0 Événements, dénombrement, conditionnement

Rappels de cours. Modélisation et codage par univers, exemples, événements, formulation ensembliste, tribus, probabilité, espaces finis et atomes, équiprobabilité et cardinal, dénombrement. Limite supérieure/inférieure. Conditionnement et indépendance, formule de Bayes.

Exercice 0.1 (Urne, boules, tirages). *On considère une urne contenant n boules numérotées de 1 à n (donc distinguables) et on effectue le tirage de r boules dans l'urne.*

1. *Montrer que le nombre de tirages sans remise et ordonnés est donné par $A_{n,r} = \frac{n!}{(n-r)!}$. Application : nombre de tiercés avec n chevaux au départ ($r = 3$)*
2. *Montrer que le nombre de tirages sans remise et non ordonnés est donné par $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. Application : nombre de binômes possibles dans une classe de n élèves ($r = 2$)*
3. *Montrer que le nombre de tirages avec remise et ordonnés est donné par n^r. Application : nombre de mots de r lettres ($n = 26$)*
4. *Montrer que le nombre de tirages avec remise et non ordonnés est donné par $\binom{n+r-1}{n-1}$. Application : nombre de «mains» de r lettres dans un Scrabble sans contraintes ($n = 26$)*

Solution.

1. Il s'agit des arrangements : $A_{n,r} = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$, c'est-à-dire du nombre de r-uplets $(b_1,\ldots,b_r)$ constitués d'éléments $b_1,\ldots,b_r$ de $\{1,\ldots,n\}$ deux à deux distincts. Pour $r = n$ on trouve le nombre de manières de permuter $\{1,\ldots,n\}$, soit $A_{n,n} = n!$.
2. Il s'agit du nombre de sous-ensembles de cardinal r de $\{1,\ldots,n\}$, autrement dit du nombre d'injections de $\{1,\ldots,r\}$ dans $\{1,\ldots,n\}$.
3. Il s'agit du nombre de r-uplets $(b_1,\ldots,b_r)$ constitués d'éléments $b_1,\ldots,b_r$ de $\{1,\ldots,n\}^r$, i.e. le nombre d'applications de $\{1,\ldots,r\}$ dans $\{1,\ldots,n\}$. Bien entendu, $n^r \geq A_{n,r} \geq \binom{n}{r}$.
4. il y en a $\binom{n+r-1}{n-1}$. Il s'agit également du nombre de manières de placer r boules indistinguables dans n urnes distinguables, ou encore le nombre d'applications $f : \{1,\ldots,n\} \to \{0,\ldots,r\}$ vérifiant $f(1) + \cdots + f(n) = r$, c'est-à-dire
$$\mathrm{card}\{(r_1,\ldots,r_n) \in \{0,\ldots,r\}^n : r_1 + \cdots + r_n = r\}.$$
Il s'agit également du nombre de termes dans la formule du multinome de taille r avec n types (ce qui n'est pas égal au coefficient multinomial ! Attention également à l'ordre de n et r ici !). Pour obtenir la formule, on aligne les r boules entre deux cloisons fixes, et on place ensuite $n-1$ cloisons supplémentaires pour délimiter les n urnes (tiroirs !). Il y a $r+1$ possibilités pour la première cloison, $r+2$ pour la seconde (car certaines urnes peuvent être vides !), . . . , $r+n-1$ possibilités pour la $n-1$ ième cloison. L'ordre de placement des $n-1$ cloisons n'étant pas pertinent, on obtient $(r+1)\cdots(r+n-1)/(n-1)! = \binom{n+r-1}{n-1}$.
Pour mémoire, la formule du multinome de taille r avec n types est
$$(a_1 + \cdots + a_n)^r = \sum_{\substack{(r_1,\ldots,r_n)\in\{0,\ldots,r\}^n \\ r_1+\cdots+r_n=r}} \frac{r!}{r_1!\cdots r_n!} a_1^{r_1}\cdots a_n^{r_n}$$
et le nombre de termes vaut
$$\sum_{\substack{(r_1,\ldots,r_n)\in\{0,\ldots,r\}^n \\ r_1+\cdots+r_n=r}} 1 = \mathrm{card}\{(r_1,\ldots,r_n) \in \{0,\ldots,r\}^n : r_1 + \cdots + r_n = r\}$$
$$= \binom{n+r-1}{n-1} = \binom{n+r-1}{r}.$$

Note : sans remise signifie non ordonnés. La combinatoire est l'art de la bijection pour compter.

Exercice 0.2 (Modélisation de tirage avec remise et sans remise). *On dispose d'une urne contenant n boules numérotées de 1 à n (donc distinguables). Proposer un univers équiprobable codant le tirage de r boules avec remise, et un univers codant le tirage de r boules sans remise.*

Solution. On modélise le tirage de r boules avec remise par la probabilité uniforme sur l'univers $\Omega = \{1,\dots,n\}^r$, dont le cardinal vaut n^r. En conséquence, lors d'un tirage avec remise de deux cartes dans un jeu de 32 cartes, la probabilité que les cartes soient de la même couleur (par *couleur* on entend ici de manière non standard *rouge* ou *noir* et pas *pique*, *carreau*, *trèfle*, *cœur*) vaut $(2\times 16^2)/32^2 = 1/2$, tandis que la probabilité d'obtenir 2 as vaut $4^2/32^2 = 1/64$.

On modélise le tirage de r boules sans remise par la probabilité uniforme sur l'univers $\Omega = \{T \subset \{1,\dots,n\} \ : \mathrm{card}(T) = r\}$, de cardinal $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ (notons que ce modèle d'équiprobabilité n'est pas celui d'un espace produit). En conséquence, lors d'un tirage sans remise de deux cartes dans un jeu de 32 cartes, la probabilité de tirer deux cartes de même couleur vaut $2\binom{16}{2}/\binom{32}{2} = 15/31$, tandis que la probabilité d'obtenir 2 as vaut $\binom{4}{2}/\binom{32}{2} = 3/(32\times 31)$. Comme on peut s'y attendre, ces deux probabilités sont plus petites que celles avec remise.

Exercice 0.3 (Jeu de pile ou face). *Qu'elle est la probabilité d'obtenir k fois pile en n parties de pile ou face avec une pièce de monnaie équilibrée ? On précisera l'univers de probabilité utilisé.*

Solution. L'univers est $\Omega = \{(a_1,\dots,a_n) \in \{0,1\}^n\}$ de cardinal 2^n, où 0 code face et 1 code pile. L'évènement d'intérêt est $A = \{(a_1,\dots,a_n) : a_1 + \cdots + a_n = k\}$, de cardinal $\binom{n}{k}$. La pièce étant équilibrée, on choisit le modèle d'équiprobabilité, et donc $\mathbb{P}(A) = \mathrm{card}(A)/\mathrm{card}(\Omega) = \binom{n}{k}2^{-n}$. Nous verrons qu'il s'agit d'un cas particulier de la loi binomiale (taille n et paramètre $1/2$).

Exercice 0.4 (Problème de modélisation). *Proposer un univers codant le résultat du jet simultané de deux dés indistinguables. Pour tous $1 \le i,j \le 6$, calculer la probabilité d'obtenir $\{i,j\}$.*

Solution. La modélisation par la probabilité uniforme sur $\{\{i,j\} : 1 \le i,j \le 6\}$ n'est pas bonne (elle donne trop de poids à l'égalité des dés). Le bon modèle est la probabilité uniforme sur $\{(i,j) : 1 \le i,j \le 6\}$, qui force à distinguer les dés, et donc à reformuler les évènements où ils ne sont pas distingués. Par exemple : pour tous $1 \le i,j \le 6$, l'évènement $\{i,j\} = \{(i,j),(j,i)\}$ a probabilité $1/21$ dans le mauvais modèle, tandis que dans le bon modèle, il a probabilité $2/36 = 1/18$ si $i \ne j$ et $1/36$ si $i = j$.

Exercice 0.5 (Anniversaires). *Si n personnes sont réunies dans une pièce, quelle est la probabilité qu'aucune paire d'individus n'ait un anniversaire commun ? Déterminer approximativement $n_{\min}$ pour que cette probabilité soit inférieure à $1/2$.*

Solution. On adopte le modèle d'équiprobabilité sur $\Omega = \{1,\dots,d\}^n$ où $d = 365$. Dans ce modèle, les individus sont numérotés. On néglige les années bisextiles et la non uniformité des dates de naissance (tous les modèles sont faux, mais certains sont utiles). La probabilité recherchée vaut[1] 1 si $n \ge d$ tandis que si $n < d$, elle vaut (et l'approximation est bonne si $n \ll d$)

$$p_n = \frac{A_{d,n}}{d^n} = \frac{d(d-1)\cdots(d-n+1)}{d^n} = \prod_{i=1}^{n-1}\left(1-\frac{i}{d}\right) = e^{\sum_{i=1}^{n-1}\log(1-i/d)} \approx e^{-\frac{1}{d}\sum_{i=1}^{n-1} i} \approx e^{-\frac{n^2}{2d}}$$

Ainsi, $p_{20} \approx 0,41$, $p_{30} \approx 0,71$, $p_{40} \approx 0,89$, et $p_{50} \approx 0,97$. Il y a un phénomène de seuil, et la courbe $n \mapsto p_n$ est décroissante d'allule sigmoïdale ressemblant au membre droit d'une courbe en cloche gaussienne. Le $n_{\min}$ vaut à peu de choses près $\sqrt{(2d)\log(2)} \approx 22$. Similairement, dans une classe de n élèves, la probabilité que deux élèves soient nés le même jour de l'année est $1-p_n$. En revanche, la probabilité qu'un des élèves soit né le même jour que le professeur est tout à fait différente : $1-((d-1)/d)^n$. Il n'y a pas de phénomène de seuil dans ce cas (notons en l'occurence que le nombre de couples est linéaire en n et non pas quadratique comme dans le cas précédent).

Exercice 0.6 (Lancer de dé). *A possède deux dés à six faces, et B possède un dé à douze faces. Le joueur qui fait le plus grand score remporte la mise (match nul si égalité). Le jeu est-il équilibré ? On calculera la probabilité que A gagne et la probabilité d'avoir un match nul.*

Solution. Le modèle d'équiprobabilité sur $\{1,\dots,6\}^2 \times \{1,\dots,12\}$ donne

$$\mathbb{P}(\text{A gagne}) = \frac{\sum_{1\le i,j\le 6}(i+j-1)}{6^2\times 12} = \frac{6\sum_{1\le i\le 6} i + 6\sum_{1\le j\le 6}(j-1)}{6^2\times 12} = \frac{\frac{(1+6)\times 6}{2} + \frac{(0+5)\times 6}{2}}{6\times 12} = \frac{1}{2}$$

et

$$\mathbb{P}(\text{match nul}) = \frac{\sum_{1\le i,j\le 6} 1}{6^2\times 12} = \frac{1}{12} \approx 0.08$$

et donc $\mathbb{P}(\text{B gagne}) = 1/2 - 1/12 = 5/12 \approx 0.42$. Notons que B peut faire 1 et pas A !

1. C'est le *principe des tiroirs* en français, et *pigeon hole principle* en anglais.

Exercice 0.7 (Formule de Poincaré)**.** *Soit $A_1, ..., A_n$ des évènements.*

1. *Montrer que $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$.*
2. *Montrer par récurrence que $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{p=1}^n (-1)^{p+1} \sum_{1 \le i_1 < ... < i_p \le n} \mathbb{P}(A_{i_1} \cap ... \cap A_{i_p})$.*
3. *Application*
   (a) *Pour fêter leur réussite à un concours, n étudiants se donnent rendez-vous dans un chalet. En entrant chaque personne dépose sa veste dans un vestiaire. Au petit matin, quand les esprits ne sont plus clairs, chacun prend au hasard une veste. Quelle est la probabilité pour qu'une personne au moins ait sa propre veste ?*
   (b) *En déduire le nombre de permutations de $\{1, \ldots, n\}$ sans point fixe ;*
   (c) *En s'inspirant de la question (a), calculer la probabilité $\pi_n(k)$ pour que k personnes exactement aient leur propre veste ;*
   (d) *Calculer la limite $\pi(k)$ de $\pi_n(k)$ quand n tend vers l'nfini. Vérifier que la famille $(\pi(k), k \in \mathbb{N})$ détermine une probabilité sur $\mathbb{N}$. Il s'agit en fait de la loi de Poisson.*

Solution. La formule (ou crible) de Poincaré s'appelle également principe d'exclusion-inclusion.

1. Faire un dessin : probabilité = mesure de surface, et utiliser l'additivité disjointe
2. On réduit le cas n au cas $n-1$ en utilisant deux fois le cas $n = 2$
3. On code cette expérience par l'équiprobabilité sur le groupe symétrique (de cardinal $n!$).
   (a) En notant A = « avoir au moins un point fixe » et A_i = « i est point fixe », il vient $\sum_{1 \le i_1 < \cdots < i_p \le n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_p}) = \sum_{1 \le i_1 < \cdots i_p \le n} \frac{(n-p)!}{n!} = \binom{n}{p} \frac{(n-p)!}{n!} = \frac{1}{p!}$ et donc
   $$\mathbb{P}(A) = \mathbb{P}(\cup_{1 \le i \le n} A_i) = \sum_{p=1}^n \frac{(-1)^{p+1}}{p!} \underset{n \to \infty}{\longrightarrow} 1 - e^{-1} \approx 0.63.$$
   (b) Une permutation sans points fixes est appelée *dérangement*. Leur nombre est
   $$\mathrm{card}(A^c) = n!\mathbb{P}(A^c) = n!(1 - \mathbb{P}(A)) = n! \sum_{p=0}^n \frac{(-1)^p}{p!}.$$
   Autre situation concrète justiciable de la même modélisation : le battage de cartes.
   (c) Soit $c_{n,k}$ le nombre de permutation à k points fixes (dérangements partiels, leurs nombres sont les *rencontres numbers* en combinatoire). Pour calculer $c_{n,k}$, on choisit la position des k points fixes, puis un dérangement des $n-k$ positions restantes, ce qui donne la probabilité recherchée assez rapidement :
   $$p_{n,k} = \frac{c_{n,k}}{n!} = \frac{1}{n!} \binom{n}{k} (n-k)! \sum_{p=0}^{n-k} \frac{(-1)^p}{p!} = \frac{1}{k!} \sum_{p=0}^{n-k} \frac{(-1)^p}{p!}.$$
   (d) On a $\lim_{n \to \infty} p_{n,k} = e^{-1}/k!$, qui correspond à la loi de Poisson de paramètre 1.

Exercice 0.8 (Test de dépistage de maladie et formule de Bayes)**.** *Des laboratoires pharmaceutiques ont mis au point un test médical pour dépister une maladie. Les experts pensent qu'une personne sur mille est malade dans la population. De plus, des expériences ont montré que le test déclare positifs 99% des malades qu'on lui soumet, et qu'il déclare malades 2% des personnes saines qu'on lui soumet. Calculer la probabilité qu'une personne soit malade sachant qu'elle a un test positif. Le test est-t-il efficace ?*

Solution. Si on définit les évènements A *«le test médical est positif»* et B *«la personne est malade»* alors les données se traduisent par $\mathbb{P}(B) = 1/1000$, $\mathbb{P}(A|B) = 99/100$, et $\mathbb{P}(A|B^c) = 2/100$. La probabilité que la personne soit malade sachant que test est positif vaut

$$\mathbb{P}(B|A) = \frac{1}{1 + \frac{\mathbb{P}(A|B^c)}{\mathbb{P}(A|B)} \frac{\mathbb{P}(B^c)}{\mathbb{P}(B)}} = \frac{1}{1 + \frac{2 \times 999}{99}} \approx \frac{1}{20} = 0.05.$$

Le test n'est vraiment pas efficace de ce point de vue ! Le paradoxe vient du fait que sur 99.9% de la population le test se trompe dans 2% des cas, et que ces cas représentent finalement une grande part des réponses positives du test. Cet exemple est l'occasion de rappeler que du point de vue statistique, un test comporte deux types d'erreur (faux positifs et faux négatifs) qui ne jouent pas un rôle symétrique du point de vue du risque modélisé.

Exercice 0.9 (Cartes). *On considère trois cartes : une avec les deux faces rouges, une avec les deux faces blanches, et une avec une face rouge et une face blanche. On tire une carte au hasard. On expose une face au hasard. Elle est rouge. Parieriez-vous que la face cachée est blanche ? Pour vous aider dans votre choix :*

1. *Déterminer l'espace de probabilité.*
2. *Calculer la probabilité que la face cachée soit blanche sachant que la face visible est rouge.*

Solution.

1. Le modèle d'équiprobabilité sur $\Omega = \{RR, BB, RB\}$ conduit à modéliser le couple (carte tirée, face visible) par l'espace de probabilité $\Omega = \{(RR, R), (BB, B), (RB, B), (RB, R)\}$ avec probabilités $1/3$, $1/3$, $1/6$, $1/6$. (On peut donc modéliser la face visible par $\Omega = \{R, B\}$ avec probabilités $1/3+1/6 = 1/2$ et $1/3 + 1/6 = 1/2$, c'est-à-dire l'équiprobabilité) ;
2. Si A=«la face cachée est blanche» et B=«la face visible rouge» alors
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{RB, R\})}{\mathbb{P}(\{(RR, R), (RB, R)\})} = \frac{1/6}{1/3 + 1/6} = \frac{1}{3}.$$

Exercice 0.10 (Lemme de Borel-Cantelli). *Soit (A_n) une suite d'évènements dans $(\Omega, \mathcal{F}, \mathbb{P})$.*

1. *Rappeler la définition de $\underline{\lim}\, A_n$ et $\overline{\lim}\, A_n$;*
2. *Montrer que si $\sum_n \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\overline{\lim}\, A_n) = 0$;*
3. *Montrer que si les (A_n) sont indépendants et si $\sum_n \mathbb{P}(A_n) = \infty$ alors $\mathbb{P}(\overline{\lim}\, A_n) = 1$. En déduire la loi du 0-1 de Borel : si les (A_n) sont indépendants alors $\mathbb{P}(\overline{\lim}\, A_n) \in \{0, 1\}$ et*
$$\mathbb{P}(\overline{\lim}\, A_n) = \mathbf{1}_{\sum_n \mathbb{P}(A_n) = \infty}.$$
4. *Application : dans un jeu de pile ou face avec probabilité de gagner p_n à la n-ième partie, donne une condition nécessaire et une condition suffisante sur la suite $(p_n)_{n\geq 1}$ pour que le jeu permette de gagner une infinité de fois ;*
5. *Loi du 0-1 de Kolmogorov : montrer que si $(\mathcal{F}_n)$ est une suite de sous-tribus de $\mathcal{F}$, indépendantes, et si $\mathcal{G}_n$ est la tribu engendrée par $\cup_{m\geq n}\mathcal{F}_m$, alors $\mathbb{P}(A) \in \{0, 1\}$ pour tout A dans la* tribu terminale *$\mathcal{G}_\infty = \cap_n \mathcal{G}_n$. Monter que si $A_n \in \mathcal{F}_n$ pour tout n alors $\overline{\lim}\, A_n \in \mathcal{G}_\infty$.*

Solution.

1. $\underline{\lim}\, A_n = \cup_n \cap_{k\geq n} A_k = \{\omega : \exists n, \forall m \geq n : \omega \in A_m\}$ est l'ensemble des $\omega \in \Omega$ qui appartiennent à A_n à partir d'un certain rang sur n (qui peut dépendre de ω), tandis que $\overline{\lim}\, A_n = \cap_n \cup_{k\geq n} A_k = \{\omega : \forall n, \exists m \geq n : \omega \in A_m\}$ est l'ensemble des $\omega \in \Omega$ qui appartiennent à A_n pour une infinité de valeurs de n (qui peuvent dépendre de ω). On a $(\underline{\lim}\, A_n)^c = \overline{\lim}\, A_n^c$ et $(\overline{\lim}\, A_n)^c = \underline{\lim}\, A_n^c$. De plus, $\mathbf{1}_{\underline{\lim}\, A_n} = \underline{\lim}\, \mathbf{1}_{A_n}$ et $\mathbf{1}_{\overline{\lim}\, A_n} = \overline{\lim}\, \mathbf{1}_{A_n}$. Il faut penser à l'analogie $\sup_n = \cup_n = \exists$ et $\inf_n = \cap_n = \forall$;
2. La suite (B_n) définie par $B_n = \cup_{m\geq n} A_m$ est décroissante, et si $\sum_n \mathbb{P}(A_n) < \infty$ alors
$$\mathbb{P}(\overline{\lim}\, A_n) = \mathbb{P}(\cap_n B_n) = \lim_n \mathbb{P}(B_n) \leq \lim_n \sum_{m\geq n} \mathbb{P}(A_m) = 0;$$
3. Il suffit d'établir que si $\sum_n \mathbb{P}(A_n) = \infty$ alors $\mathbb{P}(\underline{\lim}\, A_n^c) = 0$. Si $B_n = \cap_{m\geq n} A_m^c$ alors (B_n) est croissante et donc $\mathbb{P}(\underline{\lim}\, A_n^c) = \lim_n \mathbb{P}(B_n)$. À présent, l'indépendance des (A_n^c), l'inégalité $1 - x \leq e^{-x}$ valable pour $x \in \mathbb{R}$, et $\sum_n \mathbb{P}(A_n) = \infty$, donnent pour tout n :
$$\mathbb{P}(B_n) = \prod_{m\geq n} \mathbb{P}(A_n^c) = \prod_{m\geq n} e^{-\mathbb{P}(A_m)} = e^{-\sum_{m\geq n} \mathbb{P}(A_m)} = 0;$$
4. Soit A_n l'événement « gagner la n-ième partie ». L'événement « gagner une infinité de fois » s'écrit $\overline{\lim}_n A_n = \cap_n \cup_{m\geq n} A_m$. Les $(A_n)_{n\geq 1}$ sont indépendants. Par le lemme de Borel-Cantelli, $\mathbb{P}(\overline{\lim}\, A_n)$ vaut 0 ou 1, et 1 ssi $\sum_n p_n = +\infty$. Il y a aussi une autre application : le singe de Shakespeare.
5. Les tribus $\mathcal{F}_n$ et $\mathcal{G}_{n+1}$ sont indépendantes, et donc la tribu engendrée par $\cup_n \mathcal{F}_n$ est indépendante de la tribu $\mathcal{G}_\infty$. Comme $\mathcal{G}_\infty \subset \cup_n \mathcal{F}_n$ on en déduit que tout évènement de la tribu terminale $\mathcal{G}_\infty$ est indépendant de lui même, et est donc de probabilité 0 ou 1. En quelque sorte, on a $\mathcal{G}_\infty = \overline{\lim}\, \mathcal{F}_n$. Il est clair que si $A_n \in \mathcal{F}_n$ pour tout n alors $\underline{\lim}_n A_n$ et $\overline{\lim}_n A_n$ sont dans la tribu terminale et donc sont de probabilité 0 ou 1, ce qui montre que la loi du 0-1 de Borel découle de celle de Kolmogorov.

Exercice 0.11 (Échantillonnage)**.**

1. *Considérons une population de $N = N_1 + N_2$ individus dont N_1 de type 1 et N_2 de type 2. On effectue un tirage sans remise de $n \leq N$ individus. Calculer la probabilité d'obtenir de la sorte $n_1 \leq n$ individus de type 1 (ou de manière équivalente $n_2 = n - n_1$ individus de type 2), puis calculer la limite de cette probabilité lorsque $N \to \infty$ avec $N_1/N \to p \in [0,1]$;*
2. *Plus généralement, considérons une population de $N = N_1 + \cdots + N_d$ individus dont N_i sont de type i pour tout $1 \leq i \leq d$. On effectue un tirage sans remise de $n \leq N$ individus. Pour tous $n_1 \leq N_1, \ldots, n_d \leq N_d$ tel que $n_1 + \cdots + n_d = n$, calculer la probabilité d'obtenir n_i individus de type i pour tout $1 \leq i \leq n$. Calculer la limite de cette probabilité si $N \to \infty$ avec $(N_1/N, \ldots, N_d/N) \to (p_1, \ldots, p_d)$.*

Solution.

1. Il y a $\binom{N}{n}$ tirages possibles. Adoptons le modèle de la probabilité uniforme sur l'ensemble de ces possibilités, c'est-à-dire sur l'univers Ω des sous-ensembles de $\{1, \ldots, N\}$ à n éléments. Pour tout $0 \leq n_1 \leq n$, le nombre de tirages avec n_1 individus de type 1 est $\binom{N_1}{n_1}\binom{N_2}{n-n_1}$, et la probabilité de tirer n_1 individus de type 1 vaut donc
$$\frac{\binom{N_1}{n_1}\binom{N_2}{n-n_1}}{\binom{N}{n}}.$$
Cette formule définit la loi hypergéométrique sur les sous-populations de taille n d'une population de taille N à deux types. Cela montre au passage que
$$\binom{N}{n} = \sum_{k=0}^{n} \binom{N_1}{k}\binom{N_2}{n-k}.$$
Il s'agit de l'identité de convolution de Vandermonde, qui découle également de l'identité $(1+x)^{N_1}(1+x)^{N_2} = (1+x)^N$ en développant et en identifiant. La formule de Stirling entraîne que $\binom{N_1}{k}\binom{N_2}{n-k}/\binom{N}{n} \to p^k(1-p)^{n-k}\binom{n}{k}$ lorsque $N_1, N_2 \to \infty$ avec $N_1/N \to p$. Cette formule définit la loi binomiale de taille n et de paramètre p (les petits tirages sans remise dans une très grande population sont comme des tirages avec remise). Le modèle binomial, qui n'est pas forcément équiprobable, apparaît comme limite d'un modèle hypergéométrique équiprobable. Les modèles équiprobables sont fondamentaux ;
2. Il y a $\binom{N}{n}$ tirages possibles. Le nombre de tirages comportant n_i individus de type i pour tout $1 \leq i \leq d$ est $\binom{N_1}{n_1} \cdots \binom{N_d}{n_d}$. Avec le modèle de la probabilité uniforme sur les tirages, la probabilité du tirage $(n_1, \ldots, n_d)$ est
$$\frac{\binom{N_1}{n_1} \cdots \binom{N_d}{n_d}}{\binom{N}{n}}.$$
Cette formule définit la loi hypergéométrique « multitype ». On retrouve au passage une version multivariée de l'identité de convolution Vandermonde, qui découle également de l'identité $(1+x)^{N_1} \cdots (1+x)^{N_d} = (1+x)^N$ en développant et en identifiant :
$$\binom{N}{n} = \sum_{\substack{n_1 + \cdots + n_d = n \\ n_1 \leq N_1, \ldots, n_d \leq N_d}} \binom{N_1}{n_1} \cdots \binom{N_d}{n_d}.$$
La formule de Stirling indique que $\binom{N_1}{n_1} \cdots \binom{N_d}{n_d}/\binom{N}{n}$ converge vers $\frac{n!}{n_1! \cdots n_d!} p_1^{n_1} \cdots p_d^{n_d}$ lorsque $N_1, \ldots, N_d \to \infty$ avec $(N_1/N, \ldots, N_d/N) \to (p_1, \ldots, p_d)$. Cette formule définit la loi multinomiale de taille n et de paramètre $(p_1, \ldots, p_d)$.

Exercice 0.12 (Tirage aléatoire d'une partie)**.** *Le tirage d'une partie à k élément dans un ensemble à n éléments est modélisable par la loi uniforme sur les $\binom{n}{k}$ parties à k éléments. Il est également possible d'effectuer ce tirage élément par élément, ce qui correspond à k tirages sans remise successifs. Montrer que les deux modélisations coïncident.*

Solution.
$$k! \frac{1}{n} \frac{1}{n-1} \cdots \frac{1}{n-k+1} = \frac{1}{\binom{n}{k}}.$$
Le membre de droite provient du fait qu'il y a $k!$ manières d'ordonner les k éléments à tirer, et la probabilité de chaque suite de k tirages vaut $\frac{1}{n} \frac{1}{n-1} \cdots \frac{1}{n-k+1}$ car il faut tirer la première parmi n, puis la seconde parmi $n-1$, etc. On a donc utilisé les probabilités totales et l'indépendance.

Exercice 0.13 (Entropie de Boltzmann-Shannon)**.** *On considère un système macroscopique formé par n particules distinguables, chacune pouvant être dans l'un des r niveaux d'énergie possibles. On a $n = n_1 + \cdots + n_r$ où $n_i \in \{0, \ldots, n\}$ est le nombre de particules de type $i \in \{1, \ldots, r\}$.*

1. *Combien existe-t-il d'états macroscopiques ? Pour un état macroscopique $(n_1, \ldots, n_r)$ donné, combien existe-t-il d'états microscopiques (leur nombre est noté $\binom{n}{n_1,\ldots,n_r}$) ?;*
2. *On s'intéresse au cas où $n \to \infty$ de sorte que $\frac{n_i}{n} \to p_i$ pour tout $1 \le i \le n$. Montrer que $\frac{1}{n} \ln \binom{n}{n_1,\ldots,n_r} \to H(p_1, \ldots, p_r) := -\sum_{i=1}^r p_i \ln(p_i)$;*
3. *Trouver les extrema de $H : \{(p_1, \ldots, p_r) \in [0,1]^r : p_1 + \cdots + p_r = 1\} \to \mathbb{R}$;*
4. *Montrer que parmi les lois de probabilité P sur $\mathbb{N}$ à moyenne $m = \sum_{i\in\mathbb{N}} ip_i$ fixée, l'entropie $H(P) = -\sum_{i\in\mathbb{N}} p_i \ln(p_i)$ est maximisée par une unique loi de probabilité, à déterminer.*

Solution.

1. Le nombre d'états macroscopiques vaut (question 4 de l'exercice 0.1, cloisons-tiroirs)
$$\mathrm{card}\{(n_1, \ldots, n_r) \in \{0, \ldots, n\}^r : n_1 + \cdots + n_r = n\} = \binom{n+r-1}{r-1} = \binom{n+r-1}{n}.$$
(attention, les rôles de n et r sont inversés par rapport à la question 4 de l'exercice 0.1). Il s'agit du nombre de termes dans la formule du multinome de taille n et avec r types.
Pour un état macroscopique $(n_1, \ldots, n_r)$ fixé, avec $n_1+\cdots+n_r = n$, le nombre $\binom{n}{n_1,\ldots,n_r}$ d'état microscopiques correspond aux degrés de liberté du système. Comme les particules sont distinguables, on a
$$\binom{n}{n_1, \ldots, n_r} := \mathrm{card}\{(b_1, \ldots, b_n) : \mathrm{card}\{1 \le k \le n : b_k = i\} = n_i \text{ pour tout } 1 \le i \le r\}.$$
Pour $r = 2$ on retrouve le placement de n_1 parties pile et n_2 parties face parmi les n parties. Plus généralement, pour $r \ge 2$, il faut penser à n jets d'un dé à r faces : on choisit de positionner les n_1 jets de type 1 parmi les n disponibles, puis les n_2 jets de type 2 parmi les $n - n_1$ restants, etc. On a donc, en notant que $n_r = n - n_1 - \cdots - n_{r-1}$,
$$\begin{aligned}\binom{n}{n_1, \ldots, n_r} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r}\\ &= \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-\cdots-n_{r-1})!}{n_r!0!}\\ &= \frac{n!}{n_1!\cdots n_r!}.\end{aligned}$$
Il s'agit du coefficient multinomial. Le nombre total d'états microscopiques possibles s'obtient en sommant le nombre d'états microscopique pour chaque état macroscopique, ce qui donne (formule du multinome) :
$$\sum_{\substack{(n_1,\ldots,n_r)\in\{0,\ldots,n\}^r\\ n_1+\cdots+n_r=n}} \frac{n!}{n_1!\cdots n_r!} = (\underbrace{1+\cdots+1}_{r \text{ fois}})^n = r^n.$$
On retrouve le fait qu'on choisit le niveau d'énergie parmi r possibles pour n particules distinguables !
2. Formule de Stirling $n! \sim \sqrt{2\pi n}(n/e)^n$;
3. La fonction H est strictement concave car $p \in [0,1] \mapsto p\ln(p)$ est strictement convexe. Le minimum vaut 0 quand l'incertitude est minimale : $p \in \{\delta_i : 1 \le i \le n\}$ et le maximum $\ln(n)/n$ quand l'incertitude est maximale, c'est-à-dire pour la loi uniforme $p = (1/n, \ldots, 1/n)$;
4. Contraintes : $\sum_i p_i = 1$, $\sum_i ip_i = m$, $0 \le p_i \le 1$. L'équation de Lagrange $\partial_i H(P) = -\ln(p_i)-1 = \alpha+\beta i$ conduit à $p_i = e^{-(1+\alpha)}e^{-\beta i}$, et donc P est la loi géométrique de moyenne m.

1 V.a. et lois discrètes, espérance, variance, moments

Rappels de cours. Notion de v.a., exemple des indicatrices, du jeu de pile ou face (compter c'est sommer des indicatrices), lois discrètes associées au jeu de pile ou face (sans démo, cf. exos). Notion de loi d'une v.a. Notion d'espérance, espérance d'une indicatrice, propriétés de l'espérance. Notion d'indépendance de v.a. Notions de variance et de covariance, lien avec l'indépendance.

