Université Paris 13, Institut Galilée
Année universitaire 2013-2014 Préparation à l'agrégation

Exercices de probabilités avec éléments de correction

Memento

Fonctions associées aux lois

Pour X variable aléatoire à valeurs dans $\mathbb{R}^d$,
– Fonction de répartition (si $d=1$) : $F_X(t) = P(X \le t)$, $t \in \mathbb{R}$
– Fonction génératrice (si X à valeurs dans $\mathbb{N}$) : $G_X(s) = E[s^X] = \sum_{n=0}^{\infty} P(X=n)s^n$, $s \in |-R, R|$
– Transformée de Laplace : $\mathcal{L}_X(\lambda) = E[e^{-\langle \lambda, X\rangle}] \in]0, +\infty]$, $\lambda \in \mathbb{R}^d$
– Fonction caractéristique : $\Phi_X(t) = E[e^{i\langle t, X\rangle}] \in \mathbb{C}$, $t \in \mathbb{R}^d$

Lois discrètes

Nom	Paramètres	Support	Définition : $P(A) = \sum_{a\in A} p(a)$
Loi de Dirac δ_a	$a \in \mathbb{R}$	$\{a\}$	$p(a) = 1$
Loi de Bernoulli $\mathcal{B}(p)$	$p \in [0,1]$	$\{0,1\}$	$p(0) = 1-p$, $p(1) = p$
Loi binomiale $\mathcal{B}(n,p)$	$n \in \mathbb{N}$, $p \in [0,1]$	$\{0,\dots,n\}$	$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$
Loi géométrique $\mathcal{G}(p)$	$p \in]0,1]$	$\mathbb{N}^*$	$p(k) = (1-p)^{k-1}p$
Loi de Poisson $\mathcal{P}(\lambda)$	$\lambda \in]0,+\infty[$	$\mathbb{N}$	$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$

Lois continues

Nom	Paramètres	Support	Définition : $P(A) = \int_A f(x)dx$
Loi uniforme $\mathcal{U}([a,b])$	$a < b$	$[a,b]$	$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$
Loi exponentielle $\mathcal{E}(\lambda)$	$\lambda \in]0,\infty[$	$]0,+\infty[$	$f(x) = \lambda e^{-\lambda x}\mathbf{1}_{]0,+\infty[}(x)$
Loi de Cauchy	$a \in]0,+\infty[$	$\mathbb{R}$	$f(x) = \frac{a}{\pi(a^2+x^2)}$
Loi normale/gaussienne $\mathcal{N}(m,\sigma^2)$	$m \in \mathbb{R}$, $\sigma^2 \in]0,+\infty[$	$\mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

Déterminer des lois : exemples

Exercice 1. *Lois binomiale et géométrique*
Soit $X_1, X_2, \dots$ une suite de variables aléatoires indépendantes et de loi $\mathcal{B}(p)$ où $p \in [0,1]$.
1. On suppose $p > 0$. On définit $N = \inf\{n \ge 1 | X_n = 1\}$.
1.a) Montrer que $P(N = \infty) = 0$ et que N suit la loi géométrique de paramètre p.
1.b) Calculer l'espérance et la variance de N.
2. Soit $n \ge 1$. On définit $S_n = X_1 + \cdots + X_n$.
2.a) Montrer que S_n suit la loi binomiale de paramètres n et p, par une preuve directe puis en utilisant des fonctions génératrices.
2.b) Calculer l'espérance et la variance de S_n (utiliser la définition de S_n).

Exercice 2. *Minimum et maximum d'une famille de variables aléatoires exponentielles*
Soit X, Y deux variables aléatoires indépendantes de lois respectives $\mathcal{E}(\lambda)$ et $\mathcal{E}(\mu)$. À l'aide de fonctions de répartition, déterminer les lois de $U = \min(X,Y)$ et $V = \max(X,Y)$. On précisera leur densité (le cas échéant).

Exercice 3. *Somme de variables aléatoires*
1. Soit X, Y des variables aléatoires indépendantes de lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Déterminer la loi de $X+Y$, directement puis via les fonctions génératrices.
2. Soit X, Y des variables aléatoires indépendantes de loi de Cauchy de paramètre a et b. À l'aide des fonctions caractéristiques, déterminer la loi de $X+Y$. *Pour obtenir Φ_X, on pourra utiliser la formule de Cauchy avec un contour bien choisi, ou alors avoir l'idée de calculer la fonction caractéristique de la loi de Laplace $\frac{a}{2}e^{-a|x|}\mathrm{d}x$ et utiliser la formule d'inversion.*

Exercice 4. *Lois images*
1. Soit X une variables aléatoire de loi $\mathcal{E}(\lambda)$. Déterminer la loi de $\lfloor X \rfloor + 1$. *C'est une loi géométrique.*
2. Soit U une variable aléatoire de loi $\mathcal{U}([-1,1])$. Déterminer la loi de $\arcsin(U)$.
3. Soit X de loi $\mathcal{N}(0,1)$. Déterminer la loi de $|X|$.

4. Soit X, Y deux variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. Déterminer la loi de $\frac{X}{Y}$. En déduire la loi de $\frac{1}{Z}$ si Z suit une loi de Cauchy de paramètre 1.
5. Soit X, Y deux variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. On définit les variables aléatoires R, Θ par $(X,Y) = (R\cos\Theta, R\sin\Theta)$, $R > 0$ et $\Theta \in [0, 2\pi[$. Montrer que R et Θ sont indépendantes et déterminer leurs lois.

Exercice 5. *Loi Gamma*
Pour $a > 0$ et $\lambda > 0$, on définit la loi $\gamma_{a,\lambda}$ par sa densité relativement à la mesure de Lebesgue :

$$f_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}^+}(x).$$

1. Vérifier que cette fonction définit bien une densité.
2. Déterminer l'espérance de cette loi.

On utilise le fait que $\Gamma(a+1) = a\Gamma(a)$ pour obtenir que l'espérance de cette loi est a/λ.

3. Soit $V_1, V_2, \ldots, V_n$ des variables aléatoires réelles indépendantes de loi $\mathcal{E}(\lambda)$. Déterminer la loi du vecteur $(V_1, V_1 + V_2, \ldots, V_1 + \cdots + V_n)$ et en déduire que $V_1 + \cdots + V_n \sim \gamma_{n,\lambda}$.

Pour $n = 1$, ok. Supposons $n \geq 2$ et $S := V_1 + \ldots + V_{n-1}$ de loi $\gamma_{n-1,\lambda}$. Soit g une fonction mesurable bornée de $\mathbb{R}$ dans $\mathbb{R}$. On a

$$\mathbb{E}(g(V_1 + \ldots + V_n)) = \mathbb{E}(g(S + V_n)) = \int_{\mathbb{R}} g(x+y)dP_{(S,V_n)}(x,y)$$

et

$$\mathbb{E}(g(V_1 + \ldots + V_n)) = \int_{\mathbb{R}} g(t)dP_{V_1+\ldots+V_n}(t).$$

Comme $f(v_1, \ldots, v_{n-1}) = v_1 + \ldots + v_{n-1}$ et $g(v_n) = v_n^2$ mesurables on en déduit que S et V_n sont indépendantes car $(V_1, \ldots, V_{n-1})$ et V_n le sont,

$$\begin{aligned}
\int_{\mathbb{R}} g(x+y)dP_{(S,V_n)}(x,y) &= \int_0^\infty dx \int_x^\infty dt g(t) \frac{\lambda^{n-1}}{\Gamma(n-1)} e^{-\lambda t} x^{n-2} \\
&= \int_0^\infty g(t) \frac{\lambda^{n-1}}{\Gamma(n-1)} e^{-\lambda t} \left[x^{n-1}/(n-1)\right]_0^t dt \\
&= \int_{\mathbb{R}} g(t) \frac{\lambda^n}{\Gamma(n)} \exp(-\lambda t) t^{n-1} \mathbf{1}_{\mathbb{R}^+}(t) dt
\end{aligned}$$

4. Soit X et Y deux variables aléatoires réelles indépendantes de loi $\gamma_{a,\lambda}$.
4.a) Déterminer la loi de λX.

On peut utiliser la fonction de répartition. Avec un changement de variable on voit que $\lambda X \sim \gamma_{a,1}$.

4.b) Montrer que $X + Y$ et X/Y sont des v.a. indépendantes dont on calculera les lois.

Soit g une fonction mesurable bornée de $\mathbb{R}^2$ dans $\mathbb{R}^2$. On a

$$\mathbb{E}(g(X+Y, X/Y)) = \int_{\mathbb{R}^2} g(u,v)dP_{(X+Y,X/Y)}(u,v)$$

et

$$\mathbb{E}(g(X+Y, X/Y)) = \int_{\mathbb{R}^2} g \circ f(x,y)dP_{(X,Y)}(x,y)$$

où $f(x,y) = (x+y, x/y)$ définie de $(\mathbb{R}^{*+})^2$ vers $(\mathbb{R}^{*+})^2$. Comme les variables X et Y sont indépendantes, le couple (X,Y) a pour densité $dP_X(x)dP_Y(y)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$.
On fait alors le changement de variable $u = x + y$, $v = x/y$, pour $x > 0$ et $y > 0$;
Ceci est équivalent à $x = uv/(v+1)$ et $y = u/(v+1)$ pour $u > 0$ et $v > 0$.
On a de plus $|J(u,v)| = \begin{vmatrix} v/(v+1) & u/(v+1) \\ 1/(v+1) & -u/(v+1)^2 \end{vmatrix} = \dfrac{u}{(v+1)^2}$. Il suit

$$\mathbb{E}(g(X+Y, X/Y)) = \int_{\mathbb{R}^2} g(u,v) u^{2a-1} e^{-\lambda u} \mathbf{1}_{u>0} \frac{v^{a-1}}{(v+1)^{2a}} \mathbf{1}_{v>0} \frac{\lambda^{2a}}{\Gamma(a)^2} dudv.$$

Les variables sont indépendantes, $dP_{X+Y}(u) = \frac{\lambda^{2a}}{\Gamma(2a)} u^{2a-1} e^{-\lambda u} \mathbf{1}_{u>0} du$ et $dP_{X/Y}(v) = \frac{\Gamma(2a)}{\Gamma(a)^2} \frac{v^{a-1}}{(v+1)^{2a}} \mathbf{1}_{v>0} dv.$

4.c) Montrer que $X + Y$ et $X/(X + Y)$ sont des v.a. indépendantes. Calculer la loi de $X/(X + Y)$.

