M61B, Probabilités et intégration
Licence 3e année
Parcours mathématiques

DS1-2h

19 Mars 2022

- *La note maximale est 20, il n'y a pas besoin de traiter tout le DS pour l'obtenir. Parcourez rapidement tous les exercices avant de vous lancer, afin de vous concentrer en priorité sur ceux que vous maîtrisez le mieux.*
- *Les exercices sont indépendants entre eux, mais dans un même exercice, les questions ne le sont pas : pensez à utiliser les questions précédentes.*
- *Attention, les questions faciles valent moins de points, attaquez vous aussi à des questions difficiles !*
- *Il est possible d'admettre certaines questions pour traiter les suivantes.*

Exercice 1 : Questions de cours, 4 pts

1) Soit E un ensemble, qu'est ce qu'une tribu sur E ?
2) Qu'est ce qu'un espace de probabilité ?
3) Donner la définition d'une *fonction mesurable* de $(E, \mathscr{T})$ dans $(F, \mathscr{T}')$.
4) Donner la définition et les conditions d'existence de l'espérance, pour une variable aléatoire discrète X à valeurs dans $\mathbb{N}$ et pour une variable aléatoire Y, réelle à densité f_Y.
5) Donner la définition et la fonction de répartition d'une variable de loi exponentielle de paramètre $\lambda > 0$.

Solution : Voir poly. □

Exercice 2 : Quizz, 5 pts

Pour chaque affirmation, indiquer si elle est vraie ou fausse. Si elle est vraie, la justifier, si elle est fausse, l'infirmer, par exemple à l'aide d'un contre exemple.

1) Les ensembles

$$\mathscr{T} := \{A \subset \mathbb{N},\ A \text{ ne contient que des entiers pairs}\}$$

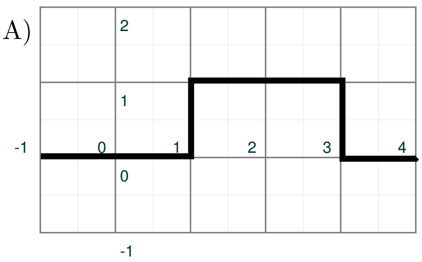
et

$$\mathscr{T}' := \{A \subset \mathbb{N},\ card(A) < \infty\}$$

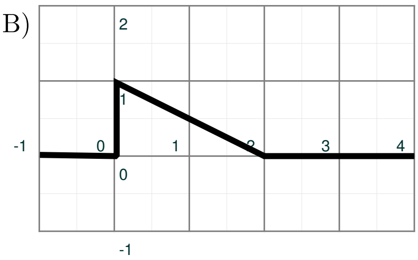
sont ils des tribus sur $\mathbb{N}$?

2) Parmi les fonctions représentées ci-dessous, indiquer lesquelles sont des densités de probabilité.

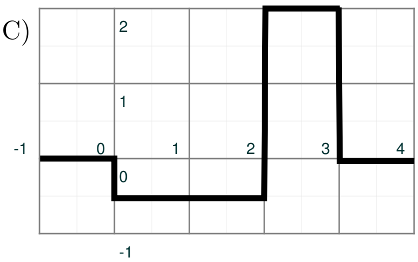
A)



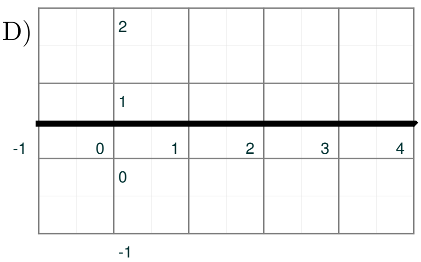
B)



C)



D)



3) Soient X, Y et Z trois variables aléatoires à valeurs dans $\mathbb{N}$. On suppose que X est indépendante de Y et de Z, et que Y et Z sont indépendantes. Alors, pour tous entiers x, y, $z \in \mathbb{N}$, on a

$$\mathbb{P}(X = x, Y = y, Z = z) = \mathbb{P}(X = x)\mathbb{P}(Y = y)\mathbb{P}(Z = z).$$

Solution :

1) L'ensemble $\mathscr{T}$ n'est pas stable par passage au complémentaire, ce n'est donc pas une tribu. L'ensemble $\mathscr{T}$ n'est stable ni par passage au complémentaire, ni par union dénombrable, ce n'en est donc pas une non plus.

2) La fonction représentée en A) est une fonction positive, continue par morceaux, mais son intégrale vaut 2, ce n'est donc pas une densité de probabilité. L'intégrale de B) par contre vaut 1, c'est une densité de probabilités. La courbe de C) ne représente pas une fonction positive, donc même si elle est d'intégrale 1, ce n'est pas une densité. Enfin, on ne voit pas toute la courbe de D), mais même en admettant qu'elle soit positive, l'intégrale de la partie représentée est plus grande que 2, ce n'est donc pas une densité non plus.