Exercice 1.1 (Jeu de pile ou face). *On modélise le jeu de pile ou face par une suite $(X_n)_{n\geq 1}$ de v.a. de Bernoulli de paramètre $p \in [0,1]$, en codant 1 pour succès (pile) et 0 pour échec (face) : $\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = 0) = p$. Soit $S_n = X_1 + \cdots + X_n$ et $T_{m+1} = \inf\{n > T_m : X_n = 1\}$, $T_0 = 0$.*

1. *Que représentent les v.a. S_n et T_m ? Expliciter leur loi en fonction de p, la moyenne et la variance (interpréter le comportement en p). Que se passe-t-il lorsque $p = 0$ et $p = 1$?*
2. *Que représente la v.a. $T_1 - 1$, qu'elle est sa loi, sa moyenne, sa variance ?*
3. *Soit $A_n = \{X_n = 1\}$ pour tout $n \geq 1$. Que représentent les événements $\underline{\lim}\, A_n$ et $\overline{\lim}\, A_n$? Qu'elle est leur probabilité ? Qu'elle est la probabilité qu'une suite finie fixée de 0 et et 1 apparaisse une infinité de fois dans la suite $(X_n)_{n\geq 1}$?*
4. *On suppose que $0 < p < 1$ et on considère le premier temps d'apparition de la séquence «01» donné par $W = \inf\{n \geq 2; X_{n-1} = 0, X_n = 1\}$. Montrer que W a la même loi que $U + V$ où U, V sont des v.a. indépendantes de loi géométrique de paramètre p et $1 - p$.*

Solution.

1. La v.a. $S_n \sim \mathrm{Binom}(n,p)$ est le nombre de succès lors des n premières parties, et $\mathbb{E}(S_n) = np$ et $\mathrm{Var}(S_n) = np(1-p)$. La v.a. T_m est le rang (ou temps d'apparition) du m^{e} succès. Elle prend ses valeurs dans $\mathbb{N}^*$. Si $p = 0$ alors $\mathbb{P}(T_m = \infty) = 1$. Si $p > 0$ alors $T_1 \sim \mathrm{Geom}(p)$, $\mathbb{E}(T) = 1/p$, $\mathrm{Var}(T) = (1-p)/p^2$. On a $T_m = G_1 + \cdots + G_m$ où $G_1, \ldots, G_m$ sont i.i.d. de même loi que T_1. On dit que T_m suit une loi de Pascal ou loi binomiale negative. On a $\mathbb{E}(T_m) = m/p$ et $\mathrm{Var}(T_m) = m(1-p)/p^2$. Si $p = 1$ alors $\mathbb{P}(S_n = 1, T_m = m) = 1$;
2. $T_1 - 1$ est le nombre d'échecs avant le premier succès. Elle suit une loi géométrique sur $\mathbb{N}$ de moyenne $\mathbb{E}(T_1 - 1) = \mathbb{E}(T_1) - 1 = (1-p)/p$ et de variance $\mathrm{Var}(T_1 - 1) = \mathrm{Var}(T_1) = (1-p)/p^2$;
3. $\underline{\lim}\, A_n$ est « ne faire que gagner à partir d'un certain moment » et $\overline{\lim}\, A_n$ est «gagner une infinité de fois». Par le lemme de Borel-Cantelli, ces événements sont de probabilité 0 ou 1 et on passe de l'un à l'autre en passant au complémentaire et en échangeant p et $1-p$. On a $\mathbb{P}(\overline{\lim}\, A_n) = 1$ ssi $p > 0$. Si la suite a une probabilité non nulle d'apparaître dès le début, alors elle a une probabilité 1 d'apparaître une infinité de fois d'après le lemme de Borel-Cantelli à nouveau, pour un jeu de pile ou face par blocs par exemple ;
4. D'abord une suite de 1 de longueur $U - 1$ puis une suite de 0 de longueur V puis un 1.

Exercice 1.2 (Lois multinomiales). *On considère un dé à $d \geq 2$ faces, et on note $p_1, \ldots, p_d$ les probabilités d'apparition de chacune des faces. On lance n fois ce dé, et on code les résultats dans un vecteur aléatoire $Z = (Z_1, \ldots, Z_d)$ à valeurs dans $\{(n_1, \ldots, n_d) \in \{0, \ldots, n\}^d : n_1 + \cdots + n_d = n\}$.*

1. *Déterminer la loi de Z. Déterminer la loi des composantes $Z_1, \ldots, Z_d$ de Z. Sont-elles indépendantes ? Calculer leur matrice de covariance lorsque $p_1 = \cdots = p_d = p$;*
2. *Pour toute partition $I_1 \cup \cdots \cup I_r = \{1, \ldots, d\}$, calculer la loi de $Z' = (\sum_{i\in I_1} Z_i, \ldots, \sum_{i\in I_r} Z_i)$;*
3. *Soient $Z_1, \ldots, Z_d$ des v.a. indépendantes de loi de Poisson de paramètres $\lambda_1, \ldots, \lambda_d$. Montrer que pour tout $n \geq 0$, sachant l'événement $\{Z_1 + \cdots + Z_d = n\}$, le vecteur X suit la loi multinomiale de taille n et de paramètre $(\lambda_1/\lambda, \ldots, \lambda_d/\lambda)$ où $\lambda = \lambda_1 + \cdots + \lambda_d$.*

Solution.

1. Le vecteur Z suit la loi multinomiale de taille n et de paramètre $(p_1, \ldots, p_d)$, qui prend ses valeurs sur $\{(n_1, \ldots, n_d) \in \{0, \ldots, n\} : n_1 + \cdots + n_d = n\}$. On a $Z = X_1 + \cdots + X_n$ avec $X_1, \ldots, X_n$ i.i.d. de loi $\sum_{i=1}^d p_i \delta_{e_i}$. Pour tout $(n_1, \ldots, n_d)$ admissible, on calcule en plaçant les n_1 faces 1, etc, n_d faces d sur les n positions possibles (nous avons déjà fait ce calcul pour l'exercice sur l'entropie de Boltzmann

de la feuille précédente !) :

$$\begin{aligned}
\mathbb{P}(Z=n) &= \sum_{\substack{(b_1,\ldots,b_n)\in\{1,\ldots,d\}^n\\ \forall i:\mathrm{card}\{k:b_k=i\}=n_i}} \mathbb{P}(X_1=b_1,\ldots,X_n=b_n)\\
&= \sum_{\substack{(b_1,\ldots,b_n)\in\{1,\ldots,d\}^n\\ \forall i:\mathrm{card}\{k:b_k=i\}=n_i}} \mathbb{P}(X_1=b_1)\cdots\mathbb{P}(X_n=b_n)\\
&= \sum_{\substack{(b_1,\ldots,b_n)\in\{1,\ldots,d\}^n\\ \forall i:\mathrm{card}\{k:b_k=i\}=n_i}} p_1^{n_1}\cdots p_d^{n_d}\\
&= p_1^{n_1}\cdots p_d^{n_d}\mathrm{card}\{(b_1,\ldots,b_n)\in\{1,\ldots,d\}^n : \forall i : \mathrm{card}\{k:b_k=i\}=n_i\}\\
&= p_1^{n_1}\cdots p_d^{n_d}\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-(n_1-\cdots-n_{r-1})}{n_r}\\
&= p_1^{n_1}\cdots p_d^{n_d}\binom{n}{n_1,\ldots,n_d}.
\end{aligned}$$

Ses composantes suivent des lois binomiales de taille n et de paramètre $p_1,\ldots,p_d$. On a la relation $Z_1+\cdots+Z_d=n$. Supposons à présent que $p_1=\cdots=p_d=p$. Dans ce cas, $Z_1,\ldots,Z_d$ ont même loi et donc $0=\mathrm{Var}(Z_1+\cdots+Z_d)=n\mathrm{Var}(Z_1)+n(n-1)\mathrm{Cov}(Z_1,Z_2)$. Comme $\mathrm{Var}(Z_1)=np(1-p)$ il vient $\mathrm{Cov}(Z_1,Z_2)=-p(1-p)\frac{n}{n-1}$, et donc $\mathrm{Cov}_Z=p(1-p)(nI_d-\frac{n}{n-1}J)$ où $J=(\mathbf{1}_{i\neq j})_{1\leq i,j\leq n}$. Elle n'est pas diagonale. Il n'y a pas indépendance ;

2. Le vecteur Z' suit la loi multinomiale de taille r et de paramètres $(\sum_{i\in I_1}p_i,\ldots,\sum_{i\in I_r}p_i)$;
3. On a $\mathbb{P}(Z_1+\cdots+Z_d=n)=e^{-\lambda}\frac{\lambda^n}{n!}$ et donc pour tous $n_1,\ldots,n_d$ vérifiant $n_1+\cdots+n_d=k$,

$$\begin{aligned}
\mathbb{P}(Z_1=n_1,\ldots,Z_d=n_d|Z_1+\cdots+Z_d=n) &= \frac{\mathbb{P}(Z_1=n_1,\ldots,Z_d=n_d)}{\mathbb{P}(Z_1+\cdots+Z_d=n)}\\
&= \frac{e^{-\lambda_1}\frac{\lambda_1^{n_1}}{n_1!}\cdots e^{-\lambda_n}\frac{\lambda_d^{n_d}}{n_d!}}{e^{-\lambda}\frac{\lambda^n}{n!}}\\
&= \frac{n!}{n_1!\cdots n_d!}\left(\frac{\lambda_1}{\lambda}\right)^{n_1}\cdots\left(\frac{\lambda_d}{\lambda}\right)^{n_d}.
\end{aligned}$$

Exercice 1.3 (Compter). *On rappelle que si X est une v.a. dans $\mathbb{N}\cup\{\infty\}$, et si $\mathbb{P}(X=\infty)>0$ alors $\mathbb{E}(X)=\infty$. La réciproque est fausse. Contraposée : si $\mathbb{E}(X)<\infty$ alors $\mathbb{P}(X=\infty)=0$.*

1. *Soit $(X_n)_{n\geq 1}$ une suite de v.a. à valeurs dans $\mathbb{N}\cup\{\infty\}$. Montrer que $\mathbb{E}(\sum_n X_n)=\sum_n\mathbb{E}(X_n)$;*
2. *En déduire une preuve de la première partie du lemme de Borel-Cantelli.*

Solution.

1. Théorème de convergence monotone ou théorème de Fubini-Tonelli positif ;
2. Avec $X_n=\mathbf{1}_{A_n}$ on a $\sum_n\mathbb{E}(X_n)=\sum_n\mathbb{P}(A_n)$ tandis que $\{\sum_n X_n=\infty\}=\overline{\lim}\,A_n$.

Exercice 1.4 (Collectionneur de coupons). *On joue une infinité de fois avec un dé équilibré à $r\geq 2$ faces. On code cette expérience par une suite $(X_n)_{n\geq 1}$ de v.a. i.i.d. à valeurs dans $\{1,\ldots,r\}$. On a $\mathbb{P}(X_1=k)=1/r$ pour tout $1\leq k\leq r$. On s'intéresse au premier temps d'observation de toutes les faces : $T=\inf\{n\geq 1 : \{X_1,\ldots,X_n\}=\{1,\ldots,r\}\}=\inf\{n\geq 1:\mathrm{card}\{X_1,\ldots,X_n\}=r\}$.*

1. *Qu'elle est la loi de T quand $r=2$? (indication : penser au jeu de pile ou face). Plus généralement, pour tout $r\geq 2$, montrer que $T=G_1+\cdots+G_r$ où $G_1,\ldots,G_r$ sont des v.a. géométriques sur $\mathbb{N}^*$, indépendantes, avec G_i de paramètre $\pi_i:=(r-i+1)/r$;*
2. *En déduire que $\mathbb{E}(T)=\gamma r+r\ln(r)+o_{r\to\infty}(r)$ et $\mathrm{Var}(T)=\frac{\pi^2}{6}r^2-r\ln(r)-r+o_{r\to\infty}(r^2)$;*
3. *En déduire que pour tout $\varepsilon>0$, $\lim_{r\to\infty}\mathbb{P}\left(\left|\frac{T}{r\ln(r)}-1\right|>\varepsilon\right)=0$.*

Solution.

1. On pose $G_1 \equiv 1$ et pour tout $1 < i \leq r$,
$$G_i = \min\{n \geq 1 : X_{G_{i-1}+n} \notin \{X_1, \ldots, X_{G_{i-1}}\}\}.$$
On a $\mathrm{card}(\{X_1, \ldots, X_{G_i}\}) = i$ pour tout $1 \leq i \leq n$. Les variables aléatoires $G_1, G_1 + G_2, \ldots, G_1 + \cdots + G_r$ sont les temps d'apparition des r premiers gains dans un jeu de pile ou face spécial dans lequel la probabilité de gagner change après chaque gain : cette probabilité vaut successivement $1, (r-1)/r, (r-2)/r, \ldots, 1/r$. Cela témoigne du fait qu'il est de plus en plus difficile d'obtenir un coupon d'un nouveau type au fil de la collection.
2. Par linéarité de l'espérance (nul besoin de l'indépendance !), en posant $\pi_i = (r-i+1)/r$,
$$\mathbb{E}(T) = \sum_{i=1}^{r} \mathbb{E}(G_i) = \sum_{i=1}^{r} \frac{1}{\pi_i} = \sum_{i=1}^{r} \frac{r}{r-i+1} = r \sum_{i=1}^{r} \frac{1}{i} = r(\ln(r) + \gamma + o_{r\to\infty}(1))$$
où $\gamma = \lim_{n\to\infty}(\sum_{i=1}^{n} 1/i - \ln(n)) \approx 0.577$ est la constante d'Euler. Comme les $G_1, \ldots, G_r$ sont indépendantes avec $\sigma^2(G_i) = (1-\pi_i)/\pi_i^2 = r(i-1)/(r-i+1)^2$, on a également
$$\sigma^2(T) = \sum_{i=1}^{r} \sigma^2(G_i) = r \sum_{i=1}^{r-1} \frac{r-i}{i^2} = \frac{\pi^2}{6} r^2 - r\ln(r) - r + o_{r\to\infty}(r^2).$$
3. L'inégalité de Markov et le comportement des deux premiers moments donnent
$$\mathbb{P}\left(\left|\frac{T}{r\ln(r)} - 1\right| > \varepsilon\right) \leq \frac{\mathbb{E}(|T - r\ln(r)|^2)}{r^2\ln(r)^2\varepsilon^2} = \frac{\sigma^2(T) + (\mathbb{E}(T) - r\ln(r))^2}{r^2\ln(r)^2\varepsilon^2} = \mathcal{O}_{r\to\infty}\left(\frac{1}{\ln(r)^2}\right).$$
Notons que cette borne ne permet pas d'appliquer le lemme de Borel-Cantelli.

Exercice 1.5 (Espérance, variance, fonction génératrice)**.**

1. *Montrer que si X est une v.a. sur $\mathbb{N}$ alors $\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$ dans $\mathbb{N} \cup \{\infty\}$;*
2. *Calculer la fonction génératrice des lois de Bernoulli, binomiale, géométrique, de Poisson, uniforme. Préciser les liens avec l'indépendance, la moyenne, la variance ;*

Solution.

1. Découle de l'identité $x = \sum_{n=0}^{\infty} \mathbf{1}_{x>n}$ pour tout $x \in \mathbb{N}$;
2. Si $G_X(s) := \mathbb{E}(s^X)$ alors $G'(1) = \mathbb{E}(X)$ et $G''(1) = \mathbb{E}(X(X-1)) = \mathrm{Var}(X) + \mathbb{E}(X) - \mathbb{E}(X^2)$. De plus, X et Y indépendantes ssi $G_{(X,Y)}(s,t) := \mathbb{E}(s^X t^Y) = G_X(s)G_Y(t)$ pour tous $0 \leq s, t \leq 1$;

Exercice 1.6 (Ruine du joueur)**.** *Un joueur possédant une fortune initiale de $x > 0$ Euros joue avec une machine à sous dans un casino. À chaque partie, le joueur peut gagner 1 Euro avec probabilité p et perdre 1 Euro avec probabilité $1-p$. Le joueur décide de quitter le casino lorsque sa fortune atteint a (ruine) ou b (fortune), avec $0 \leq a \leq x \leq b$ (les cas $x = a$ et $x = b$ sont triviaux). On note $r(x)$ (resp. $f(x)$) la probabilité que le joueur quitte le casino ruiné (resp. fortuné).*

1. *Établir une relation de récurrence en $r(x-1), r(x), r(x+1)$, puis une formule pour $r(x)$;*
2. *Le joueur va-t-il quitter le casino ?*
3. *Donner un sens au cas où $a < 0$. Que se passe-t-il quand $b \to \infty$?*

Solution.

1. Pour tout $n \in \mathbb{N}$, on pose $S_n = x + X_1 + \cdots + X_n$ où $(X_n)_{n\geq 0}$ est une suite de v.a. i.i.d. de loi de Bernoulli sur $\{-1, 1\}$ de paramètre p : $\mathbb{P}(X_k = 1) = 1 - \mathbb{P}(X_k = -1) = p$. L'instant de sortie du casino est modélisé par la variable aléatoire $T = \inf\{n \geq 0 : S_n \in \{a, b\}\}$ à valeurs dans $\mathbb{N} \cup \{\infty\}$. On a $r(x) = \mathbb{P}(T < \infty; S_T = a)$. En conditionnant par rapport à X_1, il vient
$$r(x) = r(x-1)(1-p) + r(x+1)p.$$
L'ensemble des solution de cette équation de récurrence linéaire est un espace vectoriel de dimension 2, contenant la constante 1. Si $p \neq 1/2$, alors ρ^x avec $\rho = (1-p)/p$ est une solution linéairement indépendante de 1, et l'ensemble des solution est de la forme $A + B\rho^x$. Les conditions au bord $r(a) = 1$ et $r(b) = 0$ donnent $B = 1/(\rho^a - \rho^b)$ et $A = -\rho^b B$, d'où
$$r(x) = \frac{\rho^b - \rho^x}{\rho^b - \rho^a}.$$

Si $p = 1/2$ alors la solution ρ^x se confond avec 1 mais x est une solution linéairement indépendante, ce qui donne (notons que les deux formules sont reliées quand $\rho \to 1/2$)

$$r(x) = \frac{b-x}{b-a};$$

2. Le même raisonnement pour $f(x)$ donne $f(x) = \frac{\rho^x - \rho^a}{\rho^b - \rho^a}$ si $p \neq 1/2$ et $f(x) = \frac{x-a}{b-a}$ si $p = 1/2$. Ainsi, $r(x) + f(x) = 1$ et donc $\mathbb{P}(T < \infty) = 1$;
3. Le cas $a \leq 0$ s'interprète comme une possibilité d'endètement. Si $b \to \infty$ alors $r(x) \to 1$.

Exercice 1.7 (Théorème de Weierstrass et polynômes de Bernstein)**.** *Soit $f : [0,1] \to \mathbb{R}$ continue et $S \sim \mathrm{Binom}(n,p)$. Montrer que la quantité $f_n(p) = \mathbb{E}(f(S/n))$ est un polynôme en p convergeant ponctuellement vers f sur $[0,1]$ quand $n \to \infty$. En utilisant le théorème de Heine et l'inégalité de Tchebychev pour S, montrer que la convergence est uniforme.*

Solution. On a $f(p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(k/n)$ qui est bien un polynôme. Pour tout $\eta > 0$ et tout $p \in [0,1]$, on a, par l'inégalité de Tchebychev, $\mathbb{P}(|S_{n,p}/n - p| \geq \eta) \leq \frac{p(1-p)}{n\eta^2} \leq \frac{1}{4n\eta^2}$, d'où

$$|f_n(p) - f(p)| \leq \sum_{|k/n-p|\leq\eta} |f(k/n) - f(p)| \binom{n}{k} p^k (1-p)^{n-k} + \frac{2\|f\|_\infty}{4n\eta^2}.$$

On fixe d'abord η en utilisant la continuité uniforme de f (théorème de Heine), puis n.

Exercice 1.8 (Permutations aléatoires et algorithme de Fisher-Yates-Knuth)**.** *On souhaite fabriquer une variable aléatoire σ_n qui suit la loi uniforme sur le groupe symétrique Σ_n.*

1. *On construit σ_n en tirant $\sigma_n(1)$ uniformément sur $\{1,\ldots,n\}$, puis $\sigma_n(2)$ uniformément sur $\{1,\ldots,n\} \setminus \{\sigma_n(1)\}$, etc. Montrer que cette construction par tirages uniformes sans remise fournit une permutation aléatoire σ_n de loi uniforme sur Σ_n ;*
2. *On construit σ_n en effectuant le produit de transpositions aléatoires $(1,U_1)\cdots(n,U_n)$, où $U_1,\ldots,U_n$ sont des variables aléatoires indépendantes avec U_i de loi uniforme sur $\{1,\ldots,i\}$ pour tout $1 \leq i \leq n$. Montrer que σ_n ainsi construite suit la loi uniforme sur Σ_n ;*
3. *Quel est le meilleur algorithme ?*

Solution.

1. Par conditionnements successifs on obtient, pour tout $\sigma \in \Sigma_n$,

$$\mathbb{P}(\sigma_n = \sigma) = \prod_{i=1}^n \frac{1}{i} = \frac{1}{n!} = \frac{1}{\mathrm{card}(\Sigma_n)};$$

2. On procède par récurrence sur n. La propriété est triviale pour $n = 1$. Pour tout $\sigma \in \Sigma_n$, on note encore σ l'élément de Σ_{n+1} obtenu à partir de σ en ajoutant le cycle $(n+1)$ (point fixe). Supposons que $\sigma_n = (1,U_1)\cdots(n,U_n)$ suit la loi uniforme sur Σ_n. Soit U_{n+1} une variable aléatoire indépendante de σ_n, de loi uniforme sur $\{1,\ldots,n+1\}$. Montrons que $\sigma_{n+1} = \sigma_n(n+1,U_{n+1})$ suit la loi uniforme sur Σ_{n+1}. Pour tout $\sigma \in \Sigma_{n+1}$, on a tout d'abord

$$\mathbb{P}(\sigma_{n+1} = \sigma) = \sum_{i=1}^{n+1} \mathbb{P}(\sigma_n = \sigma(n+1,i))\mathbb{P}(U_{n+1} = i) = \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{P}(\sigma_n = \sigma(n+1,i)).$$

Comme $n+1$ est point fixe de σ_n, et n'est point fixe de $\sigma(n+1,i)$ que pour une et une seule valeur de i, notée i_σ, image réciproque de $n+1$ par σ, il en découle finalement que

$$\mathbb{P}(\sigma_{n+1} = \sigma) = \frac{1}{n+1}\mathbb{P}(\sigma_n = \sigma(n+1,i_\sigma)) = \frac{1}{n+1}\frac{1}{n!} = \frac{1}{(n+1)!}.$$

Note : l'inversion étant bijective, et les transpositions étant leur propre inverse, il en découle que le produit renversé $(n,U_n)\cdots(1,U_1)$ suit également la loi uniforme sur Σ_n. Notons que la transposition $(1,U_1)$ est triviale (élément neutre de Σ_n), et il n'est pas nécessaire d'en tenir compte si $n \geq 2$ (elle rend cependant la formule valable pour $n = 1$) ;

3. Les deux algorithmes ont une complexité théorique $\mathcal{O}(n)$. Cependant, la mise en œuvre pratique du premier est pénible car il faut tenir compte des éléments déjà tirés.

Exercice 1.9 (Partitions aléatoires et algorithme de Stam). *Soit Π_n l'ensemble des partitions de $\{1,\ldots,n\}$, et $B_n = \mathrm{card}(\Pi_n)$. En combinatoire, la suite $(B_n)_{n\geq 1}$ constitue les* nombres de Bell.

1. *Montrer que $B_1 = 1$, $B_2 = 2$, et que plus généralement, en utilisant la convention $B_0 = 1$, on dispose de la formule de récurrence triangulaire*
$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k;$$
2. *Montrer que la série formelle $G(X) = \sum_{n=0}^{\infty} \frac{B_n}{n!} X^n$ vérifie $G'(X) = \exp(X)G(X)$ et en déduire que B_n est le moment d'ordre n de la loi de Poisson de moyenne 1, d'où la formule*
$$B_n = \frac{1}{e}\sum_{k=1}^{\infty} \frac{k^n}{k!};$$
3. *Soit K un entier aléatoire valant k avec probabilité $k^n/(k!eB_n)$ pour tout $k \geq 0$. Sachant K, soient $C_1,\ldots,C_n$ des variables aléatoires i.i.d. de loi uniforme sur $\{1,\ldots,K\}$ (à interpréter comme des couleurs). Soit P la partition aléatoire de $\{1,\ldots,n\}$ obtenue en décidant que i,j sont dans le même bloc ssi $C_i = C_j$. Montrer que P suit la loi uniforme sur Π_n.*

Solution.

1. Il faut interpréter k comme le nombre d'éléments qui ne sont pas dans le bloc de $n+1$;
2. La formule de récurrence des nombres de Bell donne $G(X) = \exp(\exp(X) - 1)$ (produit de Cauchy des séries). On reconnaît la transformée de Laplace de la loi de Poisson de moyenne 1. Les nombres de Bell sont donc les moments de cette loi, d'où la formule (de Dobinski) ;
3. Il suffit d'observer que pour tout $p \in \Pi_n$, en notant b son nombre de blocs,
$$\mathbb{P}(P = p) = \sum_{k=b}^{\infty} \mathbb{P}(P = p | K = k)\mathbb{P}(K = k) = \sum_{k=b}^{\infty} \frac{k(k-1)\cdots(k-b+1)}{k^n}\frac{k^n}{k!eB_n} = \frac{1}{B_n}.$$

La loi de K est bien définie grâce à la formule de Dobinski. Note : la décomposition en cycles d'une permutation aléatoire uniforme fournit une partition aléatoire, qui n'est pas uniforme.

Exercice 1.10 (Loi des petits nombres – Bernoulli – Poisson – Inégalité de Le Cam). *Si X et Y sont des v.a. sur $\mathbb{N}$, on pose $d(X,Y) = \sum_{k=0}^{\infty} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)|$ (ne dépend que des lois).*

1. *Soient $X_1,\ldots,X_n$ (et $Y_1,\ldots,Y_n$) des v.a. indépendantes sur $\mathbb{N}$. Montrer que*
$$d(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) \leq d(X_1,Y_1) + \cdots + d(X_n,Y_n);$$
2. *Montrer que si X (et Y) est de Bernoulli (et de Poisson) de moyenne p alors $d(X,Y) \leq 2p^2$;*
3. *En déduire que si $X_1,\ldots,X_n$ sont indépendantes de Bernoulli de moyennes $p_1,\ldots,p_n$ et si Y est de Poisson de moyenne $\lambda = p_1 + \cdots + p_n$ alors*
$$d(X_1 + \cdots + X_n, Y) \leq 2(p_1^2 + \cdots + p_n^2) \leq 2\lambda \max_{1\leq k\leq n} p_k;$$
Étudier le cas spécial où $p_1 = \cdots = p_n = \lambda/n$.

Solution.

1. Trivial pour $n = 1$. Pour $n = 2$, on observe tout d'abord que pour tous $0 \leq r \leq k$,
$$\begin{aligned}&\mathbb{P}(X_1 = r)\mathbb{P}(X_2 = k-r) - \mathbb{P}(Y_1 = r)\mathbb{P}(Y_2 = k-r)\\ &\quad = (\mathbb{P}(X_1 = r) - \mathbb{P}(Y_1 = r))\mathbb{P}(X_2 = k-r) - \mathbb{P}(Y_1 = r)(\mathbb{P}(Y_2 = k-r) - \mathbb{P}(X_2 = k-r))\end{aligned}$$
ce qui donne, pour tout $k \geq 0$,
$$\begin{aligned}&|\mathbb{P}(X_1 + X_2 = k) - \mathbb{P}(Y_1 + Y_2 = k)|\\ &\quad \leq \sum_{r=0}^{k} |\mathbb{P}(X_1 = r) - \mathbb{P}(Y_1 = r)|\mathbb{P}(X_2 = k-r) + |\mathbb{P}(Y_2 = k-r) - \mathbb{P}(X_2 = k-r)|\mathbb{P}(Y_1 = r)\end{aligned}$$
ce qui conduit au résultat pour $n = 2$ en sommant sur $k \geq 0$ et en commutant les sommes (qui sont à termes positifs). Pour tout $n \geq 2$, on passe de n à $n+1$ grâce au cas $n = 2$ car
$$d(X_1 + \cdots + X_{n+1}, Y_1 + \cdots + Y_{n+1}) \leq d(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) + d(X_{n+1}, Y_{n+1});$$

2. On a, par définition de $d(X,Y)$ et des lois de X et Y,

$$\begin{aligned}d(X,Y) &= |1-p-e^{-p}| + |p-e^{-p}p| + \sum_{k\geq 2} e^{-p}\frac{p^k}{k!}\\ &= (e^{-p}-(1-p)) + p(1-e^{-p}) + (1-e^{-p}(1+p))\\ &= 2p(1-e^{-p}) \leq 2p^2;\end{aligned}$$

3. Immédiat. Lorsque $p_1 = \cdots = p_n = p$ alors $X_1+\cdots+X_n \sim \mathrm{Binom}(n,p)$;

Note : on peut améliorer sans peine l'inégalité car par concavité de $t \mapsto 1-e^{-t}$,

$$d(X_1+\cdots+X_n,Y) \leq 2\sum_{k=1}^n p_i(1-e^{-p_i}) \leq 2(1-e^{-(p_1^2+\cdots+p_n^2)})(\leq 2(p_1^2+\cdots+p_n^2)).$$

Exercice 1.11 (Caractérisation des lois géométriques par absence de mémoire). *Montrer que pour toute v.a. X à valeurs dans $\{1,2,\ldots\}$, les propriétés suivantes sont équivalentes :*

1. *X suit une loi géométrique ;*
2. *La loi de $X-n$ sachant $\{X>n\}$ est identique à la loi de X pour tout entier $n \geq 0$.*

Solution. Si $G(n) = \mathbb{P}(X>n)$ la seconde propriété s'écrit $G(n)G(m) = G(n+m)$ pour tous $n,m \in \mathbb{N}$, et donc $G(n) = G(1)^n$ pour tout $n \in \mathbb{N}$, d'où X suit la loi géométrique de paramètre $1-G(1)$. Réciproquement, si X suit la loi géométrique de paramètre p alors

$$\mathbb{P}(X-n>m|X>n) = \frac{G(n+m)}{G(n)} = (1-p)^m$$

pour tous $n,m \in \mathbb{N}$, et donc la loi de $X-n$ sachant $\{X>n\}$ est égale à la loi de X.

Exercice 1.12 (Biais par la taille). *Soit p_k la fréquence des foyers de taille $k \in \mathbb{N}^*$ dans la population française. On a $\sum_{k\geq 1} p_k = 1$. Prenons à présent un français au hasard dans la population. La taille T du foyer auquel il appartient est aléatoire. Montrer que $\mathbb{P}(T=k) = \frac{k}{m}p_k$ pour tout $k \in \mathbb{N}^*$, où par définition $m := \sum_{k\geq 1} kp_k$ est la taille moyenne des foyers dans la population.*

Solution. La loi de T n'est pas la loi de départ, car les grands foyers sont sur-représentés tandis que les petits foyers sont sous-représentés. Ce phénomène est appelé *biais par la taille*[2] Il s'agit sans doute du biais d'échantillonnage le plus connu. Le biais est d'autant plus important que la taille du foyer diffère de la taille moyenne m.

On se place conditionnellement à la population, qui compte au total $\sum_{k\geq 1} N_k$ foyers où N_k désigne le nombre de foyers de taille k. La population compte au total $N_1 + 2N_2 + \cdots$ individus. On modélise le choix d'un français au hasard par le tirage d'un entier selon la loi uniforme sur l'intervalle $[\![1, N_1+2N_2+\cdots]\!]$. Il y a N_k foyers de taille k qui comptent au total kN_k individus. Par définition de la loi uniforme (formule « cas favorables sur cas totaux ») la probabilité que ce français appartiennent à un foyer de taille k est

$$\frac{kN_k}{N_1+2N_2+\cdots} = \frac{k}{m}p_k \quad \text{où} \quad p_k := \frac{N_k}{N_1+N_2+\cdots} \quad \text{et} \quad m := p_1+2p_2+\cdots.$$

Exercice 1.13 (Méthode von Neumann pour le débiaisage de pièce de monnaie). *Comment simuler une pièce de monnaie non biaisée à partir d'une pièce de monnaie biaisée ?*

Solution. Évidemment, on ne connait pas la probabilité $p \in [0,1]$ d'obtenir 1 (pile) avec la pièce biaisée. L'idée est la suivante : on la lance deux fois. Si on obtient 10 alors on décide 0 et si on obtient 01 alors on décide 1, tandis que si on obtient 00 ou 11 alors on recommence. Les résultats ainsi produits sont i.i.d. de loi de Bernoulli de paramètre $1/2$.

2. In English : $\sum_{k\geq 1} \frac{k}{m} p_k\delta_k$ is the size-biased law constructed from the initial law $\sum_{k\geq 1} p_k\delta_k$.

2 Variables aléatoires réelles, notion de densité de probabilité

Rappels de cours. Transfert, densité, fonction de répartition, moments, lois classiques.