Soit g une fonction mesurable bornée de $\mathbb{R}^2$ dans $\mathbb{R}^2$. On a

$$\mathbb{E}(g(X+Y, X/(X+Y))) = \int_{\mathbb{R}^2} g(u,v) dP_{(X+Y,X/(X+Y))}(u,v)$$

et

$$\mathbb{E}(g(X+Y, X/(X+Y))) = \int_{\mathbb{R}^2} g \circ f(x,y) dP_{(X+Y,X/(X+Y))}(x,y)$$

où $f(x,y) = (x+y, x/(x+y))$ définie de $(\mathbb{R}^{*+})^2$ vers $(\mathbb{R}^{*+})^2$. Comme les variables X et Y sont indépendantes, le couple (X,Y) a pour loi $dP_X dP_Y = f_{a,\lambda}(x) f_{a,\lambda}(y) dx dy$.
On fait alors le changement de variable $u = x + y$, $v = x/(x + y)$, pour $x > 0$ et $y > 0$;
Ceci est équivalent à $x = uv$ et $y = u(1 - v)$ pour $u > 0$ et $v \in (0,1)$.
On a de plus $|J(u,v)| = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = u$. Il suit

$$\mathbb{E}(g(X+Y, X/(X+Y))) = \int_{\mathbb{R}^2} g(u,v) \frac{\lambda^{2a}}{\Gamma(2a)} u^{2a-1} e^{-\lambda u} \mathbf{1}_{u>0} \frac{\Gamma(2a)}{\Gamma(a)^2} \left(v(1-v)\right)^{a-1} \mathbf{1}_{0<v<1} du dv.$$

Les variables sont donc indépendantes et on a de plus $dP_{X/(X+Y)}(v) = \frac{\Gamma(2a)}{\Gamma(a)^2} \left(v(1-v)\right)^{a-1} \mathbf{1}_{0<v<1} dv.$

5. Soit X et Y deux variables aléatoires réelles indépendantes de loi $\gamma_{a,\lambda}$ et $\gamma_{b,\lambda}$ respectivement. Déterminer la loi de $X + Y$.

Le seul point délicat est de calculer $\int_0^t x^{a-1}(t-x)^{b-1} dx = t^{a+b-1} \int_0^1 y^{a-1}(1-y)^{b-1} dy = t^{a+b-1} C_{a,b}$. La constante $C_{a,b}$ est forcément égale à $\Gamma(a)\Gamma(b)/\Gamma(a+b)$ en tenant compte de la normalisation.

6. Soit $Z_1, Z_2, \ldots, Z_n$ des variables aléatoires réelles indépendantes de loi $\mathcal{N}(0,1)$.
6.a) Montrer que Z_1^2 suit une loi $\gamma_{1/2,1/2}$.

Si Z_1 est de loi $\mathcal{N}(0,1)$ et g une fonction mesurable bornée de $\mathbb{R}$ dans $\mathbb{R}$, on a

$$\mathbb{E}(g(X^2)) = \int_{\mathbb{R}} g(u) dP_{X^2}(u) \qquad \mathbb{E}(g(X^2)) = \int_{\mathbb{R}} g(x^2) dP_X(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x^2) e^{-x^2/2} dx.$$

Par parité de $x \mapsto g(x^2)e^{-x^2/2}$ on a $\mathbb{E}(g(X^2)) = \frac{2}{\sqrt{2\pi}} \int_0^\infty g(x^2) e^{-x^2} dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty g(y) e^{-y/2} \frac{dy}{2\sqrt{y}}$ donc $dP_{X^2}(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{-1/2} \mathbf{1}_{\mathbb{R}^+}(y) dy.$

6.b) Montrer que $Z_1^2 + \cdots + Z_n^2$ suit une loi $\gamma_{n/2,1/2}$. *La loi $\gamma_{n/2,1/2}$ est également appelée loi du khi-deux à n degrés de liberté, notée χ_n^2.*

On le montre par récurrence. Pour $n = 1$ c'est vrai. Supposons que $S_{n-1} = Z_1^2 + \ldots + Z_{n-1}^2 \sim \gamma_{\frac{n-1}{2},\frac{1}{2}}$ et $Z_n \sim \mathcal{N}(0,1)$. On a $S_n = S_{n-1} + Z_n^2$. Comme $f(z_1, \ldots, z_{n-1}) = z_1^2 + \ldots + z_{n-1}^2$ et $g(x_n) = z_n^2$ mesurables on en déduit que S_{n-1} et Z_n^2 sont indépendantes car $(Z_1, \ldots, Z_{n-1})$ et Z_n le sont. On utilise ensuite la question 5 donnant que S_n suit une $\gamma_{\frac{n-1}{2}+\frac{1}{2},\frac{1}{2}} = \gamma_{\frac{n}{2},\frac{1}{2}}$.

Propriétés générales

Exercice 6. *Conséquences du théorème de Fubini, fonctions indicatrices*
Résoudre les questions suivantes en appliquant le théorème de Fubini(-Tonelli) de la façon suggérée.
1. Soit N une variable aléatoire à valeurs dans $\mathbb{N}$. Montrer que

$$\mathbb{E}[N] = \sum_{n \geq 1} \mathbb{P}(N \geq n).$$

On note que, comme N est à valeurs entières, $N = \sum_{k=1}^{N} 1 = \sum_{k=1}^{\infty} \mathbf{1}_{\{k \leq N\}}$. Le théorème de Fubini-Tonelli donne

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{k=1}^{\infty} \mathbf{1}_{\{k \leq N\}}\right] = \sum_{k=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{k \leq N\}}] = \sum_{k=1}^{\infty} \mathbb{P}(k \leq n).$$

Le théorème de Fubini est ici appliqué à la fonction $(n, \omega) \mapsto \mathbf{1}_{\{k \leq N(\omega)\}}$ par rapport à la mesure produit $m_{\mathbb{N}} \otimes \mathbb{P}$, où $m_{\mathbb{N}}$ est la mesure de comptage sur $\mathbb{N}$: $m_{\mathbb{N}}(A) = \mathrm{Card}(A)$ si $A \subset \mathbb{N}$ (et donc $\int f dm_{\mathbb{N}} = \sum_{n \in \mathbb{N}} f(n)$ pour $f : \mathbb{N} \to \mathbb{R}$). En l'occurrence, il est en fait plus simple de voir ceci comme une application du théorème de convergence monotone pour les séries à termes positifs.

2. Soit X une variable aléatoire à valeurs dans $\mathbb{R}_+$, et $\alpha > 0$. Montrer que

$$\mathbb{E}[X^\alpha] = \alpha \int_0^\infty t^{\alpha-1} \mathbb{P}(X > t) dt$$

et donner une généralisation de cette formule.

On note que, comme $X \geq 0$, par « intégration de la dérivée », $X^\alpha = \int_0^X \alpha t^{\alpha-1} dt = \int_0^\infty \mathbf{1}_{\{t<X\}} \alpha t^{\alpha-1} dt$. Le théorème de Fubini-Tonelli (pour la mesure $dt \otimes \mathbb{P}$) donne donc

$$\mathbb{E}[X^\alpha] = \int_0^\infty \mathbb{E}[\mathbf{1}_{\{t<X\}} t^{\alpha-1}] dt = \int_0^\infty \mathbb{P}(X > t) \alpha t^{\alpha-1} dt.$$

Le principe de la preuve s'applique par exemple à toute fonction g monotone de classe $\mathcal{C}^1$ de $]0, +\infty[$ dans $\mathbb{R}$, pour laquelle on peut écrire $g(X) = g(0) + \int_0^X g'(t) dt$, d'où de même

$$\mathbb{E}[g(X)] = g(0) + \int_0^\infty \mathbb{P}(X > t) g'(t) dt.$$

3. Soit $(A_n)_{n \geq 1}$ une suite d'événements.
3.a) On note N le nombre (aléatoire) d'événéments parmi ceux-ci qui se produisent. Calculer $\mathbb{E}[N]$ en fonction des probabilités $\mathbb{P}(A_n)$, $n \geq 1$.

On note que $N = \sum_{n=1}^{\infty} \mathbf{1}_{A_n}$. Par suite, par le théorème de Fubini-Tonelli (pour la mesure $m_{\mathbb{N}} \otimes \mathbb{P}$ où $m_{\mathbb{N}}$ est la mesure de comptage sur $\mathbb{N}$),

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} \mathbb{E}[\mathbf{1}_{A_n}] = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

3.b) On suppose que $\sum_n \mathbb{P}(A_n) < \infty$. Montrer que presque-sûrement seul un nombre fini d'événements de la suite ont lieu. *C'est le lemme de Borel-Cantelli (partie la plus facile mais néanmoins la plus utile).*

Par la question précédente, l'hypothèse équivaut à $\mathbb{E}[N] < \infty$. Or ceci implique que $N < \infty$ p.s., ce qui signifie que, presque sûrement, un nombre fini d'événement de la suite ont lieu.