3) Ce n'est pas vrai : c'est la différence entre une famille indépendante, et une famille indépendante deux à deux. On peut par exemple prendre deux variables de Bernoulli indépendantes X et Y, et définir $Z = X + Y - 2XY$, qui vaut 1 si X et Y sont différents, et zero sinon. Alors, les variables X, Y et Z sont indépendantes deux-à-deux, mais ne sont pas mutuellement indépendantes :

$$\mathbb{P}(X = Y = Z = 1) = 0 \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 1)\mathbb{P}(Z = 1) = \frac{1}{2^3}.$$

$\square$

Exercice 3 : Grippe bovine, 8 pts

On souhaite dépister, grâce à un contrôle sanguin, quels sont les bovins atteints par un certain virus dans un troupeau de n bêtes. On sait que chaque bête a, indépendamment des autres, une probabilité $p \in]0,1[$ d'être atteinte. Comme chaque test coûte cher, on considère la stratégie suivante : on forme des groupes de m bêtes (où m divise n) ; dans chaque groupe, on mélange les prélèvements sanguins des bovins en un seul « superéchantillon », dans lequel on teste la présence du virus (il suffit qu'une seule bête du groupe soit infectée pour que le test soit positif). Si on décèle la présence du virus dans un super-échantillon, on teste toutes les bêtes du groupe séparément. On appelle cette stratégie "test de super-échantillon de taille m", et on appelle "test individuel" la stratégie consistant à tester automatiquement chaque bête individuellement.

1) Quelle est la loi du nombre de bêtes atteintes dans le troupeau ? Rappeler (sans calcul) son espérance et sa variance.

2) On veut calculer l'espérance $\mathbb{E}(N)$ du nombre N de tests effectués pour la stratégies des super-échantillons.

(i) Quelle est la probabilité qu'un groupe donné soit testé positif ? Quel est alors le nombre de tests effectués sur le groupe ?

(ii) En déduire l'espérance de N en fonction de n, m et p.

3) On fixe un $m \geq 2$ quelconque, on veut comparer les deux stratégies.

(i) Que vaudrait $\mathbb{E}(N)$ pour la stratégie des tests individuels ?

(ii) Si la probabilité d'infection est très grande (p très proche de 1), est-il plus intéressant de tester des super-échantillons de taille m, plutot que d'effectuer des tests individuels ? Justifier à l'aide de l'expression de $\mathbb{E}(N)$ obtenues aux questions 2) - (ii) et 3) - (i).

(iii) Même question si p est très petit (très peu d'infections).

4) On souhaite maintenant étudier l'efficacité des tests de super-échantilons de taille n par rapport aux test individuels.

(i) On pose $q = (1-p)^m$. Que représente cette quantité ?

(ii) On note à nouveau N le nombre de tests effectués pour la stratégie des super-échantillons. Exprimer $n - \mathbb{E}(N)$ en fonction de q.

(iii) On pose $p_0(m) = 1 - \frac{1}{m^{1/m}}$. Pour quelles valeurs de p l'utilisation de super-échantillons est-elle plus intéressante que l'utilisation de tests individuels ?

<u>*Solution*</u> :

1) Le nombre I de bêtes infectées suit une loi binomiale $Bin(n,p)$. Son espérance est $\mathbb{E}(I) = np$, sa variance est $Var(I) = np(1-p)$.

2) (i) On fixe un groupe donné de taille m (le premier par exemple), il est testé négatif si tous les individus (qui sont supposés indépendants) sont sains. La probabilité qu'il soit testé négatif est donc $\mathbb{P}(Bin(m,p) = 0) = (1-p)^m$. La probabilité que le test soit positif est donc $1-(1-p)^m$, auquel cas on effectue $m+1$ tests : 1

pour le super-échantillon, m pour tester individuellement ensuite.

(ii) Si le test de super échantillon est négatif, on ne fait qu'un seul test, par conséquent par la formule des probabilités totales, en notant N_1 le nombre de tests efféctués dans le premier groupe,

$$\mathbb{E}(N_1) = (1-p)^m + (m+1)(1-(1-p)^m).$$

Les groupes étant indépendants et identiquements distribués, on peut maintenant calculer l'espérance du nombre total de tests effectués dans les n/m groupes :

$$\mathbb{E}(N) = \frac{n}{m}\mathbb{E}(N_1) = \frac{n}{m}\left((1-p)^m + (m+1)(1-(1-p)^m)\right).$$

3) (i) Pour la stratégie des tests individuels, le nombre de tests est toujours n, et donc l'espérance du nombre de tests (qui, dans ce cas, n'est pas aléatoire) est n.