Exercice 2.1 (Fonction de répartition, algorithme de simulation par inversion). *Déterminer la fonction de répartition des lois de Bernoulli, géométrique, uniforme, exponentielle, gaussienne. À quoi voit-on que la densité existe ou pas ? Comment simuler ces lois à partir de la fonction rand de la calculatrice/ordinateur ? Préciser la complexité de l'algorithme.*

Solution. La fonction de répartition des lois uniforme sur $[a,b]$, exponentielle d'intensité λ, et gaussienne $\mathcal{N}(0,1)$ sont données respectivement par $\frac{t-a}{b-a}\mathbf{1}_{t\in[a,b]}+\mathbf{1}_{t>b}$, $(1-e^{-\lambda t})\mathbf{1}_{t\geq 0}$, $t\mapsto(2\pi)^{-1/2}\int_{-\infty}^{t}e^{-x^2/2}\,dx$. On les simule en utlisant la méthode d'inversion de la fonction de répartition (sur le bon intervalle).

Les fonctions de répartitions sont constantes par morceaux, continue à droite avec limite à gauche, la position des sauts correspondent aux atomes, tandis que les amplitudes des sauts correspondent aux probabilités des atomes. Pour simuler une loi discrète d'atomes $p_1,p_2,\ldots$ on partitionne l'intervalle $[0,1]$ en $I_1=[0,p_1[, I_2=[p_1,p_1+p_2[,\ldots$ et on répond i si rand appartient à I_i. Comme rand simule une v.a.r. de loi uniforme U sur $[0,1]$, on a $\mathbb{P}(U\in I_i)=|I_i|=p_i$. Il s'agit d'une instance de la méthode d'inversion, basée sur une inverse généralisée. Il est utile d'ordonner les poids pour minimiser le nombre de tests.

Exercice 2.2 (Loi uniforme et de l'arc-sinus). *Soit X une v.a. uniforme sur $[0,1]$. Soit $Y:=\min(X,1-X)$ et $Z:=\max(X,1-X)$. Trouver les lois de Y, Z, Y/Z, YZ. Calculer $\mathbb{E}(Y)$, $\mathbb{E}(Y^2)$, $\mathbb{E}(Z)$, $\mathbb{E}(Z^2)$, $\mathbb{E}(YZ)$.*

Solution. La v.a. Y prend ses valeurs dans $[0,1/2]$ et pour tout $t\in[0,1/2]$,

$$F_Y(t)=\mathbb{P}(Y\leq t,X\leq 1/2)+\mathbb{P}(Y\leq t,X\geq 1/2)=\mathbb{P}(X\leq t)+\mathbb{P}(X\geq 1-t)=2t$$

donc Y suit la loi uniforme sur $[0,1/2]$. De même, on montre que Z suit la loi uniforme sur $[1/2,1]$. La v.a. Y/Z prend ses valeurs dans $[0,1]$ et comme $Y+Z=1$ on obtient, pour tout $t\in[0,1]$,

$$\mathbb{P}(Y/Z\leq t)=\mathbb{P}(1/Z-1\leq t)=\mathbb{P}(Z\geq 1/(1+t))=1-2\mathbf{1}_{[1/2,1]}(1/(1+t)).$$

Comme $Y+Z=1$, la v.a. $YZ=Z(1-Z)$ prend ses valeurs dans $[0,1/4]$, et pour tout $t\in[0,1]$,

$$\mathbb{P}(YZ>t)=\mathbb{P}((1-Z)Z>t)=\mathbb{P}\left(Z\in\left[\frac{1-\sqrt{1-4t}}{2},\frac{1+\sqrt{1-4t}}{2}\right]\right)=\mathbb{P}(2Z-1<\sqrt{1-4t})=\sqrt{1-4t}.$$

La densité de YZ est $f_{YZ}(t)=F'_{YZ}(t)=\frac{2}{\sqrt{1-4t}}\mathbf{1}_{[0,1/4]}(t)$. Il s'agit de la loi de l'arc-sinus sur l'intervalle $[0,1/4]$. On a bien sûr, dans un premier mouvement bien naturel, la formule

$$\mathbb{E}(YZ)=\int t f_{YZ}(t)\,dt=\int_0^{1/4}\frac{2t}{\sqrt{1-4t}}\,dt,$$

qui est calculable par changement de variable, mais mieux vaut recourir à l'astuce suivante :

$$\mathbb{E}(YZ)=\mathbb{E}(Y(1-Y))=2\int_0^{1/2}(t-t^2)\,dt=2[t^2/2-t^3/3]_0^{1/2}=1/4-1/12=1/6.$$

Exercice 2.3 (Saturation et détection). *Soit X une v.a.r. exponentielle de paramètre λ, et $a,\varepsilon\in\mathbb{R}$. Déterminer la loi des v.a.r. $Y:=\min(X,a)$ et $Z:=X\mathbf{1}_{X>\varepsilon}$. Ont-elles une densité ?*

Solution. Si $a\leq 0$ alors $\min(X,a)$ est une v.a.r. constante et égale à 0. Sa loi est la masse de Dirac en 0, notée δ_0. Sa fonction de répartition est $\mathbf{1}_{\mathbb{R}_+}$. Supposons à présent que $a>0$. On a alors $\min(X,a)\geq 0$ et pour tout $t\geq 0$,

$$F_{\min(X,a)}(t)=\mathbb{P}(\min(X,a)\leq t)=\mathbf{1}_{t\geq a}+\mathbb{P}(X\leq t)\mathbf{1}_{t<a}=\mathbf{1}_{t\geq a}+(1-e^{-\lambda t})\mathbf{1}_{t<a}.$$

Ainsi, $F_{\min(X,a)}$ a un unique saut en a d'amplitude $1-(1-e^{-\lambda a})=e^{-\lambda a}$. Comme elle n'est pas continue, elle ne possède pas de densité. La loi de $\min(X,a)$ est un mélange à deux composantes :

$$F_{\min(X,a)}=e^{-\lambda a}F_1+(1-e^{-\lambda a})F_2,$$

combinaison convexe de la fonction de répartition $F_1=\mathbf{1}_{[a,\infty[}$ de la v.a. constante et égale à a et de la fonction de répartition F_2 de la loi de densité $\frac{\lambda}{1-e^{-\lambda a}}e^{-\lambda x}\mathbf{1}_{[0,a]}(x)$. De manière générale, la fonction de répartition d'une v.a.r. quelconque Y s'écrit toujours sous la forme d'une combinaison convexe $F_Y = p_1F_1+p_2F_2+p_3F_3$ où $p_1, p_2, p_3 \in [0, 1]$ sont tels que $p_1+p_2+p_3 = 1$, où F_1 est la fonction de répartition d'une loi discrète (constante par morceaux, nombre dénombrable de valeurs), F_2 est la fonction de répartition d'une loi à densité (on dit qu'elle est absolument continue), et F_3 est continue, fonction de répartition d'une loi étrangère à la mesure de Lebesgue.

L'étude de la variable aléatoire $Z = X\mathbf{1}_{X>\varepsilon}$ est similaire mais l'atome est en 0.

Exercice 2.4 (Régularisation par bruitage, mélange de population)**.**

1. *Soit X une v.a. de loi de Rademacher $\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = -1)$ et ε une v.a. de loi gaussienne $\mathcal{N}(0, \sigma^2)$, indépendante de X. Montrer que $X + \varepsilon$ a une densité et calculer cette densité ;*
2. *Considérons une population répartie en deux sous-populations distinctes dont les proportions sont p et $1 - p$. On considère un caractère morphologique. On modélise la répartition du caractère morphologique dans les deux population par les lois gaussienne $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ respectivement. On tire un individu au hasard dans a population totale, quelle est la loi de son caractère ?*

Solution.

1. Pour tout intervalle $I \subset \mathbb{R}$,
$$\mathbb{P}(X + \varepsilon \in I) = (1 - p)\mathbb{P}(X \in I + 1) + p\mathbb{P}(X \in I - 1) = \int_I (1 - p)\gamma(x + 1) + p\gamma(x - 1)\,dx.$$
2. On modélise par Y où le couple (X, Y) vérifie $\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = -1)$ et $\mathcal{L}(Y|X = -1) = \mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{L}(Y|X = 1) = \mathcal{N}(m_2, \sigma_2^2)$. D'où $Y \sim (1 - p)\mathcal{N}(m_1, \sigma_1^2) + p\mathcal{N}(m_2, \sigma_2^2)$.

Exercice 2.5 (Lois log-normale, du chi-deux, et de Laplace)**.**

1. *Calculer la densité de e^X où X suis la loi normale $\mathcal{N}(m, \sigma^2)$;*
2. *Soit X de densité f_X, montrer que $Y = X^2$ a une densité f_Y. Expliciter le cas $X \sim \mathcal{N}(0, 1)$;*
3. *Soit X de densité $f_X(x) = \frac{\lambda}{2}e^{-\lambda|x|}$. Quelle est la loi de $Y = |X|$?*

Solution.

1. Pour toute $f : \mathbb{R} \to \mathbb{R}$ continue et à support compact, en posant $y = e^x$,
$$\mathbb{E}(f(e^X)) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} f(e^x)e^{-\frac{(x-m)^2}{2}}\,dx = \frac{1}{\sqrt{2\pi\sigma^2}}\int_0^{\infty} f(y)y^{-1}e^{-\frac{(\ln(x)-m)^2}{2}}\,dx$$
donc la densité de e^X est $\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln(x)-m)^2}{2\sigma^2}}\mathbf{1}_{\mathbb{R}_+}(x)$;
2. La v.a.r. X^2 est positive. Si $h : \mathbb{R}_+ \to \mathbb{R}$ est continue et à support compact alors,
$$\mathbb{E}(h(X^2)) = \int_0^{\infty} h(x^2)(f_X(x) + f_X(-x))\,dx \overset{y=x^2}{=} \int_0^{\infty} h(y)\frac{f_X(\sqrt{y}) + f_Y(-\sqrt{y})}{2\sqrt{y}}\,dy$$
et $Y = X^2$ a pour densité $f_Y(y) = (f_X(\sqrt{y}) + f_X(-\sqrt{y}))/(2\sqrt{y})$. Lorsque $X \sim \mathcal{N}(0, 1)$ on trouve $f_Y(y) = (2\pi ye^y)^{-1/2}\mathbf{1}_{\mathbb{R}_+}(y)$, qui est une loi $\Gamma(a = 1/2, \lambda = 1/2) = \chi^2(1)$.
3. Exponentielle de paramètre λ.

Exercice 2.6 (Lois de Cauchy)**.** *Soit X une variable aléatoire de Cauchy, de densité $(\pi(1 + x^2))^{-1}$.*

1. *Calculer et reconnaître la loi de $1/X$;*
2. *Pour quelles valeurs de $\alpha \in \mathbb{R}$ la variable aléatoire $|X|^\alpha$ est-elle intégrable ?*

Solution.

1. Pour tout $f : \mathbb{R} \to \mathbb{R}$ continue et à support compact, en posant $y = 1/x$ sur $\mathbb{R}_+$ et $\mathbb{R}_-$,
$$\mathbb{E}(f(1/X)) = \int f(1/x)\frac{1}{\pi(1 + x^2)}\,dx = \int f(y)\frac{1}{\pi(1 + 1/y^2)}\frac{dy}{y^2} = \int f(y)\frac{1}{\pi(y^2 + 1)}\,dy$$
et donc $1/X$ a la même loi que X ;
2. Comme $\mathbb{E}(|X|^\alpha) = \frac{2}{\pi}\int_0^\infty \frac{x^\alpha}{1+x^2}\,dx$ il vient $\mathbb{E}(|X|^\alpha) < \infty$ ssi $\alpha < 1$, et X n'a pas d'espérance !

Exercice 2.7 (Convolution et lois gamma)**.**

1. *Montrer que si f et g sont des densités alors $f * g$ est aussi une densité ;*
2. *Pour tout $a > -1$ et $\lambda > 0$, la loi $\Gamma(a,\lambda)$ a pour densité $\frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$. Montrer que si f et g sont les densités des lois $\Gamma(a,\lambda)$ et $\Gamma(b,\lambda)$ alors $f * g$ est la densité de la loi $\Gamma(a+b,\lambda)$. En déduire h^{*n} lorsque h est la densité de la loi exponentielle de paramètre λ ;*
3. *Calculer les moments de la loi $\Gamma(a,\lambda)$.*

Solution.

1. En utilisant le théorème de Fubini-Tonelli positif et un changement de variable, on a
$$\int (f*g)(x)\,dx = \int \left(\int f(y)g(x-y)\,dy\right) dx = \int \left(\int g(x-y)\,dx\right) f(y)\,dy = \int f(y)\,dy = 1;$$
2. On a, en concluant avec $\mathrm{Beta}(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$,
$$\begin{aligned}(f*g)(x) &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} \int y^{a-1}(x-y)^{b-1} e^{-\lambda y - \lambda(x-y)} \mathbf{1}_{\mathbb{R}_+}(y) \mathbf{1}_{\mathbb{R}_+}(x-y)\,dy \\ &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) \int_0^x y^{a-1}(x-y)^{b-1}\,dy \\ &\overset{u=y/x}{=} \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a+b-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) \int_0^1 u^{a-1}(1-u)^{b-1}\,du \\ &= \frac{\lambda^{a+b}}{\Gamma(a+b)} x^{a+b-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x);\end{aligned}$$
La loi exponentielle correspond à $\Gamma(1,\lambda)$ et donc $\mathrm{Exp}(\lambda)^{*n} = \Gamma(1,\lambda)^{*n} = \Gamma(n,\lambda)$;
3. Pour tout entier $p \geq 1$,
$$m_p = \frac{\lambda^a}{\Gamma(a)} \int x^p x^{a-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) = \frac{\lambda^a}{\Gamma(a)} \int x^{a+p-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) = \frac{\lambda^a}{\Gamma(a)} \frac{\Gamma(a+p)}{\lambda^{a+p}} = \frac{\prod_{k=0}^{p-1}(a+k)}{\lambda^p}.$$

Exercice 2.8 (Loi uniforme et jeu de pile ou face). *Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. de loi de Bernoulli de paramètre $p \in [0,1]$, et soit $Y = \sum_{n=1}^{\infty} 2^{-n} X_n$.*

1. *Montrer que si $p = 1/2$ alors Y est une v.a.r. de loi uniforme sur $[0,1]$;*
2. *Réciproquement, montrer que si U est une v.a.r. de loi uniforme sur $[0,1]$ alors les bits de son écriture en base 2 sont des v.a. i.i.d. de loi de Bernoulli de paramètre $1/2$;*
3. *Que se passe-t-il en base $b \geq 2$ avec la loi uniforme sur $\{0,\ldots,b-1\}$?*

Solution.

1. Notons que pour tout $n \geq 1$,
$$\sum_{k=n+1}^{\infty} 2^{-k} X_k \leq \sum_{k=n+1}^{\infty} 2^{-k} = 2^{-(n+1)} \frac{1}{1-1/2} = 2^{-n}.$$
Pour tout dyadique $t = \sum_{k=1}^{n} t_k 2^{-k}$ avec $n \geq 1$ et $t_1,\ldots,t_n \in \{0,1\}$,
$$\mathbb{P}(t < Y < t + 2^{-n}) = \mathbb{P}(X_1 = t_1, \ldots, X_n = t_n) = 2^{-n}.$$
Donc $F_Y(t) = t$ pour tout nombre dyadique et donc Y suit la loi uniforme sur $[0,1]$;
2. Il suffit d'utiliser la formule précédente !
3. Même chose avec $Y = \sum_{n=1}^{\infty} b^{-n} X_n$ et $(X_n)_{n\geq 1}$ i.i.d. de loi uniforme sur $\{0,\ldots,b-1\}$.

Note : il peut se produire un phénomène surprenant lorsque $b > 3$ et les X_n ne sont pas de loi uniforme sur $\{0,\ldots,b-1\}$: la loi de Y peut être continue et étrangère à la mesure de Lebesgue (dans ce cas F_Y est continue mais n'est pas à densité). Un exemple célèbre est fourni par l'escalier du diable, qui correspond à F_Y quand $b = 3$ et $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 2) = 1/2$ et $\mathbb{P}(X_1 = 1) = 0$.

Exercice 2.9 (Intégrabilité). *Établir que si X est une v.a.r. positive alors, dans $[0,\infty]$,*

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq x)\,dx = \int_0^\infty \mathbb{P}(X > x)\,dx.$$

Plus généralement, pour tout $p > 0$, établir que dans $[0,\infty]$,

$$\mathbb{E}(X^p) = p\int_0^\infty x^{p-1}\mathbb{P}(X \geq x)\,dx = p\int_0^\infty x^{p-1}\mathbb{P}(X > x)\,dx.$$

Solution. Le cas $p > 0$ est similaire au cas $p = 1$. La dernière égalité est due au fait que les fonctions $t \mapsto \mathbb{P}(X > t) = 1 - F_X(t)$ et $t \mapsto \mathbb{P}(X \geq t) = 1 - F_X^-(t)$ sont égales presque partout. Pour établir la première égalité, on observe que pour tout $x \geq 0$, on dispose de la formule déterministe

$$x = \int_0^\infty \mathbf{1}_{[0,x]}(t)\,dt.$$

En utilisant le théorème de Fubini-Tonelli positif, il vient l'égalité suivante dans $[0,\infty]$:

$$\mathbb{E}(X) = \mathbb{E}\Big(\int_0^\infty \mathbf{1}_{[0,X]}(t)\,dt\Big) = \int_0^\infty \mathbb{E}\big(\mathbf{1}_{[0,X]}(t)\big)\,dt = \int_0^\infty \mathbb{E}\big(\mathbf{1}_{[t,\infty[}(X)\big)\,dt = \int_0^\infty \mathbb{P}(X \geq t)\,dt.$$

Il est également possible de partir de $\mathbf{1}_{[0,x[}(t)$ pour obtenir $\mathbb{P}(X > t)$ en lieu et place de $\mathbb{P}(X \geq t)$.

Alternativement, lorsque X est intégrable et possède un densité f_X, on peut écrire, pour tout $r \geq 0$, par intégration par parties,

$$\mathbb{E}(X\mathbf{1}_{X\leq r}) = \int_0^r x\,f_X(x)\,dx = \int_0^r x\,(F_X - 1)'(x)\,dx = [x(F_X-1)(x)]_0^r + \int_0^r \mathbb{P}(X > x)\,dx,$$

et comme $r\mathbf{1}_{X>r} \leq X$ et X est intégrable, on a $r(1 - F_X(r)) = \mathbb{E}(r\mathbf{1}_{X>r}) \to 0$ par convergence dominée, ce qui donne finalement, par convergence monotone,

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x)\,dx.$$

Exercice 2.10 (Inégalité de Paley-Zygmund). *Montrer que si X est de carré intégrable alors*

$$\mathbb{P}(X > \theta\mathbb{E}(X)) \geq (1-\theta)^2\frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)} = (1-\theta)^2\frac{\mathbb{E}(X)^2}{\mathbb{E}(X)^2 + \mathrm{Var}(X)},$$

pour tout réel $\theta \in [0,1]$. Indication : $X \leq \theta\mathbb{E}(X) + X\mathbf{1}_{\{X>\theta\mathbb{E}(X)\}}$.

Solution. Inégalité de Cauchy-Schwarz.

Exercice 2.11 (Caractérisation de la loi exponentielle par l'absence de mémoire). *Montrer que pour toute v.a.r. X positive telle que $\mathbb{P}(X > 0) > 0$, les propriétés suivantes sont équivalentes :*

1. *$\mathcal{L}(X)$ est une loi exponentielle ;*
2. *La loi de $X - t$ sachant $\{X > t\}$ est égale à la loi de X, pour tout $t \geq 0$;*
3. *Est-il raisonnable de modéliser une durée de vie par une loi exponentielle ?*

Solution. Si X suit la loi exponentielle de paramètre λ alors pour tous $s,t \geq 0$,

$$\mathbb{P}(X - t > s | X > t) = \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > t)} = e^{-\lambda s}$$

et donc $X - t$ sachant $\{X > t\}$ suit la même loi que X. Réciproquement, si $G(t) = \mathbb{P}(X > t)$, alors la seconde propriété s'écrit $G(t)G(s) = G(t+s)$ pour tous $s,t \geq 0$. Par continuité inférieure, $G(\varepsilon) = \mathbb{P}(X > \varepsilon) \to \mathbb{P}(X > 0) > 0$ lorsque $\varepsilon \to 0^+$. Ainsi, $G(\varepsilon) > 0$ pour $\varepsilon > 0$ assez petit. Si $t > 0$ alors $t \leq n\varepsilon$ avec $n \in \mathbb{N}$ et donc $G(t) = \mathbb{P}(X > t) \geq \mathbb{P}(X > n\varepsilon) = G(\varepsilon)^n$. Ainsi, $G(t) > 0$ pour tout $t \geq 0$. D'autre part, les solutions non identiquement nulles de l'équation fonctionnelle $G(t+s) = G(t)G(s)$ pour tous $s,t \in \mathbb{R}_+$ sont de la forme $G(t) = G(1)^t$ (considérer d'abord les t rationnels puis utiliser la décroissance de G lorsque $t \in \mathbb{R}_+$).

L'absence de mémoire des lois exponentielles les rend assez peu réalistes pour modéliser des durées de vie car elles ne tiennent pas compte du phénomène de rodage et de vieillissement. Il est plus raisonnable d'utiliser par exemple des lois de Weibull, dont le λ varie en quelque sorte.

Exercice 2.12 (Simulation par la méthode d'inversion)**.**

1. *Si U suit une loi uniforme sur $[0,1]$ quelle est la loi de $X = \ln(1/U)$ et $Y = \tan(\pi(U - 1/2))$?*
2. Méthode de simulation par inversion. *On considère une fonction de répartition F sur $\mathbb{R}$ dont on connaît explicitement l'inverse généralisée*
$$p \in [0,1] \mapsto F^{-1}(p) = \inf\{x;\, F(x) \geq p\} \in \overline{\mathbb{R}}.$$
Si U est une v.a.r. de loi uniforme sur $[0,1]$, montrer que la v.a. $X := F^{-1}(U)$ suit la loi de fonction de répartition F. Retrouver les résultats de la première question, et en déduire plus généralement une méthode de simulation de v.a.r. (discuter le cas des v.a.r. discrètes).
3. *Qu'elle est la loi de $F(X)$ lorsque F est continue et lorsque F ne l'est pas ?*

Solution.

1. On a $F_X(t) = \mathbb{P}(\ln(1/U) \leq t) = \mathbb{P}(1/U \leq e^t) = \mathbb{P}(U \geq e^{-t}) = 1 - e^{-t}$ donc $X \sim \mathrm{Exp}(\lambda)$. De même, $\mathbb{P}(Y \leq t) = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$ et Y suit la loi de Cauchy ;
2. Comme $\lim_{x\to\infty} F_X(x) = 1$ et $\lim_{x\to-\infty} F_X(x) = 0$ il vient $F_X^{-1}(p) \in \mathbb{R}$ pour tout $0 < p < 1$. Si $F_X(x) < p$ alors $F_X^{-1}(p) > x$, et comme F_X est continue à droite,
$$F_X(F_X^{-1}(p)) \geq p.$$
Donc $F_X(x) \geq p$ ssi $F_X^{-1}(p) \leq x$ pour tous x et $0 < p < 1$. Comme $\mathbb{P}(0 < U < 1) = 1$,
$$\mathbb{P}(F_X^{-1}(U) \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x)$$
pour tout $x \in \mathbb{R}$ ce qui montre que $F_X^{-1}(U)$ et X ont même loi. Le fait que $F_X^{-1}(p)$ puisse prendre les valeurs $\pm\infty$ pour $p = 0$ et $p = 1$ n'a pas d'importance car
$$\mathbb{P}(U = 0) = \mathbb{P}(U = 1) = 0.$$
D'autre part, le fait que $F_X^{-1}(U)$ soit une v.a.r. découle du fait que
$$\{F_X^{-1}(U) \leq x\} \cap \{0 < U < 1\} = \{U \leq F_X(x)\} \cap \{0 < U < 1\}.$$
Note : dans le cas des v.a.r. discrètes, on retrouve la méthode de la partition de $[0,1]$.
Note : on sait que $F_X(F_X^{-1}(p)) \geq p$. Comme $F_X(x) < p$ pour $F_X^{-1}(p) > x$, il en découle que si F_X est continue en $x \in \mathbb{R}$ alors $F_X^{-1}(F_X(x)) \leq x$ et donc $F_X^{-1}(F_X(x)) = x$. En particulier, si F_X est continue et strictement croissante alors c'est une bijection de $\mathbb{R}$ sur $]0,1[$ et F_X^{-1} est sa fonction réciproque. On peut aussi établir que si F_X est continue alors $F_X(X)$ suit la loi uniforme sur $[0,1]$.

Exercice 2.13 (Loi de l'arc-sinus)**.** *Donner un sens mathématique à la loi uniforme sur le cercle unité du plan $\{z \in \mathbb{C} : |z| = 1\}$. Si Z est une v.a. suivant cette loi, déterminer la loi de $|Z|$, de $\mathfrak{Re}Z$, et de $\mathfrak{Im}z$.*

Solution. On peut identifier le cercle unité à l'intervalle $[0, 2\pi[$, ce qui revient à identifier Z et $\mathfrak{Arg}(Z)$, et poser $\mathbb{P}(Z \in I) := \mathbb{P}(\mathfrak{Arg}(Z) \in I) = \frac{|I|}{2\pi}$ pour tout intervalle $I \subset [0, 2\pi[$. La loi de $|Z|$ est la masse de Dirac en 1. Les v.a.r. $\mathfrak{Re}Z$ et $\mathfrak{Im}z$ ont même loi, car la loi de Z est invariante par rotation, i.e. $e^{i\theta}Z$ et Z ont même loi quelque soit θ. Pour tout intervalle $J \subset [-1,1]$, on a
$$\mathbb{P}(\mathfrak{Re}Z \in J) = \mathbb{P}(\cos(\mathfrak{Arg}(Z)) \in J) = 2\mathbb{P}(\mathfrak{Arg}(Z) \in \arccos(J), 0 \leq \mathfrak{Arg}(Z) \leq \pi) = 2\frac{|\arccos(J)|}{2\pi}.$$
En prenant $J = [-1, x]$ avec $x \in [-1,1]$ il vient $|\arccos(J)| = |[\arccos(x), \pi]| = \pi - \arccos(x)$ et donc $\mathbb{P}(\mathfrak{Re}Z \in J) = 1 - \frac{\arccos(x)}{\pi}$. Comme $\arcsin(x) + \arccos(x) = \frac{\pi}{2}$ il vient $\mathbb{P}(\mathfrak{Re}Z \in J) = \frac{1}{2} + \frac{\arcsin(x)}{\pi}$. C'est la fonction de répartition de $\mathfrak{Re}Z$. Ainsi, la densité de $\mathfrak{Re}Z$ est $x \mapsto \frac{\mathbf{1}_{[-1,1]}(x)}{\pi\sqrt{1-x^2}}$ (loi de l'arc-sinus).

Exercice 2.14 (Entropie)**.** *Soit X une v.a.r. de densité f. Son entropie de Boltzmann-Shannon est définie par $H(X) := -\int f(x)\ln(f(x))\,dx$ si l'intégrale converge, et $H(X) = +\infty$ sinon.*

1. *Calculer l'entropie d'une v.a.r. X de loi normale $\mathcal{N}(m, \sigma^2)$;*
2. *Montrer que si Y est une v.a.r. de même variance que X alors $H(Y) \leq H(X)$;*
3. *Calculer l'entropie d'une v.a.r. X de loi uniforme sur $[0,1]$;*
4. *Montrer que si Y est une v.a.r. de même support que X alors $H(Y) \leq H(X)$;*

5. *Calculer l'entropie d'une v.a.r. X de loi exponentielle de paramètre λ ;*
6. *Montrer que si Y est une v.a.r. sur $\mathbb{R}_+$ de même moyenne que X alors $H(Y) \leq H(X)$;*
7. *Montrer que si X est une v.a.r. de densité $f(x) = e^{-\beta V}$ avec $\beta > 0$ et $V : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$, et si Y est une v.a.r. telle que $\mathbb{E}(V(Y)) = \mathbb{E}(V(X))$ alors $H(Y) \leq H(X)$.*

Solution. Remarque : $H(aX+b) = H(aX) = -\int(\ln(f(x/a)) - \ln(a))f(x/a)/a\,dx = H(X) + \ln(a)$.

1. $H(\mathcal{N}(m,\sigma^2)) = H(\mathcal{N}(0,1)) + \ln(\sigma)$ et $H(\mathcal{N}(0,1)) = \ln\sqrt{2\pi} + \frac{1}{2} = \frac{1}{2}\ln(2\pi e)$.
2. Comme $\ln(f_X(x)) = a + bx^2$, son espérance est identique pour X et Y, d'où

$$\begin{aligned}
H(X) - H(Y) &= -\int \ln(f_X(x)) f_X(x)\,dx + \int \ln(f_Y(x)) f_Y(x)\,dx \\
&= -\int \ln(f_X(x)) f_Y(x)\,dx + \int \ln(f_Y(x)) f_Y(x)\,dx \\
&= -\int_{\{f_Y>0\}} \ln(f_X(x)/f_Y(x))\, f_Y(x)\,dx \\
&\geq -\ln \int_{\{f_Y>0\}} f_X(x)\,dx \\
&\geq -\ln \int f_X(x)\,dx = 0.
\end{aligned}$$

   où l'avant dernière inégalité provient de l'inégalité de Jensen pour $-\ln$. Cas d'égalité...
3. $H(X) = \int_0^\infty (\lambda x - \ln(\lambda))\lambda e^{-\lambda x}\,dx = 1 - \ln(\lambda) = H(\mathrm{Exp}(1)) + \ln(\frac{1}{\lambda})$.
4. Comme $\ln(f_X(x)) = a+bx$, son espérance est identique pour X et Y, d'où un raisonnement identique au cas gaussien.
5. Comme $\ln(f_X(x)) = -\beta V(x)$, son espérance est identique pour X et Y, d'où un raisonnement identique au cas gaussien.

Exercice 2.15 (Calculs de moments)**.**

1. *Montrer que si X suit la loi exponentielle de paramètre $\lambda > 0$ alors $\mathbb{E}(X^n) = \frac{n!}{\lambda^n}$;*
2. *Montrer que si X suit la loi $\mathcal{N}(0,1)$ alors $\mathbb{E}(X^{2n}) = \prod_{k=1}^n (2k-1) = \frac{(2n)!}{2^n n!}$.*

Solution. Intégration par parties.

Exercice 2.16 (Caractérisation par les moments)**.** *Montrer que si deux v.a.r.* bornées *X et Y ont les mêmes moments, c'est-à-dire que $\mathbb{E}(X^n) = \mathbb{E}(Y^n)$ pour tout $n \in \mathbb{N}$, alors X et Y ont même loi. Indication : utiliser le théorème de Weierstrass de densité des polynômes.*

Solution. Soit K un compact de $\mathbb{R}$ tel que $\mathbb{P}(X \in K) = \mathbb{P}(Y \in K) = 1$. Pour toute fonction continue $f : \mathbb{R} \to \mathbb{R}$ à support dans K, et tout $\varepsilon > 0$, il existe un polynôme P tel que $\|P - f\|_\infty \leq \varepsilon$ (théorème de Weierstrass), et donc $|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq |\mathbb{E}(P(X)) - \mathbb{E}(P(Y))| + 2\varepsilon = 2\varepsilon$.

Exercice 2.17 (Analycité de la transformée de Fourier et moments)**.** *Soit X une variable aléatoire possédant des moments de tout ordre. Pour tout réel $t \in \mathbb{R}$ et entier $n \geq 0$ on pose $\varphi(t) := \mathbb{E}(e^{itX})$ et $\kappa_n := \mathbb{E}(X^n)$. Montrer que les propositions suivantes sont équivalentes :*

1. *φ est analytique sur un voisinage de 0*
2. *φ est analytique sur $\mathbb{R}$*
3. $\overline{\lim}_n \left(\frac{1}{n!}|\kappa_n|\right)^{\frac{1}{n}} < \infty$.

Si ces conditions (suffisantes) sont vérifiées alors la loi de X est caractérisée par ses moments. La dernière condition montre qu'une loi à support compact est toujours caractérisée par ses moments, et que les lois exponentielles et normales sont toutes caractérisées par leurs moments.

Solution. La formule de Stirling $n! = \sqrt{2\pi n}(n/e)^n(1+\mathcal{O}(1/n))$ donne $(1/n!)^{1/n} = \mathcal{O}(1/n)$ et donc si $\overline{\lim}_{n\to\infty} \frac{1}{n}|\kappa_n|^{1/n}$ alors μ est caractérisée par ses moments.

Pour tout n, on a $\mathbb{E}(|X|^n) < \infty$ et donc φ est n fois dérivable sur $\mathbb{R}$. De plus, $\varphi^{(n)}$ est continue sur $\mathbb{R}$ et pour tout $t \in \mathbb{R}$,

$$\varphi^{(n)}(t) = \mathbb{E}((ix)^n e^{itx}).$$

En particulier, $\varphi^{(n)}(0) = i^n \kappa_n$, et la série de Taylor de φ en 0 est déterminée par la suite $(\kappa_n)_{n\geq 1}$. Le rayon de convergence r de la série entière $\sum_n a_n z^n$ associée à la suite de nombres complexes $(a_n)_{n\geq 0}$ est donné par la formule de Hadamard $r^{-1} = \overline{\lim}_n |a_n|^{\frac{1}{n}}$. Ainsi, 1$\Leftrightarrow$ 3 (prendre $a_n = i^n\kappa_n/n!$). D'autre part, comme pour tout $n \in \mathbb{N}$, $s,t \in \mathbb{R}$,

$$\left|e^{isx}\left(e^{itx} - 1 - \frac{itx}{1!} - \cdots - \frac{(itx)^{n-1}}{(n-1)!}\right)\right| \leq \frac{|tx|^n}{n!},$$

on a pour tout $n \in \mathbb{N}$ pair et tous $s,t \in \mathbb{R}$,

$$\left|\varphi(s+t) - \varphi(s) - \frac{t}{1!}\varphi'(s) - \cdots - \frac{t^{n-1}}{(n-1)!}\varphi^{(n-1)}(s)\right| \leq \kappa_n \frac{|t|^n}{n!},$$

qui montre que 3 $\Rightarrow$ 2. Comme 2 $\Rightarrow$ 1, on a bien équivalence de 1$\Leftrightarrow$2$\Leftrightarrow$3. Si ces propriétés ont lieu, alors les arguments précédents donnent un $r > 0$ tel que φ est développable en série entière en tout $x \in \mathbb{R}$ avec un rayon de convergence $\geq r$. De proche en proche, on obtient que φ est caractérisée par ses dérivées en zéro. Si μ est à support compact, alors $\sup_n |\kappa_n|^{\frac{1}{n}} < \infty$ et donc 3 a lieu (via Stirling $n! = (n/e)^n\sqrt{2\pi n}(1+\mathcal{O}(1/n))$).

