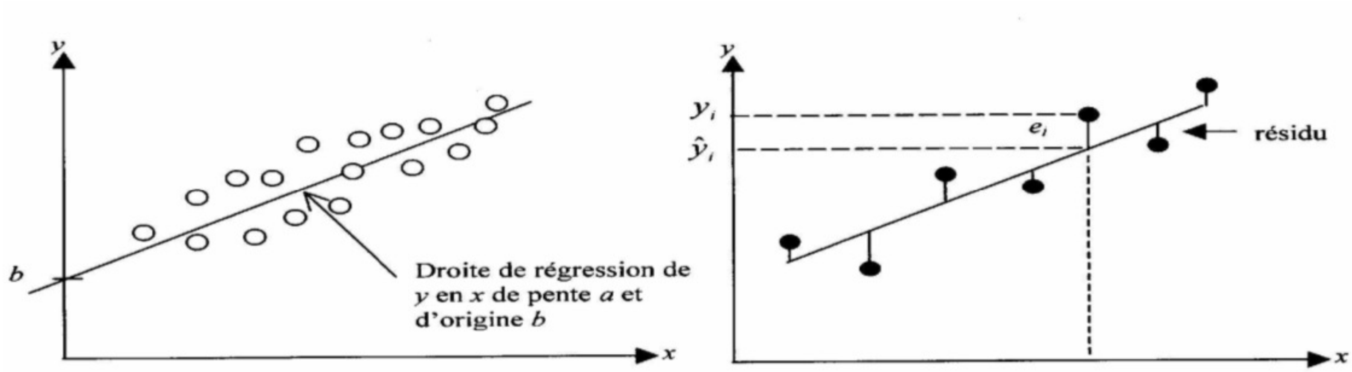
Exercice 14 ☆☆☆ *Modèle d'assurance*

Des sinistres surviennent sur une année avec une probabilité $p \in]0, 1[$. Chaque sinistre à un coût fixe de C pour l'assureur. Chaque client paye une prime annuelle d'assurance π.

1. Pourquoi peut-on modéliser par une famille $(X_i)_{1\leq i\leq N}$ de N variables aléatoires de loi $b(p)$ les N clients d'un assureur ?
2. Quelle est l'espérance de S le solde annuel des sinistres pour l'assureur ?
3. Quelle est la variance de S ?
4. Donner une valeur de π la prime annuelle d'un assuré pour que l'assureur soit rentable une année avec une probabilité supérieure à 99%.

Exercice 15 ☆☆☆☆ *Modélisation par moindres carrés*

$$\sum_{i=0}^{n} \varepsilon_i^2 = \sum_{i=0}^{n} (y_i - (ax_i + b))^2.$$



Etant donné un n échantillon de mesures physiques $((x_i, y_i))_{1\le i\le n}$, on cherche à savoir si cet échantillon peut-être assimilé à des réalisations aléatoires de couples $((x_i, ax_i+b))_{1\le i\le n}$, pour des paramètres a et b de régression linéaire, selon une relation linéaire de la forme

$$y = a\ x + b$$

On note $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ et $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, et $U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathfrak{M}_{n,1}(\mathbb{R})$