(ii) Si p est très proche de 1, on fait $n + n/m$ tests au lieu de n, il vaut donc mieux tester individuellement.

(iii) Si au contraire p est très faible, on ne fait que n/m tests au lieu de n, il vaut donc mieux tester les superéchantillons.

4) (i) C'est la probabilité qu'un groupe n'ait aucun individu malade, et que donc le test de groupe soit négatif.

(ii)
$$n - \mathbb{E}(N) = n - \frac{n}{m}\left(q + (m+1)(1-q)\right) = n\left(q - \frac{1}{m}\right).$$

(iii) Par conséquent, si $q > 1/m$, on fait en moyenne plus de tests si on teste individuellement. Il vaut donc mieux tester par super-échantillons. Si au contraire $q \leq 1/m$ il vaut mieux tester individuellement. Comme $q = (1-p)^m$, on en déduit que pour $p < p_0(m)$ il vaut mieux tester par super-échantillons, et pour $p > p_0(m)$ il vaut mieux tester individuellement.

□

Problème : Théorème de Scheffé, 14 pts

Dans tout le problème, on fixe un espace de probabilités $(\Omega, \mathscr{F}, \mathbb{P})$ sur lequel seront définies toutes les variables aléatoires considérées.

Partie I : Théorème de Scheffé

1) Étant donnée une fonction $g : \mathbb{R} \to \mathbb{R}$, on note

$$g^+ : x \mapsto g(x)\mathbf{1}_{\{g(x)>0\}}$$

la partie positive de g, et

$$g^- : x \mapsto -g(x)\mathbf{1}_{\{g(x)<0\}}$$

sa partie négative. Exprimer les fonctions g et $|g|$ en fonction de g^+ et g^-.

2) Soit $(f_n)_{n\in\mathbb{N}}$ une suite de densités de probabilités sur $\mathbb{R}$. Soit f une autre densité de probabilités sur $\mathbb{R}$, on suppose que f_n converge simplement vers f.

(i) Justifier les inégalités

$$(f-f_n)^+ \leq |f-f_n| \quad \text{et} \quad (f-f_n)^- \leq |f-f_n|,$$

et montrer que $(f-f_n)^+$ et $(f-f_n)^-$ convergent simplement vers 0.

(ii) Montrer que

$$0 \leq (f-f_n)^+(x) \leq f(x).$$

(iii) Justifier que $(f-f_n)^+$ et $(f-f_n)^-$ sont toutes les deux intégrables. Montrer que

$$\int_{\mathbb{R}} (f-f_n)^+(x)dx = \int_{\mathbb{R}} (f-f_n)^-(x)dx = \frac{1}{2}\int_{\mathbb{R}} |f-f_n|(x)dx.$$

(iv) Déduire de ce qui précède la valeur de $\lim_{n\to\infty} \int_{\mathbb{R}} (f-f_n)^+(x)dx$, puis déterminer celle de $\lim_{n\to\infty} \int_{\mathbb{R}} |f-f_n|(x)dx$.

3) On note F_n et F les fonctions de répartition associées à f_n et f respectivement.

(i) Rappeller l'expression de F_n et F en fonction de f_n et f. Soit $x \in \mathbb{R}$, montrer que

$$|F_n(x) - F(x)| \leq \int_{\mathbb{R}} |f_n - f|\,(y)dy.$$

(ii) Déduire de ce qui précède le Théorème de Scheffé, qui affirme que si une suite de densités $(f_n)_{n\in\mathbb{N}}$ converge simplement vers une densité f, alors les fonctions de répartitions correspondantes (F_n) convergent simplement vers la fonction de répartition F associée à f.

Partie II : Réciproque du théorème de Scheffé

4) Pour tout $n \in \mathbb{N}$, on définit

$$f_n(x) = (1+\cos(2\pi n x))\,\mathbf{1}_{[0,1]}(x)$$

(i) Montrer que f_n est une densité de probabilités.

(ii) Pour $n \in \mathbb{N}$ fixé, et pour $x \in \mathbb{R}$, calculer $F_n(x)$.

(iii) Montrer que $F_n(x)$ converge simplement vers une fonction F que l'on déterminera. La fonction F est-elle une fonction de répartition ?

(iv) Soit $x \in \mathbb{R}$ fixé. La suite $(f_n(x))_{n\in\mathbb{N}}$ converge t-elle lorsque $n \to \infty$? La réciproque du théorème de Scheffé est-elle vraie ?

__Solution__ :

Partie I : Théorème de Scheffé

1) On a les identités

$$g = g^+ - g^- \quad \text{et} \quad |g| = g^+ + g^-.$$

2) (i) Les fonctions $(f - f_n)^{\pm}$ et $|f - f_n|$ sont toutes positives, et par ailleurs,

$$(f - f_n)^+ + (f - f_n)^- = |f - f_n|,$$

ce qui donne les bornes voulues. Par ailleurs, comme on a supposé que f_n converge simplement vers f, on a pour tout x que $|f - f_n|(x) \to 0$, et donc en particulier les deux fonctions $(f - f_n)^{\pm}$ convergent simplement vers 0.