Exercice 2.18 (Contre exemple de Heyde pour la loi log-normale)**.** *Soit f la densité de $X := e^Y$ où Y suit la loi $\mathcal{N}(0,1)$. Pour tout réel $-1 < a < 1$ fixé, on définit $f_a : \mathbb{R}_+ \to \mathbb{R}_+$ par $f_a(x) := f(x)(1 + a\sin(2\pi\ln(x)))$. Montrer que f_a est une densité possédant les mêmes moments que f, et en déduire que la loi log-normale n'est pas caractérisée par ses moments.*

Solution. On a $f_a \geq 0$. On a $f(y) = \frac{1}{\sqrt{2\pi}y}e^{-\frac{\ln(y)^2}{2}}$. Pour tout $n \in \mathbb{N}$, il vient, en posant $\ln(x) = y + n$,

$$\begin{aligned}\int_0^\infty x^n \sin(2\pi\ln(x)) f(x)\,dx &= \int_{-\infty}^{+\infty} e^{(n+1)(y+n)}\sin(2\pi y) f(e^{y+n})\,dy \\ &= \int_{-\infty}^{+\infty} e^{(n+1)(y+n)}\sin(2\pi y)\frac{1}{\sqrt{2\pi}e^{y+n}}e^{-\frac{(y+n)^2}{2}}\,dy \\ &= e^{\frac{n^2}{2}}\int_{-\infty}^{+\infty}\sin(2\pi y)\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy \\ &= 0\end{aligned}$$

par (im)parité. Ceci montre que f_a est une densité, et qu'elle a les mêmes moments que f.

3 Vecteurs aléatoires à densité

Rappels de cours. Transfert ; densité ; fonction de répartition ; marginales ; indépendance ; lois conditionnelles ; lois gaussiennes.

Exercice 3.1 (Lois exponentielles en série et en parallèle)**.** *On considère n composants pouvant être montés en série ou en parallèle. On modélise leur durées de vie par des v.a.r. $T_1, \ldots, T_n$ indépendantes de loi exponentielle de paramètres $\lambda_1, \ldots, \lambda_n$.*

1. *On suppose que le système est monté en parallèle : il fonctionne ssi au moins un des composants fonctionne. Déterminer la loi de la durée de vie S du système ;*
2. *On suppose que le système est monté en série : il fonctionne ssi tous les composants fonctionnent. Déterminer la loi de la durée de vie S du système ;*

Solution. Modéliser des durées de vie par des lois exponentielles est trop idéal car l'absence de mémoire de ces lois ne permet pas de tenir compte des phénomène de rodage et de vieillissement.

1. Montage en parallèle : $S = \max(T_1, \ldots, T_n)$ et donc $F_S(s) = \prod_{i=1}^n (1 - e^{-\lambda_i s})$. Cette loi n'est pas exponentielle et a donc de la mémoire (vieillissement) ;
2. Montage en série : $S = \min(T_1, \ldots, T_n)$ et donc $F_S(s) = 1 - \prod_{i=1}^n e^{-\lambda_i s} = 1 - e^{-s \sum_{i=1}^n \lambda_i}$. Ainsi S suit la loi exponentielle de paramètre $\sum_{i=1}^n \lambda_i$, qui n'a pas de mémoire !

Exercice 3.2 (Couples de variables aléatoires, densité, et indépendance)**.**

1. *Soit X une v.a.r. de densité f_X. Le couple (X, X) a-t-il une densité ?*
2. *Donner un exemple de v.a.r. X et Y sans densités et telles que $X + Y$ possède une densité ;*
3. *Soit (X, Y) un couple de v.a.r. de loi uniforme sur le disque $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Montrer que X et Y ont même loi, et calculer leur densité. Sont-elles indépendantes ?*
4. *Déterminer la constante c pour que la fonction $f(x, y) = c(x^2 + y^2) \exp(-\frac{x^2+y^2}{2})$ soit une densité sur $\mathbb{R}^2$. Si (X, Y) est un couple qui suit cette densité, déterminer ses lois marginales, ainsi que la covariance de X et Y. Les variables X et Y sont-elles indépendantes ?*

Solution.

1. Non car $\mathbb{P}((X, X) \in \Delta) = 1$ où $\Delta = \{(x, x) : x \in \mathbb{R}\}$ (de mesure de Lebesgue nulle) ;
2. Exemple : $(X, Y) = (U\mathbf{1}_{U \leq 1/2}, U\mathbf{1}_{U > 1/2})$ avec $U \sim \mathrm{Unif}([0, 1])$ (X et Y sont dépendantes) ;
3. Si $h : \mathbb{R} \to \mathbb{R}$ continue et bornée, $\mathbb{E}(h(X)) = \iint h(x) \frac{\mathbf{1}_{x^2+y^2 \leq 1}}{\pi} \, dxdy = \int_{-1}^{1} h(x) \frac{2\sqrt{1-x^2}}{\pi} \, dx$ et donc X suit la loi du demi-cercle sur $[-1, 1]$. Idem pour Y par symétrie. Les v.a.r. X et Y ne sont pas indépendantes car $X^2 + Y^2 \leq 1$ et cela se confirme : $f_{(X,Y)} \neq f_X \otimes f_Y$.
4. $\int f(x, y) \, dxdy = 2c\sqrt{2\pi} \int x^2 e^{-\frac{x^2}{2}} \, dx = 4\pi c$ donc $c = \frac{1}{4\pi}$. Par symétrie, X et Y ont même loi de densité $x \mapsto \int f(x, y) \, dy = \frac{1+x^2}{2\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. D'où $f_{(X,Y)} \neq f_X \otimes f_Y$, donc X et Y sont dépendantes. Néanmoins $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) = \int xy f_{(X,Y)}(x, y) dxdy = 0$ et donc X et Y ne sont pas correllées.

Exercice 3.3 (Lois uniformes)**.**

1. *Soit (X, Y) un couple de variables aléatoires réelles. Montrer que (X, Y) suit la loi uniforme sur le carré $[0, 1]^2$ si et seulement si X et Y sont indépendantes et de loi uniforme sur $[0, 1]$;*
2. *On coupe un bâton au hasard en trois morceaux en utilisant deux v.a. indépendantes et uniformes sur $[0, 1]$ pour déterminer les points de coupe. Vérifier que les longueurs des trois morceaux ainsi obtenus sont des v.a. de même loi. Sont-elles indépendantes ? Quelle est la probabilité de pouvoir fabriquer un triangle avec ces trois morceaux ?*
3. *Deux amis se donnent rendez vous entre 12h et 13h, et arrivent indépendamment uniformément entre ces deux horaires. Calculer le temps moyen d'attente du premier arrivé.*

Solution.

1. $\mathbf{1}_{[0,1]^2} = \left(\mathbf{1}_{[0,1]}\right)^{\otimes 2}$;

2. $[0,1]=[0,U]\cup[U,V]\cup[V,1]$, $(U,V):=(\min(X,Y),\max(X,Y))$. Pour tout $t\in[0,1]$ (dessin)

$$\mathbb{P}(\min(X,Y)\leq t)=\iint_{[0,1]^2}\mathbf{1}_{\min(x,y)\leq t}\,dxdy=1-(1-t)^2$$

(on peut alternativement utiliser l'indépendance de X et Y) et

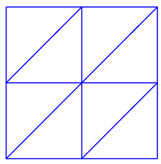
$$\mathbb{P}(1-\max(X,Y)\leq t)=\iint_{[0,1]^2}\mathbf{1}_{\max(x,y)\geq 1-t}\,dxdy=1-(1-t)^2$$

et comme $\max(x,y)-\min(x,y)=|x-y|$,

$$\mathbb{P}(\max(X,Y)-\min(X,Y)\leq t)=\iint_{[0,1]^2}\mathbf{1}_{|x-y|\leq t}\,dxdy=1-(1-t)^2.$$

Ces trois longueurs $a:=\min(X,Y)$, $b:=\max(X,Y)-\min(X,Y)$, et $c:=1-\max(X,Y)$ sont de même loi (Bêta), mais ne sont pas indépendantes (leur somme fait 1). De manière générale, un dessin montre qu'on peut former un triangle avec trois bâtons de longueurs $(\min=)\ell_1\leq\ell_2\leq\ell_3(=\max)$ lorsque $\ell_3-\ell_1\leq\ell_2$. Si nous savons que $\ell_1+\ell_2+\ell_3=1$ alors la condition devient $\ell_3\leq\frac{1}{2}$. On a ici $\ell_3=\max(a,b,c)$. Pour tout $t\in[0,1]$, on a

$$\begin{aligned}\mathbb{P}(\max(a,b,c)\leq t)&=\mathbb{P}(a\leq t,b\leq t,c\leq t)\\&=\mathbb{P}(\min(X,Y)\leq t,|X-Y|\leq t,\max(X,Y)\geq 1-t)\\&=\iint_{[0,1]^2}\mathbf{1}_{\min(x,y)\leq t,|x-y|\leq t,\max(x,y)\geq 1-t}\,dxdy.\end{aligned}$$



Pour $t=\frac{1}{2}$ on trouve $\frac{1}{4}$ (dessin, papillon). On peut donc former un triangle dans 25% des cas !

3. On a $\mathbb{E}(\max(X,Y)-\min(X,Y))=\int_0^1 t(1-(1-t)^2)'\,dt=\frac{1}{3}$ (d'heure, soit vingt minutes). Astuce pour aller plus vite, comme $1=a+b+c$ et que a,b,c ont même loi il vient $\mathbb{E}(a)=\mathbb{E}(b)=\mathbb{E}(c)=1/3$.

Exercice 3.4 (Lois marginales et loi de couple). *Soit (X,Y) et (X',Y') des couples de densités*

$$f_{(X,Y)}(x,y)=\frac{1}{4}(1+xy)\mathbf{1}_{[-1,1]^2}(x,y).\quad\textit{et}\quad f_{(X',Y')}(x',y')=\frac{1}{4}\mathbf{1}_{[-1,1]^2}(x',y').$$

1. *Vérifier qu'il s'agit bien de densités ;*
2. *Montrer que (X,Y) et (X',Y') ne suivent pas la même loi ;*
3. *Montrer que (X,Y) et (X',Y') ont les mêmes lois marginales, c'est-à-dire que X et X' sont de même loi, et que Y et Y' sont de même loi (en fait X,X',Y,Y' sont de même loi !).*

Solution.

1. $\iint_{[-1,1]^2}xy\,dxdy=0$ par (im)parité !
2. On a par exemple $\mathbb{E}(X'Y')=0$ mais

$$\mathbb{E}(XY)=\mathbb{E}(X'Y')+\frac{1}{4}\iint_{[-1,1]^2}x^2y^2\,dxdy=0+\mathbb{E}(X'^2Y'^2)=\mathbb{E}(X'^2)^2=\frac{1}{9}\neq 0.$$

3. X' et Y' i.i.d. uniforme sur $[-1,1]$. X et Y sont de même loi par symétrie. Densité de X :

$$\frac{1}{2}\mathbf{1}_{[-1,1]}(x)+\frac{x}{4}\mathbf{1}_{[-1,1]}(x)\underbrace{\int_{-1}^{1}y\,dy}_{=0}.$$

Exercice 3.5 (Statistique d'ordre et permutations aléatoires). *Montrer que si $X_1,\ldots,X_n$ sont des v.a.r. i.i.d. de densité f, alors p.s. il existe une unique permutation aléatoire σ telle que $X_{\sigma(1)}\leq\cdots\leq X_{\sigma(n)}$, et que σ suit la loi uniforme sur le groupe symétrique Σ_n.*

Solution. Le vecteur aléatoire $X := (X_1, \ldots, X_n)$ a pour densité $x \in \mathbb{R}^n \mapsto \prod_{1\leq i\leq n} f(x_i)$, qui est symétrique (invariante par permutation des coordonnées). Par conséquent, les vecteurs aléatoires $(X_1, \ldots, X_n)$ et $(X_{\sigma'(1)}, \ldots, X_{\sigma'(n)})$ ont même loi pour tout $\sigma' \in \Sigma_n$ (on dit que la loi de X est échangeable). D'autre part, comme X a une densité, p.s. ses coordonnées sont deux à deux distinctes et donc σ est définie de manière unique. À présent, la permutation aléatoire σ vérifie

$$\mathbb{P}(\sigma = \sigma') = \mathbb{P}(X_{\sigma'(1)} \leq \cdots \leq X_{\sigma'(n)}) = \mathbb{P}(X_1 \leq \cdots \leq X_n),$$

et cette quantité ne dépend pas de σ', et donc σ est uniforme sur Σ_n : $\mathbb{P}(\sigma = \sigma') = \frac{1}{\mathrm{card}(\Sigma_n)} = \frac{1}{n!}$.

4 Vecteurs aléatoires à densité (bis)

Rappels de cours. Caractérisation de la loi et de la densité ; formule du changement de variable.

Exercice 4.1 (Des lois normales à la loi de Cauchy)**.** *Montrer que si X et Y sont des v.a. indépendantes de loi normale $\mathcal{N}(0,1)$ alors $Z = X/Y$ est de loi de Cauchy de densité $f_Z(z) = (\pi(1+z^2))^{-1}$.*

Solution. En posant $z = x/y$ on a $(x,y) = (zy,y)$ qui est un difféomorphisme de $]0,\infty[^2$ dans lui même. On a $dxdy = \left|\det\begin{pmatrix} y & z \\ 0 & 1\end{pmatrix}\right| dxdu = ydzdy$ (on dérive les anciennes variables (x,y) en fonction des nouvelles variables (z,y), penser aux coordonnées polaires !) et pour toute $h : \mathbb{R}^2 \to \mathbb{R}$ continue bornée paire, grâce au changement de variable et au théorème de Fubini-Tonelli,

$$\begin{aligned}
\mathbb{E}(h(Z)) &= \frac{4}{2\pi}\int_0^\infty\int_0^\infty h(x/y)e^{-\frac{x^2+y^2}{2}}\,dxdy \\
&= \frac{2}{\pi}\int_0^\infty\int_0^\infty h(z)e^{-\frac{(zy)^2+y^2}{2}}y\,dzdy \\
&= \frac{2}{\pi}\int_0^\infty h(z)\Big(\int_0^\infty ye^{-\frac{(1+z^2)y^2}{2}}\,dy\Big)dz \\
&= \frac{2}{\pi}\int_0^\infty h(z)\frac{1}{1+z^2}\,dz \\
&= \frac{1}{\pi}\int_{-\infty}^\infty \frac{h(z)}{1+z^2}\,dz.
\end{aligned}$$

Note : si h est impaire alors l'intégrale est nulle par symétrie de la densité, et donc, comme toute fonction est somme d'une fonction paire et d'une impaire, on a une formule pour tout h.

Exercice 4.2 (Loi exponentielle)**.** *Soit $E_1,\ldots,E_n$ des v.a.r. indépendantes avec $E_k \sim \mathrm{Exp}(\lambda_k)$, et soit $\min(E_1,\ldots,E_n) = E_{(1)} \leq \cdots \leq E_{(n)} = \max(E_1,\ldots,E_n)$ la suite réordonnée.*

1. *Établir que $M := E_{(1)} = \min(E_1,\ldots,E_n) \overset{\text{loi}}{=} \frac{\lambda_1}{\lambda_1+\cdots+\lambda_n}E_1 \sim \mathrm{Exp}(\lambda_1+\cdots+\lambda_n)$. Montrer de plus que le $\min$ est atteint en un unique entier aléatoire N dans $\{1,\ldots,n\}$, indépendant de la valeur M du $\min$, et tel que $\mathbb{P}(N=k) = \lambda_k/(\lambda_1+\cdots+\lambda_n)$ pour tout $1 \leq k \leq n$.*
2. *Si $\lambda_1 = \cdots = \lambda_n$, établir que*

$$(E_{(1)},\ldots,E_{(n)}) \overset{\text{loi}}{=} \left(\frac{E_1}{n},\ldots,\frac{E_1}{n}+\cdots+\frac{E_n}{1}\right)$$

Solution.

1. On a $F_{\min(X_1,\ldots,X_n)}(t) = 1-\mathbb{P}(X_1>t)\cdots\mathbb{P}(X_n>t) = 1-e^{-t(\lambda_1+\cdots+\lambda_n)} = F_{\mathrm{Exp}(\lambda_1+\cdots+\lambda_n)}(t)$ pour tout $t \in \mathbb{R}_+$ d'où $M \sim \mathrm{Exp}(\lambda_1+\cdots+\lambda_n)$. Comme $(X_1,\ldots,X_n)$ est à densité, il donne une probabilité nulle à toute hypersurface, et en particulier à celle qui correspond à un min atteint en plusieurs points. Ainsi, l'$\arg\min$ est unique p.s.

   Le raisonnement qui suit fournit à la fois l'unicité p.s. de l'$\arg\min$, l'indépencence de M et N, et les lois de ces variables ! Soit N l'élément aléatoire de $\{1,\ldots,n\}$ valant k si $E_{k'} > E_k$ pour tout $k' \neq k$, et ∞ sinon. Déterminons la loi du couple (M,N). On a pour tout pour $k \in \{1,\ldots,n\}$ et tout réel $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(M>t \text{ et } N=k) &= \mathbb{P}(X_k>t, X_{k'}>X_k\ \forall k'\neq k) \\
&= \mathbb{E}(\mathbf{1}_{x_k>t,x_{k'}>x_k\ \forall k'}(E_1,\ldots,E_n)) \\
&= \int_t^\infty \lambda_k e^{-x_k\lambda_k}\left(\prod_{k'\neq k}\int_{x_{k'}>x_k}\lambda_{k'}e^{-\lambda_{k'}x_{k'}}\,dx_{k'}\right)dx_k \\
&= \int_t^\infty \lambda_k e^{-x_k\lambda_k}e^{-x_k\sum_{k'\neq k}\lambda_k}\,dx_k \\
&= e^{-t\sum_{k'=1}^n\lambda_{k'}}\frac{\lambda_k}{\sum_{k'=1}^n\lambda_{k'}}.
\end{aligned}$$

En particulier, on en déduit que M et N sont indépendantes, que M suit la loi exponentielle de paramètre $\lambda_1+\cdots+\lambda_n$ (en sommant sur k), et que N suit la loi discrète sur $\{1,\dots,n\}$ définie par $p_k=\lambda_k/(\lambda_1+\cdots+\lambda_n)$ (en posant $t=0$). En particulier $\mathbb{P}(N<\infty)=\sum_{k=1}^n\mathbb{P}(N=k)=1$ et donc le min est atteint p.s. en un unique k. Ce calcul est donc auto-suffisant !

2. Pour toute fonction test $\varphi:\mathbb{R}^n\to\mathbb{R}$ mesurable et bornée, par changements de variables,

$$\begin{aligned}\mathbb{E}(\varphi(E_{(1)},\dots,E_{(n)})) &= \int_{[0,\infty)^n}\varphi(x_{(1)},\dots,x_{(n)})\,\lambda^n e^{-\lambda(x_1+\cdots+x_n)}\,dx_1\cdots dx_n\\ &= n!\int_{0\le y_1<\cdots<y_n}\varphi(y_1,\dots,y_n)\,\lambda^n e^{-\lambda(y_1+\cdots+y_n)}\,dy_1\cdots dy_n\\ &= \int_{[0,\infty)^n}\varphi(z_1,\dots,z_1+\cdots+z_n)\,n!\lambda^n e^{-\lambda(nz_1+\cdots+1z_n)}\,dz_1\cdots dz_n.\end{aligned}$$

Il en découle que les variables aléatoires $E_{(1)},E_{(2)}-E_{(1)},\dots,E_{(n)}-E_{(n-1)}$ sont indépendantes avec $E_{(k)}-E_{(k-1)}$ de loi exponentialle de moyenne $1/((n-k+1)\lambda)$.

Le volume du cube $[0,1]^n$ est $n!$ le volume du simplexe $\{(x_1,\dots,x_n)\in[0,1]^n:x_1\le\cdots\le x_n\}$. Plus généralement, pour toute fonction $\varphi:\mathbb{R}_+^n\to\mathbb{R}$ symétrique on a

$$\begin{aligned}n!\int_{0\le x_1\le\cdots\le x_n}\varphi(x_1,\dots,x_n)\,dx_1\cdots dx_n &= \sum_{\sigma\in S_n}\int_{0\le x_{\sigma(1)}\le\cdots\le x_{\sigma(n)}}\varphi(x_{\sigma(1)},\dots,x_{\sigma(n)})\,dx_1\cdots dx_n\\ &= \int_{\mathbb{R}_+^n}\varphi(x_1,\dots,x_n)\,dx_1\cdots dx_n.\end{aligned}$$

Exercice 4.3 (Somme de variables indépendantes et convolution)**.**

1. *Soient X et Y deux v.a. discrètes à valeurs dans $\mathbb{Z}$, indépendantes. Exprimer la loi de (X,Y) puis de $X+Y$ en fonction de la loi de X et de la loi de Y ;*
2. *Soient X et Y deux v.a. réelles de densités f_X et f_Y, indépendantes. Montrer que (X,Y) et $X+Y$ ont des densités, notées $f_{(X,Y)}$ et f_{X+Y}, qu'on exprimera en fonction de f_X et f_Y ;*
3. *Calculer la densité de $X+Y$ lorsque X et Y sont indépendantes de loi uniforme sur $[0,1]$;*
4. *Soient X et Y deux v.a.r. indépendantes de loi exponentielle de paramètre λ. Montrer que $X+Y$ suit la loi Gamma de densité $\lambda^2x^{-1}e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$;*
5. *Établir par récurrence sur n que si $X_1,\dots,X_n$ sont des v.a.r. indépendantes de loi exponentielle de paramètre λ alors $X_1+\cdots+X_n$ suit la loi Gamma de densité $\frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$;*
6. *Établir par récurrence sur n que si $X_1,\dots,X_n$ sont des v.a.r. indépendantes avec $X_k\sim\mathcal{N}(m_k,\sigma_k^2)$ pour tout $1\le k\le n$, alors $X_1+\cdots+X_n\sim\mathcal{N}(m_1+\cdots+m_n,\sigma_1^2+\cdots+\sigma_n^2)$.*

Solution.

1. Pour tous $x,y\in\mathbb{Z}$, $\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\mathbb{P}(Y=y)$ par indépendance. On a aussi, pour tout $z\in\mathbb{Z}$, $\mathbb{P}(X+Y=z)=\sum_{x+y=z}\mathbb{P}(X=x,Y=y)=\sum_x\mathbb{P}(X=x)\mathbb{P}(Y=z-x)$;
2. $f_{(X,Y)}=f_X\otimes f_Y$: par indépendance et par le théorème de Fubini-Tonelli, pour tous $I,J\subset\mathbb{R}$,

$$\mathbb{P}((X,Y)\in I\times J)=\mathbb{P}(X\in I)\mathbb{P}(Y\in J)=\left(\int_I f_X(x)\,dx\right)\left(\int_J f_Y(y)\,dy\right)=\iint_{I\times J}f_X(x)f_Y(y)\,dxdy.$$

$f_{X+Y}=f_X*f_Y$: pour toute fonction test $h:\mathbb{R}\to\mathbb{R}$ continue et à support compact, par les théorèmes du changement de variable et de Fubini-Tonelli,

$$\begin{aligned}\mathbb{E}(h(X+Y)) &= \iint h(x+y)f_{(X,Y)}(x,y)\,dxdy\\ &= \iint h(x+y)f_X(x)f_Y(y)\,dxdy\\ &= \int h(z)\left(\int f_X(x)f_Y(z-x)\,dx\right)dz\\ &= \int h(z)(f_X*f_Y)(z)\,dz;\end{aligned}$$

3. Loi triangulaire sur $[0,2]$ car pour tout $z \in [0,2]$,

$$f_{X+Y}(z) = \int \mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,1]}(z-x)\,dx = \int_{\max(0,z-1)}^{\min(1,z)} dz = \min(1,z) - \max(0,z-1) = 1 - |z-1|.$$

4. Pour tout $z \in \mathbb{R}_+$,

$$\int \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty[}(x)\lambda e^{-\lambda(z-x)}\mathbf{1}_{[0,\infty[}(z-x)\,dx = \lambda^2 e^{-\lambda z}\int_0^z dx = \lambda^2 z e^{-\lambda z}.$$

5. Pour tout $z \in \mathbb{R}_+$,

$$\int \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}\mathbf{1}_{[0,\infty[}(x)\lambda e^{-\lambda(z-x)}\mathbf{1}_{[0,\infty[}(z-x)\,dx = \frac{\lambda^{n+1}}{(n-1)!}e^{-\lambda z}\int_0^z z^{n-1}dx = \frac{\lambda^{n+1}}{n!}z^n e^{-\lambda z}.$$

6. Par translation on se ramène à $m_1 = \cdots = m_n = 0$. Pour tous $a, b > 0$ et $z \in \mathbb{R}$,

$$\int e^{-ax^2}e^{-b(z-x)^2}\,dx = e^{-\frac{ab}{a+b}z^2}\int e^{-(x\sqrt{a+b}-\frac{b}{\sqrt{a+b}}z)^2}\,dx = c_{a,b}e^{-\frac{z^2}{\frac{1}{a}+\frac{1}{b}}}.$$

Exercice 4.4 (Simulation des lois gaussiennes avec l'algorithme de Box-Muller)**.**

1. *Soient (X,Y) un couple de v.a.r. et son écriture en coordonnées polaires $(r\cos(\theta), r\sin(\theta))$. Montrer que X et Y sont indépendantes et de loi $\mathcal{N}(0,1)$ si et seulement si r et θ sont indépendantes avec r^2 de loi exponentielle de paramètre $1/2$ et θ de loi uniforme sur $[0,2\pi]$;*
2. *Soient U et V deux v.a.r. uniformes sur $[0,1]$. Montrer que les v.a.r. $X = \sqrt{-2\ln(U)}\cos(2\pi V)$ et $Y = \sqrt{-2\ln(U)}\sin(2\pi V)$ sont indépendantes et de loi $\mathcal{N}(0,1)$.*

Note en passant : $\mathrm{Gamma}(1/2,n) = (\mathrm{Exp}(1/2))^{*n} = (\chi^2(2))^{*n} = \chi^2(2n)$, *pour tout* $n \in \mathbb{N}^*$.

Solution.

1. Difféomorphisme $(r,\theta) \in \mathbb{R}_+ \times [0,2\pi[\mapsto (x,y) = (r\cos(\theta), r\sin(\theta)) \in \mathbb{R}\backslash\{(0,0)\}$. Jacobien : $\det\begin{pmatrix}\cos(\theta) & \sin(\theta)\\ -r\sin(\theta) & r\cos(\theta)\end{pmatrix}$ $r(\cos(\theta)^2+\sin(\theta)^2) = r$. Cela donne :

$$\left(\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right)\left(\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\right)dxdy = \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}dxdy = \left(\frac{\mathbf{1}_{[0,2\pi]}(\theta)}{\pi}\right)\left(re^{-\frac{r^2}{2}}\mathbf{1}_{[0,\infty[}(r)\right)drd\theta.$$

2. $2\pi V$ et $-2\ln(U)$ sont indépendantes de loi uniforme sur $[0,2\pi]$ et de loi exponentielle de paramètre $1/2$ si et seulement si U et V sont indépendantes et de loi uniforme sur $[0,1]$.

Exercice 4.5 (Simulation des lois de Poisson)**.** *Soit $(U_n)_{n\geq 1}$ une suite de v.a.r. indépendantes de loi uniforme, et soit λ un réel strictement positif.*

1. *Qu'elle est la loi de $T_n := E_1 + \cdots + E_n$ où $E_k = -\ln(U_k)/\lambda$?*
2. *Montrer que la variable discrète $N = \min\{n \geq 0 : U_1 \cdots U_{n+1} < e^{-\lambda}\}$ suit la loi de Poisson de paramètre λ, et en déduire un algorithme de simulation de cette loi. Étudier sa complexité.*

Solution.

1. $E_k \sim \mathrm{Exp}(1/\lambda)$ et $E_1 + \cdots + E_n \sim \mathrm{Gamma}(n,\lambda)$
2. On pose $(x_1,\ldots,x_{n+1}) \mapsto (y_1,\ldots,y_{n+1}) = (x_1,\ldots,x_1+\cdots+x_{n+1})$ (linéaire et triangulaire) :

$$\mathbb{P}(N=n) = \mathbb{P}(T_1 \leq 1,\ldots,T_n \leq 1, T_{n+1} > 1) = \int_{0\leq y_1\leq\cdots\leq y_n\leq 1, y_{n+1}>1}\lambda^{n+1}e^{-\lambda y_{n+1}}\,dy_1\cdots dy_{n+1} = \frac{\lambda^n}{n!}e^{-\lambda}.$$

Le nombre de « produit et test » effectués par l'algorithme est aléatoire et suit la loi de Poisson de paramètre λ (de moyenne et de variance λ).

Exercice 4.6 (Simulation par méthode du rejet)**.** *Soit $D \subset \mathbb{R}^d$ un morceau compact de $\mathbb{R}^d$ égal à l'adhérence de son intérieur, de volume $|D| > 0$. Soit C un hypercube contenant D et $(X_n)_{n\geq 1}$ des v.a. i.i.d. de loi uniforme sur C. Montrer que $T := \inf\{n \geq 1 : X_n \in D\}$ est géométrique $\mathrm{Geom}(\frac{|D|}{|C|})$, que T et X_T sont indépendantes, et que X_T est uniforme sur D de densité $\frac{1}{|D|}\mathbf{1}_D$.*

Solution. Pile ou face : $T = \inf\{n \geq 1 : Y_n = 1\}$ avec $Y_n := \mathbf{1}_{\{X_n \in D\}}$ et $\mathbb{P}(Y_n = 1) = \frac{|D|}{|C|}$ (densité). Pour l'indépendance de T et de X_T on écrit, pour tout $n \in \mathbb{N}^*$ et tout $A \subset C$,

$$\mathbb{P}(X_T \in A, T = n) = \mathbb{P}(X_1 \notin D, \ldots, X_{n-1} \notin D, X_n \in A \cap D) = \left(1 - \frac{|D|}{|C|}\right)^{n-1} \frac{|D|}{|C|} \frac{|A \cap D|}{|D|}.$$

En sommant sur $n \in \mathbb{N}^*$ on obtient $\mathbb{P}(X_T \in A) = \frac{|A \cap D|}{|D|}$ et donc X_T suit bien la loi uniforme sur D. En prenant $A = C$ on retrouve bien que $\mathbb{P}(T = n) = \left(1 - \frac{|D|}{|C|}\right)^{n-1} \frac{|D|}{|C|}$ et donc que $T \sim \mathrm{Geom}(\frac{|D|}{|C|})$.

Exercice 4.7 (Simulation par méthode du rejet)**.** *Soit f une densité sur $\mathbb{R}$. Supposons qu'il existe une densité g sur $\mathbb{R}$ telle que $f \leq cg$ pour une constante $c \geq 1$. Soit (U_n) une suite de v.a.r. i.i.d. de loi uniforme sur $[0,1]$, et soit (X_n) une suite de v.a.r. i.i.d. de densité g, indépendantes de la suite (U_n). Montrer que $T := \inf\{n \geq 1 : cg(X_n)U_n \leq f(X_n)\}$ est géométrique $\mathrm{Geom}(1/c)$, que T et X_T sont indépendantes, et que X_T a pour densité f. Étudier le cas où g est uniforme sur $[0,1]$.*

Solution. Posons $Y_n = (X_n, U_n)$ et $D := \{(x,u) \in \mathbb{R} \times [0,1] : cg(x)u \leq f(x)\}$. Les v.a. $(Y_n)_{n \geq 1}$ sont i.i.d. de Bernoulli et $T := \inf\{n \in \mathbb{N}^* : Y_n \in D\}$ (pile ou face). Maintenant $T \sim \mathrm{Geom}(1/c)$ car

$$\mathbb{P}(Y_1 \in D) = \iint \mathbf{1}_{cg(x)u \leq f(x)} \mathbf{1}_{0 \leq u \leq 1}\, g(x)dxdu = \iint \mathbf{1}_{0 \leq u \leq \frac{f(x)}{cg(x)}}\, g(x)dxdu = \int \frac{f(x)}{c}\, dx = \frac{1}{c}.$$

Pour établir l'indépendance de T et X_T et que X_T a pour densité f, on pose et on écrit, pour tout intervalle $I \subset \mathbb{R}$ et tout entier $n \in \mathbb{N}^*$, en posant $D_I := \{(x,u) \in D : x \in I\} \subset D$,

$$\mathbb{P}(X_T \in I, T = n) = \mathbb{P}(Y_1 \notin D)^{n-1} \mathbb{P}(Y_1 \in D_I) = \left(1 - \frac{1}{c}\right)^{n-1} \frac{1}{c} \int_I f(x)\, dx.$$

5 LGN, convergences p.s., en moyenne, et en probabilité

Rappels de cours. Convergences presque sûre, en probabilité, dans L^p, lemme de Borel-Cantelli, théorème de convergence dominée, inégalité de Markov, lois faible et forte des grands nombres, méthode de Monte Carlo pour le calcul d'intégrales.