1. Par définition, la droite de régression linéaire associée à ces mesures est la droite d'équation $y = \hat{a}x + \hat{b}$, où $(\hat{a}, \hat{b}) \in \mathbb{R}^2$ telle que $\delta_n(\hat{a}, \hat{b}) = \sum_{i=1}^{n} |y_i - \hat{a}\ x_i - \hat{b}|^2$ réalise le minimum de la fonction $(a, b) \mapsto \sum_{i=1}^{n} |y_i - a\ x_i - b|^2$.
2. Justifier que résoudre ce problème revient à minimiser une quantité de la forme
$$f(V) = \|MV - Y\|_2^2 \qquad (\star)$$
avec $M = [X, U] \in \mathfrak{M}_{n,2}(\mathbb{R})$ et $V = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathfrak{M}_{2,1}(\mathbb{R})$, et $Y \in \mathfrak{M}_{n,1}(\mathbb{R})$ à préciser.
3. On note $\hat{Y}$ le projeté orthogonal de Y sur le sous-espace vectoriel $\mathrm{Im}(M)$.
   (a) Pourquoi a-t-on $\mathrm{Im}(M) = \mathrm{Vect}(X, U)$?
   (b) Justifier que ce maximum est atteint et est unique.
   on utilisera la formule de projection sur un sous-espace vectoriel de dimension finie, muni d'une base orthonormée
   (c) Pourquoi a-t-on $\hat{Y} - Y \perp MV,\ \forall V \in \mathfrak{M}_{2,1}(\mathbb{R})$?
   (d) On pose $\hat{V}$ tel que $M\hat{V} = \hat{Y}$. En déduire que : $\forall\ V \in \mathfrak{M}_{2,1}(\mathbb{R}),\ \ V^T M^T (Y - M\hat{V}) = 0$
   (e) En déduire que : $M^T(Y - M\hat{V}) = 0$
   (f) Montrer que $\det(M^T M) = n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i^2)^2$
   (g) Etablir que $\hat{V} = (M^T M)^{-1}\ M^T\ Y$.
4. On rappelle que l'espérance de X (resp. Y) est $\mathbb{E}(X) = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ (resp. $\mathbb{E}(Y) = \dfrac{1}{n}\sum_{i=1}^{n} y_i$).
   On rappelle que la variance de X (resp. Y) est $\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \mathbb{E}(X))^2$ (resp. $\mathbb{V}(Y) = \mathbb{E}(Y - \mathbb{E}(Y))^2 = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - \mathbb{E}(Y))^2$).
   On appelle covariance de X et Y le nombre : $\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \mathbb{E}(X))(y_i - \mathbb{E}(Y))$
   On vérifie que :
$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \left(\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots x_n & \end{pmatrix} Y$$

$$= \frac{1}{n^2\mathbb{V}(X)} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.$$

En déduire que l'équation de la droite des moindres carrées est obtenue pour les paramètres :
$\hat{b} = \mathbb{E}(Y) - \hat{a}\ \mathbb{E}(Y)$ et $\hat{a} = \dfrac{\mathrm{Cov}(X,Y)}{\mathbb{V}(X)}$.

5. La droite des moindres carrées et obtenue pour des paramètres $\hat{a}$ et $\hat{b}$ et a pour équation cartésienne

$$y = \hat{a}x + \hat{b}$$

Interpréter l'alignement des points sur cette droites dans le cas d'une covariance nulle.

Notes

[1] correction :

$\mathbf{P}_X(0) = 1/8 + 2/8 = 3/8$, $\mathbf{P}_X(1) = 3/8 + 2/8 = 5/8$, $\mathbf{P}_Y(0) = 1/8 + 3/8 = 1/2$, $\mathbf{P}_Y(1) = 2/8 + 2/8 = 1/2$

Non, X et Y ne sont pas indépendantes car $\mathbf{P}((0,0)) = 1/8 \neq 3/8 * 1/2 = 3/16$

[3] correction : a)S_n soit une loi de Bernoulli de paramètres n et p.

b) Puisque S_n suit une loi de Bernoulli, $\mathrm{E}(S_n) = np$ et $\mathrm{V}(S_n) = np(1-p)$.

Par conséquent $\mathrm{E}(S_n/n) = p$ et $\mathrm{V}(S_n/n) = p(1-p)/n$

c) Par l'inégalité de Bienaymé-Tchebychev $P(|Sn/n - \mathrm{E}(Sn/n)| > \varepsilon) \leq \mathrm{V}(Sn/n)/\varepsilon^2$

et donc $P(|Sn/n - p| > \varepsilon) \leq p(1-p)/(n\varepsilon^2)$

Enfin, l'inégalité classique $p(1-p) \leq 1/4$ permet de conclure.

d) On choisit n de sorte que $1/(4n\varepsilon^2) \leq 0,05$

La valeur $n = 2000$ est convenable.

