Colles semaine 32

En bref

- Étude des homothéties et projections vectorielles.
- Étude des symétries vectorielle.
- Loi conditionnelle d'une variable sachant un évènement.
- Couple de variables aléatoire : Loi conjointe, loi marginale.
- Covariance de deux variables, formule $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$.
- Théorème de transfert pour les couples de variable. Application notable au calcul de $\mathbb{E}\left(XY\right)$.
- Indépendance d'un couple ou d'une suite de variables aléatoires. Lemme des coalitions.
- Inégalité de Markov et de Bienaymé-Tchebychev.

Exemples non exhaustifs de questions de cours

Les suggestions suivantes restent des exemples d'illustrations, les colleurs ont toute liberté pour poser une question de cours.

- Soit p la projection sur $F := \mathrm{Vect}((1,2,0))$ parallèlement à $G := \{(x,y,z) \in \mathbb{R}^3 \mid x - y + 2z = 0\}$. Déterminer la matrice de p dans la base canonique, puis dans une base adaptée à la somme directe $F \oplus G$.
- Montrer par récurrence la formule $\mathbb{E}\left(X_1 \cdots X_n\right) = \mathbb{E}\left(X_1\right) \cdots \mathbb{E}\left(X_n\right)$ si la famille $(X_1, \dots, X_n)$ est indépendante.
- Calculer $Cov(X,Y)$ lorsque la loi conjointe est donnée par $X(\Omega) = Y(\Omega) = [\![0\,;\,n]\!]$ et $\forall (i,j) \in ([\![1\,;\,n]\!])^2,\ \mathbb{P}\left(X = i \cap Y = j\right) = \begin{cases} \frac{1}{n(n+1)} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$. .
- Citer et prouver les inégalités de Markov et de Bienaymé-Tchebychev.

Note aux colleurs

- La formule exprimant la variance d'une somme de plus que trois variable avec les covariance n'est pas au programme. On se limite à la somme de deux variables. On peut en revanche calculer la variance d'une somme indépendante.
- Si un étudiant a à sa disposition la loi conjointe d'un couple (X,Y) ; alors il n'est censé présenter aucune hésitation si on lui demande de calculer la loi de $\min(X,Y)$, de $\max(X,Y)$ ou encore de $X+Y$.

En détails :

Endomorphismes remarquables : homothéties, projections et symétries

Reprise du programme précédent

Couple et suite de variables aléatoires réelles

1 Lois conditionnelles

On commence par définir une notion, qui sera utile pour la suite.

Définition 1. Soit X un variable aléatoire réelle sur Ω et A un évènement de **probabilité non nulle.** On appelle *loi conditionnelle de X sachant A*, la loi de la variable X pour la probabilité $\mathbb{P}_A$. Concrètement, pour les univers finis, il s'agit de la donnée de :
— L'ensemble-image $X(\Omega)$, (inchangé par rapport à la loi classique).
— Pour tout $k \in X(\Omega)$ la valeur de $\mathbb{P}_A(X = k)$.

Exercice 2. Soit X une variable aléatoire dont la loi est donnée par :

$$X(\Omega) = \{-1, 0, 2\}$$

$$\mathbb{P}(X = -1) = \frac{1}{2} \quad , \quad \mathbb{P}(X = 0) = \frac{1}{3} \quad , \quad \mathbb{P}(X = 2) = \frac{1}{6}$$

Déterminer la loi conditionnelle de X sachant $[X \geqslant 0]$.

Exercice 3. Soit D_1 et D_2 deux variables aléatoires réelles modélisant le résultat d'un jet de deux dés équilibré à 6 faces (D_1 pour le premier dé et D_2 pour le second). On note $S = D_1 + D_2$.
— Déterminer la loi de S.
— Déterminer la loi conditionnelle de S sachant l'évènement $[D_1 = 1]$.
— Chercher une formule permettant de calculer la loi de S à partir des lois conditionnelles sachant $[D_1 = i]$ pour tout $i \in D_1(\Omega)$.

2 Rappels sur l'indépendance

2.1 Indépendance de deux variables

2.1.1 Définitions et propriétés élémentaires

Définition 4 (Variables aléatoires indépendantes)**.** Soient X et Y deux variables aléatoires réelles. On dit que les variables aléatoires X et Y sont **indépendantes** si, pour toute partie A de $\mathbb{R}$ et toute partie B de $\mathbb{R}$ les évènements $[X \in A]$ et $[Y \in B]$ sont indépendants, c'est-à-dire vérifient

$$\mathbb{P}([X \in A] \cap [Y \in B]) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

Définition 5. Soit X et Y deux variables aléatoires réelles sur un univers fini. Ces deux variables sont indépendantes si et seulement si

$$\forall x \in X(\Omega), \forall y \in Y(\Omega),\ \mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$$

2.1.2 Liens avec les lois conditionnelles

Lemme 6. *Soit X et Y deux variables aléatoires réelles sur un univers **fini**.*

Alors X et Y sont indépendantes si et seulement si pour tout $y \in Y(\Omega)$ vérifiant $\mathbb{P}(Y=y) > 0$, la loi conditionnelle de X sachant $[Y=y]$ est identique à la loi de X.