(ii) Toutes les fonctions étant positives, l'identité est trivialement vérifiée si $f(x) - f_n(x) \le 0$. Si au contraire $f(x) - f_n(x) > 0$, on a $(f(x) - f_n(x))^+ = f(x) - f_n(x) < f(x)$ car $f_n(x) \ge 0$. Dans tous les cas, on a donc $(f - f_n)^+ \le f$.

(iii) Comme la fonction f est intégrable, par la question précédente, $(f - f_n)^+$ l'est aussi par monotonie de l'intégrale. Par linéarité de l'intégrale, on déduit de la question 1) et du fait que f et f_n sont des densités de probabilité que

$$0 = \int f(x)dx - \int f_n(x)dx = \int (f - f_n)(x)dx = \int (f - f_n)^+(x)dx - \int (f - f_n)^-(x)dx.$$

Par conséquent, $\int (f - f_n)^+(x)dx = \int (f - f_n)^-(x)dx$, et donc en utilisant l'autre identité, on obtient

$$\int (f - f_n)(x)dx + \int (f - f_n)^+(x)dx = \int |f - f_n|(x)dx,$$

ce qui donne le résultat souhaité.

(iv) Comme f est intégrable (car c'est une densité de probabilité), et que $(f - f_n)^+$ converge simplement vers 0, on peut appliquer le théorème de convergence dominée avec domination $g = f$, et on obtient que $(f - f_n)^+$ converge dans $\mathcal{L}^1$ vers la fonction constante égale à 0. En particulier,

$$\lim_{n \to \infty} \int (f - f_n)^+(x)dx = 0.$$

On peut procéder exactement de même pour montrer que $(f - f_n)^-$ converge également vers 0 dans $\mathcal{L}^1$, et donc

$$|f - f_n| = (f - f_n)^+ + (f - f_n)^-$$

converge également dans $\mathcal{L}^1$ vers 0.

3) (i)

$$F_n(x) = \int_{-\infty}^{x} f_n(y)dy \quad \text{and} \quad F(x) = \int_{-\infty}^{x} f(y)dy.$$

Par ailleurs, par monotonie de l'intégrale, et comme pour tout réel u, on a $-|u| \le u \le |u|$

$$-\int_{-\infty}^{x} |f_n - f|(y)dy \le F_n(x) - F(x) = \int_{-\infty}^{x} (f_n - f)(y)dy \le \int_{-\infty}^{x} |f_n - f|(y)dy.$$

On en déduit, toujours par monotonie de l'intégrale, que pour tout $x \in \mathbb{R}$ fixé,

$$|F_n(x) - F(x)| \le \int_{-\infty}^{x} |f_n - f|(y)dy \le \int_{-\infty}^{+\infty} |f_n - f|(y)dy$$

(ii) On a montré dans la question 2) - (iii) que $|f_n - f|$ converge dans L^1 vers 0. En particulier, par théorème des gendarmes, $|F_n(x) - F(x)|$ converge vers 0 pour tout $x \in \mathbb{R}$, ce qui revient exactement à dire que F_n converge simplement vers F.

Partie II : Réciproque du théorème de Scheffé

4) (i) f_n est une fonction positive, mesurable car continue. Par ailleurs, son intégrale sur $\mathbb{R}$ peut être calculée :

$$\int_{\mathbb{R}} f_n(x)dx = \int_{[0,1]} f_n(x)dx = 1 + \int_0^1 \cos(2\pi nx)dx = 1 + \left[\frac{\sin(2\pi nx)}{2\pi n}\right]_0^1 = 1.$$

(ii) Pour $x \leq 0$, $F_n(x)$ vaut 0. De même, pour $x \geq 1$, $F_n(x)$ vaut 1. Pour $x \in [0, 1]$

$$F_n(x) = x + \left[\frac{\sin(2\pi ny)}{2\pi n}\right]_0^x = x + \frac{\sin(2\pi nx)}{2\pi n}.$$

(iii) La fonction F_n converge simplement vers $F(x) := x\mathbf{1}_{[0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x)$, qui est la fonction de répartition d'une variable uniforme sur $[0, 1]$.

(iv) Pour tout x irrationnel fixé, la suite $\cos(2\pi nx)$ ne converge pas, par conséquent, presque-partout sur $[0, 1]$, la suite $f_n(x)$ ne converge pas. En particulier, la réciproque du théorème de Scheffé n'est pas vraie : la convergence simple des fonctions de répartition ne guarantit pas la convergence simple des densités.

$\square$