4. Calculer $C = \int_{\mathbb{R}} e^{-x^2} dx$ sans utiliser de coordonnées polaires. (Écrire C^2 comme une intégrale double puis, dans l'intégrale, $e^{-(x^2+y^2)} = \int_{x^2+y^2}^{\infty} e^{-t} dt$)

Par le théorème de Fubini-Tonelli,

$$C^2 = \int_0^\infty e^{-x^2} dx \int_0^\infty e^{-y^2} dy = \int_{]0,\infty[^2} e^{-(x^2+y^2)} dx dy.$$

En écrivant (par une intégration immédiate) $e^{-(x^2+y^2)} = \int_{x^2+y^2}^{\infty} e^{-t} dt = \int_0^\infty \mathbf{1}_{\{t > x^2+y^2\}} e^{-t} dt$ pour $x, y > 0$, on a, à nouveau par le théorème de Fubini-Tonelli,

$$C^2 = \int_{]0,\infty[^2} \int_0^\infty \mathbf{1}_{\{t > x^2+y^2\}} e^{-t} dt dx dy = \int_0^\infty e^{-t} \left(\int_{]0,\infty[^2} \mathbf{1}_{\{t > x^2+y^2\}} dt \right) dx dy$$

L'intégrale entre parenthèses est l'aire du quart de disque de rayon $\sqrt{t}$, donc vaut $\frac{\pi t}{4}$ (certes, s'il fallait le redémontrer, ceci se ferait le plus directement en passant en coordonnées polaires...). On a donc

$$C^2 = \frac{\pi}{4} \int_0^\infty t e^{-t} dt = \frac{\pi}{4}$$

(intégration par parties, ou reconnaître l'espérance de la loi $\mathcal{E}(1)$ pour éviter de refaire le calcul), d'où $C = \frac{\sqrt{\pi}}{2}$. On rappelle la preuve classique : partant de $C^2 = \int_{]0,+\infty[^2} e^{-(x^2+y^2)}dxdy$ comme ci-dessus, il suffit d'effectuer un changement de coordonnées pour passer en coordonnées polaires ($(r,\theta) \mapsto (r\cos\theta, r\sin\theta)$ est bien un difféomorphisme de $]0,+\infty[\times]0,\frac{\pi}{2}[$ sur $]0,+\infty[^2$, de jacobien $r \mapsto r$) :

$$C^2 = \int_0^{\pi/2}\int_0^{\infty} e^{-r^2} r dr d\theta = \frac{\pi}{2}\left[-\frac{1}{2}e^{-r^2}\right]_{r=0}^{\infty} = \frac{\pi}{4}.$$

5. Soit $A_1, \ldots, A_n$ des événements. Montrer la *formule du crible*, où $|S|$ désigne le cardinal de S :

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n}(-1)^{k+1}\sum_{1\le i_1<\cdots<i_k\le n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}) = \sum_{S\subset\{1,\ldots,n\},S\neq\emptyset}(-1)^{|S|+1}\mathbb{P}\Big(\bigcap_{i\in S} A_i\Big).$$

Pour tous événements A et B, on a $\mathbf{1}_{A\cap B} = \mathbf{1}_A\mathbf{1}_B$ et $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$, en revanche il n'y a pas de formule aussi simple pour la réunion, mais on peut se ramener à une intersection en passant au complémentaire :

$$\mathbf{1}_{A\cup B} = 1 - \mathbf{1}_{A^c\cap B^c} = 1 - \mathbf{1}_{A^c}\mathbf{1}_{B^c} = 1 - (1-\mathbf{1}_A)(1-\mathbf{1}_B) = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A\mathbf{1}_B = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B},$$

ce qui, en intégrant chaque membre, donne $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cap B)$. On généralise maintenant cette formule à n événements. On a

$$\begin{aligned}\mathbf{1}_{A_1\cup\cdots\cup A_n} &= 1 - \mathbf{1}_{A_1^c\cap\cdots\cap A_n^c} = 1 - \prod_i (1-\mathbf{1}_{A_i})\\ &= 1 - \sum_{S\subset\{1,\ldots,n\}}(-1)^{1+|S|}\prod_{i\in S}\mathbf{1}_{A_i}.\end{aligned}$$

(où, en développant, S désigne l'ensemble des indices du produit où l'on garde le terme $-\mathbf{1}_{A_i}$ au lieu du terme 1). En prenant l'espérance de chaque terme, on obtient l'égalité demandée (l'expression avec les indices i_j est une réécriture en utilisant l'injection $S \mapsto (i_1, \ldots, i_k)$ où $k = |S|$, $S = \{i_1, \ldots, i_k\}$ avec $i_1 < \ldots < i_k$, cette écriture étant unique). Par exemple,

$$\mathbb{P}(A\cup B\cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A\cap B) - \mathbb{P}(A\cap C) - \mathbb{P}(B\cap C) + \mathbb{P}(A\cap B\cap C).$$

Exercice 7. *Autour de l'indépendance*

1. Soit X, Y, Z des variables aléatoires indépendantes, de loi $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. On note $X' = YZ$, $Y' = XZ$ et $Z' = XY$. Montrer que X', Y', Z' sont des variables aléatoires de loi μ indépendantes deux à deux, mais non indépendantes (dans leur ensemble).

On note que X', Y' et Z' sont à valeurs dans $\{-1, 1\}$ et sont définies de manière symétrique (cyclique), donc pour montrer l'indépendance 2 à 2 il va suffire de vérifier que

$$\mathbb{P}(X' = 1, Y' = 1) = \mathbb{P}(X' = 1)\mathbb{P}(Y' = 1).$$

En effet on a ensuite $\mathbb{P}(X' = 1, Y' = -1) = \mathbb{P}(X' = 1) - \mathbb{P}(X' = 1, Y' = -1) = \mathbb{P}(X' = 1) - \mathbb{P}(X' = 1)\mathbb{P}(Y' = 1) = \mathbb{P}(X' = 1)\mathbb{P}(Y' = -1)$, puis on a passe de même à $\mathbb{P}(X' = -1, Y' = -1) = \mathbb{P}(X' = -1) - \mathbb{P}(X' = -1, Y' = 1)$, et la symétrie induit les autres vérifications. Or on a

$$\mathbb{P}(X' = 1, Y' = 1) = \mathbb{P}(YZ = 1, XZ = 1) = \mathbb{P}(X = Y = Z) = \mathbb{P}(X = 1)^3 + \mathbb{P}(X = -1)^3 = \frac{1}{4}$$

et

$$\mathbb{P}(X' = 1) = \mathbb{P}(YZ = 1) = \mathbb{P}(Y = 1, Z = 1) + \mathbb{P}(Y = -1, Z = -1) = \mathbb{P}(Y = 1)^2 + \mathbb{P}(Y = -1)^2 = \frac{1}{2},$$

d'où l'égalité voulue. X, Y, Z sont donc indépendantes 2 à 2. En revanche,

$$\mathbb{P}(X' = 1, Y' = 1, Z' = -1) = \mathbb{P}(YZ = 1, XZ = 1, XY = -1) = 0 \neq \mathbb{P}(X' = 1)\mathbb{P}(Y' = 1)\mathbb{P}(Z' = -1),$$

car $YZ = 1$ et $XZ = 1$ impliquent $X = Z = Y$ donc $XY = 1$. Ainsi, X, Y, Z ne sont pas indépendantes dans leur ensemble.

2. Est-ce qu'un événement A peut être indépendant de lui-même ? Même question pour une variable aléatoire X.

A est indépendant de lui-même si, et seulement si $\mathbb{P}(A\cap A)=\mathbb{P}(A)\mathbb{P}(A)$, c'est-à-dire $\mathbb{P}(A)=\mathbb{P}(A)^2$, ce qui équivaut à $\mathbb{P}(A)=0$ ou 1 : les événements négligeables et presque sûrs sont les seuls à être indépendants d'eux-même (et de n'importe quel autre événement, d'ailleurs).
Si une v.a. X est indépendante d'elle-même, à valeurs dans $(E,\mathcal{E})$, pour toute partie $A\in\mathcal{E}$ on a $\mathbb{P}(X\in A)=\mathbb{P}(X\in A, X\in A)=\mathbb{P}(X\in A)^2$ car l'événement $\{X\in A\}$ est indépendant de lui-même, d'où $\mathbb{P}(X\in A)\in\{0,1\}$. Si X est entière, on peut noter que $1=\sum_{n\in\mathbb{N}}\mathbb{P}(X=x)$ et chaque terme vaut 0 ou 1, donc un seul vaut 1 : il existe $n\in\mathbb{N}$ tel que $X=n$ presque sûrement (on dit que X est presque sûrement constante). Si, plus généralement, X est à valeurs dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, par le même argument, pour tout n, il existe un unique intervalle de la forme $[k2^{-n},(k+1)2^{-n}[$, où $k\in\mathbb{Z}$, qui contient X presque sûrement ; ces intervalles doivent être emboîtés, et leur largeur tend vers 0, ce qui détermine X, donc X est constante presque sûrement. On peut généraliser le raisonnement à $\mathbb{R}^d$ (muni des boréliens), et à des espaces satisfaisant quelques hypothèses naturelles, mais pas à n'importe quel espace $(E,\mathcal{E})$. Entre autres, la tribu doit être assez fine (si $\mathcal{E}=\{\emptyset,E\}$ est la tribu grossière, toutes les fonctions à valeurs dans $(E,\mathcal{E})$ sont des variables aléatoires, et elles sont toutes indépendantes d'elles-mêmes). En revanche, si E est un espace métrique complet séparable (*espace polonais*), et $\mathcal{E}$ est la tribu borélienne, l'argument s'étend (il existe une suite $(x_k)_k$ dense dans E, donc pour tout n les boules $B(x_k,2^{-n})$ $(k\in\mathbb{N})$ recouvrent E d'où $1=\mathbb{P}(X\in E)\le\sum_{k\in\mathbb{N}}\mathbb{P}(d(X,x_k)<2^{-n})$, or ces probabilités valent 0 ou 1, donc il existe $x^{(n)}$ tel que $d(X,x^{(n)})<2^{-n}$ presque sûrement ; l'inégalité triangulaire implique que $(x^{(n)})_n$ est une suite de Cauchy, et X est presque sûrement égal à la limite de cette suite).
Inversement, toute variable aléatoire X constante presque sûrement (c'est-à-dire qu'il existe $x\in E$ tel que $\mathbb{P}(X=x)=1$) est indépendante d'elle-même (et de n'importe quelle variable aléatoire).