Exercice 5.1 (Convergences). *Soit $(X_n)_{n\geq 1}$ des v.a.r. telles que $X_n \sim \mathcal{B}_{0,1}(p_n)$ avec $p_n \to 0$.*

1. *Montrer que $X_n \overset{\mathbb{P}}{\to} 0$ et même que $X_n \overset{\mathrm{L}^1}{\to} 0$;*
2. *Montrer que $X_n \overset{p.s.}{\to} 0$ si $\sum_n p_n < \infty$;*
3. *Si les $(X_n)_{n\geq 1}$ sont indépendantes, montrer que $\mathbb{P}(X_n \not\to 0) = 1$ si $\sum_n p_n = \infty$;*
4. *Montrer enfin que si les $(X_n)_{n\geq 1}$ sont dépendantes, alors on peut avoir $\mathbb{P}(X_n \to 0) = 1$ et $\sum_n p_n = \infty$. Indication : trouver un contre exemple du type $X_n = f_n(U)$ avec U uniforme.*

Solution.

1. On a $\mathbb{E}(|X_n|) = \mathbb{E}(X_n) = \mathbb{P}(X_n = 1) = p_n \to 0$, donc $X_n \to 0$ L^1 et donc en $\mathbb{P}$ (Markov). Pour la convergence $\mathbb{P}$, on peut alternativement remarquer que $\mathbb{P}(|X_n| \geq \varepsilon) = p_n \mathbf{1}_{\varepsilon \leq 1} \to 0$.
2. Premier lemme de Borel-Cantelli : si $\sum_n p_n < \infty$ alors $\mathbb{P}(\overline{\lim}_n \{|X_n| = 1\}) = 0$ i.e. $\mathbb{P}(\underline{\lim}_n \{|X_n| = 0\}) = 1$. Donc p.s. $|X_n| = 0$ pour n assez grand.
3. Second lemme de Borel-Cantelli (nécessite l'indépendance des événements) : si $\sum_n p_n = \infty$ alors $\mathbb{P}(\overline{\lim}_n \{|X_n| = 1\}) = 1$ i.e. p.s. $|X_n| = 1$ pour une infinité de valeurs de n, donc p.s. $X_n \not\to 0$.
4. En prenant $X_n = \mathbf{1}_{[0,p_n]}(U)$ et $p_n = 1/n$, on a $X_n \to 0$ p.s. car $p_n \to 0$, mais $\sum_n p_n = \infty$.

Exercice 5.2 (LGN). *En considérant une suite de v.a.r. i.i.d. de loi uniforme sur $[0,1]$, et en utilisant la loi faible des grands nombres, établir que pour toute $f : \mathbb{R} \to \mathbb{R}$ continue et bornée,*

$$\lim_{n\to\infty} \int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n = f(1/2).$$

Solution. Si $S_n = \frac{X_1+\cdots+X_n}{n}$ avec $X_1, X_2, \ldots$ i.i.d. uniformes sur $[0,1]$ alors

$$\int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n = \mathbb{E}(f(S_n)).$$

Comme f est continue en $1/2$, pour tout $\varepsilon > 0$, il existe $\eta > 0$ dépendant de $\varepsilon > 0$ tel que $|f(x) - f(1/2)| \leq \varepsilon$ pour tous $x \in [0,1]$ tel que $|x - 1/2| \leq \eta$. D'autre part, la LGN faible donne $\mathbb{P}(|S_n - 1/2| \geq \eta) \to 0$ quand $n \to \infty$. Par conséquent :

$$\begin{aligned}
|\mathbb{E}(f(S_n)) - f(1/2)| &= |\mathbb{E}(f(S_n) - f(1/2))| \\
&\leq \mathbb{E}|f(S_n) - f(1/2)| \\
&= \mathbb{E}(|f(S_n) - f(1/2)|\mathbf{1}_{|S_n - 1/2| \geq \eta}) + \mathbb{E}(|f(S_n) - f(1/2)|\mathbf{1}_{|S_n-1/2|<\eta}) \\
&\leq 2\|f\|_\infty \mathbb{P}(|S_n - 1/2| \geq \eta) + \varepsilon.
\end{aligned}$$

On choisit donc ε, qui donne η, puis n assez grand. Note : la LGN forte permet d'obtenir le résultat beaucoup plus rapidement grâce au théorème de convergence dominée.

Exercice 5.3 (LGN). *Soit $(X_i)_{i\geq 1}$ des v.a.r. i.i.d. > 0 avec $\mathbb{E}(X_1) = 1$ et $\mathbb{P}(X_1 = 1) < 1$. Soit $Y_n := \prod_{i=1}^n X_i$.*

1. *En utilisant la LGN et l'inégalité de Jensen, montrer que $(Y_n)_{n\geq 1}$ converge p.s. vers une limite que l'on précisera. Un joueur joue tous les jours le dixième de sa fortune à pile ou face avec un ami. Quelle est l'évolution de cette fortune au bout d'un temps très long ?*
2. *Calculer $\mathbb{E}(Y_n)$. La suite $(Y_n)_{n\geq 1}$ converge-t-elle en moyenne ?*

Solution.

1. Jensen : $\mathbb{E}(\ln(X_i)) \leq \ln \mathbb{E}(X_i) = \ln(1) = 0$ et le cas d'égalité donne $m := \mathbb{E}(\ln(X_i)) < 0$ car $\mathbb{P}(X_i = 1) < 1$. À présent, par la LGN : p.s. $Y_n = e^{\sum_{i=1}^n \ln(X_i)} = e^{n(m+o(1))} \to 0$ car $m < 0$. Dans l'exemple d'application financier : $1 - \mathbb{P}(X_n = 9/10) = \mathbb{P}(X_n = 11/10) = 1/2$. Note : rien n'indique que $Z_i := \ln(X_i)$ est intégrable, en fait Z_i^+ est intégrable par Jensen mais pas forcément Z_i^-, mais on peut tronquer $Z_i^- = \max(Z_i^-, c) + Z_i^- - \max(Z_i^-, c)$ avec c assez grand pour que la moyenne de $Z_i - \max(Z_i^-, c)$ reste < 0, et jeter le reste $-(Z_i^- - \max(Z_i^-, c))$ en majorant Y_n ;

2. $\mathbb{E}(Y_n) = \mathbb{E}(X_1)^n = 1$. Si la suite $(Y_n)_{n\geq 1}$ convergeait en moyenne vers une v.a.r. Y_∞, alors une sous-suite converge p.s. vers Y_∞, or cela donnerait $Y_\infty = 0$ car $(Y_n)_{n\geq 1}$ converge p.s. vers 0, ce qui est impossible car $\mathbb{E}(Y_\infty) = \lim_{n\to\infty} \mathbb{E}(Y_n) = 1$.

Exercice 5.4 (Réduction de variance dans une méthode de Monte Carlo)**.** *Soit $g : [0,1] \to [0,1]$ une fonction mesurable et bornée. On souhaite calculer $m = \int_0^1 g(x)\,dx$. On pose $\sigma^2 := \int_0^1 g^2(x)\,dx - m^2$. Soient X et Y des v.a.r. i.i.d. de loi uniforme sur $[0,1]$. On pose*

$$U = \mathbf{1}_{Y\leq g(X)}, \quad V = g(X), \quad et \quad W = \frac{g(X)+g(1-X)}{2}.$$

1. *Calculer l'espérance et la variance de U, V, W, et comparer les variances ;*
2. *Proposer trois méthodes de type Monte-Carlo pour calculer m ;*
3. *On suppose dans la suite que g est monotone. Vérifier que $\mathbb{E}(g(X)g(1-X)) \leq m^2$. Indication : on pourra d'abord montrer que $(g(x)-g(y))(g(1-x)-g(1-y)) \leq 0$ pour tous x, y.*
4. *Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. i.i.d. de loi uniforme sur $[0,1]$, et les estimateurs de m*

$$A_n = \frac{1}{2n}\sum_{i=1}^{2n} g(X_i) \quad et \quad B_n = \frac{1}{2n}\sum_{i=1}^{n}(g(X_i)+g(1-X_i)).$$

   Lequel possède la plus petite erreur quadratique moyenne ?
5. *Dans le cas où $g(x) = x^2$, déterminer le nombre n de simulations nécessaires garantissant une précision relative de 1% sur le calcul de m en erreur quadratique avec A_n et B_n.*

Solution.

1. Comme $U^2 = U$ (Bernoulli) on a $\mathrm{Var}(U) = \mathbb{E}(U)(1-\mathbb{E}(U))$. À présent,

$$\mathbb{E}(U) = \mathbb{P}(Y \leq g(X)) = \iint_{[0,1]^2} \mathbf{1}_{y\leq g(x)}\,dxdy = \int_0^1 g(x)\,dx = m \quad \text{et} \quad \mathrm{Var}(U) = m(1-m).$$

   Par le théorème du transfert

$$\mathbb{E}(V) = \int_0^1 g(x)\,dx = m \quad \text{et} \quad \mathbb{E}(V^2) = \int_0^1 g^2(x)\,dx \quad \text{donc} \quad \mathrm{Var}(V) = \sigma^2.$$

   Comme X et $1-X$ ont même loi on a par linéarité de l'espérance

$$\mathbb{E}(W) = m \quad \text{et} \quad \mathbb{E}(W^2) = \frac{(\sigma^2+m^2)+\int_0^1 g(x)g(1-x)\,dx}{2}$$

   et donc

$$\mathrm{Var}(W) = \frac{(\sigma^2-m^2)+\int_0^1 g(x)g(1-x)\,dx}{2}$$

   Comme g prend ses valeurs dans $[0,1]$, il vient que V a une plus petite variance que U car

$$\mathrm{Var}(U) - \mathrm{Var}(V) = \int_0^1 g(x)(1-g(x))\,dx \geq 0.$$

   D'autre part, la comparaison des variances de V et W donne :

$$\mathrm{Var}(V) - \mathrm{Var}(W) = \frac{\int_0^1 g(x)(g(x)-g(1-x))\,dx}{2}.$$

2. Par la LGN forte on a p.s.

$$m = \lim_{n\to\infty} \frac{\mathrm{card}\{1 \leq i \leq n : Y_i \leq g(X_i)\}}{n} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(X_i) = \lim_{n\to\infty} \frac{1}{2n}\sum_{i=1}^{n}(g(X_i)+g(1-X_i)).$$

3. Comme g est monotone, on a $(g(x)-g(y))(g(1-x)-g(1-y)) \leq 0$ pour tous x, y et donc $\mathbb{E}(g(X)g(1-X)) + \mathbb{E}(g(Y)g(1-Y)) - 2\mathbb{E}(g(X))\mathbb{E}(g(Y)) \leq 0$, c'est-à-dire $\mathbb{E}(g(X)g(1-X)) \leq m^2$. Il en découle que $\mathrm{Var}(W) \leq \frac{\sigma^2}{2} = \frac{\mathrm{Var}(V)}{2}$.

4. $\mathrm{EQM}(\hat{m}) := \mathbb{E}((\hat{m}-m)^2) = \mathrm{Var}(\hat{m}) + (\mathbb{E}(\hat{m})-m)^2$. (EQM=Variance+CarréDuBiais). Les estimateurs A_n et B_n sont sans biais. On a $\mathrm{EQM}(A_n) = \frac{\sigma^2}{2n}$ et $\mathrm{EQM}(B_n) = \frac{\mathrm{Var}(W)}{n} \leq \frac{\sigma^2}{2n}$. Mais A_n nécessite un échantillon deux fois plus gros que B_n.
5. $\frac{\mathrm{EQM}(A_n)}{m^2} = \frac{1}{100}$ donne $n = \frac{50\sigma^2}{m^2}$. Pour $g(x) = x^2$ il vient $m = \frac{1}{3}$ et $\sigma^2 = \frac{4}{45}$ et donc $n \approx 40$ avec A_n, et en fait plus de 16 fois moins avec B_n car il s'avère que $\mathrm{Var}(W) = \frac{1}{180}$.

Exercice 5.5 (Loi uniforme sur les sphères $\ell_n^1(\mathbb{R})$ et $\ell_n^2(\mathbb{R})$)**.**

1. *Montrer que si $X_1,\ldots,X_n$ sont i.i.d. de loi $\mathcal{N}(0,1)$ alors $(X_1,\ldots,X_n)/\sqrt{X_1^2+\cdots+X_n^2}$ suit la loi uniforme sur la sphère unité de $\mathbb{R}^n$. Que se passe-t-il lorsque $n\to\infty$?;*
2. *Montrer que si $X_1,\ldots,X_n$ sont i.i.d. de loi $\mathrm{Exp}(1)$ alors $(X_1,\ldots,X_n)/(X_1+\cdots+X_n)$ suit la loi uniforme sur $\{(x_1,\ldots,x_n)\in\mathbb{R}_+^n : x_1+\cdots+x_n = 1\}$. Que se passe-t-il lorsque $n\to\infty$?*

Solution.

1. La loi de $X = (X_1,\ldots,X_n)$ est invariante par rotation. En effet, pour tout $R\in\mathbb{O}(\mathbb{R}^n)$ et toute $h:\mathbb{R}^n\to\mathbb{R}$ continue et bornée, comme la densité $x\mapsto f(x)$ de X ne dépend que de $\|x\|$, on a, en utilisant le fait que $\|Rx\| = \|x\|$ le changement de variable linéaire $y = Rx$ dont le jacobien $\mathrm{Jac}(x) = R$ est constant (ne dépend pas de x) et de déterminant ±1,
$$\mathbb{E}(h(RX)) = \int_{\mathbb{R}^n} h(Rx)f(x)\,dx = \int_{\mathbb{R}^n} h(y)f(R^{-1}y)\,dy = \int_{\mathbb{R}^n} h(y)f(y)\,dy = \mathbb{E}(h(X)).$$
Il en découle que la loi de $\Theta := X/\|X\|$ est également invariante par rotation, c'est-à-dire que pour tout borélien A de la sphère unité $\mathcal{S}(\mathbb{R}^n) := \{x\in\mathbb{R}^n : \|x\| = 1\}$, on a $\mathbb{P}(Z\in A) = \mathbb{P}(RZ\in A)$, ce qui est une manière de dire que Z suit la loi uniforme sur $\mathcal{S}(\mathbb{R}^n)$. La loi uniforme sur $\mathcal{S}(\mathbb{R}^n)$ apparaît également dans le passage en coordonnées sphériques $dx = r^{n-1}drd\sigma$ qui montre également l'indépendance de $\|X\|$ et $X/\|X\|$, comme pour le cas polaire $n=2$ (Box-Muller). On a $\|X\|^2\sim\chi^2(n)$. La LGN donne $\sqrt{X_1^2+\cdots+X_n^2} = \sqrt{n}(1+o_{n\to\infty}(1))$ p.s. et donc par convergence dominée, la loi image de la loi uniforme de la sphère de rayon $\sqrt{n}$ de $\mathbb{R}^n$, par la projection sur $\mathbb{R}^k$, tend vers $\mathcal{N}(0,I_k)$ quand $n\to\infty$.
2. Si $Y := X/\|X\|_1$ alors $Y_n = 1-(Y_1+\cdots+Y_{n-1})$. Par linéarité, il suffit de montrer que le vecteur aléatoire $(Y_1,\ldots,Y_{n-1})$ suit la loi uniforme sur le simplexe $\{y\in[0,1]^{n-1} : y_1+\cdots+y_{n-1}\leq 1\}$. Le changement de variable «projection conique» $(x_1,\ldots,x_n)\mapsto(y_1,\ldots,y_{n-1},r)$ avec $x_i = ry_i$ pour $1\leq i\leq n-1$ et $r := x_1+\cdots+x_n$ a pour jacobien
$$\begin{pmatrix} r & 0 & \cdots & 0 & y_1 \\ 0 & r & \ddots & \vdots & y_2 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & r & y_{n-1} \\ -r & \cdots & \cdots & -r & 1-(y_1+\cdots+y_{n-1}) \end{pmatrix}.$$
Son déterminant ne dépend que de r, et ceci montre que la densité de $(Y_1,\ldots,Y_{n-1})$ est constante : c'est donc bien une loi uniforme. Il s'agit d'un cas spécial de loi de Dirichlet. La LGN donne ensuite $X_1+\cdots+X_n = n(1+o_{n\to\infty}(1))$ p.s. et donc par convergence dominée, la loi image de la loi uniforme de la sphère ℓ_n^1 de rayon n de $\mathbb{R}^n$, par la projection sur $\mathbb{R}^k$, tend vers $\mathrm{Exp}(1)^{\otimes k}$ quand $n\to\infty$.

Exercice 5.6 (Biais par la taille)**.** *On considère une population comportant un grand nombre n de foyers. On modélise la taille de ces foyers par une suite de v.a. i.i.d. $X_1,\ldots,X_n$ sur $\mathbb{N}^*$, de moyenne $m := \mathbb{E}(X_1) = \sum_{k\geq1}p_k < \infty$ où $p_k := \mathbb{P}(X_1 = k)$. Soit T la taille du foyer d'un individu pris au hasard dans la population. Montrer que $\mathbb{P}(T=k)\approx\frac{k}{m}p_k$, pour tout $k\in\mathbb{N}^*$.*

Solution. La population compte au total $X_1+\cdots+X_n$ individus. On modélise le choix d'un français au hasard par le tirage, conditionnellement à $X_1,\ldots,X_n$ (c'est-à-dire sachant $X_1,\ldots,X_n$), d'un entier selon la loi uniforme sur l'intervalle $[\![1,X_1+\cdots+X_n]\!]$. Pour tout $k\geq1$, il y a $N_k := \mathbf{1}_{\{X_1=k\}}+\cdots+\mathbf{1}_{\{X_n=k\}}$ foyers de taille k qui comptent au total kN_k individus. Par conséquent, par définition de la loi uniforme (formule « cas favorables sur cas totaux ») et à une double application de la loi forte des grands nombres :
$$\mathbb{P}(T=k|X_1,\ldots,X_n) = \frac{k\sum_{i=1}^n\mathbf{1}_{\{X_i=k\}}}{\sum_{i=1}^n X_i} = \frac{k\frac{1}{n}\sum_{i=1}^n\mathbf{1}_{\{X_i=k\}}}{\frac{1}{n}\sum_{i=1}^n X_i} \xrightarrow[n\to\infty]{\text{p.s.}} \frac{k}{m}p_k.$$

À présent, par convergence dominée, on obtient enfin le résultat souhaité :

$$\mathbb{P}(T=k) = \mathbb{E}(\mathbb{P}(T=k|X_1,\dots,X_n)) \underset{n\to\infty}{\longrightarrow} \frac{k}{m}p_k.$$

Exercice 5.7 (Processus de Poisson)**.** *Soit $(N_t)_{t\geq 0}$ un processus de Poisson simple d'intensité $\lambda > 0$: pour tout $n \in \mathbb{N}^*$ et tous $0 := t_0 < t_1 < \dots < t_n$, les v.a. $N_{t_1} - N_{t_0}, \dots, N_{t_n} - N_{t_{n-1}}$ sont indépendantes avec $N_{t_k - t_{k-1}} \sim \mathrm{Poi}(\lambda(t_k - t_{k-1}))$ pour tout $1 \leq k \leq n$. Montrer que p.s.*

$$\lim_{t\to\infty} \frac{N_t}{t} = \lambda.$$

Solution. Les trajectoires de $(N_t)_{t\geq 0}$ sont croissantes car à incréments ≥ 0. Par la LGN, p.s.

$$\underbrace{\frac{N_0}{t}}_{\to 0} + \underbrace{\frac{\lfloor t\rfloor}{t}}_{\to 1}\underbrace{\frac{\sum_{k=1}^{\lfloor t\rfloor} N_k - N_{k-1}}{\lfloor t\rfloor}}_{\to\lambda} = \frac{N_{\lfloor t\rfloor}}{t} \leq \frac{N_t}{t} \leq \frac{N_{\lfloor t\rfloor+1}}{t} = \underbrace{\frac{N_0}{t}}_{\to 0} + \underbrace{\frac{\lfloor t\rfloor+1}{t}}_{\to 1}\underbrace{\frac{\sum_{k=1}^{\lfloor t\rfloor+1} N_k - N_{k-1}}{\lfloor t\rfloor+1}}_{\to\lambda}.$$

Note : $N_t \sim \mathrm{Poi}(\lambda t)$ donc $\mathbb{E}(N_t) = \mathrm{Var}(N_t) = \lambda t$. Tchebychev : $\mathbb{P}(|\frac{N_t}{t} - \lambda| \geq \varepsilon) \leq \frac{\lambda}{t\varepsilon^2}$ ce qui donne $\frac{N_t}{t} \underset{t\to\infty}{\overset{\mathbb{P}}{\to}} \lambda$ en quelque sorte (t est continu). Cette approche ne donne pas la convergence p.s.

Exercice 5.8 (Marche aléatoire simple sur $\mathbb{R}$)**.** *On modélise la position d'une particule sur $\mathbb{R}$ à l'instant n par $X_{n+1} = X_n + \varepsilon_{n+1}$ où $(\varepsilon_n)_{n\geq 1}$ est une suite de v.a.r. i.i.d. indépendantes de X_0, et de moyenne m. Montrer que $\lim_{n\to\infty} |X_n| = +\infty$ presque sûrement si $m \neq 0$. Que peut-on dire que lorsque $m = 0$? Peut-on calculer la loi de X_n lorsque les $(\varepsilon_n)_{n\geq 1}$ sont gaussiens ou de Bernoulli ?*

Solution. LGN : $X_n - X_0 = \varepsilon_1 + \cdots + \varepsilon_n = n(m + o_{n\to\infty}(1)) \to \pm\infty$ p.s. si $m \neq 0$. Si $m = 0$, la LGN ne suffit plus. Si les $(\varepsilon_n)_{n\geq 0}$ sont $\mathcal{N}(m, \sigma^2)$ alors $X_n - X_0 \sim \mathcal{N}(nm, n\sigma^2)$ tandis que si les $(\varepsilon_n)_{n\geq 1}$ sont Bernoulli $\mathcal{B}_{\pm 1}(p)$ alors $\frac{X_n - X_0 + n}{2} \sim \mathrm{Binom}(n, p)$.

Exercice 5.9 (Marche aléatoire simple sur $\mathbb{Z}^d$)**.** *On modélise la position d'une particule sur la grille $\mathbb{Z}^d$ à l'instant n par $X_{n+1} = X_n + \varepsilon_{n+1}$ où $(\varepsilon_n)_{n\geq 1}$ est une suite de v.a. i.i.d. indépendantes de X_0, et de loi uniforme sur $\{\pm e_1, \dots, \pm e_d\}$ où $e_1, \dots, e_d$ est la base canonique de $\mathbb{R}^d$. Montrer que $\lim_{n\to\infty} \frac{1}{n} X_n \to 0$ presque sûrement. Peut-on calculer la loi de X_n ? Quel est son support ?*

Solution. $\frac{X_n}{n} \to 0$ p.s. par LGN. Si $(Z_1, \dots, Z_{2d}) \sim \mathrm{Multinom}(n, \{\frac{1}{2d}, \dots, \frac{1}{2d}\})$ alors le vecteur aléatoire $(Z_1 - Z_2, \dots, Z_{2d-1} - Z_{2d})$ a la même loi que X_n (penser à un dé à $2d$ faces).

$$\mathbb{P}(X_n = (x_1, \dots, x_n)) = \sum_{\substack{0\leq n_1,\dots,n_{2d}\leq n \\ n_1+\cdots+n_{2d}=n \\ n_1-n_2=x_1,\dots,n_{2d-1}-n_{2d}=x_n}} \frac{n!}{n_1!\cdots n_{2d}!}\frac{1}{(2d)^n}.$$

Exercice 5.10 (Matrices aléatoires)**.** *Soit $(X_{i,j})_{i,j\geq 1}$ un tableau infini de v.a.r. i.i.d. de moyenne μ et de variance σ^2. Soit $M := (X_{i,j})_{1\leq i\leq m, 1\leq j\leq n}$ la matrice aléatoire $m \times n$ obtenue en tronquant le tableau dans son coin supérieur gauche. Soient $\lambda_1, \dots, \lambda_m$ les valeurs propres de la matrice symétrique aléatoire $\frac{1}{n}MM^\top$ (ce sont des v.a.r. dépendantes à valeurs dans $\mathbb{R}_+$). Montrer que*

$$\frac{\lambda_1 + \cdots + \lambda_m}{m} \underset{mn\to\infty}{\overset{p.s.}{\longrightarrow}} \sigma^2 + \mu^2.$$

Solution. LGN :

$$\frac{\lambda_1 + \cdots + \lambda_m}{m} = \frac{\mathrm{Tr}(MM^\top)}{nm} = \frac{\sum_{i=1}^m \sum_{j=1}^n |M_{i,j}|^2}{mn} \underset{mn\to\infty}{\overset{\text{p.s.}}{\longrightarrow}} \mathbb{E}(|M_{1,1}|^2) = \sigma^2 + m^2.$$

Exercice 5.11 (Loi forte des grands nombres dans L^4)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et centrées, bornée dans L^4, et $S_n := \frac{1}{n}(X_1 + \cdots + X_n)$. Montrer que $\lim_{n\to\infty} S_n = 0$ p.s.*

Solution. On a $\mathbb{E}(S_n^4) = \mathcal{O}(n^{-2})$ d'où $\sum_n \mathbb{E}(S_n^4) < \infty$ d'où par le théorème de Fubini-Tonelli $\mathbb{E}(\sum_n S_n^4) < \infty$ d'où $\sum_n S_n^4 < \infty$ p.s. d'où $S_n \to 0$ p.s. Alternativement, on peut également utiliser le lemme de Borel-Cantelli car l'inégalité de Markov donne $\mathbb{P}(S_n \geq \varepsilon) \leq \varepsilon^{-4}\mathbb{E}(S_n^4) = \mathcal{O}(n^{-2})$ pour tout $\varepsilon > 0$.

Exercice 5.12 (Convergence complète)**.**

1. *Soit $(\mu_n)_{n\geq 1}$ une suite de lois de probabilités sur $\mathbb{R}$, et soit $c \in \mathbb{R}$ un réel fixé. Montrer que les propriétés suivantes sont équivalentes (on dit dans ce cas que $\mu_n \to c$ complètement) :*
   - *(a) Pour tout espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et pour toute suite $(X_n)_{n\geq 1}$ de v.a.r. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ avec $X_n \sim \mu_n$ pour tout $n \geq 1$, on a $\lim_{n\to\infty} X_n = c$ presque sûrement ;*
   - *(b) Pour tout réel $\varepsilon > 0$ on a $\sum_{n\geq 1} \mu_n([c-\varepsilon, v+\varepsilon]^c) < \infty$*
2. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes, centrées, et bornée dans L^4. Montrer que $\mathrm{Loi}(\frac{1}{n}(X_1 + \cdots + X_n)) \to 0$ complètement. Que se passe-t-il si les v.a. sont bornées dans L^r avec $r < 4$.*

Solution.

1. Borel-Cantelli (les deux parties !)
2. Convergence monotone ou Fubini-Tonelli, Markov et Borel-Cantelli.

6 TCL, convergence en loi, fonctions caractéristiques

Rappels de cours. Vitesse dans la loi des grands nombres. Théorème limite centrale. Espérance et variance fixée. Convergence en loi : fonctions tests continues et bornées, fonctions test indicatrices (fonctions de répartition), fonctions test trigonométriques fonctions caractéristiques). Théorème de Paul Lévy. Liens entre les modes de convergence. Limite constante, Lemme de Slutsky, utilisation en statistique.

Exercice 6.1 (Lois géométriques et loi exponentielles)**.**

1. *Soit X_k de loi géométrique de paramètre $0 < p_k < 1$. Montrer que si $\lim_{k\to\infty} kp_k = \lambda > 0$ alors $(k^{-1}X_k)_{k\geq 1}$ converge en loi vers la loi exponentielle de paramètre λ ;*
2. *Montrer que si Y suit la loi exponentielle de paramètre λ alors $\lfloor Y \rfloor$ et $Y - \lfloor Y \rfloor$ sont indépendantes et $1 + \lfloor Y \rfloor$ suit la loi géométrique de paramètre $e^{-\lambda}$.*

Solution.

1. On a $\varphi_{X_k}(t) = \mathbb{E}(e^{itX_k}) = p_k \sum_{n=1}^{\infty}(1-p_k)^{n-1}e^{itn} = \frac{p_k}{e^{-it}-(1-p_k)}$ et donc

$$\varphi_{k^{-1}X_k}(t) = \mathbb{E}(e^{itk^{-1}X_k}) = \varphi_{X_k}(k^{-1}t) = \frac{p_k}{e^{itk^{-1}} - (1-p_k)} = \frac{kp_k}{it + ko(1/k) + kp_k} \to \frac{\lambda}{it+\lambda}.$$

   La même méthode permet d'établir que $\Gamma(m,\lambda) = \mathrm{Exp}(\lambda)^{*m}$ est limite de lois binomiales négatives $\mathrm{Geom}(p_n)^{*m}$ contractées (du jeu de pile ou face au processus de Poisson sur $\mathbb{R}_+$) ;

2. Pour tout $n \geq 0$ et tout $u \in [0,1]$,

$$\mathbb{P}(\lfloor Y \rfloor = n, Y - \lfloor Y \rfloor \leq u) = \mathbb{P}(n \leq Y \leq n+u) = e^{-\lambda n} - e^{-\lambda(n+u)} = e^{-\lambda n}(1-e^{-\lambda u}).$$

   En posant $u = 1$ on obtient que $\lfloor Y \rfloor$ suit la loi géométrique sur $\mathbb{N}$ de paramètre $e^{-\lambda}$ (donc $1 + \lfloor Y \rfloor$ suit la loi géométrique sur $\mathbb{N}^*$ de paramètre $e^{-\lambda}$), tandis qu'en sommant sur $n \in \mathbb{N}$ on obtient la loi de $Y - \lfloor Y \rfloor$, celle d'une exponentielle tronquée sur $[0,1]$. L'indépendance de $\lfloor Y \rfloor$ et $Y - \lfloor Y \rfloor$ découle du fait que la formule est le produit des deux lois obtenues.

Exercice 6.2 (Convergence en probabilité et convergence en loi)**.**

1. *Montrer que si $(X_n)_{n\geq 1}$ est une suite de v.a.r. et si X est une v.a.r. constante, alors $(X_n)_{n\geq 1}$ converge en loi vers X si et seulement si $(X_n)_{n\geq 1}$ converge en probabilité vers X ;*
2. *Montrer que si $(X_n)_{n\geq 1}$ converge en loi vers une v.a.r. X et si $(Y_n)_{n\geq 1}$ converge en loi vers une constante c alors $(X_n, Y_n)_{n\geq 1}$ converge en loi vers (X,c) (lemme de Slutsky) ;*
3. *Montrer que si $(X_n)_{n\geq 1}$ converge en loi vers une v.a.r. X et si $(Y_n)_{n\geq 1}$ converge en probabilité vers une v.a.r. Y indépendante des X_n alors $(X_n, Y_n)_{n\geq 1}$ converge en loi vers (X, Y') où Y' est indépendante de X et de même loi que Y (version du lemme de Slutsky) ;*
4. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. de carré intégrable, de moyenne m et de variance $\sigma^2 > 0$. En notant $\hat{m}_n = (X_1 + \cdots + X_n)/n$ et $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \hat{m}_n)^2$, montrer que*

$$\sqrt{n}\frac{\hat{m}_n - m}{\hat{\sigma}_n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

Solution.