Remarque 7. Ce résultat se généralisera sans peine l'an prochain pour des variables discrètes sur des univers finis. Pour ce qui est des variables continues, c'est une autre paire de manches...

2.1.3 Application pour l'espérance, la variance, la covariance

Définition 8. Soit X et Y deux variables aléatoires réelles. On appelle *covariance* de X et Y le réel suivant :

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$$

Remarque 9. Dans le cas $X = Y$, on a $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

Lemme 10. *Deux variables indépendantes ont une covariance nulle. On dit aussi qu'elles sont décorrélées.*

Attention ! La réciproque est évidemment fausse. le lecteur en cherchera un contre-exemple explicite.

Proposition 11 (Variance d'une somme non indépendante)**.** *Pour deux variables aléatoires réelles X et Y, on a* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$.

Lemme 12 (La covariance est bilinéaire)**.** *Pour X un variable aléatoire réelle fixé, l'application $Y \mapsto \mathrm{Cov}(X,Y)$ est une forme linéaire sur l'espace vectoriel des variables aléatoires réelles sur Ω. De même, pour l'application $Y \mapsto \mathrm{Cov}(Y,X)$*

2.1.4 Sommes, minimums, maximums

Tout élève de PTSI doit connaître les méthodes suivantes permettant de calculer la loi d'une somme, d'un minimum, d'un maximum de variables aléatoires.

Aucune des formules n'apparaît explicitement au programme donc il convient de refaire les preuves associées.

Méthode 13 (Loi d'une somme)**.** Soit X et Y deux variables aléatoires réelles et posons $S := X+Y$. On a alors $S(\Omega) = \{x+y \mid x \in X(\omega), y \in Y(\Omega)\}$. pour déterminer la loi de S, on utilise la formule suivante, conséquence de la formule des probabilités totales appliquée au système complet d'évènements associé à la variable Y (ou X). (On suppose ici que $\mathbb{P}(X=x) > 0$ pour chaque élément $x \in X(\Omega)$ et on adapte facilement sinon).

$$\begin{aligned}\forall k \in S(\Omega),\ \mathbb{P}(S=k) &= \sum_{i\in X(\Omega)} \mathbb{P}(S=k \,|\, X=i)\,\mathbb{P}(X=i)\\ &= \sum_{i\in X(\Omega)} \mathbb{P}(Y=k-i \,|\, X=i)\,\mathbb{P}(X=i) = \cdots\end{aligned}$$

$$= \sum_{j\in Y(\Omega)} \mathbb{P}(X=k-j \,|\, Y=j)\,\mathbb{P}(Y=j)$$

Si de plus les variables aléatoires réelles X et Y sont indépendantes, alors :

$$\forall k \in S(\Omega),\ \mathbb{P}(S=k) = \sum_{i\in X(\Omega)} \mathbb{P}(Y=k-i)\,\mathbb{P}(X=i) = \sum_{j\in Y(\Omega)} \mathbb{P}(X=k-j)\,\mathbb{P}(Y=j)$$

Méthode 14 (Loi de minimum)**.** Soit X et Y deux variables aléatoires réelles. On note $M = \min(X, Y)$. Alors $M(\Omega) \subset X(\Omega) \cap Y(\Omega)$. Par ailleurs, la loi se calcule à l'aide des fonctions de répartitions.

On note pour toute variable aléatoire réelle Z la fonction $F_Z : \begin{array}{ccc} \mathbb{R} & \to & [0\,;1] \\ x & \mapsto & \mathbb{P}\left(Z \leqslant x\right) \end{array}$.

On peut calculer la fonction de répartition de $M := \min(X, Y)$ à partir de celles de X et Y par la méthode suivante :

$$\forall k \in \mathbb{R},\ F_M(k) = \mathbb{P}\left(\min(X, Y) \leqslant k\right) = \mathbb{P}\left(X \leqslant k \cap Y \leqslant k\right) = F_X(k)F_y(k).$$

Ensuite on revient à la loi en notant $M(\Omega) = \{x_1, x_2, \ldots, x_n\}$ et en ordonnant ces valeurs **par ordre strictement croissant**. Pour un x_i, on utilise alors le lemme suivant $[M \leqslant x_i] = [M < x_i] \sqcup [X = x_i]$. On en déduit :

$$\forall x_i \in M(\Omega),\ \mathbb{P}\left(M = x_i\right) = \mathbb{P}\left(X \leqslant x_i\right) - \mathbb{P}\left(X \leqslant x_{i-1}\right) = F_M(x_i) - F_M(x_{i-1})$$

Remarque 15. La locution *fonction de répartition* est absente du programme pour une raison qui m'échappe. Il faut alors à chaque fois la redéfinir. Souvent le sujet, le fait pour vous.