Problèmes (simples) classiques

Exercice 8. *Formule de Bayes.*
1. Soit $(\Omega,\mathbb{P})$ un espace de probabilité discret, et $(H_1,\ldots,H_n)$ une partition de Ω en n événements de probabilité non nulle. Montrer que, pour $i=1,\ldots,n$, si A est un événement de probabilité non nulle :

$$\mathbb{P}(H_i|A)=\frac{\mathbb{P}(A|H_i)\mathbb{P}(H_i)}{\sum_{j=1}^n\mathbb{P}(A|H_j)\mathbb{P}(H_j)}.$$

2. Une maladie M affecte une personne sur 1000 dans une population donnée. On dispose d'un test sanguin qui détecte M avec une fiabilité de 99% lorsque cette maladie est effectivement présente. Cependant, on obtient aussi un résultat faussement positif pour 0,2% des personnes saines testées. Quelle est la probabilité qu'une personne soit réellement malade lorsque son test est positif ?

La population est notée Ω, on la munit de la mesure uniforme $\mathbb{P}$. On note M l'ensemble des personnes malades, et T l'ensemble des personnes dont le test est positif. L'énoncé donne donc $\mathbb{P}(M)=0.001$, $\mathbb{P}(T|M)=0.99$ et $\mathbb{P}(T|M^c)=0.002$, et on cherche $\mathbb{P}(M|T)$. Comme la famille (M,M^c) partitionne la population, on peut appliquer la formule de Bayes :

$$\mathbb{P}(M|T)=\frac{\mathbb{P}(T|M)\mathbb{P}(M)}{\mathbb{P}(T)}=\frac{\mathbb{P}(T|M)\mathbb{P}(M)}{\mathbb{P}(T|M)\mathbb{P}(M)+\mathbb{P}(T|M^c)\mathbb{P}(M^c)}=\frac{\frac{99}{100}\frac{1}{1000}}{\frac{99}{100}\frac{1}{1000}+\frac{2}{1000}\frac{99}{100}}=\frac{1}{3}\simeq 33\%$$

Ainsi, bien que le résultat soit rarement faussement positif, la rareté plus grande encore de la maladie fait que la plupart (66%) des résultats positifs sont en fait faussement positifs, ce qui peut apparaître de prime abord comme un paradoxe.

Exercice 9. *Problèmes « avec des mots »*
1. Un secrétaire vient de mettre cent lettres dans des enveloppes comportant des adresses avant de se rendre compte que les lettres étaient nominatives. Quelle est la probabilité que pas une seule des lettres ne soit dans la bonne enveloppe ? En donner une valeur approchée. *Penser à la formule du crible.*

Posons $n=100$. Un espace de probabilités possible est $(\mathfrak{S}_n,\mathcal{P}(\mathfrak{S}_n),P)$ où P est la loi de probabilité uniforme P sur $\mathfrak{S}_n$. Pour $\sigma\in\mathfrak{S}_n$ et $i\in\{1,\ldots,n\}$, $\sigma(i)$ donne le numéro de la personne qui recevra la lettre destinée à la i-ième personne. On note $A_i=\{\sigma(i)=i\}$ l'événement {la personne i recevra sa lettre}, pour $i=1,\ldots,n$. On cherche $\mathbb{P}((A_1\cup\cdots\cup A_n)^c)=1-\mathbb{P}(A_1\cup\cdots\cup A_n)$. En vue d'appliquer la formule du crible, on calcule, pour $1\le k\le n$, et $1\le i_1<\cdots<i_k\le n$,

$$\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})=\mathbb{P}(\sigma(i_1)=i_1,\ldots,\sigma(i_k)=i_k)$$

et pour cela on constate que le nombre de façons de réaliser l'événement ci-dessus est $(n-k)!$ car, en-dehors de $\{i_1,\ldots,i_k\}$, la restriction de σ est une bijection quelconque. On a donc

$$\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})=\frac{(n-k)!}{n!}.$$

Par la formule du crible,

$$\begin{aligned}\mathbb{P}(A_1\cup\cdots\cup A_n)&=\sum_{k=1}^n(-1)^{k+1}\sum_{1\le i_1<\cdots<i_k\le n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_n})=\sum_{k=1}^n(-1)^{k+1}\sum_{1\le i_1<\cdots<i_k\le n}\frac{(n-k)!}{n!}\\&=\sum_{k=1}^n(-1)^{k+1}\binom{n}{k}\frac{(n-k)!}{n!}=\sum_{k=1}^n(-1)^{k+1}\frac{1}{k!}\end{aligned}$$

d'où finalement la probabilité qu'aucune lettre n'arrive à son destinataire :

$$\mathbb{P}((A_1\cup\cdots\cup A_n)^c)=\sum_{k=0}^n\frac{(-1)^k}{k!}.$$

Remarquons que cette suite n'est ni croissante, ni décroissante (ce qui peut étonner), et qu'elle converge très rapidement vers $e^{-1}\simeq 0.368$ (le fait que la probabilité d'envoyer une lettre au bon destinataire ne dépend presque pas du nombre de lettres peut aussi surprendre).

2. Quelles est la probabilité que, dans un groupe de 25 personnes, deux au moins aient leur anniversaire le même jour ? (On négligera les années bissextiles, on supposera les dates de naissances équiprobables et indépendantes) En donner une valeur approchée.

Supposons qu'il y a N personnes, dont les dates de naissances $X_1,\ldots,X_N$ sont indépendantes et de loi uniforme dans $\{1,\ldots,A\}$ où $A=365$. Alors on cherche

$$\begin{aligned}\mathbb{P}(2\text{ anniversaires le même jour})&=1-\mathbb{P}(X_1,\ldots,X_n\text{ distincts})\\&=1-\sum_{a_1,\ldots,a_N\text{ distincts}}\mathbb{P}(X_1=a_1,\ldots,X_N=a_N)=1-\sum_{a_1,\ldots,a_N\text{ distincts}}\left(\frac{1}{A}\right)^N\\&=1-\frac{A(A-1)\cdots(A-N+1)}{A^N}.\end{aligned}$$

On veut montrer que cette valeur est grande (supérieure à 0.5) ; donnons donc une minoration, qui donne une idée de l'ordre de grandeur et se prête mieux par exemple à une évaluation sur une calculatrice : en utilisant l'inégalité $1-x\le e^{-x}$ pour $x\in\mathbb{R}$,

$$\begin{aligned}\mathbb{P}(2\text{ anniversaires le même jour})&=1-\left(1-\frac{1}{A}\right)\left(1-\frac{2}{A}\right)\cdots\left(1-\frac{N-1}{A}\right)\\&\ge 1-e^{-\frac{1}{A}}e^{-\frac{2}{A}}\cdots e^{-\frac{N-1}{A}}\\&=1-\exp\left(-\frac{1}{A}\sum_{k=1}^{N-1}k\right)=1-\exp\left(-\frac{N(N-1)}{2A}\right).\end{aligned}$$

Pour $N=25$, on obtient une minoration proche de 0.56 et, pour $N=23$, proche de 0.500002. Dès 23 personnes, il devient avantageux de parier sur le fait que deux personnes ont leur anniversaire le même jour. Le calcul de la valeur exacte confirmerait cette valeur « critique »de $N=23$. Cela étonne a priori, car 23 est très faible comparé à 365. Cependant, il est plus pertinent de comparer non pas le nombre de personnes, mais le nombre de *paires de personnes*, soit $\frac{N(N-1)}{2}$, avec 365 (on remarque d'ailleurs que cette quantité apparaît dans le calcul). Or, pour $N=23$, il y a déjà 253 paires, ce qui est bien plus comparable à 365.
On pourrait vouloir aussi une majoration. Par exemple,

$$\begin{aligned}\mathbb{P}(2\text{ anniversaires le même jour})&=\mathbb{P}\left(\bigcup_{1\le i\ne j\le N}\{X_i\ne X_j\}\right)\\&\le\sum_{1\le i\ne j\le N}\mathbb{P}(X_i\ne X_j)=\frac{N(N-1)}{2}\cdot\frac{1}{A}.\end{aligned}$$

(Ceci donne 0.82 pour $N=25$ et 0.69 pour $N=23$). Dans la limite $A\to\infty$, elle coïnciderait avec la minoration.

3. Sachant que chaque paquet de céréales contient une vignette à collectionner, que l'on ne découvre qu'à l'ouverture du paquet, combien en moyenne faut-il ouvrir de paquets pour posséder au moins un exemplaire de chacune des n vignettes ? En donner une expression approchée.