1. Si $(X_n)_{n\geq 1}$ converge en probabilité vers X alors elle converge en loi vers X. La réciproque est fausse en général. Cependant, si X est constante et égale à c, alors $F_X = \mathbf{1}_{[c,\infty[}$ et

$$\begin{aligned}\mathbb{P}(|X_n - c| > \varepsilon) &= \mathbb{P}(X_n \in [c-\varepsilon, c+\varepsilon]^c) = 1 - (F_{X_n}(c+\varepsilon) - F_{X_n}^-(c-\varepsilon)) \\ &\leq 1 - (F_{X_n}(c+\varepsilon) - F_{X_n}(c-\varepsilon)) \\ &\xrightarrow[n\to\infty]{} 1 - (F_X(c+\varepsilon) - F_X(c-\varepsilon)) = 1 - (1-0) = 0.\end{aligned}$$

2. Comme $Y_n \xrightarrow{\mathbb{P}} c$, et comme pour tout $t \in \mathbb{R}$, la fonction $y \mapsto e^{ity}$ est uniformément continue sur $\mathbb{R}$, on a que pour tous $s,t \in \mathbb{R}$ et tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour n assez grand,

$$\left|\mathbb{E}(e^{isX_n+itY_n}) - \mathbb{E}(e^{isX_n+itc})\right| \leq \mathbb{E}(\left|e^{itY_n} - e^{itc}\right| \mathbf{1}_{|Y_n-c|\leq\eta}) + 2\mathbb{P}(|Y_n - c| > \eta) \leq \varepsilon + 2\varepsilon.$$

Alternativement on peut utiliser le caractère Lipschitz plutôt que l'uniforme continuité ce qui est au bout du compte plus simple encore :

$$\left|\mathbb{E}(e^{isX_n+itY_n})-\mathbb{E}(e^{isX_n+itc})\right|\le\mathbb{E}(\left|e^{itY_n}-e^{itc}\right|\mathbf{1}_{|Y_n-c|\le\eta})+2\mathbb{P}(|Y_n-c|>\eta)\le|t|\eta+2\varepsilon.$$

D'autre part, comme $X_n\xrightarrow{\text{loi}}X$, on a, pour tous $s,t\in\mathbb{R}$,

$$\mathbb{E}(e^{isX_n+itc})=e^{itc}\mathbb{E}(e^{isX_n})\underset{n\to\infty}{\longrightarrow}e^{itc}\mathbb{E}(e^{isX})=\mathbb{E}(e^{isX+itc}).$$

3. On peut réutiliser la preuve précédente.
4. On peut utiliser le lemme de Slutsky car $\sqrt{n}(\hat{m}_n-m)$ converge en loi vers $\mathcal{N}(0,\sigma^2)$ par le TCL tandis que $\hat{\sigma}_n^2$ converge en probabilité (et même p.s.) vers σ^2 par la LGN.

Exercice 6.3 (Méthode Delta)**.** *Supposons que $A_n(Z_n-B_n)\xrightarrow{\text{loi}}L$ où $(A_n)_{n\ge1}$ et $(B_n)_{n\ge1}$ sont des suites déterministes telles que $A_n\to\infty$ et $B_n\to B$. Montrer que si $f:\mathbb{R}\to\mathbb{R}$ est une fonction $\mathcal{C}^1$ telle que $f'(B)\neq0$, alors $a_n(f(Z_n)-b_n)\to L$ en loi, où $a_n=A_n/f'(B)$ et $b_n=f(B_n)$. Application aux fluctuations en statistique : $A_n(Z_n-B_n)=\sqrt{n}(\widehat{\theta}_n-\theta)$. Que se passe-t-il si $f'(B)=0$? Et au delà ?*

Solution. Par une formule de Taylor ou par le théorème des valeurs intermédiaires, on a

$$f(Z_n)-f(B_n)=f'(W_n)(Z_n-B_n)$$

où W_n est une variable aléatoire située entre B_n et Z_n. Comme $A_n\to\infty$, le lemme de Slutsky donne $Z_n-B_n\to0$ en loi, et donc en $\mathbb{P}$ car la limite est déterministe. Par conséquent $W_n-B_n\to0$ en $\mathbb{P}$ et donc $W_n\to B$ en $\mathbb{P}$. La continuité de f' en B assure que $f'(W_n)\to f'(B)$ en $\mathbb{P}$, d'où $f'(W_n)/f'(B)\to1$ en $\mathbb{P}$, et à nouveau grâce au lemme de Slutsky, on a enfin

$$a_n(f(Z_n)-b_n)=\frac{A_n}{f'(B)}(f(Z_n)-f(B_n))=\frac{f'(W_n)}{f'(B)}A_n(Z_n-B_n)\xrightarrow{\text{loi}}P.$$

Note : la vitesse et la loi de fluctuation ne sont pas modifiées !

Note : si $f'(B)=0$ c'est la première dérivée non-nulle qui compte.

Exercice 6.4 (Convergences de marginales)**.** *Soient $(X_n)_{n\ge1}$, $(Y_n)_{n\ge1}$, X, Y des v.a.r.*

1. *Montrer que si $X_n\to X$ p.s. et $Y_n\to Y$ p.s. alors $(X_n,Y_n)\to(X,Y)$ p.s.*
2. *Montrer que la propriété reste vraie si on remplace la convergence p.s. par la convergence en probabilités. Montrer que cela ne fonctionne pas pour la convergence en loi.*

Solution.

1. $\|(X_n,Y_n)-(X,Y)\|^2=|X_n-X|^2+|Y_n-Y|^2\to0$ p.s.
2. $\mathbb{P}(\|(X_n,Y_n)-(X,Y)\|>\varepsilon)\le\mathbb{P}(|X_n-X|>\varepsilon/2)+\mathbb{P}(|Y_n-Y|>\varepsilon/2)\to0$
3. Contre exemple : X et Y indépendantes et de même loi et $(X_n,Y_n)=(X,X)$. On a $\mathcal{L}(X_n)=\mathcal{L}(X)$ donc $X_n\to X$ en loi, et $\mathcal{L}(Y_n)=\mathcal{L}(X)=\mathcal{L}(Y)$ donc $Y_n\to Y$ en loi, et $(X_n,Y_n)\to(X,X)$ en loi, mais $\mathcal{L}((X,X))\neq\mathcal{L}(X)\otimes\mathcal{L}(Y)=\mathcal{L}((X,Y))$ si X et Y ne sont pas constantes.

Exercice 6.5 (Critère de densité et d'inversion)**.** *Soit X une v.a.r. de loi μ et de fonction caractéristique φ donnée par $\varphi(t)=\mathbb{E}(e^{itX})=\int e^{itx}\,d\mu(x)$. On se propose d'établir que si φ est Lebesgue intégrable alors μ possède une densité continue et bornée f, qui vérifie, pour tout $x\in\mathbb{R}$,*

$$f(x)=\frac{1}{2\pi}\int e^{-itx}\varphi(t)\,dt.$$

1. *Montrer que pour tous $y,\sigma>0$, en notant γ_σ la densité de $\mathcal{N}(0,\sigma^2)$,*

$$2\pi\int\gamma_\sigma(y-x)\,d\mu(x)=\int e^{-ity}\varphi(t)e^{-\frac{\sigma^2t^2}{2}}\,dt;$$

2. *Montrer que si $Y\sim\mathcal{N}(0,1)$ est une v.a.r. indépendante de X et alors $X+\sigma Y$ a pour densité*

$$y\mapsto\int\gamma_\sigma(y-x)\,d\mu(x);$$

3. *En déduire le résultat escompté en faisant tendre σ vers 0.*

Solution.

1. Pour tous $t, y \in \mathbb{R}$, $e^{-ity}\varphi(t) = \int e^{it(x-y)}\,d\mu(x)$, ce qui donne, en exploitant le théorème de Fubini-Tonelli et la formule pour la fonction caractéristique de $\mathcal{N}(0, 1/\sigma^2)$ au point $(x-y)$,
$$\int e^{-ity}\varphi(t)e^{-\frac{\sigma^2t^2}{2}}\,dt = \iint e^{it(x-y)}e^{-\frac{\sigma^2t^2}{2}}\,dt\,d\mu(y) = \sqrt{\frac{2\pi}{\sigma^2}}\int e^{-\frac{(x-y)^2}{2\sigma^2}}\,d\mu(y) = 2\pi\int \gamma_\sigma(y-x)\,d\mu(y);$$
2. Il suffit d'utiliser une fonction test et le théorème de Fubini-Tonelli ;
3. La formule de la question 1 et le théorème de convergence dominée entraînent que la densité f_σ de $X+\sigma Y$ converge uniformément sur tout compact vers $f = \int e^{-it\cdot}\varphi(t)\,dt$. Or $X+\sigma Y$ converge p.s. et donc en loi vers X quand $\sigma \to 0$ et donc $\int_I f_\sigma(y)\,dy \to \mu(I)$ pour tout intervalle I quand $\sigma \to 0$. Il en découle que $\mu(I) = \int_I f(y)\,dy$ pour tout intervalle I.

Exercice 6.6 (Convergence en loi et fonction caractéristique)**.**

1. *Soit $(X_k)_{k\geq 1}$ une suite de v.a.r. i.i.d. gaussiennes centrées réduites. Montrer que*
$$Y_n = \frac{1}{n}\sum_{k=1}^n \sqrt{k}X_k \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, 1/2);$$
2. *Soit X et Y deux v.a.r. i.i.d. de loi uniforme sur $[-1,1]$. Déterminer la loi et la fonction caractéristique de $X+Y$. En déduire que $f(x) = \pi^{-1}(\sin(x)/x)^2$ est une densité de probabilité ;*
3. *Déterminer la fonction caractéristique de la loi de densité $p(x) = \frac{1}{2}e^{-|x|}$. En déduire celle de la loi de Cauchy, puis celle de la loi de la moyenne arithmétique de n v.a.r. i.i.d. de loi de Cauchy. En déduire un contre exemple pour la LGN.*

Solution.

1. On peut observer que Y_n est gaussienne de moyenne 0 et variance $\frac{1}{n^2}\sum_{k=1}^n k \to \frac{1}{2}$ et donc, en utilisant les fonctions caractéristiques, $(Y_n)_{n\geq 1}$ converge en loi vers $\mathcal{N}(0, 1/2)$;
2. La loi de $X+Y$ a pour densité $f_{X+Y}(x) = (\frac{1}{2}\mathbf{1}_{[-1,1]} * \frac{1}{2}\mathbf{1}_{[-1,1]})(x) = \frac{1}{4}(2-|x|)\mathbf{1}_{[-2,2]}(x)$. D'autre part, $\varphi_{X+Y}(t) = \varphi_X(t)^2 = (\sin(x)/x)^2$. Comme φ_{X+Y} est Lebesgue intégrable, on a, par inversion, $f(x) = \frac{1}{2\pi}\int e^{-itx}\varphi_{X+Y}(t)\,dt$. En prenant $x=0$ on obtient $\int (\sin(x)/x)^2\,dx = \pi$;
3. On a $\int e^{itx}\frac{1}{2}e^{-|x|}\,dx = 1/(1+t^2)$. Comme cette fonction est intégrable, on obtient, par inversion, $\frac{1}{2}e^{-|x|} = \frac{1}{2\pi}\int e^{-itx}\frac{1}{1+t^2}\,dt$ et donc $\varphi_{\text{Cauchy}}(x) = e^{-|x|}$. Si $C_1, \ldots, C_n$ dont des v.a.r. i.i.d. de loi de Cauchy, alors $\varphi_{(C_1+\cdots+C_n)/n}(t) = (\varphi_{C_1}(t/n))^n = e^{-|t|} = \varphi_{C_1}(t)$, et donc $(C_1+\cdots+C_n)/n$ suit la loi de Cauchy ! Pas de LGN ici car la loi de Cauchy n'a pas de moyenne.

Exercice 6.7 (Lois de Cauchy et moyenne harmonique empirique)**.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. de densité f bornée, paire (i.e. symétrique), continue en 0, telle que $f(0) > 0$.*

1. *Montrer que $Y_n = \frac{1}{n}\sum_{k=1}^n \frac{1}{X_k}$ converge en loi vers une loi de Cauchy. On rappelle que*
$$\int_0^\infty \frac{1-\cos(u)}{u^2}\,du = \lim_{r\to\infty}\int_0^r \frac{\sin(u)}{u}\,du = \frac{\pi}{2};$$
2. *En déduire la convergence en loi de la moyenne harmonique empirique $H_n = \left(\frac{1}{n}\sum_{k=1}^n \frac{1}{X_k}\right)^{-1}$;*
3. *Que se passe-t-il si $f(0) = 0$? Si f est discontinue en 0 ou si f n'est pas symétrique ?*

Solution.

1. Par indépendance, équidistribution, et symétrie de f, on a, pour tout $t \in \mathbb{R}$,
$$\mathbb{E}(e^{itY_n}) = \mathbb{E}(e^{i\frac{t}{nX_1}})^n = \left(\int \cos\left(\frac{t}{nx}\right)f(x)\,dx\right)^n = \left(1 + 2\int_0^\infty \left(\cos\left(\frac{t}{nx}\right) - 1\right)f(x)\,dx\right)^n.$$
Le changement de variable $u = \frac{t}{nx}$ donne $dx = \frac{|t|}{nu^2}du$, et par convergence dominée,
$$\mathbb{E}(e^{itY_n}) = \left(1 + \frac{2|t|}{n}\int_0^\infty \frac{\cos(u)-1}{u^2}f\left(\frac{t}{nu}\right)du\right)^n \xrightarrow[n\to\infty]{} e^{-\pi f(0)|t|}.$$
On reconnaît la fonction caractéristique d'une loi de Cauchy ;

2. Si C suit une loi de Cauchy standard de densité $x \mapsto 1/(\pi(1+x^2))$ alors pour toute fonction continue et à support compact $f : \mathbb{R} \to \mathbb{R}$, il vient, en posant $y = 1/x$ sur $\mathbb{R}_+$ et sur $\mathbb{R}_-$,

$$\mathbb{E}(f(1/C)) = \int \frac{f(1/x)}{\pi(1+x^2)}\,dx = \int \frac{f(y)}{\pi(1+1/y^2)}\,\frac{dy}{y^2} = \int \frac{f(y)}{\pi(y^2+1)}\,dy = \mathbb{E}(f(C)),$$

et donc $1/C$ a la même loi que C. Il en découle que H_n converge en loi vers la loi de Cauchy de fonction caractéristique $t \mapsto e^{-c|t|}$ où $1/c = \pi f(0)$. Contrôle : $f(x) = 1/(\pi(1+x^2))$! Rappel : si $Y_n \to Y$ en loi alors $h(Y_n) \to h(Y)$ en loi pour toute fonction continue h (dans le cas d'espèce $f(y) = 1/y$ mais on peut séparer $\mathbb{R}_+$ et $\mathbb{R}_-$ pour éviter le problème en 0) ;

3. Si $f(0) = 0$ alors $Y_n \to 0$ en loi (fonction caractéristique $\equiv 1$ et on pourrait utiliser la première dérivée non nulle en zéro pour f et adapter la vitesse). Si f n'est pas continue en 0 alors c'est $f^-(0) = f^+(0)$ qui surgira en lieu de place de $f(0)$. Si f n'est pas symétrique alors le terme en sinus ne disparaît pas.

Exercice 6.8 (Produits et convergences). *Soit $(U_k)_{k\geq 1}$ une suite de v.a.r. i.i.d. de loi uniforme sur $[0,1]$, et soit $\alpha > 0$ un réel fixé.*

1. *Montrer que $(X_n)_{n\geq 1}$ définie par $X_n = (U_1 \cdots U_n)^{\alpha/n}$ converge p.s. et préciser la limite ;*
2. *Montrer que $(Y_n)_{n\geq 1}$ définie par $Y_n = (X_n e^{\alpha})^{\sqrt{n}}$ converge en loi et déterminer la limite.*

Solution.

1. On a $-\ln(U_k) \sim \mathrm{Exp}(1)$ et donc, par la LGN, p.s. $X_n = e^{\alpha \frac{\sum_{k=1}^n \ln(U_k)}{n}} \to e^{\alpha}$;
2. TCL et continuité de $\exp$: $(X_n e^{\alpha})^{\sqrt{n}} = \exp\left(\alpha\sqrt{n}\left(\frac{\sum_{k=1}^n \ln(U_k)}{n} + 1\right)\right) \to e^{\alpha Z}$ avec $Z \sim \mathcal{N}(0,1)$ (stabilité par fonction continue de la convergence en loi).

Exercice 6.9 (Processus auto-régressif). *Soit X_0 une v.a.r. et $X_{n+1} = Q_{n+1}(X_n + E_{n+1})$ pour tout $n \geq 0$, où $(Q_n)_{n\geq 1}$ et $(E_n)_{n\geq 1}$ sont des suites indépendantes de v.a.r. i.i.d. de loi uniforme sur $[0,1]$ et exponentielle de paramètre λ, indépendantes de X_0. Montrer que quelque soit X_0, la suite $(X_n)_{n\geq 0}$ converge en loi vers la loi de $\sum_{n=1}^{\infty} Q_1 \cdots Q_n E_n$. La convergence a-t-elle lieu p.s. ?*

Solution. En utilisant la formule auto-regressive et les hypothèses d'indépendance et de stationnarité, on obtient, par récurrence sur n, la formule et l'identité en loi

$$X_n = Q_n \cdots Q_1 X_0 + \sum_{k=1}^{n} Q_n \cdots Q_{n-k+1} E_{n-k+1} \overset{\text{loi}}{=} Q_1 \cdots Q_n X_0 + \sum_{k=1}^{n} Q_1 \cdots Q_k E_k.$$

On a $\mathbb{E}(Q_1 \cdots Q_n) = \mathbb{E}(Q)^n$. Or $\mathbb{E}(Q) < 1$ car $\mathbb{P}(0 \leq Q < 1) = 1$ et donc par convergence monotone $\mathbb{E}(\sum_n Q_1 \cdots Q_n) = \sum_n \mathbb{E}(Q)^n < \infty$, donc $\mathbb{P}(\sum_n Q_n < \infty) = 1$ et donc $\mathbb{P}(Q_1 \cdots Q_n \to 0) = 1$. Par conséquent, $Q_1 \cdots Q_n X_0 \to 0$ p.s. D'autre part, par convergence monotone, $\sum_{k=1}^n Q_1 \ldots Q_k E_k$ converge p.s. vers la série aléatoire $\sum_{n=1}^{\infty} Q_1 \cdots Q_n E_n$. Cependant, il n'y a pas convergence p.s. de X_n (on rajoute de l'aléatoire non négligeable à chaque étape) ! La loi limite a pour moyenne $\lambda^{-1}\mathbb{E}(Q_1)/\mathbb{E}(1-Q_1)$. L'oubli de la condition initiale apparaît clairement sur la formule suivante :

$$\mathbb{E}(X_n) = \mathbb{E}(Q)^n \mathbb{E}(X_0) + \mathbb{E}(Q_1)\frac{1-\mathbb{E}(Q_1)^n}{\lambda(1-\mathbb{E}(Q_1))}.$$

La loi limite vérifie l'équation en loi $F \overset{\text{loi}}{=} Q_1(F + E_1)$ où F est indépendante de Q_1, E_1.

Exercice 6.10 (Extrêmes). *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. de loi L, de fonction de répartition F. On s'intéresse au comportement asymptotique de $M_n = \max(X_1, \ldots, X_n)$.*

1. *Montrer que pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}$ on a $F_{M_n}(x) = F^n(x)$;*
2. *Weibull : Montrer que si $L = \mathrm{Unif}([0,\theta])$ alors $n\big(\theta^{-1}M_n - 1\big)$ converge en loi quand $n \to \infty$.*
3. *Gumbel : montrer que si $L = \mathrm{Exp}(\lambda)$ alors $\lambda M_n - \ln(n)$ converge en loi quand $n \to \infty$.*
4. *Fréchet : montrer que si $L = \mathrm{Cauchy}$ alors $\pi n^{-1} M_n$ converge en loi quand $n \to \infty$.*
5. *Cas dégénéré : montrer que si $L = \mathrm{Bern}(p)$ avec $0 < p < 1$ alors $M_n \to 1$ p.s.*
6. *Montrer que $M_n \to x_F$ p.s. où $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} \in \mathbb{R} \cup \{\infty\}$.*

Solution.

1. $F_{M_n}(x) = \mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x) \cdots \mathbb{P}(X_n \leq x) = F^n(x)$. Cette quantité tend vers 0 ou 1 selon que $F(x) < 1$ ou $F(x) = 1$. Cela suggère que le comportement asymptotique de M_n dépend du comportement de la loi de L au bord droit de son support. On cherche donc $(a_n, b_n)_{n\geq 1}$ de sorte que $\mathbb{P}(a_n M_n + b_n \leq x)$ tende vers une fonction de répartition d'une loi, si possible non dégénérée c'est-à-dire non Dirac (pas toujours possible : cas Bernoulli).

2. Pour tout $x \leq 0$ on a $n^{-1}x + 1 \leq 1$ et $\mathbb{P}(M_n \leq \theta(n^{-1}x+1)) = (n^{-1}x+1)^n \to e^x$. Ainsi, $\lim_{n\to\infty} F(x) = e^x \mathbf{1}_{\mathbb{R}_-}(x) + \mathbf{1}_{\mathbb{R}_+}(x)$ pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition de $-E$ où $E \sim \mathrm{Exp}(1)$ (on dit parfois qu'il s'agit de la loi de Weibull des extrêmes, à ne pas confondre avec les lois de Weibull utilisées pour modéliser les durées de vie).

   Remarque : cela donne la vitesse et la loi de fluctuation (non gaussienne) de l'estimateur $\widehat{\theta}_n = \max\{U_1, \ldots, U_n\}$ de θ où $U_1, \ldots, U_n$ sont i.i.d. de loi uniforme sur $[0, \theta]$ (il s'agit d'un modèle statistique non régulier).

3. Pour tout $x \in \mathbb{R}$ tel que $x + \ln(n) \geq 0$ (toujours vrai si n assez grand)

   $$\mathbb{P}(\lambda M_n - \ln(n) \leq x) = (1 - n^{-1}e^{-x})^n \to e^{-e^{-x}}.$$

   Loi de Gumbel. Application : on peut approcher $\mathbb{P}(M_n \leq x)$ par $F(\lambda x - \ln(n))$ où F est la fonction de répartition de la loi de Gumbel. Par exemple si $\lambda = 1/10$, $x = 50$, $n = 100$, on a

   $$\mathbb{P}(M_{100} \geq 50) \approx e^{-e^{-(\lambda x - \ln(n))}} = 0,49023$$

   tandis que le calcul exact donne

   $$\mathbb{P}(M_{100} \geq 50) = 1 - (1 - e^{-\lambda x})^n = 0,49139.$$

4. Comme $\arctan(x) = \pi/2 - 1/x + \mathcal{O}_{x\to\infty}(1/x^2)$, pour tout $x \geq 0$,

   $$\mathbb{P}\Big(M_n \leq \frac{nx}{\pi}\Big) = \left(\int_{-\infty}^{nx\pi^{-1}} \frac{dy}{\pi(1+x^2)}\right)^n = \Big(1 - \frac{1}{nx} + \mathcal{O}_{x\to\infty}(n^{-2})\Big)^n \to e^{-1/x}.$$

   Donc on a

   $$\lim_{n\to\infty} F_{\pi n^{-1} M_n}(x) = e^{-1/x} \mathbf{1}_{\mathbb{R}_+^*}(x)$$

   pour tout $x \in \mathbb{R}$. La limite est la fonction de répartition de la loi de Fréchet.

5. On a $M_n = 1$ sur $\{n \geq T\}$ où $T = \inf\{n \geq 1 : X_n = 1\}$, et comme $T \sim \mathrm{Geom}(1/p)$, on a $\mathbb{P}(T < \infty) = 1$, d'où $\mathbb{P}(M_n \to 1) = 1$.

6. Pour tout $x < x_F$, on a $F(x) < 1$ et donc

   $$\mathbb{P}(M_n \leq x) = F^n(x) \to 0.$$

   De plus, dans le cas où $x_F < \infty$, on a pour tout $x \geq x_F$, $F(x) = 1$, et donc

   $$\mathbb{P}(M_n \leq x) = F^n(x) \to 1.$$

   Ainsi, la suite $(M_n)_{n\geq 1}$ converge en probabilité vers x_F, et comme elle est croissante, elle converge presque sûrement vers la même limite x_F.

Exercice 6.11 (Théorème de Cramèr-Wold)**.** *Montrer que la loi d'un vecteur aléatoire X de $\mathbb{R}^n$ est caractérisée par les lois de ses projections de dimension 1, c'est-à-dire $\langle X, v\rangle$ pour tout $v \in \mathbb{R}^n$.*

Solution. Fonctions caractéristiques : $\varphi_{\langle X,v\rangle}(t) = \varphi_X(tv)$ pour tout $v \in \mathbb{R}^n$ et tout $t \in \mathbb{R}$.

7 Vecteurs gaussiens, intervalles de confiance, TCL multivarié

Rappels de cours. Intervalles de confiance ; vecteurs gaussiens (définition, densité, indépendance, simulation) ; TCL multivarié.

Exercice 7.1 (Sondage)**.** *Dans une population de très grande taille, un sondage, effectué par tirage uniforme sans remise, sur la popularité du premier ministre indique que 51% des personnes interrogées sont favorables à sa politique. Proposer une modélisation avec une loi hypergéométrique, puis avec des v.a.r. i.i.d. de loi de Bernoulli. Donner un intervalle bilatéral symétrique de niveau de confiance asymptotique* 95% *pour la proportion* p *de personnes favorables au premier ministre. Le sondage a été réalisé auprès de* $n = 1000$ *personnes. Même question si* $n = 10000$.

Solution. Le véritable sondage consiste en un tirage sans remise d'un nombre fixe $n = n_1 + n_2$ d'individus dans une population finie à deux types de taille $N = N_1 + N_2$. On a bien sûr $n_1 \le N_1$ et $n_2 \le N_2$. Il se modélise donc avec une loi hypergéométrique à deux types grâce à la loi uniforme sur les parties à n éléments : la probabilité d'observer (n_1, n_2) est

$$\frac{\binom{N_1}{n_1}\binom{N_2}{n_2}}{\binom{N}{n}}.$$

Lorsque la population est de grande taille ($N \to \infty$), on peut approcher cette loi hypergéométrique par une loi binomiale : pour tout (n_1, n_2) avec $n = n_1 + n_2$ et $n_1 \le N_1$, $n_2 \le N_2$ on a

$$\lim_{N\to\infty, \frac{N_1}{N}\to p} \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}}{\binom{N}{n}} = p^{n_1}(1-p)^{n_2}\frac{n!}{n_1!n_2!}.$$

Ainsi, les n tirages sans remise dans la population de taile N deviennent en quelque sorte, dans l'asymptotique $N \to \infty$, des tirages i.i.d. avec remise, qui ne font intervenir que la proportion p. On peut aussi voir cette loi binomiale comme la somme n v.a.r. i.i.d. de loi de Bernoulli de paramètre p, qui codent pour le résultat du sondage sur chaque individus de l'échantillon de taille n. Le théorème central limite permet d'approcher cette loi binomiale par la loi normale $\mathcal{N}(np, np(1-p))$, et l'approximation est d'autant meilleure que la variance de la binomiale est élevée. Au bout du compte, on modélise donc le sondage par une suite $X_1, \dots, X_n$ de v.a.r. i.i.d. de loi de Bernoulli sur $\{0,1\}$ paramètre inconnu $p \in [0,1]$, avec $\mathbb{P}(X_i = 1) = p$. La LGN indique que p est estimable avec la moyenne empirique $\hat{p} = \frac{1}{n}(X_1 + \cdots + X_n)$. Le texte nous donne sa valeur : $\hat{p} = .51$. D'après le TCL et le lemme de Slutsky, pour tout $a \ge 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(-a \le \sqrt{\frac{n}{\hat{p}(1-\hat{p})}}(\hat{p}-p) \le a\right) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-\frac{x^2}{2}}\,dx = 1 - 2\Phi(-a)$$

où Φ est la fonction de répartition de $\mathcal{N}(0,1)$ (il est éclairant de faire un dessin). Cela donne

$$\lim_{n\to\infty} \mathbb{P}\left(p \in \left[\hat{p} - a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]\right) = 1 - 2\Phi(-a).$$

Posons $\alpha = .05$ de sorte que $1 - \alpha = 0.95$. On a $\Phi(-a) = \alpha/2$ lorsque $-a$ est le quantile $\alpha/2$ de $\mathcal{N}(0,1)$, c'est-à-dire $\Phi(-a) = \alpha/2$ (on a aussi $\Phi(a) = 1 - \alpha/2$). Cela donne $a = 1.96$. L'intervalle de confiance asymptotique de niveau $.95$ avec $n = 1000$ et $\hat{p} = .51$ vaut $[.48, .54]$. Pour $n = 10000$ il vaut $[.50, .52]$. Ces intervalles de confiance asymptotiques *de Wald* sont souvent imprécis en pratique, à cause de l'erreur commise en approchant la probabilité par la valeur asymptotique gaussienne fournie par le TCL. Il est alternativement possible d'utiliser des intervalles exacts non asymptotiques, comme ceux de Clopper-Pearson (chercher sur Internet !).

On peut utiliser un autre estimateur de σ, ou même majorer σ par $1/2$. On peut aussi utiliser une inégalité de concentration de Hoeffding, qui donne un intervalle exact (pessimiste).

Exercice 7.2 (Téléphonie)**.** *Un central téléphonique dessert* 5000 *abonnés. À tout instant, un abonné a une probabilité égale à 0.02 d'utiliser son téléphone. Les appels des abonnés sont supposés indépendants entre eux. Quel est le nombre d'abonnés que le central doit être capable de traiter simultanément pour qu'à tout instant, la probabilité que tous les abonnés ne puissent être satisfaits soit inférieure a 2,5% ?*

Solution. On se place à un instant donné. Soit X_i une variable aléatoire de Bernoulli, de paramètre $p = 0.02$, qui indique si l'abonné i utilise sont téléphone. Les v.a.r. $(X_i)_{1\le i\le n}$ sont i.i.d. et $n = 5000$. Le nombre d'appels simultanés est $\sum_{i=1}^n X_i \sim \mathrm{Binom}(n,p) \approx \mathcal{N}(np, np(1-p))$ (d'après TCL). On cherche donc un nombre m tel que $\mathbb{P}(np + \sqrt{np(1-p)}Z \ge m) \le 0.025$, où $Z \sim \mathcal{N}(0,1)$. Le quantile $1 - 0.025$ de la loi $\mathcal{N}(0,1)$ vaut $q = 1.96$ et on prend alors $m = \lceil np + q\sqrt{np(1-p)}\rceil = 120$.

Exercice 7.3 (Vecteurs gaussiens : double contre exemple)**.** *Soit X et Y deux v.a.r. indépendantes avec $X \sim \mathcal{N}(0,1)$ et $\mathbb{P}(Y=-1)=\mathbb{P}(Y=1)=1/2$. Déterminer la loi de $Z=XY$, et la covariance de X et Z. Les variables X et Z sont-elles indépendantes ? Le vecteur (X,Z) est-il gaussien ?*

Solution. Les v.a.r. Z et X ont même loi car pour toute fonction mesurable et bornée $f:\mathbb{R}\to\mathbb{R}$,

$$\mathbb{E}(f(Z)) = \mathbb{E}(f(X)\mathbf{1}_{Y=-1}) + \mathbb{E}(f(-X)\mathbf{1}_{Y=1}) = \mathbb{E}(f(X))\frac{1}{2} + \mathbb{E}(f(-X))\frac{1}{2} = \mathbb{E}(f(X))$$

où on a utilisé l'indépendante de X et de Y, la loi de Y, et $X \overset{\text{loi}}{=} -X$. Les v.a.r. X et Z sont décorrélées car $\mathrm{Cov}(X,Z)=\mathbb{E}(XZ)-\mathbb{E}(X)\mathbb{E}(Z)=\mathbb{E}(X^2Y)=\mathbb{E}(X^2)\mathbb{E}(Y)=0$, mais pas indépendantes car $\mathbb{E}(|Z||X|)=\mathbb{E}(X^2)=1\neq\frac{2}{\pi}=\mathbb{E}(|X|)^2=\mathbb{E}(|Z|)\mathbb{E}(|X|)$. Les composantes du vecteur aléatoire (X,Z) sont gaussiennes mais il ne s'agit pas d'un vecteur gaussien car la combinaison linéaire $X+Z=X(1+Y)$ n'est pas gaussienne car $\mathbb{P}(X+Z=0)=\mathbb{P}(1+Y=0)=1/2\notin\{0,1\}$.

Exercice 7.4 (Moyenne et matrice de covariance empiriques)**.** *Soient $X_1,\ldots,X_n$ des vecteurs aléatoires i.i.d. de $\mathbb{R}^d$, de moyenne $m\in\mathbb{R}^d$ et de matrice de covariance Σ. Trouver des estimateurs de m et de Σ sans biais et fortement consistants (i.e. qui convergent p.s. quand $n\to\infty$). On étudiera d'abord le cas univarié $d=1$ puis le cas général $d\geq 1$.*

Solution. Truc du statisticien pour fabriquer un estimateur : exprimer la quantité à estimer au moyen d'une espérance, et remplacer par l'espérance empirique. Un estimateur sans biais et fortement convergent (LGN !) de m est la moyenne empirique :

$$\hat{m}_n := \frac{1}{n}\sum_{k=1}^n X_k.$$

Pour la matrice de covariance, on dispose de deux estimateurs sans biais et convergents, tous deux appelés matrice de covariance empirique :

$$\hat{\Sigma}_n := \frac{1}{n}\sum_{k=1}^n (X_k-m)\otimes(X_k-m) \quad\text{et}\quad \hat{\Sigma}'_n := \frac{1}{n-1}\sum_{k=1}^n (X_k-\hat{m}_n)\otimes(X_k-\hat{m}_n).$$

Le premier nécessite de connaître m. La normalisation par $n-1$ plutôt que par n dans le second permet de garantir l'absence de biais et vient d'une sorte de perte de degré de liberté due au remplacement de m par son estimation empirique $\hat{m}_n$. Le fait que $\hat{\Sigma}_n$ soit sans biais est aussi évident que pour $\hat{m}_n$. La convergence p.s. $(\hat{\Sigma}_n)_{i,j}\to(\Sigma_n)_{i,j}$ découle tout simplement de la LGN appliquées aux v.a.r. i.i.d. $Y_k := (X_k-m_i)(X_k-m_j)$. Pour la convergence $(\hat{\Sigma}'_n)_{i,j}\to(\Sigma_n)_{i,j}$, les v.a.r. Y_k dépendent à présent de n, et il vaut alors mieux développer :

$$\begin{aligned}(n-1)\hat{\Sigma}_{i,j} &= \sum_{k=1}^n (X_k-\hat{m})_i(X_k-\hat{m})_j\\ &= \sum_{k=1}^n X_{k,i}X_{k,j} - \frac{2}{n}\left(\sum_{k=1}^n X_{k,i}\right)\left(\sum_{k'=1}^n X_{k',j}\right) + n\hat{m}_i\hat{m}_j\end{aligned}$$

et donc par LGN $(\hat{\Sigma}'_n)_{i,j}\to(\Sigma_n)_{i,j}-2\hat{m}_i\hat{m}_j+\hat{m}_i\hat{m}_j=(\Sigma_n)_{i,j}$ p.s. Pour établir ensuite que $(\hat{\Sigma})'$ est sans biais, on utilise, grâce à la décomposition ci-dessus,

$$(n-1)\mathbb{E}(\hat{\Sigma}_{i,j}) = n(\Sigma_{i,j}+m_im_j) - \frac{2}{n}(n(\Sigma_{i,j}+m_im_j)+n(n-1)m_im_j) + n\mathbb{E}(\hat{m}_i\hat{m}_j)$$

et

$$n\mathbb{E}(\hat{m}_i\hat{m}_j) = \frac{1}{n}\sum_{k,k'}\mathbb{E}(X_{k,i}X_{k',j}) = \frac{1}{n}(n(\Sigma_{i,j}+m_im_j)+n(n-1)m_im_j) = \Sigma_{i,j}+nm_im_j,$$

ce qui donne bien au total $(n-1)\mathbb{E}(\hat{\Sigma}_{i,j})=(n-2+1)\Sigma_{i,j}+(n-2n+n)m_im_j=(n-1)\Sigma_{i,j}$.