Méthode 16 (Loi d'un maximum)**.** Soit X et Y deux variables aléatoires réelles. On note $N := \min(X, Y)$. Alors $N(\Omega) \subset X(\Omega) \cap Y(\Omega)$. Par ailleurs, la loi se calcule à l'aide des queues de fonctions de répartitions

$$\forall k \in \mathbb{R},\ 1 - F_N(k) = \mathbb{P}\left(\max(X, Y) > k\right) = \mathbb{P}\left(X \leqslant k \cap Y \leqslant k\right) = (1 - F_X(k))(1 - F_y(k)).$$

Ensuite on revient à la loi en notant $N(\Omega) = \{x_1, x_2, \ldots, x_n\}$ et en ordonnant ces valeurs **par ordre strictement croissant**. Pour un x_i, on utilise alors le lemme suivant $[M \geqslant x_i] = [M > x_i] \sqcup [X = x_i]$. On en déduit :

$$\forall x_i \in M(\Omega),\ \mathbb{P}\left(M = x_i\right) = \mathbb{P}\left(X > x_{i-1}\right) - \mathbb{P}\left(X > x_{i-1}\right) = (1 - F_N(x_{i-1})) - (1 - F_N(x_i)) = F_N(x_i) - F_N($$

2.2 Indépendance mutuelle

2.2.1 Brève généralisation des résultats précédents

Définition 17. Soit $n \in \mathbb{N}^*$ et $(X_1, \ldots X_n)$ une famille de variables aléatoires réelles. Cette famille est dite mutuellement indépendante si pour tout $(A_1, \ldots, A_n) \in (\mathcal{P}(\mathbb{R}))^n$, la famille $([X_1 \in A_1], \ldots, [X_n \in A_n])$ est mutuellement indépendante.

Lemme 18. *Soit $n \in \mathbb{N}^*$ et $(X_1, \ldots X_n)$ une famille de variables aléatoires réelles sur un univers fini. Cette famille est mutuellement indépendante si et seulement si pour tout $(x_1, \ldots, x_n) \in (X_1(\omega), \ldots, X_n(\Omega))$, la famille $([X_1 = x_1], \ldots, [X_n = x_n])$ est mutuellement indépendante.*

2.2.2 Lemme des coalitions...

Lemme 19 (lemme des coalitions avec deux coalitions)**.** *Soit $X_1, \ldots X_r, X_{r+1}, \ldots, X_{r+n})$ une famille de variables aléatoires réelles mutuellement indépendantes ainsi que f et g deux fonctions définies respectivement sur $\mathbb{R}^r$ et $\mathbb{R}^n$. On pose $C_1 = f(X_1, \ldots, X_r)$ et $C_2 = g(X_{r+1}, \ldots, X_{r+n})$.*

Alors les variables C_1 et C_2 sont indépendantes.

Lemme 20 (lemme des coalitions très général)**.** *Le lemme est long à écrire, le lecteur s'assurera qu'il comprend qu'il s'agit d'une généralisation très naturelle du précédent.*
Soit $(X_1, \ldots X_{n_1}, X_{n_1+n_2}, \ldots, X_{n_1+n_2}, X_{n1+n_2+1}, \ldots, X_{n_1+n_2+\cdots n_r})$ une famille de variables aléatoires réelles mutuellement indépendantes. Ainsi que $(f_1, \ldots, f_r)$ une famille de fonctions à valeurs réelles. Pour tout $j \in [\![1\,; r]\!]$, la fonction f_j et définie sur $\mathbb{R}^{n_j}$.
On pose pour tout $j \in [\![1\,; r]\!]$, $C_j := f_j(X_{n_1+\cdots n_{j-1}+1}, \cdots, X_{n_1+\cdots+n_{j-1}+n_j})$.

Alors la famille $(C_1, C_2, \ldots, C_r)$ est mutuellement indépendante.

2.2.3 ... et ses conséquences

Les trois assertions suivantes sont à connaître et leur preuve utilise un récurrence où l'hérédité fait appel au lemme des coalitions.

Théorème 21. *Soit $n \in \mathbb{N}^*$ et $(X_1, X_2, \ldots, X_n)$ une famille de variables aléatoires réelles mutuellement indépendantes. Alors :*

i) $\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i)$.

ii) $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$.