On décompose ce nombre en $N_n = \tau_1+\tau_2+\cdots+\tau_n$ où τ_k est le nombre de paquets supplémentaires nécessaires pour obtenir k vignettes différentes quand on en a déjà $k-1$ différentes.
Lorsque l'on dispose de $k-1$ vignettes différentes, pour en avoir k il faut effectuer de nouveaux tirages de vignettes, indépendants, jusqu'à ce qu'une vignette soit différente des $k-1$ modèles déjà possédés, ce qui arrive avec probabilité $\frac{n-(k-1)}{n}$. On en déduit donc (voir exercice 1) que τ_k suit la loi géométrique de paramètre $\frac{n-(k-1)}{n}$. *Il y a en fait une subtilité, en cela que l'on regarde les tirages après un temps aléatoire (la première fois où il y a $k-1$ vignettes différentes), mais ce temps ne dépend que des tirages précédents, donc n'influe pas sur la loi des tirages suivants ; cf. la notion de temps d'arrêt pour les chaînes de Markov.*
Par suite, par linéarité de l'espérance, le temps moyen pour avoir toutes les vignettes est

$$\mathbb{E}[N_n] = \mathbb{E}[\tau_1] + \cdots + \mathbb{E}[\tau_n] = \sum_{k=1}^{n} \frac{n}{n-(k-1)} = n\sum_{j=1}^{n}\frac{1}{j}$$

(en effectuant le changement d'indice $k \mapsto j = n-(k-1)$). En particulier, on a classiquement

$$\mathbb{E}[N_n] \underset{n\to\infty}{\sim} n\ln n.$$

(comparaison série-intégrale)

Exercice 10. *Paradoxes probabilistes pour tous*
1. Hier, je discutais avec ma nouvelle voisine :

MOI – Combien avez-vous d'enfants ?
ELLE – Deux.
MOI – Y a-t-il une fille parmi eux ?
ELLE – Oui.
MOI – Et l'autre enfant, est-ce une fille également ?

1.a) Quelle est la probabilité que ma voisine réponde « oui » ?

Il est important de préciser le modèle. On suppose que les sexes des enfants sont indépendants et uniformes : la probabilité d'avoir un garçon vaut 1/2, de même pour une fille.
En particulier, pour deux enfants, il y a quatre possibilités : FF, FG, GF et GG, où chacune a pour probabilité 1/4 (attention, il y a bien FG et GF).
Sachant que l'un des enfants est une fille, on se restreint au sous-ensemble {FF, FG, GF} à trois éléments dont un seul contient deux filles, donc la probabilité que l'autre enfant soit une fille aussi est $\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{4}+\frac{1}{4}} = \frac{1}{3}$.

1.b) Qu'en est-il si à ma deuxième question j'avais demandé si l'*aîné(e)* était une fille ?

Sachant que le premier enfant est une fille, on se restreint au sous-ensemble {FF, FG} seulement, donc la probabilité que l'autre enfant soit un garçon est $\frac{1}{2}$. On peut aussi dire que, par hypothèse, le sexe du deuxième enfant est indépendant du premier, donc (toujours par hypothèse) le deuxième, quel que soit le premier, est une fille avec probabilité 1/2.

2. *(Problème de Monty Hall)* Un jeu télévisé se déroule à chaque fois de la façon suivante : on présente trois boîtes fermées à un candidat, dont l'une contient 10 000 euros, et seul le présentateur sait laquelle ; le candidat choisit une boîte mais, avant qu'il ait pu l'ouvrir, le présentateur l'interrompt, et ouvre l'une des deux autres boîtes, qu'il sait vide. Le candidat peut alors maintenir son choix, ou ouvrir la boîte restante. L'une de ces options est-elle meilleure que l'autre ?

L'intuition est ici mise à l'épreuve. Voir par exemple l'article wikipedia pour les faux raisonnements habituels.
La difficulté vient du conditionnement induit par l'information nouvelle.
Le plus convaincant est de s'abstraire de ce conditionnement en considérant toutes les situations (sans conditionnement, elles sont équiprobables). Une façon simple de voir les choses est la suivante. Appelons les boîtes 1, 2 et 3, et par symétrie on peut supposer que le candidat choisit la boîte 1. S'il maintient son choix, il gagne lorsque le gain est en 1, ce qui arrive avec probabilité 1/3. S'il change, il gagne lorsque le gain est en 2 ou en 3 (s'il est en 2, le présentateur ouvre 3, et vice-versa), ce qui arrive avec probabilité 2/3. La probabilité de gain est donc doublée en changeant de boîte !

Preuves du cours : caractérisation des lois

Exercice 11. *Fonction de répartition inverse*
Soit X une variable aléatoire réelle. On rappelle que sa fonction de répartition $F_X : t \mapsto F_X(t) = P(X \leq t)$ est croissante, continue à droite, et admet pour limites 0 et 1 en $-\infty$ et $+\infty$.
On définit, pour tout $u \in]0,1[$ l'inverse généralisée continue à gauche de F_X :

$$F_X^{-1}(u) = \inf\{t \mid F_X(t) \geq u\}(\in \mathbb{R}).$$

1. Montrer que, pour tous $t \in \mathbb{R}, u \in]0,1[$, $(F_X^{-1}(u) \leq t) \Leftrightarrow (u \leq F_X(t))$.

> Soit $u \in]0,1[$. Pour tous $s < t \in \mathbb{R}$, si $F_X(s) \geq u$, alors $F_X(t) \geq F_X(s) \geq u$ par croissance de F_X, donc l'ensemble $\{t \,|\, F_X(t) \geq u\}$ est un intervalle de la forme $[\alpha, +\infty[$ ou $]\alpha, +\infty[$, avec par définition $\alpha = F_X^{-1}(t)$. Pour tout $t > F_X^{-1}(u)$, on a donc $F_X(t) \geq u$. Par continuité de F_X à gauche, ceci implique $F_X(F_X^{-1}(u)) \geq u$. Ainsi, $F_X^{-1}(u)$ appartient à l'ensemble précédent, qui est donc
>
> $$\{t \,|\, F_X(t) \geq u\} = [F_X^{-1}(u), +\infty[.$$
>
> C'est ce que l'on voulait démontrer.

2. En déduire que, si U suit la loi uniforme sur $[0,1]$, alors $\widetilde{X} = F_X^{-1}(U)$ suit la même loi que X. Expliciter F_X^{-1} dans les cas d'une loi exponentielle et d'une loi de Cauchy.

> Calculons la fonction de répartition de $\widetilde{X}$: pour tout $t \in \mathbb{R}$,
>
> $$\mathbb{P}(\widetilde{X} \leq t) = \mathbb{P}(F_X^{-1}(U) \leq t) = \mathbb{P}(U \leq F_X(t)) = F_X(t)$$
>
> (en utilisant la question 1, puis le fait que U suit la loi uniforme et que $F_X(t) \in [0,1]$)
> Ceci montre que X et $\widetilde{X}$ ont même fonction de répartition, donc ont même loi.
> Pour la loi $\mathcal{E}(\lambda)$, $F_X(t) = (1 - e^{-\lambda t})\mathbf{1}_{]0,+\infty[}(t)$, et pour tout $u \in]0,1[$, $F_X(t) = u$ équivaut à $t = -\frac{1}{\lambda}\ln(1-u)$. On en déduit que $\widetilde{X} = -\frac{1}{\lambda}\ln(1-U)$ suit la loi $\mathcal{E}(\lambda)$. En fait, $1-U$ a même loi que U, donc $-\frac{1}{\lambda}\ln U$ suit aussi la loi $\mathcal{E}(\lambda)$.
> Si X suit la loi de Cauchy de paramètre a, $F_X(t) = \frac{1}{\pi}\arctan\frac{t}{a} + \frac{1}{2}$, d'où $F_X^{-1}(u) = a\tan\left(\pi\left(u - \frac{1}{2}\right)\right)$. On en déduit que $\widetilde{X} = a\tan\left(\pi\left(U - \frac{1}{2}\right)\right)$ suit la loi de Cauchy de paramètre a.

3. À l'aide de ce qui précède, montrer que toute fonction croissante F continue à droite telle que $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$ est la fonction de répartition d'une loi de probabilité sur $\mathbb{R}$.

> On définit F^{-1} comme ci-dessus à partir de F. On a $F_X^{-1}(u) \in \mathbb{R}$ pour $0 < u < 1$ vu les limites, et on pose $X = F^{-1}(U)$ où U est une variable aléatoire de loi uniforme sur $[0,1]$. Alors, comme F vérifie les mêmes propriétés que F, on peut reprendre la preuve ci-dessus et obtenir que $F_X = F$. On a donc construit une variable aléatoire dont la fonction de répartition est F.

Exercice 12. *Fonction caractéristique* Soit X, Y des variables aléatoires à valeurs dans $\mathbb{R}$. On suppose que leurs fonctions caractéristiques coïncident sur $\mathbb{R}$: pour tout $t \in \mathbb{R}$,

$$E[e^{itX}] = E[e^{itY}].$$

On souhaite conclure que X et Y ont même loi.
Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue à support compact.
1. Soit $A > 0$, assez grand pour que le support de f soit inclus dans $[-A, A]$. Justifier l'existence de $f_A : \mathbb{R} \to \mathbb{R}$ périodique de période $2A$ qui coïncide avec f sur $[-A, A]$. Montrer que

$$\mathbb{E}[f_A(X)] \underset{A\to\infty}{\longrightarrow} \mathbb{E}[f(X)].$$

On a $f_A(X) = f(X) + (f_A(X) - f(X))\mathbf{1}_{\{|X| \geq A\}}$, $f_A - f$ *est bornée et* $P(|X| > A) \to 0$*...*
2. Soit $\varepsilon > 0$. Montrer qu'il existe $K \in \mathbb{N}$, et $\alpha_0, \ldots, \alpha_K \in \mathbb{C}$ tels que, pour tout $x \in \mathbb{R}$,

$$\left| f_A(x) - \sum_{k=0}^{K} \alpha_k e^{\frac{2i\pi}{2A}kx} \right| < \varepsilon$$

(c'est un théorème classique). En déduire que $E[f_A(X)] = E[f_A(Y)]$.
3. Conclure : pour cela, il suffit de montrer que $E[f(X)] = E[f(Y)]$.