Exercice 7.5 (Maximum de vraisemblance)**.** *Soit $X_1,\ldots,X_n$ un échantillon de la loi $\mathrm{Exp}(\lambda)$. Trouver l'estimateur de maximum de vraisemblance de λ et calculer son biais.*

Solution. On a $L(X_1,\ldots,X_n;\lambda)=e^{-\lambda(x_1+\cdots+x_n)+n\log(\lambda)}$, qui donne $\mathrm{EML}_n=n/(X_1+\cdots+X_n)$. L'inégalité de Jensen donne $\mathbb{E}(\mathrm{EML}_n)\geq n/\mathbb{E}(X_1+\cdots+X_n)=\lambda$ et le biais est donc positif. La LGN indique que l'estimateur converge p.s. vers λ. Comme $X_1+\cdots+X_n\sim\mathrm{Gamma}(n,\lambda)$ on a

$$\mathbb{E}\left(\frac{1}{X_1+\cdots+X_n}\right)=\int_0^\infty\frac{\lambda^n}{(n-1)!}s^{(n-1)-1}e^{-\lambda s}\,ds=\frac{\lambda}{n-1}\int_0^\infty\frac{\lambda^{n-1}}{(n-1)!}s^{(n-1)-1}e^{-\lambda s}\,ds=\frac{\lambda}{n-1}.$$

Le biais vaut donc $(n/(n-1)-1)\lambda=\lambda(1/(n-1))\to 0$. Note : $n>1$ car $1/X_1$ n'est pas intégrable.

Exercice 7.6 (Mieux que la LGN et le TCL). *Soit $(X_n)_{n\geq 1}$ des v.a.r. i.i.d. de loi uniforme sur $[0,\theta]$.*

1. *Étudier la convergence, le biais, la variance, l'écart quadratique moyen, et la fluctuation asymptotique de l'estimateur $m_n = 2n^{-1}\sum_{i=1}^n X_i$ de θ. Construire un intervalle de confiance asymptotique ;*
2. *Montrer que l'estimateur $M_n = \max(X_1,\ldots,X_n)$ de θ a pour densité $(n/\theta)(x/\theta)^{n-1}\mathbf{1}_{[0,\theta]}(x)$ puis calculer $\mathbb{E}(M_k)$ et $\mathrm{Var}(M_k)$. Comparer avec l'estimateur moyenne empirique m_n ;*
3. *Montrer que $M_n \xrightarrow{\mathbb{P}} \theta$. Montrer de deux manières différentes que la convergence est p.s. ;*
4. *Montrer que $W_n = n(M_n/\theta - 1)$ converge en loi quand $n\to\infty$, et déterminer la limite. Construire un intervalle de confiance asymptotique.*
5. *Montrer que M_n est l'estimateur de maximum de vraisemblance de θ.*

Solution. Pour un estimateur $\hat\theta_n$ de θ, l'écart quadratique moyen possède toujours une décomposition (carré-du-)biais–variance : $\mathbb{E}((\hat\theta_n-\theta)^2) = (\mathbb{E}(\hat\theta_n)-\theta)^2 + \mathrm{Var}(\hat\theta_n)$. Cela découle du théorème de Pythagore dans L^2 : $\mathbb{E}((Z-\theta)^2) = \mathrm{Var}(Z) + (\mathbb{E}(Z)-\theta)^2$. Note : $\mathrm{Var}(Z) = \arg\min_{c\in\mathbb{R}}\mathbb{E}((Z-c)^2)$.

1. La LGN entraîne que $m_n\to\theta$ p.s. L'estimateur n'est pas biaisé : $\mathbb{E}(m_n)=\theta$. La variance est égale à l'écart quadratique moyen, et vaut $\frac{4}{n}\mathrm{Var}(X_1) = \frac{\theta^2}{3n} = \mathcal{O}(\frac{1}{n})$. La fluctuation asymptotique est gaussienne, de vitesse $\sqrt{n}$, car par le TCL, $\sqrt{n}(m_n-\theta) \xrightarrow{\text{loi}} \mathcal{N}(0,\frac{\theta^2}{3})$. L'intervalle de confiance asymptotique s'obtient avec le résultat de fluctuation. Nul besoin du lemme de Slutsky car l'inversion en θ est facile ici : si $Z\sim\mathcal{N}(0,1)$ et $J\subset\mathbb{R}$,
$$\mathbb{P}(Z\in J) \approx \mathbb{P}\left(\frac{\sqrt{3n}}{\theta}(m_n-\theta)\in J\right) = \mathbb{P}(\theta\in m_n/(1+(3n)^{-1/2}J))$$
Tout intervalle J tel que $\mathbb{P}(Z\in J) = 1-\alpha$ fournit l'intervalle de confiance $I = m_n/(1+(3n)^{-1/2}J)$ de niveau $1-\alpha$ pour θ. On peut chercher à choisir J de sorte que I soit petit ;
2. $F_{M_n}(x) = (x/\theta)^n\mathbf{1}_{[0,\theta]} + \mathbf{1}_{]\theta,\infty[}$ et $f_{M_n}(x) = F'_{M_n}(x) = (n/\theta)(x/\theta)^{n-1}\mathbf{1}_{[0,\theta]}(x)$. Par conséquent, $\mathbb{E}(M_n) = \frac{n}{n+1}\theta$ et $\mathrm{Var}(M_n) = \frac{n}{(n+1)^2(n+2)}\theta^2 = \mathcal{O}(\frac{1}{n^2})$. L'estimateur M_n est baisé (et asymptotiquement sans biais), mais il est plus rapide que m_n. L'écart quadratique moyen de M_n vaut $(\mathbb{E}(M_n)-\theta)^2 + \mathrm{Var}(M_n) = \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{2}{(n+1)(n+2)}\theta^2 = \mathcal{O}(\frac{1}{n^2})$;
3. Pour tout $\varepsilon>0$ assez petit, $\mathbb{P}(|M_n-\theta|>\varepsilon) = \mathbb{P}(M_n<\theta-\varepsilon) = ((\theta-\varepsilon)/\theta)^n$. Cela donne la convergence en probabilité de M_n vers θ, et même la convergence p.s. en utilisant le lemme de Borel-Cantelli pour ε fixé (il faut ensuite poser $\varepsilon = 1/k$ et faire $\cap_k$). Alternativement, comme p.s. $(M_n)_{n\geq 1}$ est positive croissante et majorée par θ, elle converge p.s. et dans L^1 vers une v.a.r. $\leq\theta$ de moyenne $\lim_{n\to\infty}\mathbb{E}(M_n) = \theta$, qui est forcément égale p.s. à θ ;
4. $F_{W_n}(x) = \mathbb{P}(M_n\leq\theta(1+x/n))$. Ainsi, $F_{W_n}(x)\to 1$ si $x>0$, tandis que si $x\leq 0$ alors, pour $n\gg 1$, $F_{W_n}(x) = (1+x/n)^n\to e^x$. Ainsi, $W_n\to -W$ en loi quand $n\to\infty$, où $W\sim\mathrm{Exp}(1)$. L'intervalle de confiance asymptotique s'obtient avec la fluctuation : Pour tout $J\subset\mathbb{R}$, on a
$$\mathbb{P}(-W\in J)\approx\mathbb{P}(W_n\in J) = \mathbb{P}(\theta\in M_n/(1+J/n)).$$
N'importe quel intervalle J tel que $\mathbb{P}(-W\in J) = 1-\alpha$ fournit l'intervalle de confiance asymptotique $I = M_n/(1+J/n)$ pour θ. On peut choisir J tel que I soit petit.
5. Pour tout réel $\theta\geq 0$, on a $L(\theta;X_1,\ldots,X_n) = \theta^{-n}\prod_{k=1}^n\mathbf{1}_{[0,\theta]}(X_k)$, qui vaut 0 si $\theta<M_n$ et qui décroît comme θ^{-n} si $\theta\geq M_n$. Donc $M_n = \arg\max_{\theta\geq 0}L(\theta;X_1,\ldots,X_n)$.

Exercice 7.7 (Graphes aléatoires). *On considère une population constituée de $n\geq 1$ individus, et la matrice symétrique $A_n = (A_{n,i,j})_{1\leq i,j\leq n}$ où $A_{n,i,j} := 1$ si les individus i et j sont amis et $A_{n,i,j} = 0$ sinon. On convient que $A_{n,i,i} = 0$. On se donne $(p_{n,i,j})_{n\geq 1;1\leq i,j\leq n}$ dans $[0,1]$ et des v.a.r. i.i.d. $(U_{i,j})_{j>i\geq 1}$ de loi uniforme sur $[0,1]$, et on pose $A_{n,i,j} = \mathbf{1}_{\{U_{i,j}\leq p_{n,i,j}\}}$ pour tous $1\leq i<j\leq n$.*

1. *Que représente le nombre $(A_n+I_n)^k_{i,j}$?*
2. *Montrer que s'il existe $p\in[0,1]$ tel que $p_{n,i,j} = p$ pour tous $1\leq i\neq j\leq n$ alors pour tout $1\leq i\leq n$, la v.a. $c_{n,i} := \sum_{j=1}^n A_{n,i,j}$ suit la loi binomiale $\mathrm{Binom}(n-1,p)$ et*
$$\frac{c_{n,i}}{n}\xrightarrow[n\to\infty]{\text{p.s.}} p \quad\text{et}\quad \frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\frac{c_{n,i}}{n}-p\right)\xrightarrow[n\to\infty]{\text{loi}}\mathcal{N}(0,1);$$
Comment se comportent $\mathbb{E}(c_{n,i})$ et $\mathrm{Var}(c_{n,i})$ quand $n\to\infty$?

3. *Montrer que pour tout i fixé, si on dispose des propriétés suivantes*
   (a) *stabilisation :* $\lim_{n\to\infty}\sum_{j=1}^n p_{n,i,j}=\lambda<\infty$
   (b) *dispersion :* $\lim_{n\to\infty}\max_{1\leq j\leq n} p_{n,i,j}=0$

   alors $c_{n,i}\overset{\text{loi}}{\to}\mathrm{Poi}(\lambda)$ *quand* $n\to\infty$*. Comment se comportent* $\mathbb{E}(c_{n,i})$ *et* $\mathrm{Var}(c_{n,i})$ *? Étudier le cas où* $p_{n,i,j}=p$ *et* $p_{n,i,j}=\lambda/(n-1)$ *pour tous* $1\leq i\neq j\leq n$*;*
4. *Si* μ *et* ν *sont des lois de probabilité sur* $\mathbb{N}$*, on définit leur distance en variation par*

$$d_V(\mu,\nu):=\frac{1}{2}\sum_{x\in\mathbb{N}}|\mu(x)-\nu(x)|\in[0,1].$$

   Montrer que d_V *est une distance sur l'ensemble des lois sur* $\mathbb{N}$*. Montrer que si* $(S_n)_{n\geq 1}$ *sont des v.a. sur* $\mathbb{N}$ *et si* μ *est une loi sur* $\mathbb{N}$ *alors les propriétés suivantes sont équivalentes :*
   (a) $(S_n)_{n\geq1}$ *converge en loi vers* μ *quand* $n\to\infty$*;*
   (b) $\lim_{n\to\infty}\mathbb{P}(S_n=x)=\mu(x)$ *pour tout* $x\in\mathbb{N}$*;*
   (c) $\lim_{n\to\infty}d_V(\mathrm{Loi}(S_n),\mu)=0$*;*
5. *Si* $\xi_1,\ldots,\xi_n$ *sont des v.a. indépendantes de lois de Bernoulli de paramètres respectifs* $p_1,\ldots,p_n$ *dans* $]0,1]$*, et si* $S_n:=\xi_1+\cdots+\xi_n$*, montrer l'inégalité de Le Cam :*

$$d_V(\mathrm{Loi}(S_n),\mathrm{Poi}(p_1+\cdots+p_n))\leq p_1^2+\cdots+p_n^2.$$

   En déduire une version quantitative de la convergence en loi de $c_{n,i}$ *vue précédemment;*
6. *Dans cette question, on lève l'hypothèse de symétrie :* $(A_{n,i,j})_{1\leq i\neq j\leq n}$ *sont des v.a. indépendantes de loi de Bernoulli et on pose* $p_{n,i,j}:=\mathbb{P}(A_{n,i,j}=1)$*. Les matrices* A_n *et* $(p_{n,i,j})_{1\leq i,j\leq n}$ *ne sont plus symétriques. La probabilité que l'individu* i *n'aime personne vaut*

$$\mathbb{P}(c_{n,i}=0)=(1-p_{n,i,1})\cdots(1-p_{n,i,n}).$$

   Montrer que si $p_{n,i,j}$ *ne dépend pas de* $i\neq j$ *et vérifie* $\sum_n n(1-p_n)^{n-1}<\infty$ *alors presque sûrement, pour* n *assez grand,* $c_{n,i}>0$ *pour tout* $1\leq i\leq n$*.*
7. *Dans cette question, on lève l'hypothèse d'indépendance. On considère un ensemble de* $n\geq 3$ *gènes numérotés* $1,\ldots,n$*. L'expression du gène* i *est codée par une v.a.r.* X_i*, de sorte que* $\mathbf{X}:=(X_1,\ldots,X_n)^\top\sim\mathcal{N}(0,C)$*. On note* $\hat{C}:=\frac{1}{N}\sum_{k=1}^N\mathbf{X}_k\mathbf{X}_k^\top$ *la matrice de covariance empirique d'un échantillon* $\mathbf{X}_1,\ldots,\mathbf{X}_N$ *de même loi que* $\mathbf{X}$*. Comment estimer* C *à partir de* $\hat{C}$ *? On s'intéresse aux deux modèles de graphes aléatoires suivants, construits avec* C *:*
   (a) *Les gènes* i *et* $j\neq i$ *sont amis ssi* $C_{i,j}\neq 0$*. Montrer que*

$$C_{i,j}=0\quad\text{ssi}\quad X_i\text{ et }X_j\text{ sont indépendantes};$$

   (b) *Les gènes* i *et* $j\neq i$ *sont amis ssi* $(C^{-1})_{i,j}\neq 0$ *(*C *est inversible). Montrer que*

$$(C^{-1})_{i,j}=0\quad\text{ssi}\quad X_i\text{ et }X_j\text{ sont conditionnellement indépendantes sachant }(X_k)_{k\neq i,j}.$$

Solution.

1. Nombre de chemins de longueur k allant de i à j;
2. LGN et TCL;
3. La fonction génératrice de $c_{n,i}$ s'écrit, pour $0<u<1$,

$$\mathbb{E}(u^{c_{n,i}})=\prod_{j=1}^n(1-p_{n,i,j}+p_{n,i,j}u)=\exp\left(\sum_{j=1}^n\ln(1-p_{n,i,j}(1-u))\right).$$

   Comme $-x^2\leq x+\ln(1-x)\leq 0$ valable pour $0\leq x\leq 1/2$, on a, pour n assez grand :

$$\left|\sum_{j=1}^n\ln(1-p_{n,i,j}(1-u))+\sum_{j=1}^n p_{n,i,j}(1-u)\right|\leq\sum_{j=1}^n(p_{n,i,j})^2\leq\left(\max_{1\leq j\leq n}p_{n,i,j}\right)\sum_{j=1}^n p_{n,i,j}.$$

   Le membre de droite de l'inégalité est équivalent à $\lambda\max_{1\leq j\leq n}p_{n,i,j}$, et par conséquent

$$\lim_{n\to\infty}\sum_{j=1}^n\ln(1-p_{n,i,j}(1-u))=\lim_{n\to\infty}-\sum_{j=1}^n p_{n,i,j}(1-u)=-\lambda(1-u),$$

   et donc $\lim_{n\to\infty}\mathbb{E}(u^{c_{n,i}})=\exp(-\lambda(1-u))$ comme attendu.

4. Donnons par exemple une preuve de (2)⇒(3). Pour tout $A \subset \mathbb{N}$ fini,

$$\sum_{x\in\mathbb{N}} |\mu_n(x) - \mu(x)| = \sum_{x\in A} |\mu_n(x) - \mu(x)| + \sum_{x\in A^c} |\mu_n(x) - \mu(x)|$$

où $\mu_n = \mathcal{L}(S_n)$. Par hypothèse, pour tout $\varepsilon > 0$ il existe un entier $N = N(A,\varepsilon)$ tel que le premier terme du membre de droite est majoré par $\mathrm{card}(A)\varepsilon$ pour tout $n \geq N$. Pour le second terme du membre de droite, on écrit

$$\sum_{x\in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x\in A^c} \mu_n(x) + \sum_{x\in A^c} \mu(x).$$

Puisqu'on a

$$\sum_{x\in A^c} \mu_n(x) = \sum_{x\in A} \mu(x) - \sum_{x\in A} \mu_n(x) + \sum_{x\in A^c} \mu(x)$$

on obtient

$$\sum_{x\in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x\in A} |\mu_n(x) - \mu(x)| + 2\sum_{x\in A^c} \mu(x).$$

Or μ est une probabilité, donc pour tout $\varepsilon > 0$, on peut choisir A fini vérifiant $\mu(A^c) \leq \varepsilon$.

5. Par récurrence sur n, si $X_1, \ldots, X_n$ est une suite de v.a. indépendantes sur $\mathbb{N}$ et $Y_1, \ldots, Y_n$ est une autre suite de v.a. indépendantes sur $\mathbb{N}$ alors

$$d_V(\mathrm{Loi}(X_1 + \cdots + X_n), \mathrm{Loi}(Y_1 + \cdots + Y_n)) \leq d_V(\mathrm{Loi}(X_1), \mathrm{Loi}(Y_1)) + \cdots + d_V(\mathrm{Loi}(X_n), \mathrm{Loi}(Y_n)).$$

Ensuite on montre que pour tout $p \in [0,1]$ on a $d_V(\mathrm{Poi}(p), \mathrm{Bern}(p)) \leq p^2$.

6. En définissant l'événement $\mathcal{D}_n := \{c_{n,1} \cdots c_{n,n} > 0\}$, il vient

$$\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{D}_n^c) = \sum_{n=1}^{\infty} (1 - (1 - (1-p_n)^{n-1})^n) \leq \sum_{n=1}^{\infty} n(1-p_n)^{n-1} < \infty.$$

Il ne reste plus qu'à utiliser le lemme de Borel-Cantelli.

7. La LGN donne $\lim_{N\to\infty} \hat{C} = C$ p.s. On peut estimer C à partir de $\hat{C}$, en utilisant un seuillage (même si C possède des coefficients nuls, la matrice aléatoire $\hat{C}$ n'en possède pas en général).
   (a) Soient $i < j$. Les composantes X_i et X_j du vecteur aléatoire (X_i, X_j) de $\mathbb{R}^2$ sont indépendantes ssi la densité est le produit des densités marginales. Or (X_i, X_j) suit la loi $\mathcal{N}(0, ((C_{i,i}, C_{i,j}), (C_{j,i}, C_{j,j})))$ dont la densité a une structure spéciale. On vérifie qu'elle est produit des densités marginales ssi la matrice de covariance (2×2) est diagonale, c'est-à-dire ssi $C_{i,j} = 0$. On peut aussi utiliser les fonctions caractéristiques ;
   (b) On peut supposer que $i = n-1$ et $j = n$. Décomposons la matrice C par blocs :

$$C = \begin{pmatrix} K & V \\ V^\top & L \end{pmatrix}$$

   où K est la matrice de covariance de $(X_1, \ldots, X_{n-2})$ et L la matrice de covariance de (X_{n-1}, X_n). Comme C est inversible, les matrices K et S le sont également. Comme $\mathbf{X}$ suit une loi gaussienne, la loi conditionnelle $\mathcal{L}((X_{n-1}, X_n) \,|\, X_1, \ldots, X_{n-2})$ est gaussienne de matrice de covariance $S = L - V^\top K^{-1} V$ (complément de Schur de K dans C). Par conséquent, X_{n-1} et X_n sont conditionnellement indépendantes sachant $X_1, \ldots, X_{n-2}$ ssi S est diagonale, c'est à dire ssi S^{-1} est diagonale. Or par inversion par blocs :

$$C^{-1} = \begin{pmatrix} K^{-1} + K^{-1} V S^{-1} V^\top K^{-1} & -K^{-1} V S^{-1} \\ -S^{-1} V^\top K^{-1} & S^{-1} \end{pmatrix}.$$

8 Adéquation à une loi, tests, estimations avec données i.i.d.

Exercice 8.1 (Théorème de Cochran[3]). *Soit X un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(m,\sigma^2 \mathrm{I}_n)$, et $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_p$ une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions $d_1,\ldots,d_p$ avec $d_1+\cdots+d_p = n$. Soit $\mathbf{P}_k$ la matrice du projecteur orthogonal sur E_k et $Y_k := \mathbf{P}_k X$ la projection orthogonale de X sur E_k.*

1. *Montrer que les projections $(Y_1,\ldots,Y_p)$ sont indépendantes et $Y_k \sim \mathcal{N}(\mathbf{P}_k m, \sigma^2 \mathbf{P}_k)$;*
2. *Monter que $\|Y_1 - \mathbf{P}_1 m\|^2,\ldots,\|Y_p - \mathbf{P}_p m\|^2$ sont indépendantes et $\sigma^{-2}\|Y_k - \mathbf{P}_k m\|^2 \sim \chi^2(d_k)$.*

Solution.

1. On se ramène tout d'abord au cas où $m=0$ par translation. Le vecteur aléatoire $Y = (Y_1,\ldots,Y_p)^\top$ de $\mathbb{R}^{np}$ s'écrit $Y = AX$ où A est la matrice de dimension $np \times n$
$$A = \begin{pmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_p \end{pmatrix}.$$
Il en découle que Y suit la loi $\mathcal{N}(0,\sigma^2 AA^\top)$. Pour tout $1 \leq i \leq p$, on a $\mathbf{P}_i = \mathbf{P}_i^\top = \mathbf{P}_i^2$. De plus, $\mathbf{P}_i\mathbf{P}_j = 0$ si $1 \leq i \neq j \leq p$ car $E_i \perp E_j$. Par conséquent, $AA^\top = \mathrm{Diag}(\mathbf{P}_1,\ldots,\mathbf{P}_p)$ est diagonale par blocs. Il en découle que $Y_1,\ldots,Y_p$ sont des vecteurs gaussiens indépendants avec $Y_k \sim \mathcal{N}(0,\sigma^2\mathbf{P}_k)$ pour tout $1 \leq k \leq p$.
2. Les v.a.r. $\|Y_1\|^2,\ldots,\|Y_p\|^2$ sont indépendantes. Trouvons leur loi. Comme $Y_k = P_k(\frac{X-m}{\sigma})$ on se ramène au cas $m=0$ at $\sigma=1$ par translation et dilatation. Ensuite le théorème spectral donne $P_k = O_k D_k O_k^\top$ avec O_k orthogonale et $D_k = \mathrm{diag}(1,\ldots,1,0,\ldots,0)$ où 1 est répété d_k fois, d'où
$$\|Y_k\|_2^2 = \|O_k D_k O_k^\top X\|_2^2 = \|D_k O_k^\top X\|_2^2 \overset{\text{loi}}{=} \|D_k X\|_2^2 = X_1^2 + \cdots + X_{d_k}^2 \sim \chi^2(d_k).$$

Exercice 8.2 (Échantillons gaussiens). *Soit $(X_1,\ldots,X_n)$ un échantillon de loi $\mathcal{N}(m,\sigma^2)$. On lui associe la* moyenne empirique *et la* variance empirique *définies respectivement par*
$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k \qquad \textit{et} \qquad \sigma_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \overline{X}_n)^2.$$
Monter que les variables aléatoires $\overline{X}_n$ et V_n^2 sont indépendantes avec
$$\overline{X}_n \sim \mathcal{N}\Big(m, \frac{\sigma^2}{n}\Big) \qquad \textit{et} \qquad \frac{(n-1)}{\sigma^2}\sigma_n^2 \sim \chi^2(n-1).$$
En particulier, la moyenne empirique studentisée *$T_n := \sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma_n}\right)$ vérifie $T_n \sim t(n-1)$. S'agit-il des estimateurs du maximum de vraisemblance de la moyenne et de la variance ?*

Solution. On se ramène tout d'abord à $m=0$ et $\sigma=1$ par translation et dilatation. Soit 1_n le vecteur de $\mathbb{R}^n$ dont toutes les coordonnées sont égales à 1, et soit $E_1 = \mathbb{R}1_n$ le sous-espace vectoriel de $\mathbb{R}^n$ engendré. La matrice de la projection orthogonale sur E_1 est donnée par
$$\mathbf{P}_1 = \frac{1_n 1_n^\top}{\|1_n\|^2} = \frac{1}{n}1_n 1_n^\top.$$
Le sous-espace $E_2 = E_1^\perp$ est de dimension $n-1$ et la matrice de la projection orthogonale sur E_2 est $\mathbf{P}_2 = \mathrm{I}_n - \mathbf{P}_1$. On a $Y_1 = \mathbf{P}_1 X = \overline{X}_n 1_n$ et $Y_2 = \mathbf{P}_2 X = (X_1 - \overline{X}_n, \ldots, X_n - \overline{X}_n)^\top$, ce qui entraîne $\|Y_2\|^2 = (n-1)V_n^2$. Le théorème de Cochran permet de conclure.

Exercice 8.3 (Test d'aquéquation du chi-deux). *Il s'agit d'un test non-paramétrique classique.*

1. *Soient $p=(p_1,\ldots,p_k)$, $q=(q_1,\ldots,q_k)$ des lois discrètes avec $p_i > 0$ pour tout $1 \leq i \leq k$, et*
$$D(p,q) := \sum_{i=1}^k \frac{(p_i - q_i)^2}{p_i}.$$
Montrer que $D(p,q) \geq 0$ avec égalité ssi $p=q$. S'agit-il d'une distance ?

3. Prononcer «Kok-rane»

2. *Soit $X_1,\dots,X_n$ un échantillon d'une loi discrète inconnue $q=(q_1,\dots,q_k)$ sur $\{1,\dots,k\}$. On souhaite savoir si cette loi q est égale à une loi discrète de référence $p=(p_1,\dots,p_k)$ (qui est connue). Pour cela, on introduit les hypothèses statistiques antagonistes :*

$$H_0 : p = q, \quad H_1 : p \neq q.$$

On introduit la statistique de test $S_n = S(X_1,\dots,X_n)$ definie par

$$S_n := nD(p,\widehat{q}) = n\sum_{i=1}^{k}\frac{(p_i-\widehat{q}_i)^2}{p_i} = \sum_{i=1}^{k}\frac{(n_i-N_i)^2}{n_i}$$

où
— $\widehat{q}=(N_1/n,\dots,N_k/n)$ est l'estimateur empirique de q ;
— $N_i := \mathrm{card}\{1\le m\le n : X_m = i\} = \sum_{m=1}^n \mathbf{1}_{\{X_m=i\}}$ est l'effectif de i dans l'échantillon ;
— $n_i := np_i$ est l'effectif théorique sous H_0.
En pratique, on calcule les n_i et les N_i (comptage) puis S_n. Montrer que
— si H_0 est fausse alors $S_n\to+\infty$ presque sûrement quand $n\to\infty$;
— si H_0 est vraie alors $S_n\to\chi^2(k-1)$ en loi quand $n\to\infty$.
Indication : établir que

$$S_n = \left\|\frac{1}{\sqrt{n}}\sum_{m=1}^{n}V_m\right\|_2^2 \quad \text{où} \quad V_{m,i} := \frac{\mathbf{1}_{\{X_m=i\}}-p_i}{\sqrt{p_i}}.$$

3. *Fixons un seuil de tolérance $0<\alpha<1$, typiquement $\alpha=5/100$, appelé* niveau du test, *et considérons la région $\mathcal{R}_\alpha=[0,a]$ où a est le quantile $1-\alpha$ de la loi $\chi^2(k-1)$, appelée* région d'acceptation du test. *Au vu de $X_1,\dots,X_n$, on décide comme suit :*
— Si $S_n\in\mathcal{R}_\alpha$, on accepte l'hypothèse H_0 ;
— Si $S_n\in\mathcal{R}_\alpha$, on rejette l'hypothèse H_0.
De manière résumée la décision prise avec le test est $T_n := H_{\mathbf{1}_{S_n\notin\mathcal{R}_\alpha}}$. Montrer que
— Si H_0 est vraie alors la probabilité de rejeter à tort H_0 tend vers α quand $n\to\infty$;
— Si H_0 est fausse, alors la probabilité d'accepter à tort H_0 tend vers 0 quand $n\to\infty$.
*On parle d'*erreurs *de première et de seconde espèce du test ($\to$ niveau et puissance).*

4. *Comment se comportent $\mathcal{R}_\alpha$ et T_n dans les cas extrèmes $\alpha=0$ et $\alpha=1$. Existe-t-il un meilleur choix pour la région $\mathcal{R}_\alpha$, parmi toutes les régions de probabilité α pour la loi χ^2 ?*

5. *Comment adapter le test du chi-deux au cas des échantillons à densité ? À quoi correspond la terminologie « test non-paramétrique » ?*

Solution.

1. Évident. Ce n'est pas une distance, car elle n'est pas symétrique. Cependant, $\sqrt{D(p,q)}$ est la distance euclidienne de p à q pondérée par p ;
2. D'après la loi forte des grands nombres,

$$\frac{S_n}{n}\xrightarrow[n\to\infty]{\text{p.s.}}\sum_{i=1}^{k}\frac{(p_i-q_i)^2}{p_i}=D(p,q).$$

Si H_0 est fausse alors $D(p,q)>0$ et donc $S_n=n(S_n/n)\sim_{n\to\infty} nD(p,q)\to+\infty$ presque sûrement quand $n\to\infty$. Si H_0 est vraie, alors $q=p$ et $D(p,q)=0$, et comportement de S_n quand $n\to\infty$ peut-être décrit par le théorème central limite multivarié. On a

$$S_n = n\sum_{i=1}^{k}\frac{\left(p_i-\frac{1}{n}\sum_{m=1}^n\mathbf{1}_{\{X_m=i\}}\right)^2}{p_i} = \left\|\frac{V_1+\cdots+V_n}{\sqrt{n}}\right\|_2^2$$

où $V_1,\dots,V_n$ sont les vecteurs aléatoires de $\mathbb{R}^k$ définis par

$$V_{m,i} := \frac{1}{\sqrt{p_i}}(\mathbf{1}_{\{X_m=i\}}-p_i)$$

pour tout $1\le m\le n$ et $1\le i\le k$. Or les vecteurs aléatoires $V_1,\dots,V_n$ sont i.i.d. de matrice de covariance $\Sigma=I_k-\sqrt{p}\sqrt{p}^\top$ où $\sqrt{p}^\top := (\sqrt{p_1},\dots,\sqrt{p_k})$. Par conséquent, le TCL donne

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}V_m\xrightarrow[n\to\infty]{\text{loi}}\mathcal{N}(0,\Sigma).$$

Comme la norme est continue, on obtient que

$$S_n = \left\| \frac{1}{\sqrt{n}} \sum_{m=1}^{n} V_m \right\|_2^2 \xrightarrow[n\to\infty]{\text{loi}} \|Z\|_2^2$$

où $Z \sim \mathcal{N}(0,\Sigma)$. La matrice Σ est de rang $k-1$. La matrice $\sqrt{p}\sqrt{p}^\top$ est la matrice de projection orthogonale sur $\mathrm{Vect}(\sqrt{p})$, tandis que $\Sigma = I_k - \sqrt{p}\sqrt{p}^\top$ est la matrice de projection orthogonale sur $\mathrm{vect}(\sqrt{p})^\perp$ dans $\mathbb{R}^k$. Le théorème de Cochran donne $Z \sim \chi^2(k-1)$. On peut aussi procéder directement en utilisant une décomposition de Choleski de Σ ou encore une diagonalisation de Σ. En effet, la matrice $\sqrt{p}\sqrt{p}^\top$ est de rang 1 et ses valeurs propres sont 0 avec multiplicité $k-1$ et 1 avec multiplicité 1, de sorte que les valeurs propres de Σ sont 0 avec multiplicité 1 et 1 avec multiplicité $k-1$. Ainsi $\Sigma = ODO^\top$ où O est orthogonale et $D = \mathrm{diag}(1,\dots,1,0)$. Ainsi, si $Z \sim \mathcal{N}(0,I_k)$ alors $O\sqrt{D}Z = ODZ \sim \mathcal{N}(0,\Sigma)$ et $\left\|O\sqrt{D}Z\right\|_2^2 = \|DZ\|_2^2 = Z_1^2 + \cdots + Z_{k-1}^2 \sim \chi^2(k-1)$.