Théorème 22. *Soit $p \in [0\,;\,1]$; $n \in \mathbb{N}^*$ et $(X_1, X_2, \ldots, X_n)$ une famille de variables aléatoires réelles mutuellement indépendantes. On suppose de plus que pour tout $i \in [\![1\,;\,n]\!]$, la variable X_i suit la loi de Bernoulli de paramètre p (p est donc constant, il ne dépend pas de i).*

Alors la variable $S_n := \sum_{i=1}^{n} X_i$ suit la loi binomiale de paramètres n et p.

Définition 23. Une famille $(X_1, \ldots, X_n)$ de variables aléatoires réelles mutuellement indépendante et dont chaque variable X_i suit la même loi est dite *mutuellement indépendantes et de même loi* (parfois abrégé en iid). On prendra garde que l'indépendance est bien supposée mutuelle dans cette notion même si la terminologie ne le rappelle pas explicitement.

Proposition 24 (Variance d'une somme non indépendante)**.** *Soit $n \in \mathbb{N}^*$ et $(X_1, X_2, \ldots, X_n)$ une famille de variables aléatoires réelles. Alors :*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j).$$

3 Inégalités de concentration

3.1 Inégalité de Markov

Théorème 25. *Soit X une variable aléatoire réelle* ***à valeur positives.*** *Soit également a un réel strictement positif. Alors :*

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}(X)}{a}.$$

Indication de démonstration : Pour réaliser une preuve élégante ne nécessitant pas d'introduire l'ensemble $X(\Omega)$, on pourra considérer la variable aléatoire réelle $\tilde{X}_a$ définie par :

$$\tilde{X}_a = \begin{cases} 0 & \text{si } X < a \\ a & \text{si } X \geqslant a \end{cases}$$

puis écrire des majorations naturelles.

3.2 L'inégalité de Bienaymé-Tchebychev

Théorème 26. *Soit X une variable aléatoire réelle* ***admettant une variance.*** *Soit également ε un réel strictement positif. L'inégalité suivante est vérifiée :*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geqslant \varepsilon) \;\leqslant\; \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

4 Une application fondamentale : statistique et théorie du sondage

4.1 Estimation statistique d'espérance

On introduit une suite infinie $(X_i)_{i\in\mathbb{N}^*}$ de variables aléatoires réelles mutuellement indépendantes et de même loi.

On note alors

$$\forall n \in \mathbb{N}^*,\ S_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1) Montrer sous ce cadre que $\forall n \in \mathbb{N}^*, \mathbb{E}(S_n) = \mathbb{E}(X_1)$.

2) Montrer à l'aide d'une inégalité de Bienaymé-Tchebychev que si la variable X_i admet une variance finie[1], alors pour tout $\varepsilon \in]0\,;+\infty[$, on a :

$$\mathbb{P}\left(|S_n - \mathbb{E}(X_1)| > \varepsilon\right) \xrightarrow{n\to\infty} 0. \tag{2}$$

Ainsi, la probabilité que la valeur moyenne statistique s'écarte d'une valeur non nulle de l'espérance tend vers zéro à mesure que la taille de l'échantillon tend vers l'infini. Un probabiliste dira que la suite des moyennes statistiques *converge en probabilité* vers l'espérance. En fait, il y a même un résultat plus fort mais qui n'est même pas formalisable dans le programme de PT qui affirme que pour si l'on observe une réalisation de cette suite infinie. Alors la probabilité que la suite aléatoire $(S_n(\omega))_{n\in\mathbb{N}}$ converge vers $\mathbb{E}(X_1)$ est exactement égale à 1.

C'est ce résultat crucial qui justifie que l'on puisse estimer une espérance inconnue en observant une moyenne statistique.

4.2 Estimation statistique de probabilité

On considère une famille infinie $(A_i)_{i\in\mathbb{N}^*}$ d'évènements qui est mutuellement indépendante. On suppose que chaque évènement a la même probabilité p d'être réalisée et l'on souhaite estimer cette probabilité p à priori inconnue.

Pour ce faire, on introduit la suite de variable aléatoire $(X_i)_{i\in\mathbb{N}^*}$ définie par :

$$\forall i \in \mathbb{N}^*,\ X_i = \mathbb{1}_{A_i} := \begin{cases} 1 & \text{si l'évènement } A_i \text{ est réalisé} \\ 0 & \text{sinon} \end{cases} \tag{3}$$

3) En appliquant le principe de l'approximation d'espérance des variable X_i par la suite des moyennes statistiques, montrer que :

$$\forall \varepsilon > 0,\ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0 \tag{4}$$

Ainsi, ceci justifie que l'on puisse estimer une probabilité à l'aide d'une fréquence statistique.

1. hypothèse non toujours vérifiée si on commence à travailler avec des univers infinis