Exercice 13. *Transformée de Laplace* Soit X, Y des variables aléatoires à valeurs dans $[0, +\infty[$. On suppose que leurs transformées de Laplace coïncident sur $\mathbb{R}_+$: pour tout $\lambda \geq 0$,

$$E[e^{-\lambda X}] = E[e^{-\lambda Y}].$$

On souhaite en conclure que X et Y ont même loi.
Soit $f : [0, +\infty[\to \mathbb{R}$ une fonction continue à support compact.
1. Soit $\varepsilon > 0$. Montrer qu'il existe $K \in \mathbb{N}$ et $\alpha_0, \ldots, \alpha_K \in \mathbb{R}$, $\lambda_0, \ldots, \lambda_K \in \mathbb{R}_+$ tels que

$$\text{pour tout } x \in \mathbb{R}_+, \qquad \left| f(x) - \sum_{k=0}^{K} \alpha_k e^{-\lambda_k x} \right| < \varepsilon.$$

Considérer $g(u) = f(-\ln u)$ sur $[0, 1]$ et appliquer le théorème de Stone-Weierstrass. On a en fait $\lambda_k = k$
2. En déduire que $E[f(X)] = E[f(Y)]$, et conclure.
Une autre preuve consisterait à se ramener à l'égalité entre fonctions caractéristiques : la fonction $\mathcal{L}_X : z \mapsto E[e^{-zX}]$ est holomorphe dans l'ouvert $\{\mathrm{Re}(\cdot) > 0\}$ et continue sur son adhérence (par théorèmes de régularité sous l'intégrale) ; par unicité du prolongement analytique, l'égalité entre $\mathcal{L}_X$ et $\mathcal{L}_Y$ sur $\mathbb{R}_+$ s'étend à l'ouvert précédent, et par continuité elle s'étend à son bord $i\mathbb{R}$, soit $\Phi_X = \Phi_Y$.

Convergence de variables aléatoires, exemples

Exercice 14. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, de loi $\mathcal{E}(\lambda)$.
1. Montrer la convergence en probabilité suivante :

$$\frac{1}{\ln n} \max_{1\leq k\leq n} X_k \xrightarrow[n]{(p)} \frac{1}{\lambda}.$$

Soit $\delta > 0$. Il s'agit de montrer

$$\mathbb{P}\left(\left| \frac{1}{\ln n} \max_{1\leq k\leq n} X_k - \frac{1}{\lambda} \right| > \delta \right) \xrightarrow[n]{} 0,$$

ce qui équivaut à (en découpant l'événement en deux)

$$\mathbb{P}\left(\frac{1}{\ln n} \max_{1\leq k\leq n} X_k < \frac{1}{\lambda} - \delta \right) \xrightarrow[n]{} 0 \qquad \text{et} \qquad \mathbb{P}\left(\frac{1}{\ln n} \max_{1\leq k\leq n} X_k > \frac{1}{\lambda} + \delta \right) \xrightarrow[n]{} 0.$$

La première probabilité se réécrit

$$\mathbb{P}\left(\text{pour } k = 1, \ldots, n, \frac{1}{\ln n} X_k < \frac{1}{\lambda} - \delta \right) = \mathbb{P}\left(X_1 < \left(\frac{1}{\lambda} - \delta \right) \ln n \right)^n = \left(1 - e^{-\lambda\left(\frac{1}{\lambda} - \delta\right)\ln n} \right)^n.$$

(La dernière égalité suppose $\delta < \frac{1}{\lambda}$, sans quoi la probabilité est nulle et la convergence est démontrée). On conclut facilement.
La deuxième probabilité se réécrit

$$\begin{aligned} \mathbb{P}\left(\text{il existe } k \in \{1, \ldots, n\} \text{ tel que } \frac{1}{\ln n} X_k > \frac{1}{\lambda} + \delta \right) &\leq n\mathbb{P}\left(X_1 > \left(\frac{1}{\lambda} + \delta \right) \ln n \right) \\ &= n e^{-\lambda\left(\frac{1}{\lambda} + \delta\right)\ln n} = n^{-\lambda\delta} \xrightarrow[n]{} 0. \end{aligned}$$

(Ici, cette majoration par sous-additivité de $\mathbb{P}$ suffit ; autrement, il aurait fallu passer au complémentaire pour utiliser l'indépendance : $\mathbb{P}(\exists k, X_k > \cdots) = 1 - \mathbb{P}(\forall k, X_k < \cdots) = 1 - \mathbb{P}(X_1 < \cdots)^n$, etc.)

2. Démontrer que la suite de terme général

$$\max_{1\leq k\leq n} X_k - \frac{\ln n}{\lambda}$$

converge en loi vers une limite à déterminer.

On utilise la caractérisation de la convergence en loi à l'aide des fonctions de répartition : montrons que $t \mapsto \mathbb{P}\left(\max_{1\le k\le n} X_k - \frac{\ln n}{\lambda} \le t\right)$ admet une limite qui est une fonction de répartition. Pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned}\mathbb{P}\left(\max_{1\le k\le n} X_k - \frac{\ln n}{\lambda} \le t\right) &= \mathbb{P}\left(\text{pour } k=1,\dots,n,\ X_k \le t + \frac{\ln n}{\lambda}\right)\\ &= \mathbb{P}\left(X_1 \le t + \frac{\ln n}{\lambda}\right)^n = \left(1 - e^{-\lambda\left(t+\frac{\ln n}{\lambda}\right)}\right)^n = \left(1 - \frac{e^{-t}}{n}\right)^n\\ &\xrightarrow[n]{} e^{-e^{-t}}.\end{aligned}$$

(La 3[e] égalité est vraie si $\ln n > -\lambda t$, ce qui est vrai pour n grand) La fonction $t \mapsto e^{-e^{-t}}$ est continue sur $\mathbb{R}$, et admet pour limites 0 et 1 en $-\infty$ et $+\infty$, c'est donc la fonction de répartition d'une loi de probabilité (appelée loi de Gumbel). On en déduit

$$\max_{1\le k\le n} X_k - \frac{\ln n}{\lambda} \xrightarrow[n]{\text{(loi)}} Z$$

où Z suit la loi de Gumbel : pour tout $t \in \mathbb{R}$, $\mathbb{P}(Z \le t) = e^{-e^{-t}}$.

Exercice 15. Soit $(X_n)_n$ une suite de variables aléatoires. On suppose qu'il existe une suite $(c_n)_n$ de réels telle que, pour tout $\delta > 0$.

$$P(|X_n - c_n| > \delta) \xrightarrow[n]{} 0.$$

On suppose de plus que la suite $(c_n)_n$ admet une limite c. Montrer que X_n converge vers c en probabilité.

Soit $\delta > 0$. Il existe n_0 tel que, pour $n \ge n_0$, $|c_n - c| < \frac{\delta}{2}$. Alors, pour $n \ge n_0$, $|X_n - c| > \delta$ implique $|X_n - c_n| \ge |X_n - c| - |c - c_n| > \delta - \frac{\delta}{2} = \frac{\delta}{2}$, de sorte que

$$P\left(|X_n - c| > \delta\right) \le P\left(|X_n - c_n| > \frac{\delta}{2}\right),$$

et le terme de droite converge vers 0 par l'hypothèse. Ceci montre que $(X_n)_n$ converge vers c en probabilité.

Exercice 16. On reprend le problème du collectionneur de vignettes (exercice 9) : Soit $n \ge 1$. Soit $(X_k)_{k\ge1}$ une suite de v.a. i.i.d. de loi uniforme sur $\{1,\dots,n\}$. Pour $1 \le k \le n$, on définit

$$\tau_k^n = \inf\left\{m \ge 1 \,\middle|\, \mathrm{Card}(\{X_1,\dots,X_m\}) = k\right\}.$$

1. Montrer que les variables aléatoires $(\tau_k^n - \tau_{k-1}^n)_{2\le k\le n}$, sont **indépendantes**, de lois respectives $\mathcal{G}\left(\frac{n-k+1}{n}\right)$ pour $2 \le k \le n$. *(Cette question peut être difficile à rédiger ; se convaincre intuitivement du résultat d'abord)*

Une preuve naturelle utiliserait la propriété de Markov forte ; preuve élémentaire à rédiger... **

2. En déduire l'espérance et la variance de $N_n = \tau_n^n$.

On a donc

$$E[N_n] = E[\tau_1^n + (\tau_2^n - \tau_1^n) + \cdots + (\tau_n^n - \tau_{n-1}^n)] = 1 + \sum_{k=2}^{n} \frac{n}{n-k+1} = n\sum_{j=1}^{n}\frac{1}{j}$$

(en intégrant le 1 à la somme et en inversant l'ordre de sommation). En particulier, $E[N_n] \sim n\ln n$ quand $n \to \infty$.
Et, comme les variables de la question 1 sont indépendantes, (et $\mathrm{Var}(\tau_1^n) = \mathrm{Var}(1) = 0$)

$$\begin{aligned}\mathrm{Var}(N_n) = \mathrm{Var}(\tau_2^n - \tau_1^n) + \cdots + \mathrm{Var}(\tau_n^n - \tau_{n-1}^n) &= \sum_{k=2}^{n} \frac{1 - \frac{n-k+1}{n}}{\left(\frac{n-k+1}{n}\right)^2}\\ &= \sum_{j=1}^{n-1} \frac{1-\frac{j}{n}}{\left(\frac{j}{n}\right)^2} = n^2\sum_{j=1}^{n-1}\frac{1}{j^2} + n\sum_{j=1}^{n-1}\frac{1}{j}.\end{aligned}$$

3. Donner un équivalent de l'espérance, et montrer que $\mathrm{Var}(N_n) = O_n(n^2)$.

On a $E[N_n] \sim_n n \ln n$. Et on rappelle que

$$\mathrm{Var}(N_n) = n^2 \sum_{j=1}^{n-1} \frac{1}{j^2} + n \sum_{j=1}^{n-1} \frac{1}{j}.$$

Comme $C = \sum_{j=1}^{\infty} \frac{1}{j^2}$ converge, le premier terme est équivalent à Cn^2, tandis que le second est équivalent à $n \ln n = o(n^2)$. Par suite le premier domine : $\mathrm{Var}(N_n) \sim_n Cn^2$.