3. Si H_0 est vraie alors $\mathbb{P}(T_n = 1) = \mathbb{P}(S_n > a) = 1 - \mathbb{P}(S_n \le a) \to 1-(1-\alpha) = \alpha$ car $S_n \to \chi^2(k-1)$. Si H_0 est fausse, alors $\mathbb{P}(T_n = 0) = \mathbb{P}(S_n \le a) \to 0$ car $S_n \to +\infty$ p.s.
4. Dessiner la densité du chi-deux. Puissance maximale ;
5. Intervalles ou classes.

Exercice 8.4 (Test de Kolmogorov-Smirnov non asymptotique)**.** *Soit $(X_n)_{n\ge1}$ des v.a.r. i.i.d. de fonction de répartition F, d'inverse généralisé F^{-1}. Soit $(U_n)_{n\ge1}$ des v.a.r. i.i.d. de loi uniforme sur $[0,1]$, de fonction de répartition G donnée par $G(x) = x$ pour tout $x \in [0,1]$. Pour tout $n \ge 1$, les fonctions de répartition empiriques F_n et G_n des échantillons $X_1,\dots,X_n$ et $U_1,\dots,U_n$ sont des fonctions aléatoires de $\mathbb{R}$ dans $[0,1]$ définies pour tout $x \in \mathbb{R}$ par*

$$F_n(x) := \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k \le x\}} \quad \text{et} \quad G_n(x) := \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k \le x\}}.$$

1. *Montrer que pour tout ω, F_n est la fonction de répartition de la mesure empirique $\frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ (mesure de probabilité discrète aléaoire !). Idem pour G_n et $\frac{1}{n}\sum_{k=1}^n \delta_{U_k}$;*
2. *Montrer que pour tout $x \in \mathbb{R}$, $H_n(x) := \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{F^{-1}(U_k)\le x\}} = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{U_k \le F(x)\}}$, et que les suites $(\|F_n - F\|_\infty)_{n\ge1}$ et $(\|H_n - F\|_\infty)_{n\ge1}$ ont même loi ;*
3. *Montrer que $\|H_n - F\|_\infty \le \|G_n - G\|_\infty$, avec égalité si F est continue ;*
4. *Soit $\alpha \in [0,1]$ et soit $q_{1-\alpha}$ le quantile $1-\alpha$ de la loi de $\|G_n - G\|_\infty$. On suppose que F est continue. Proposer un test d'adéquation à la loi de fonction de répartition F, non asymptotique et de niveau $1-\alpha$, en utilisant le quantile $q_{1-\alpha}$.*

Solution.

1. Immédiat. Attention aux deux couches d'aléa ;
2. La seconde formule pour $H_n(x)$ découle du fait que $u \le F(x)$ ssi $F^{-1}(u) \le x$. Pour l'égalité en loi des suites, on utilise, comme pour la méthode de simulation par inversion, le fait que les suites $(X_n)_{n\ge1}$ et $(F^{-1}(U_n))_{n\ge1}$ ont même loi.
3. Comme $F(\mathbb{R}) \subset [0,1]$, on a

$$\|H_n - F\|_\infty = \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{U_k \le F(x)\}} - F(x) \right| \le \sup_{u\in[0,1]} \left| \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{U_k \le u\}} - u \right| = \|G_n - G\|_\infty.$$

   Si F est continue alors $F(\mathbb{R}) = [0,1]$ et on a égalité ;
4. Si F est continue alors $\|F_n - F\|_\infty$ a la même loi que $\|G_n - G\|$, qui ne dépend pas de la loi des X_k ! Si H_0 :«les X_k sont de loi F» alors on accepte H_0 si $\|F_n - F\|_\infty \le q_{1-\alpha}$ et on rejette sinon. C'est le test d'adéquation, non asymptotique, de Kolmogorov-Smirnov.

Exercice 8.5 (Théorème de Glivenko-Cantelli et test de Kolmogorov-Smirnov asymptotique)**.** *On reprend les notations de l'exercice 8.4.*

1. *Montrer que pour tout $x \in \mathbb{R}$ on a $\mathbb{E}(F_n(x)) = F(x)$ et $F_n(x) - F(x) \to 0$ p.s. quand $n \to \infty$;*

2. *On veut établir que p.s. la convergence est uniforme. D'après l'exercice 8.4, les suites* $(\|F_n - F\|_\infty)_{n\geq 1}$ *et* $(\|H_n - F\|_\infty)_{n\geq 1}$ *ont même loi, et* $\|H_n - F\|_\infty \leq \|G_n - G\|_\infty$. *Montrer que* $\|G_n - G\|_\infty \to 0$ *p.s. quand* $n \to \infty$, *et en déduire le théorème de Glivenko-Cantelli :*

$$\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{p.s.}} 0;$$

3. *On admet que les convergences en loi suivantes on lieu lorsque* F *est continue :*

$$\sqrt{n}\,\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{loi}} K_1 \quad \text{et} \quad \sqrt{n}\,\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{loi}} K_2,$$

où K_1 *et* K_2 *sont les lois sur* $\mathbb{R}_+$, *appelées* lois de Kolmogorov-Smirnov, *dont les fonctions de répartition sont données pour tout* $u > 0$ *par*

$$F_{K_1}(]-\infty, u]) = 1 - e^{-2u^2} \quad \text{et} \quad F_{K_2}(]-\infty, u]) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2u^2}.$$

Montrer comment exploiter ces convergences vers les lois de Kolmogorov-Smirnov, ainsi que le théorème de Glivenko-Cantelli, pour fabriquer un test d'adéquation asymptotique, appelé test de Kolmogorov-Smirnov, dont on contrôle le comportement sous H_0 *et* H_1.

Solution.

1. LGN pour $Y_i = \mathbf{1}_{]-\infty,x]}(X_i)$ et $\mathbb{E}(F_n(x)) = F(x)$ (somme de n v.a.r. i.i.d. Bernoulli !) ;
2. Le problème est donc réduit au cas de la loi uniforme. Soit $I_1, \ldots, I_N$ une partition de $[0,1]$ en N intervalles de longueur $1/N$. On a $G_n(u_{k-1}) \leq G_n(u) \leq G_n(u_k)$ avec $u_k = k/N$ pour tout $u \in I_k$ car G_n est croissante. D'autre part, $G_n(0) = 0$ et $G_n(1) = 1$, et donc

$$\begin{aligned}\|G_n - G\|_\infty &= \sup_{1\leq k\leq N}\sup_{u\in I_k}|G_n(u) - G(u)| \\ &\leq \sup_{0\leq k\leq N}|G_n(u_k) - u_k| + \frac{1}{N} \\ &\leq \sup_{1\leq k\leq N-1}|G_n(u_k) - u_k| + \frac{1}{N}.\end{aligned}$$

On considère maintenant l'évènement

$$A_N = \cap_{k=1}^{N-1}\{\omega \in \Omega \text{ tel que } \lim_{n\to\infty} G_n(u_k,\omega) = u_k\}.$$

On a $\mathbb{P}(A_N) = 1$ car en vertu de la LGN, A_N est une intersection dénombrable d'évènements de probabilité 1. À présent, pour tout $\omega \in A_N$,

$$\overline{\lim_{n\to+\infty}}\,\|G_n - G\|_\infty \leq \frac{1}{N}.$$

Si $A = \cap_{N=1}^{\infty} A_N$, alors $\mathbb{P}(A) = 1$, d'où $\lim_{n\to\infty}\|G_n - G\|_\infty = 0$ p.s.

3. Pour fabriquer le test, on fixe $\alpha \in [0,1]$ et on considère le quantile $1-\alpha$ de la loi F_{K_2}, noté $q_{1-\alpha}$. On pose H_0 :«la loi des X_k est F et H_1 l'hypothèse antagoniste. On accepte H_0 si $\sqrt{n}\|F_n - F\|_\infty \leq q_{1-\alpha}$ et on rejette sinon. Si H_0 est vrai alors la probabilité de rejeter H_0 à tort tend vers α quand $n \to \infty$. Le niveau asymptotique du test est donc bien $1-\alpha$ (erreur de première espèce). Pour l'erreur de seconde espèce (puissance), on se place dans la situation où H_0 est faux. Alors en notant $F_\#$ la vraie fonction de répartition des X_k, on a $F_\# \neq F$ et grâce au théorème de Glivenko-Cantelli, on a $\|F_n - F_\#\|_\infty \to 0$ p.s. et donc que $\|F_n - F\|_\infty \to \|F_\# - F\|_\infty > 0$ p.s. et donc $\sqrt{n}\|F_n - F\|_\infty \to +\infty$ p.s. Il en découle que p.s. pour n assez grand $\sqrt{n}\|F_n - F\|_\infty > q_{1-\alpha}$, ce qui signifie que la probabilité d'accepter H_0 (c'est-à-dire accepter H_0 à tort) tend vers 0 quand $n \to \infty$.
Voici enfin les grandes lignes de la preuve de la convergence en loi vers les lois de Kolmogorov-Smirnov. La preuve fait appel au mouvement brownien. Comme F est continue, on a $F(\mathbb{R}) = [0,1]$, et par conséquent, la majoration devient une égalité. On est ainsi ramené à l'étude de la loi limite des suites

$$\sqrt{n}\sup_{u\in\mathbb{R}}(G_n(u) - G(u)) \quad \text{et} \quad \sqrt{n}\sup_{u\in\mathbb{R}}|G_n(u) - G(u)|.$$

Désignons par $U^n_{(1)} \leq \cdots \leq U^n_{(n)}$ le réarrangement croissant de $U_1, \ldots, U_n$, et posons $U^n_{(0)} = 0$ et $U^n_{(n+1)} = 1$. La fonction aléatoire G_n est croissante, et vérifie $G_n(U^n_{(k)}) = k/n$ pour tout $0 \leq k \leq n+1$.

D'autre part, G est également croissante et vérifie $G(u) = u$ pour tout $u \in \mathbb{R}$. Par conséquent, on a pour tout $0 \leq k \leq n$ et tout $U^n_{(k)} \leq u < U^n_{(k+1)}$,

$$\left(\frac{k+1}{n} - U^n_{(k+1)}\right) - \frac{1}{n} = \frac{k}{n} - U^n_{(k+1)} \leq G_n(u) - G(u) \leq \frac{k}{n} - U^n_{(k)}.$$

Ainsi, à une translation d'ordre $n^{-1/2}$ près, cet encadrement ramène le problème à l'étude de la loi limite des suites

$$\sqrt{n} \max_{1\leq k\leq n} \left(\frac{k}{n} - U^n_{(k)}\right) \quad \text{et} \quad \sqrt{n} \max_{1\leq k\leq n} \left|\frac{k}{n} - U^n_{(k)}\right|.$$

La fonction $k \in \{0, 1, \ldots, n+1\} \mapsto f(k) = k/n - U^n_{(k)}$ dessine des boucles enracinées en 0 et ses accroissements sont de loi Beta recentrée. La suite de la preuve montre que cette intuition est pertinente. Soit (E_n) une suite de variables i.i.d. de loi exponentielle $\mathcal{E}(1)$, et $T_n = E_1 + \cdots + E_n$. On peut vérifier que la suite $(T_k/T_{n+1})_{1\leq k\leq n}$ a la même loi que la suite $(U^n_{(k)})_{1\leq k\leq n}$ (loi de Dirichlet !). En particulier, $U^n_{(k)} - k/(n+1)$ a la même loi que

$$\frac{T_k}{T_{n+1}} - \frac{k}{n+1} = \frac{1}{T_{n+1}}\left(T_k - k - \frac{k}{n+1}(T_{n+1} - (n+1))\right).$$

La suite de la preuve est peu rigoureuse mais donne les idées directrices. Le principe d'invariance de Donsker (TCL sur les trajectoires de la marche aléatoire simple) suggère la convergence en loi (de processus) suivante :

$$\frac{1}{\sqrt{n}}\left(T_{\lfloor nt\rfloor} - \lfloor nt\rfloor - \frac{\lfloor nt\rfloor}{n+1}(T_{n+1} - (n+1))\right)_{t\in[0,1]} \overset{\text{loi}}{\to} \mathcal{L}\big((B_t - tB_1)_{t\in[0,1]}\big)$$

où $(B_t)_{t\geq 0}$ est un mouvement brownien standard issu de 0. Le processus $(P_t)_{t\in[0,1]}$ défini par $P_t = B_t - tB_1$ pour tout $t \in [0,1]$ est appelé *pont brownien sur* $[0,1]$. Il doit son nom au fait que ses trajectoires sont nulles aux temps 0 et 1 car $P_0 = P_1 = 0$. Ces trajectoires forment des « boucles browniennes enracinée en 0». Comme la suite (T_n/n) converge presque sûrement vers 1 en vertu de la LGN, on obtient

$$S^\pm_n = \sqrt{n} \max_{1\leq k\leq n} \pm\left(\frac{k}{n} - U_{(k)}\right) \overset{\text{loi}}{\to} \mathcal{L}\left(\sup_{t\in[0,1]} (\pm P_t)\right).$$

Comme $(P_t)_{t\in[0,1]}$ et $(-P_t)_{t\in[0,1]}$ ont la même loi, on en déduit que S^+_n et S^-_n ont la même loi limite quand $n \to \infty$. Cette loi est notée K_1. De même, on a par un raisonnement similaire

$$\sqrt{n} \max_{1\leq k\leq n} \left|\frac{k}{n} - U_{(k)}\right| \overset{\text{loi}}{\to} \mathcal{L}\left(\sup_{t\in[0,1]} |P_t|\right),$$

et la loi limite est notée K_2.

9 Quelques modèles dynamiques aléatoires

Exercice 9.1 (Processus de branchement avec immigration)**.** *Si Z_n est la taille de la population à l'instant n, on pose $Z_{n+1} = Y_{n+1} + \sum_{k=1}^{Z_n} X_{n+1,k}$ où Y_{n+1} correspond au nombre d'individus immigrés et où $X_{n+1,k}$ est le nombre d'enfants de l'individu k de la génération n, avec la convention $\sum_{\varnothing} = 0$. On suppose que les $(X_{n,k})_{n,k\geq 1}$ sont des v.a. sur $\mathbb{N}$, i.i.d. de carré intégrable, de fonction génératrice g, de moyenne α et de variance σ^2. On pose $a = \mathbb{E}(X_{1,1}(X_{1,1}-1))$. On suppose que les $(Y_n)_{n\geq 1}$ sont des v.a. sur $\mathbb{N}$, i.i.d. et de carré intégrable, de fonction génératrice h, de moyenne β, et de variance τ^2. On pose $b = \mathbb{E}(Y_1(Y_1-1))$. On suppose que Z_0 est une v.a. sur $\mathbb{N}$, de carré intégrable. Les v.a. $X_{n,k}, Y_n, Z_0$ sont indépendantes. On note G_n la fonction génératrice de Z_n.*

1. *Déterminer une relation de récurrence sur $(G_n)_{n\geq 0}$ et en déduire une expression pour G_n ;*
2. *Déterminer une relation de récurrence satisfaite par $e_n = \mathbb{E}(Z_n)$ et $d_n = \mathbb{E}(Z_n(Z_n-1))$ en fonction de α, β, a, b, puis une relation de récurrence pour $v_n = \mathrm{Var}(Z_n)$ en fonction de $\alpha, \beta, \sigma^2, \tau^2$. En déduire l'expression de e_n et v_n. Comment évoluent ces quantités avec n ?*

Solution.

1. On écrit, en utilisant l'indépendance,
$$\begin{aligned}G_{n+1}(s) &= \mathbb{E}(s^{Y_{n+1}+\sum_{k=1}^{Z_n} X_{n+1,k}})\\ &= h(s)\sum_{z\geq 0}\mathbb{E}(s^{\sum_{k=1}^{z} X_{n+1,k}})\mathbb{P}(Z_n = z)\\ &= h(s)\sum_{z\geq 0} g(s)^z\mathbb{P}(Z_n = z)\\ &= h(s)G_n(g(s))\end{aligned}$$
d'où $G_{n+1} = h\cdots(h\circ g^{\circ n})(G_0\circ g^{\circ(n+1)})$;
2. Si $S(s) = \mathbb{E}(s^X)$ alors $S(1) = 1$ et $S'(1) = \mathbb{E}(X)$. Ce la donne donc
$$e_{n+1} = G'_{n+1}(1) = h'(1)G_n(g(1)) + h(1)G'_n(g(1))g'(1) = \beta + \alpha e_n,$$
formule de récurrence que nous aurions pu obtenir directement à partir de l'expression de Z_{n+1} en fonction de Z_n et Y_{n+1}. On en déduit que $e_n = \beta\sum_{k=0}^{n-1}\alpha^k + \alpha^n e_0$.
De même, pour calculer d_{n+1}, on utilise le fait que $S''(1) = \mathbb{E}(X(X-1))$, ce qui donne
$$d_{n+1} = G''_{n+1}(1) = b + (a + 2\alpha\beta)e_n + \alpha^2 d_n.$$
Par conséquent, on a
$$\begin{aligned}v_{n+1} &= d_{n+1} + e_{n+1} - e_{n+1}^2\\ &= b + (a+2\alpha\beta)e_n + \alpha^2 d_n + \beta + \alpha e_n - \beta^2 - 2\alpha\beta e_n - \alpha^2 e_n^2\\ &= b + (a + \alpha - \alpha^2)e_n + \beta - \beta^2 + \alpha^2 v_n\\ &= \tau^2 + \sigma^2 e_n + \alpha^2 v_n.\end{aligned}$$
Supposons que $\beta, \sigma^2 > 0$. Si $\alpha \geq 1$ alors $(e_n)_{n\geq 0}$ et $(v_n)_{n\geq 0}$ divergent. Si $\alpha < 1$ alors $(e_n)_{n\geq 0}$ converge vers $\beta/(1-\alpha)$ tandis que $(v_n)_{n\geq 0}$ converge vers $(\tau^2 + \sigma^2\beta/(1-\alpha))/(1-\alpha^2)$.

Exercice 9.2 (Modèle de Fisher-Wright)**.** *On considère l'évolution d'une population de taille fixe $N > 1$ dont chaque individu est de type A ou B. La génération $t+1$ s'obtient en tirant pour chaque individu de la génération $t+1$, de manière i.i.d. uniforme, un parent dans la génération t. Soit X_t le nombre d'individus de type A dans la génération t. On s'intéresse au comportement en temps long de la suite récurrente aléatoire $(X_t)_{t\geq 0}$, en fonction du point de départ $X_0 = x_0$ (fixé).*

1. *Montrer que pour tout $t \geq 0$ et tous $0 \leq x, y, x_1, \ldots, x_{t-1} \leq N$*
$$P_{x,y} = \mathbb{P}(X_{t+1} = y | X_t = x, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = \binom{N}{y}\left(\frac{x}{N}\right)^y\left(1 - \frac{x}{N}\right)^{N-y};$$
2. *Montrer que $\mathbb{P}(\cap_{s\geq t}\{X_s = x\} | X_t = x) = 1$ pour tout $t \geq 0$ et tout $x \in \{0, N\}$;*
3. *Montrer que $\mathbb{P}(X_{t+1} \in \{0, N\} | X_t = x) \geq 2^{1-N}$ pour tout $0 \leq x \leq N$ et tout $t \geq 0$;*

4. *En déduire que* $\mathbb{P}(T<\infty)=1$ *et que* $\mathbb{E}(T)\leq 2^{N-1}$ *où* $T=\inf\{t>0 : X_t\in\{0,N\}\}$ *;*
5. *En déduire que* $(X_t)_{t\geq 0}$ *converge p.s. et dans* $\mathbb{L}^1$ *vers la v.a.* X_T *à valeurs dans* $\{0,N\}$ *;*
6. *Montrer que* $\mathbb{E}(X_t)=x_0$ *pour tout* $t\geq 0$*, puis que* $\mathbb{P}(X_T=N)=x_0/N$ *;*

Avec ce modèle, presque sûrement, l'un des types finit par disparaître complètement de la population en un temps fini (aléatoire). On modifie à présent le modèle en introduisant des mutations : après avoir choisi un parent, on transforme A *en* B *(resp.* B *en* A*) avec probabilité* u *(resp.* v*).*

1. *Montrer que pour tout* $t\geq 0$ *et tous* $0\leq x,y,x_1,\ldots,x_{t-1}\leq N$*, avec* $p_x=\frac{x}{N}(1-u)+\big(1-\frac{x}{N}\big)v$*,*
$$P_{x,y}=\mathbb{P}(X_{t+1}=y|X_t=x,X_{t-1}=x_{t-1},\ldots,X_0=x_0)=\binom{N}{y}p_x^y(1-p_x)^{N-y};$$
2. *Montrer que si* $0<u,v<1$ *alors* $P_{x,y}>0$ *pour tous* $0\leq x,y\leq N$ *;*
3. *On pose* $m_t=\mathbb{E}(X_t)$*. Montrer que* $m_0=x_0$ *et* $m_{t+1}=m_t(1-u-v)+Nv$ *pour tout* $t\geq 0$*. En déduire que* $m_t\to Nv/(u+v)$ *quand* $t\to\infty$ *(notez que cette limite ne dépend pas de* x_0*) ;*

Solution.

1. Sachant $\{X_0=x_0,\ldots,X_{t-1}=x_{t-1},X_t=x\}$, la loi de X_{t+1} est $\mathrm{Binom}(N,x/N)$;
2. Car $P_{0,x}=0$ si $x\neq 0$ et $P_{N,x}=0$ si $x\neq N$;
3. $\mathbb{P}(X_{t+1}\in\{0,N\}|X_t=x)=(x/N)^N+(1-x/N)^N\geq\min_{p\in[0,1]}(p^N+(1-p)^N)=1/2^{N-1}$;
4. La suite $(X_t)_{t\geq 0}$ est absorbée par $\{0,N\}$ avant qu'un joueur de pile ou face avec probabilité de gagner 2^{1-N} ne gagne pour la première fois. En formule, cela donne, pour tout $t\geq 1$,
$$\begin{aligned}\mathbb{P}(T>t)&=\mathbb{P}(X_1\notin\{0,N\},\ldots,X_t\notin\{0,N\})\\&=\sum_{x_1,\ldots,x_t\notin\{0,N\}}P_{x_0,x_1}\cdots P_{x_{t-1},x_t}\\&=\sum_{x_1\notin\{0,N\}}P_{x_0,x_1}\cdots\sum_{x_t\notin\{0,N\}}P_{x_{t-1},x_t}\\&\leq(1-2^{1-N})^t\\&=\mathbb{P}(T'>t)\end{aligned}$$
où T' suit la loi géométrique de paramètre 2^{1-N} (T' domine stochastiquement T). La seconde égalité est typique des chaînes de Markov et découle par récurrence de la première question. Il en découle que $\mathbb{P}(T=\infty)\leq\mathbb{P}(T'=\infty)=0$ (faire $t\to\infty$) et $\mathbb{E}(T)\leq\mathbb{E}(T')=2^{1-N}$ (faire $\sum_t$) ;
5. $X_t=X_T\mathbf{1}_{t\geq T}+X_t\mathbf{1}_{t<T}\underset{t\to\infty}{\to}X_T$ p.s. car $\mathbb{P}(T<\infty)=1$. Converge dans L^1 car bornée (CVD). Alternativement, pour la convergence dans L^1 : $\mathbb{E}(|X_t-X_T|)=\mathbb{E}(|X_t-X_T|\mathbf{1}_{T>t})\leq N\mathbb{P}(T>t)\to 0$;
6. Par récurrence sur $t\geq 0$ on a
$$\begin{aligned}\mathbb{E}(X_{t+1})&=\sum_{y=0}^N y\mathbb{P}(X_{t+1}=y)\\&=\sum_{y=0}^N y\sum_{x=0}^N\mathbb{P}(X_t=x)P_{x,y}\\&=\sum_{x=0}^N\mathbb{P}(X_t=x)\sum_{y=0}^N yP_{x,y}\\&=\sum_{x=0}^N\mathbb{P}(X_t=x)x\\&=\mathbb{E}(X_t)=\cdots=\mathbb{E}(X_0)=x_0.\end{aligned}$$
Donc $x_0=\mathbb{E}(X_t)\to\mathbb{E}(X_T)=0\mathbb{P}(X_T=0)+N\mathbb{P}(X_T=N)$ d'où $\mathbb{P}(X_T=N)=x_0/N$;

1. Sachant $\{X_t=x\}$, si G est le type du premier parent tiré au hasard, si M est l'indicateur de mutation, et si G' est le type du parent après mutation, alors $p_x=\mathbb{P}(G'=A)$ vaut
$$\mathbb{P}(M=0|G=A)\mathbb{P}(G=A)+\mathbb{P}(M=1|G=B)\mathbb{P}(G=B)=\frac{x}{N}(1-u)+\Big(1-\frac{x}{N}\Big)v.$$
Cela est valable pour les N individus i.i.d. de la génération $t+1$, d'où $X_{t+1}\sim\mathrm{Binom}(N,p_x)$;

2. Découle du fait que si $0 < u, v < 1$ alors $0 < p_x < 1$ pour tout $0 \leq x \leq N$;
3. On a $m_0 = x_0$ par hypothèse sur X_0. Ensuite, pour tout $t \geq 0$,

$$\begin{aligned} m_{t+1} &= \mathbb{E}(X_{t+1}) \\ &= \sum_{y=0}^{N} y\mathbb{P}(X_{t+1} = y) \\ &= \sum_{x=0}^{N} \mathbb{P}(X_t = x) \sum_{y=0}^{N} yP_{x,y} \\ &= \sum_{x=0}^{N} \mathbb{P}(X_t = x)((1-u)x + (N-x)v) \\ &= (1-u)m_t + (N-m_t)v = (1-u-v)m_t + Nv. \end{aligned}$$

La suite $(m_t)_{t\geq 0}$ est bornée (appartient à $\{0, 1, \ldots, N\}$) et si m est une valeur d'adhérence alors $m = (1-u-v)m + Nv$, d'où $m_t \to Nv/(u+v)$ quand $t \to \infty$. (formule pour m_t aussi).

Exercice 9.3 (Processus TCP window-size en informatique)**.** *Le débit instantané maximal sortant d'un ordinateur connecté à un réseau TCP/IP comme Internet est régulé par l'algorithme suivant : le débit maximal est augmenté de manière déterministe d'une unité à chaque pas de temps, et en cas de signal de congestion, il est multiplié par un facteur entre $[0, 1[$, typiquement $1/2$. Ce mécanisme simple, parfois qualifié de AIMD (Additive Increase, Multiplicative Decrease) permet à la fois une bonne exploitation du réseau et une réaction efficace en cas de congestion.*

1. ***Modélisation des temps de congestion par un processus de Poisson simple.*** *On modélise les temps d'arrivée des signaux de congestion par une suite $T_1 \leq T_2 \leq \cdots$ de variables aléatoires réelles positives séparées par des durées aléatoires $T_1 - 0, T_2 - T_1, T_3 - T_2, \ldots$ indépendantes et identiquement distribuées de loi exponentielle de paramètre λ. Le nombre de congestions dans l'intervalle $[0, t]$ est donné par la variable aléatoire de comptage (à valeurs dans $\mathbb{N} \cup \{\infty\}$)*

$$N_t := \sum_{n=1}^{\infty} \mathbf{1}_{[0,t]}(T_n) = \sup\{n \in \mathbb{N}^* : T_n \leq t\}.$$

On a $N_0 = 0$ et la courbe aléatoire $t \in \mathbb{R}_+ \mapsto N_t$ est croissante.

   (a) *Montrer que pour tout $t \geq 0$, la variable aléatoire N_t suit la loi de Poisson de moyenne λt, et en déduire que presque sûrement, pour tout $t \geq 0$, on a $N_t < \infty$;*

2. ***Modélisation de la taille de fenêtre TCP et comportement en temps long.*** *On pose $T_0 = 0$. Soit $Q = (Q_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. à valeurs dans $[0, 1[$ de même loi $\mathcal{Q}$. Soit X_0 une v.a.r. positive. On suppose que X_0, N, et Q sont indépendants. On définit le processus $X = (X_t)_{t\in\mathbb{R}_+}$ à temps continu et à espace d'états $\mathbb{R}_+$ de la manière suivante :*

$$X_t = \begin{cases} X_{T_n} + t - T_n & \text{si } T_n \leq t < T_{n+1}, \\ Q_{n+1}(X_{T_n} + T_{n+1} - T_n) & \text{si } t = T_{n+1}, \end{cases}$$

où $n = N_t \in \mathbb{N}$ est le nombre de sauts avant l'instant t. Les trajectoires de X sont affines de pente 1 par morceaux, continues à droite avec limite à gauche, et chaque saut correspond à une multiplication par un nombre entre $[0, 1]$. Il se trouve que le processus oublie exponentiellement vite sa condition initiale. Pour le voir, nous procédons par couplage.

   (a) *Soit (X, Y) un couple où X et Y partent de X_0 et Y_0 mais utilisent les mêmes temps de saut N et coefficients multiplicateurs Q. Montrer que pour tous $t \geq 0$ et $p \geq 1$,*

$$\mathbb{E}\left(|X_t - Y_t|^p\right) = \mathbb{E}(|X_0 - Y_0|^p)e^{-\lambda t(1-\mathbb{E}(Q_1^p))}.$$

Solution.

1. (a) Ce calcul figure déjà dans les feuilles précédentes : le changement de variable linéaire et triangulaire $(x_1,\ldots,x_{n+1}) \mapsto (t_1,\ldots,t_{n+1}) = (x_1,\ldots,x_1+\cdots+x_{n+1})$ donne en $(*)$:

$$\begin{aligned}
\mathbb{P}(N_t = n) &= \mathbb{P}(T_n \le t, T_{n+1} > t)\\
&= \mathbb{E}(\mathbf{1}_{x_1+\cdots+x_n \le t < x_1+\cdots+x_{n+1}}(E_1,\ldots,E_1+\cdots+E_{n+1}))\\
&= \int_{\mathbb{R}_+^{n+1}} \mathbf{1}_{x_1+\cdots+x_n \le t < x_1+\cdot+x_{n+1}} \lambda^{n+1} e^{-\lambda(x_1+\cdots+x_{n+1})}\,dx_1\cdots dx_{n+1}\\
&\overset{(*)}{=} \int_{0\le t_1\le\cdots\le t_n \le t < t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}}\,dt_1\cdots dt_{n+1}\\
&= \int_{0\le t_n \le t< t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}} \left(\int_{0\le t_1\le\cdots\le t_{n-1}\le t_n} dt_1\cdots dt_{n-1}\right) dt_n dt_{n+1}\\
&= \int_{0\le t_n \le t< t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}} \frac{t_n^{n-1}}{(n-1)!}\, dt_n dt_{n+1}\\
&= \frac{\lambda^{n+1}}{(n-1)!}\left(\int_0^t t_n^{n-1}\,dt_{n-1}\right)\left(\int_t^\infty e^{-\lambda t_{n+1}}\,dt_{n+1}\right)\\
&= \frac{t^n}{n!}\lambda^n e^{-\lambda t}.
\end{aligned}$$

On peut atteindre directement les dernières étapes et éviter le recours au volume du simplexe en utilisant la loi du couple $(T_n, T_{n+1}) = (T_n, T_n + E_{n+1})$ car T_n et E_{n+1} sont indépendantes avec $T_n \sim \mathrm{Gamma}(n,\lambda)$ de densité $t \mapsto \frac{\lambda^n}{(n-1)!} e^{-\lambda t}\,\mathbf{1}_{\mathbb{R}_+}(t)$ et $E_{n+1} \sim \mathrm{Exp}(\lambda)$.

2. (a) Au court du temps, la quantité $|X_t - Y_t|$ reste constante entre deux sauts et au k^{e} saut elle est multipliée par Q_k. Grâce à l'indépendance des ingrédients, il vient

$$\begin{aligned}
\mathbb{E}(|X_t - Y_t|^p) &= \sum_{n=0}^\infty \mathbb{E}\big(|X_t - Y_t|^p \mathbf{1}_{\{N_t = n\}}\big)\\
&= \sum_{n=0}^\infty \mathbb{E}\big(|X_0 - Y_0|^p Q_1^p\cdots Q_n^p \mathbf{1}_{\{N_t = n\}}\big)\\
&= \mathbb{E}(|X_0 - Y_0|^p)\sum_{n=0}^\infty \mathbb{E}(Q_1^p)^n \mathbb{P}(N_t = n)\\
&= \mathbb{E}(|X_0 - Y_0|^p)e^{-\lambda t}\sum_{n=0}^\infty \mathbb{E}(Q_1^p)^n \frac{(\lambda t)^n}{n!}\\
&= \mathbb{E}(|X_0 - Y_0|^p)\exp\left(-\lambda t(1-\mathbb{E}(Q_1^p))\right).
\end{aligned}$$