Colles semaine 30

En bref

- Intégration sur un segment :
  - Reprise du programme précédent.
  - Application de tout ceci à l'étude des suites et fonctions définies par des intégrales.
- Couple et suite de variables aléatoires réelles :
  - Loi conditionnelle d'une variable aléatoire réelle sachant un évènement de probabilité non nulle.
  - Application à la détermination de loi marginale d'un couple connaissant la loi conjointe.
  - Notion de covariance, variance d'une somme de variables aléatoires réelles.
  - Méthode d'étude de la loi d'une somme, d'un maximum ou d'un minimum de deux variable connaissant la loi conjointe. Cas particulier lorsque les variables sont indépendantes.
  - Indépendance mutuelle d'une famille de variables aléatoires réelles. C'est plus fort que l'indépendance deux à deux !
  - Inégalités de Markov et de Bienaymé-Tchebychev.
  - Application fondamentale au problème de l'estimation statistique (exemple des sondages politiques choisis pour coller à l'actualité électorale).

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.

i) En admettant l'énoncé du théorème des sommes de Riemann sur $[0\,;1]$, démontrer le résultat analogue sur un segment quelconque.

ii) Déterminer $\lim\limits_{n\to\infty} \sum\limits_{k=0}^{n} \frac{1}{n+k}$.

iii) Pour une fonction continue et T périodique, montrer que $a \mapsto \int\limits_{a}^{a+T} f(t)\mathrm{d}t$ est constante.

iv) Pour une fonction paire (respectivement impaire), énoncer et démontrer une formule liant $\int\limits_{-a}^{a} f$ et $\int\limits_{0}^{a} f$.

v) Pour deux variables X et Y définir la covariance et en déduire une formule pour le calcul de $\mathrm{Var}(X+Y)$.

vi) Citer et démontrer l'inégalité de Markov.

vii) Citer et démontrer l'inégalité de Bienaymé-Tchebychev.

Notes aux colleurs

- Nous avons encore à peine effleuré le lemme des coalitions.

En détail

Intégration sur un segment

Reprise du programme précédent avec accent mis sur l'étude des fonctions/suites définies par des intégrales.

Couple et suites de variables aléatoires réelles

1 Lois conditionnelles

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 $\mathbb{P}_A(X = k)$.

2 Rappels sur l'indépendance

2.1 Indépendance de deux variables

2.1.1 Définitions et propriétés élémentaires

Définition 2 (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

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

Proposition 3 (définition alternative pour les variables discrètes)**.** *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 4. *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.

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

Théorème 5. *Soit X et Y deux variables aléatoires réelles indépendantes. Alors :*

i) $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$

ii) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Définition 6. 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 7. Dans le cas $X = Y$, on a $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

Lemme 8. *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 9 (Variance d'une somme non indépendante)**.** *Pour deux variables aléatoires réelles X et Y, on a* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

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 10 (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.

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

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 11 (Loi de maximum)**.** Soit X et Y deux variables aléatoires réelles. On note $M = \max(X, Y)$. 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 $F_Z : \begin{array}{rcl} \mathbb{R} & \to & [0\,;1] \\ x & \mapsto & \mathbb{P}(Z \leqslant x) \end{array}$.

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

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

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 :

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

Remarque 12. 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 13 (Loi d'un minimum)**.** Soit X et Y deux variables aléatoires réelles. On note $N = \min(X, Y)$. Alors $M(\Omega) \subset X(\Omega) \cap Y(\Omega)$. Par ailleurs, la loi se calcule à l'aide des queues de fonctions de répartitions

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

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

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

2.2 Indépendance mutuelle

2.2.1 Brève généralisation des résultats précédents

Définition 14. 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.

Proposition 15 (définition alternative pour les variables discrètes)**.** *Soit $n \in \mathbb{N}^*$ et $(X_1, \ldots X_n)$ une famille de variables aléatoires réelles. 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 16 (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.

On a même une généralisation avec une nombre arbitraire de coalitions.

Lemme 17 (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 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

Théorème 18. *Soit $n \in \mathbb{N}^*$ et $(X_1, X_2, \ldots, X_n)$ une famille de variables aléatoires réelles mutuellement indépendantes. Alors :*

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

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

Théorème 19. *Soit $p \in [0\,;\,1]$; $n \in \mathbb{N}^*$ et $(X_1, X_2, \dots, X_n)$ 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 20. Une famille $(X_1, \dots, X_n)$ 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 21 (Variance d'une somme non indépendante)**.** *Soit $n \in \mathbb{N}^*$ et $(X_1, X_2, \dots, X_n)$ une famille de variables aléatoires réelles. Alors :*

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

3 Inégalités de concentration

3.1 Inégalité de Markov

Théorème 22. *Soit X une variable aléatoire réelle* ***à valeur positives.*** *Soit également a un réel strictement positif. Alors :*

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

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

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