4. En déduire, pour tout $\varepsilon > 0$,

$$P\Big(\big|N_n - E[N_n]\big| > \varepsilon n \log n\Big) \underset{n}{\longrightarrow} 0,$$

puis

$$\frac{N_n}{n \log n} \xrightarrow[n]{(p)} 1.$$

On applique l'inégalité de Tchebychev :

$$P\Big(\big|N_n - E[N_n]\big| > \varepsilon n \ln n\Big) \le \frac{\mathrm{Var}(N_n)}{\varepsilon^2 n^2 \ln^2 n} \sim \frac{Cn^2}{\varepsilon^2 n^2 \ln^2 n} = \frac{C}{\varepsilon^2 \ln^2 n} \underset{n}{\longrightarrow} 0.$$

Ceci se réécrit

$$P\Big(\big|\frac{N_n}{n \ln n} - \frac{E[N_n]}{n \ln n}\big| > \varepsilon\Big) \underset{n}{\longrightarrow} 0.$$

Par ailleurs, on a vu que $\frac{N_n}{n \ln n} \to 1$. En application de l'exercice 15, on conclut que

$$\frac{N_n}{n \ln n} \xrightarrow[n]{(p)} 1.$$

Modes de convergence

Exercice 17. *Convergence des images*
Soit $(X_n)_n$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$, et X une autre variable aléatoire à valeurs dans $\mathbb{R}^d$. Soit φ une fonction continue de $\mathbb{R}^d$ dans $\mathbb{R}^m$. Montrer que $X_n \underset{n}{\longrightarrow} X$ implique $\varphi(X_n) \underset{n}{\longrightarrow} \varphi(X)$ pour les convergences p.s., en probabilité et en loi.

Pour la convergence p.s., c'est immédiat : si, pour $\omega \in \Omega$, la suite (de réels) $(X_n(\omega))_n$ converge vers $X(\omega)$, alors par continuité $\varphi(X_n(\omega)))_n$ converge vers $\varphi(X(\omega))$. Donc si la première convergence a lieu avec probabilité 1, la deuxième aussi.
Supposons que $X_n \xrightarrow[n]{(\mathrm{loi})} X$. Pour toute fonction $f : \mathbb{R}^m \to \mathbb{R}$ continue bornée, $f \circ \varphi : \mathbb{R}^d \to \mathbb{R}$ est continue bornée donc $E[f(\varphi(X_n))] \to_n E[f(\varphi(X))]$, ce qui montre que $\varphi(X_n) \xrightarrow[n]{(\mathrm{loi})} \varphi(X)$.
Supposons que $X_n \xrightarrow[n]{(p)} X$. On souhaite contrôler la probabilité que $\varphi(X_n)$ soit loin de $\varphi(X)$, en sachant que X_n est proche de X avec grande probabilité. Comme X est variable, ceci suggère l'utilisation d'une continuité *uniforme.* A priori, φ n'est pas uniformément continue. Par contre, avec grande probabilité X est inférieure à une constante donnée assez grande, ce qui permet de se ramener à un compact et d'utiliser le théorème de Heine.
Soit $\delta > 0$. Soit $\varepsilon > 0$. Il existe $A > 0$ tel que $\mathbb{P}(|X| > A) < \varepsilon$ (en effet, $\lim_{n\downarrow} \mathbb{P}(|X| > n) = \mathbb{P}(\bigcap \{|X| > n\}) = 0$). On a

$$\begin{aligned} P(|\varphi(X_n) - \varphi(X)| > \delta) &\le P(|\varphi(X_n) - \varphi(X)| > \delta, |X| > A) + P(|\varphi(X_n) - \varphi(X)| > \delta, |X| \le A) \\ &\le P(|X| > A) + P(|\varphi(X_n) - \varphi(X)| > \delta, |X| \le A) \end{aligned}$$

Par le théorème de Heine, φ est uniformément continue sur $B(0, 2A)$ donc il existe $\alpha > 0$ tel que, si $|x| < 2A$, $|y| < 2A$ et $|x - y| < \alpha$, alors $|\varphi(x) - \varphi(y)| < \delta$. Notamment, si $|\varphi(X_n) - \varphi(X)| > \delta$ et $|X| \le A$, alors $|X_n| \ge 2A$ (donc $|X_n - X| \ge 2A - A = A$) ou $|X_n - X| \ge \alpha$, ce qui donne

$$P(|\varphi(X_n) - \varphi(X)| > \delta, |X| \le A) \le P(|X_n - X| \ge A) + P(|X_n - X| \ge \alpha).$$

Vu l'hypothèse, chaque terme de droite converge vers 0 quand $n \to \infty$, donc est inférieur à ε pour n grand. Finalement, pour n assez grand, on a donc

$$P(|\varphi(X_n) - \varphi(X)| > \delta) \leq 3\varepsilon.$$

Ceci montre que $\varphi(X_n) \underset{n}{\longrightarrow} \varphi(X)$.
Une autre preuve aurait consisté à utiliser le fait que la convergence en probabilité de $(X_n)_n$ vers X est équivalente à ce que toute sous-suite de $(X_n)_n$ admette une sous-(sous-)suite qui converge presque sûrement vers X, ce qui ramène la question à celle de la convergence presque sûre, qui est immédiate.

Exercice 18.
1. Donner un exemple de suite $(X_n)_{n\geq 0}$ qui converge en loi mais pas en probabilité. *Le construire par exemple à l'aide de U de loi uniforme sur $[0,1]$ et de $1-U$, qui a même loi.*

Supposons donnée une variable aléatoire U de loi uniforme sur $[0,1]$. On pose alors $X_n = U$ si n est pair et $X_n = 1-U$ si n est impair. Pour tout n, X_n a la même loi, donc converge en loi (vers la loi uniforme sur $[0,1]$). En revanche, si on avait $X_n \underset{n}{\overset{(p)}{\longrightarrow}} X$, alors $P(|X_n - X| > \delta)$ devrait converger vers 0, or cette suite ne prend que deux valeurs, selon la parité de n, et ces valeurs ne peuvent être égales à 0 vu que $P(|X_n - X| > \delta) + P(|X_{n+1} - X| > \delta) \geq P(|X_n - X_{n+1}| > 2\delta) = P(2U - 1 > 2\delta) > 0$ dès que $\delta \in]0,1[$ (si $|X_n - X_{n+1}| > 2\delta$, alors vu que $|X_n - X| + |X_{n+1} - X| \geq |X_n - X_{n+1}|$, l'un des deux premiers termes doit être supérieur à δ). On pourrait aussi dire que les sous-suites des termes pairs et impairs devraient avoir des sous-suites qui convergent p.s. vers X, d'où $X = U = 1-U$, ce qui a une probabilité nulle.

2. Donner un exemple de suite $(X_n)_{n\geq 0}$ qui converge en probabilité mais pas p.s. *Prendre X_n indépendantes, de loi $\mathcal{B}(p_n)$ où p_n tend vers 0 mais $\sum_n p_n = \infty$, et appliquer le lemme de Borel-Cantelli. Ou : Soit U de loi uniforme sur $[0,1]$. Prendre X_n égal à 1 ssi $U \in [\frac{p}{2^k}, \frac{p+1}{2^k}[$ où $n = 2^k + p$ et $0 \leq p < 2^k$.*
3. Pour $p > 0$, donner un exemple de suite $(X_n)_{n\geq 0}$ qui converge en probabilité mais pas dans L^p *Prendre X_n à valeurs dans $\{0, n\}$ et ajuster les probabilités respectives de 0 et n.*

Si la loi de X_n est définie par $P(X_n = n) = \frac{1}{n^\alpha}$ et $P(X_n = 0) = 1 - \frac{1}{n^\alpha}$, où $\alpha > 0$, on note que $X_n \underset{n}{\overset{(p)}{\longrightarrow}} 0$ car $P(|X_n| > \delta) = \frac{1}{n^\alpha}$ pour n assez grand ; et $E[|X_n|^p] = n^{p-\alpha}$ donc $(X_n)_n$ ne converge vers 0 dans L^p que si $p < \alpha$.

Exercice 19. *Lemme de Slutzky*
Montrer le lemme de Slutzkt : Soit $(X_n)_n, (Y_n)_n$ deus suites de variables aléatoires telles que $(X_n)_n$ converge en loi vers X et Y_n converge en loi vers une variable aléatoire égale p.s. à une constante c. Alors $(X_n, Y_n)_n$ converge en loi vers (X, c).

On rappelle la définition de la fonction caractéristique $\Phi_{(X,Y)}(s,t) = E[e^{i(sX+tY)}]$ pour $s, t \in \mathbb{R}$. On découpe, pour tous $s, t \in \mathbb{R}$,

$$\begin{aligned}|\Phi_{(X_n,Y_n)}(s,t) - \Phi_{(X,Y)}(s,t)| &\leq |\Phi_{(X_n,Y_n)}(s,t) - \Phi_{(X_n,Y)}(s,t)| + |\Phi_{(X_n,Y)}(s,t) - \Phi_{(X,Y)}(s,t)| \\ &\leq E[|e^{itY_n} - e^{itc}|] + |E[e^{isX_n}] - E[e^{isX}]|\end{aligned}$$

en appliquant l'inégalité triangulaire (intégrale) au premier terme seulement (puis en factorisant par e^{isX_n} et en utilisant $|e^{isX_n}| = 1$), et en mettant en facteur la constante e^{itc} dans le second terme (et en utilisant $|e^{itc}| = 1$). Les deux termes convergent vers 0, car $X_n \overset{\text{(loi)}}{\longrightarrow} X$ et $Y_n \overset{\text{(loi)}}{\longrightarrow} c$ et les fonctions $x \mapsto e^{isx}$ et $y \mapsto |e^{ity} - e^{itc}|$ sont continues bornées : $E[e^{isX_n}] \to E[e^{isX}]$ et $E[|e^{itYn} - e^{itc}|] \to E[|e^{itc} - e^{itc}|] = 0$.