[3] correction : $\mathbf{P}_X(0) = 1/8 + 2/8 = 3/8$, $\mathbf{P}_X(1) = 3/8 + 2/8 = 5/8$, $\mathbf{P}_Y(0) = 1/8 + 3/8 = 1/2$, $\mathbf{P}_Y(1) = 2/8 + 2/8 = 1/2$

Non, X et Y ne sont pas indépendantes car $\mathbf{P}((0,0)) = 1/8 \neq 3/8 * 1/2 = 3/16$

[5] correction :

Pour $n = 2000$, la loi suivie par la variable aléatoire N est une loi de Poisson de paramètre 2 : alors $\mathbf{P}[N \leq 3] \approx 0.85$

[6] correction : Pour $\lambda \in \mathbb{R}$, introduisons $Z = \lambda X + Y$. On a $\mathbb{V}(Z) \geq 0$ avec $\mathbb{V}(Z) = \lambda 2\mathbb{V}(X) + 2\lambda Cov(X,Y) + \mathbb{V}(Y)$

Si $\mathbb{V}(X) = 0$, on a nécessairement $Cov(X,Y) = 0$ pour que $\mathbb{V}(Z)$ soit positif pour tout $\lambda \in \mathbb{R}$.

Si $\mathbb{V}(X) \neq 0$, on a nécessairement $\Delta = 4Cov(X,Y)2 - 4\mathbb{V}(X)\mathbb{V}(Y) \leq 0$ pour que $\mathbb{V}(Z)$ soit positif pour tout $\lambda \in \mathbb{R}$.

Dans les deux cas, on obtient $|Cov(X,Y)| \leq \mathbb{V}(X)\mathbb{V}(Y)$

[10] correction :

1. Pour $k \geq 2$: $P[X = k] = (k-1)2^{-k}$: choix du pile parmi les $(k-1)$ premiers lancers.
   $E[X] = \sum_{k=2}^{+\infty} k(k-1)2^{-k} = \frac{1}{k}\frac{2}{(1-1/2)^3} = 4$
   La série $\sum_{k\geq 2} k^2(k-1)2^{-k}$ CV et $E[X^2] = \sum_{k=2}^{+\infty} k^2(k-1)2^{-k} = \sum_{k=3}^{+\infty} k(k-1)(k-2)2^{-k} + 2\sum_{k=2}^{+\infty} k(k-1)2^{-k} = \frac{1}{2^3}\frac{3!}{(1-1/2)^4} + 2\frac{1}{2^2}\frac{2!}{(1-1/2)^3} = 20$
   $V(X) = 20 - 4^2 = 4$.
2. sachant $\{X = k\}$, Y prend les valeurs $1\ldots, k-1$, donc$Y$ est à valeurs dans $\mathbb{N}$.
   Pour $n \geq 1$, on a $P[Y = n] = \sum_{k=2}^{+\infty} P[X = k] \times P_{X=k}[Y = n] = \sum_{k=2}^{+\infty}(k-1)2^{-k} \times \frac{1}{k-1} = \frac{1}{2^{n+1}} \times 11 - 1/2 = (1-p)^{n-1}p$ pour $p = 1/2$
   Donc (loi géométrique) $E[Y] = 2$, $V[Y] = 2$
3. Non $P[(X = 2) \cap (Y = 3)] = 0$: pas d'indépendance !
4. $E[XY] = \sum_{k=2}^{+\infty}\sum_{n=1}^{+\infty} knP[(X,Y) = (k,n)] = \sum_{k=2}^{+\infty}\sum_{n=1}^{k-1} kn(k-1)\frac{1}{2^k}\frac{1}{k-1} = \sum_{k=2}^{+\infty}\frac{k}{2^k}\frac{k(k-1)}{2} = \frac{1}{2}E[X^2] = 10$
   $cov(X,Y) = E[XY] - E[X]E[Y] = 10 - 4 * 2 = 2$
   $\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V[X]V[Y]}} = \frac{2}{\sqrt{8}} = \frac{1}{\sqrt{2}}$