Théorèmes-limites

Exercice 20. *Loi des grands nombres pour des variables non-intégrables*
Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. de même loi que X, où $\mathbb{E}[X_+] = \infty$ et $\mathbb{E}[X_-] < \infty$, de sorte que cela a un sens d'écrire $\mathbb{E}[X] = +\infty$. On pose $S_n = X_1 + \cdots + X_n$. Montrer que, p.s., $\frac{S_n}{n} \underset{n}{\longrightarrow} +\infty$.

Comme on peut appliquer la loi forte des grands nombres à la suite des parties négatives $((X_n)_-)_n$, en découpant $X_n = (X_n)_+ - (X_n)_-$ on constate qu'il suffit de montrer $\frac{1}{n}((X_1)_+ + \cdots + (X_n)_+) \to +\infty$ p.s.. Autrement dit, on peut supposer $X_n \geq 0$ pour tout n à partir de maintenant.

Soit $A > 0$. Comme la suite $(\min(X,n))_n$ est positive, croissante et converge (en fait stationne) vers X, le théorème de convergence monotone montre que $\mathbb{E}[\min(X,n)] \to \mathbb{E}[X] = +\infty$. Il existe donc $m \in \mathbb{N}$ tel que $\mathbb{E}[\min(X,m)] \geq A$. Or on peut appliquer la loi forte des grands nombres à la suite de terme général $Y_n = \min(X_n, m)$: presque sûrement,

$$\frac{\min(X_1,m) + \cdots + \min(X_n,m)}{n} \underset{n}{\longrightarrow} \mathbb{E}[\min(X,m)].$$

Et, pour tout n,

$$\frac{X_1 + \cdots + X_n}{n} \geq \frac{\min(X_1,m) + \cdots + \min(X_n,m)}{n}$$

d'où, presque sûrement,

$$\liminf_n \frac{X_1 + \cdots + X_n}{n} \geq \mathbb{E}[\min(X,m)] \geq A.$$

Ceci vaut pour tout $A > 0$, en particulier tout $A \in \mathbb{N}$. En « intervertissant p.s. et $\forall A \in \mathbb{N}$ »(opération justifiée car $\mathbb{N}$ est dénombrable), c'est-à-dire en utilisant le fait qu'une intersection dénombrable d'événements presque sûrs est presque sûre, on obtient que presque sûrement la limite inférieure ci-dessus est supérieure à tout $A \in \mathbb{N}$ et est donc égale à $+\infty$. Autrement dit, presque sûrement, $\frac{X_1+\cdots+X_n}{n} \underset{n}{\longrightarrow} +\infty$.

Exercice 21. *Calculs de limites à l'aide des théorèmes-limites*
1. Appliquer le théorème central limite à une suite $(X_n)_{n\geq 1}$ de variables aléatoires indépendantes de loi de Poisson de paramètre 1 pour trouver la limite de la suite de terme général

$$e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, de loi $\mathcal{P}(1)$. Cette loi est de carré intégrable, de moyenne 1 et de variance 1, donc le théorème central limite donne, en particulier,

$$P\left(\frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \leq 0\right) \underset{n}{\longrightarrow} P(Z \leq 0) = \frac{1}{2}.$$

Vu que $X_1 + \cdots + X_n$ suit la loi $\mathcal{P}(n)$, la probabilité ci-dessus est aussi

$$P(X_1 + \cdots + X_n \leq n) = \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!},$$

d'où la limite de l'énoncé.

2. Soit f une fonction continue $[0,1] \to \mathbb{R}$. Déterminer la limite de la suite de terme général

$$\int_0^1 \cdots \int_0^1 f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n.$$

Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, de loi uniforme sur $[0,1]$. L'intégrale multiple de l'énoncé est $\mathbb{E}[f(\frac{U_1+\cdots+U_n}{n})]$. Par la loi faible des grands nombres, $\frac{U_1+\cdots+U_n}{n} \underset{n}{\overset{(loi)}{\longrightarrow}} \mathbb{E}[U_1] = \frac{1}{2}$ donc comme f est continue sur $[0,1]$ donc bornée aussi, $\mathbb{E}[f(\frac{U_1+\cdots+U_n}{n})] \to \mathbb{E}[f(\frac{1}{2})] = f(\frac{1}{2})$.
On pourrait évidemment partir de la loi forte (convergence p.s.) : $f(\frac{U_1+\cdots+U_n}{n}) \to f(\frac{1}{2})$ presque sûrement (LFGN et continuité de f), et cette suite est majorée par $\|f\|_\infty$, qui est intégrable sur $(\Omega, \mathbb{P})$, donc le théorème de convergence dominée montre que $\mathbb{E}[f(\frac{U_1+\cdots+U_n}{n})] \to \mathbb{E}[f(\frac{1}{2})] = f(\frac{1}{2})$.
En tout cas, dès que f est bornée et continue en $1/2$,

$$\int_0^1 \cdots \int_0^1 f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n \underset{n}{\longrightarrow} f\left(\frac{1}{2}\right).$$

Exercice 22. *Polynômes de Bernstein*
On veut donner une preuve probabiliste (et constructive) du théorème de (Stone-)Weierstrass sur $[0,1]$. Soit f

une fonction continue de $[0,1]$ dans $\mathbb{R}$. On définit, pour $n \geq 0$, la fonction polynomiale sur $[0,1]$

$$P_n : x \mapsto P_n(x) = \sum_{k=0}^{n} f\left(\tfrac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k}.$$

1. En écrivant $P_n(x) = E\big[f(\frac{S_n}{n})\big]$, où S_n suit la loi binomiale de paramètres (n,x), justifier la convergence simple de P_n vers f sur $[0,1]$. *(cf. Exercice 1)*

> Soit $x \in [0,1]$. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, de loi $\mathcal{B}(x)$. Pour tout n, la variable aléatoire $S_n = X_1 + \cdots + X_n$ a pour loi $\mathcal{B}(n,x)$ (par l'exercice 1), et en particulier $E[f(\frac{X_1+\cdots+X_n}{n})] = P_n(x)$ est le polynôme de l'énoncé en la variable x. Or, par la loi faible des grands nombres, $\frac{X_1+\cdots+X_n}{n}$ converge vers $E[X_1] = x$ en loi, ce qui implique, f étant continue bornée, que $E[f(\frac{X_1+\cdots+X_n}{n})] \to E[f(x)] = f(x)$ (voir aussi exercice précédent, question 2, pour utiliser la loi forte). Par suite, $E[f(\frac{S_n}{n})] \to f(x)$. Ceci montre la convergence simple de $(P_n)_n$ vers f en x. Et donc sur $[0,1]$.

2. Soit $\delta > 0$. On note $\omega(f,\delta) = \sup\big\{|f(x)-f(y)| \,\big|\, |x-y| < \delta\big\}$ le module de continuité de f. Justifier l'inégalité

$$|f(x) - P_n(x)| \leq \omega(f,\delta) + 2\|f\|_\infty P\left(\left|\frac{S_n}{n} - x\right| > \delta\right),$$

et en déduire une majoration de $\|f - P_n\|_\infty$. Conclure.

> Soit $\delta > 0$. On a
>
> $$\begin{aligned}|f(x) - P_n(x)| &= \left|f(x) - E\left[f\left(\frac{S_n}{n}\right)\right]\right| = \left|E[f(x)] - E\left[f\left(\frac{S_n}{n}\right)\right]\right| \leq E\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \\ &\leq E\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|,\ \left|x - \frac{S_n}{n}\right| \leq \delta\right] + E\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|,\ \left|x - \frac{S_n}{n}\right| > \delta\right] \\ &\leq \omega(f,\delta)\mathbb{P}\left(\left|x - \frac{S_n}{n}\right| \leq \delta\right) + 2\|f\|_\infty P\left(\left|\frac{S_n}{n} - x\right| > \delta\right) \\ &\leq \omega(f,\delta) + 2\|f\|_\infty P\left(\left|\frac{S_n}{n} - x\right| > \delta\right),\end{aligned}$$
>
> et on peut utiliser l'inégalité de Tchebychev pour majorer la dernière probabilité : comme $E[\frac{S_n}{n}] = \frac{nE[X_1]}{n} = x$,
>
> $$P\left(\left|\frac{S_n}{n} - x\right| > \delta\right) \leq \frac{\operatorname{Var}(\frac{S_n}{n})}{\delta^2} = \frac{\operatorname{Var}(S_n)}{n\delta^2} = \frac{n\operatorname{Var}(X_1)}{n\delta^2} = \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2},$$
>
> d'où une majoration uniforme :
>
> $$\|f - P_n\|_\infty \leq \omega(f,\delta) + \frac{\|f\|_\infty}{2n\delta^2}.$$
>
> Ceci vaut pour tout $\delta > 0$. Or, par continuité (uniforme) de f, $\omega(f,\delta) \to 0$ quand $\delta \to 0^+$. Soit $\varepsilon > 0$. Il existe donc δ tel que le premier terme est inférieur à $\frac{\varepsilon}{2}$. Puis, pour $n \geq \frac{\|f\|_\infty}{\delta^2}$, le second est inférieur à $\frac{\varepsilon}{2}$ donc $\|f - P_n\|_\infty \leq \varepsilon$. Ceci montre la convergence uniforme de P_n vers f sur $[0,1]$.