# 0 Événements, dénombrement, conditionnement

**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 ($n = 26$)*

**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.*

**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.*

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

**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 $1/2$.*

**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.*

**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 $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{p=1}^n (-1)^{p+1} \sum_{1 \leq i_1 < \ldots < i_p \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_p})$.*
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.*

**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 ?*

**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.*

**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* $\overline{\lim}\, A_n \in \mathcal{G}_\infty$.

**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* $(N_1/N, \ldots, N_d/N) \to (p_1, \ldots, p_d)$.

**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.*

**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* $1 \leq i \leq n$. *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.*

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

**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 $1 - p$.*

**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 $\lambda = \lambda_1 + \cdots + \lambda_d$.*

**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.*

**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 $\varepsilon > 0$, $\lim_{r\to\infty} \mathbb{P}\left(\left|\frac{T}{r\ln(r)} - 1\right| > \varepsilon\right) = 0$.*

**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 ;*

**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 $b \to \infty$ ?*

**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.*

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

1. *On construit $\sigma_n$ 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 $\sigma_n$ de loi uniforme sur $\Sigma_n$ ;*
2. *On construit $\sigma_n$ 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 $1\le i\le n$. Montrer que $\sigma_n$ ainsi construite suit la loi uniforme sur $\Sigma_n$ ;*
3. *Quel est le meilleur algorithme ?*

**Exercice 1.9** (Partitions aléatoires et algorithme de Stam). *Soit $\Pi_n$ l'ensemble des partitions de $\{1,\ldots,n\}$, et $B_n=\mathrm{card}(\Pi_n)$. En combinatoire, la suite $(B_n)_{n\ge1}$ 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 $k\ge0$. 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 $\Pi_n$.*

**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*
$$d(X_1+\cdots+X_n,Y_1+\cdots+Y_n)\le d(X_1,Y_1)+\cdots+d(X_n,Y_n);$$
2. *Montrer que si $X$ (et $Y$) est de Bernoulli (et de Poisson) de moyenne $p$ alors $d(X,Y)\le 2p^2$ ;*
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*
$$d(X_1+\cdots+X_n,Y)\le 2(p_1^2+\cdots+p_n^2)\le 2\lambda\max_{1\le k\le n}p_k;$$
*Étudier le cas spécial où $p_1=\cdots=p_n=\lambda/n$.*

**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 $n\ge0$.*

**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 $\sum_{k\ge1}p_k=1$. 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 $m:=\sum_{k\ge1}kp_k$ est la taille moyenne des foyers dans la population.*

**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 ?*

## 2 Variables aléatoires réelles, notion de densité de probabilité

**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.*

**Exercice 2.2** (Loi uniforme). *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 $\mathbb{E}(Y)$, $\mathbb{E}(Y^2)$, $\mathbb{E}(Z)$, $\mathbb{E}(Z^2)$, $\mathbb{E}(YZ)$.*

**Exercice 2.3** (Saturation et détection). *Soit $X$ une v.a.r. exponentielle de paramètre $\lambda$, 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 ?*

**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 $\varepsilon$ 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 ?*

**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 $Y = |X|$ ?*

**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 ?*

**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 $\lambda$ ;*
3. *Calculer les moments de la loi $\Gamma(a, \lambda)$.*

**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 $\{0, \dots, b - 1\}$ ?*

**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 $[0, \infty]$,*

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

**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 : $X \leq \theta\mathbb{E}(X) + X\mathbf{1}_{\{X>\theta\mathbb{E}(X)\}}$.*

**Exercice 2.11** (Caractérisation de la loi exponentielle par l'absence de mémoire). *Montrer que si $X$ est une v.a.r. positive telle que $\mathbb{P}(X > 0) > 0$, alors $X$ suit une loi exponentielle si et seulement si pour tout $t \geq 0$ la loi conditionnelle de $X - t$ sachant $\{X > t\}$ est égale à la loi de $X$. Est-il raisonnable de modéliser une durée de vie par une loi exponentielle ?*

**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 ?*

**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 $\mathfrak{Im} z$.*

**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 $\lambda$ ;*
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 $H(Y) \leq H(X)$.*

**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 $\mathbb{E}(X^{2n}) = \prod_{k=1}^{n}(2k-1) = \frac{(2n)!}{2^n n!}$.*

**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.*

**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. *$\varphi$ est analytique sur un voisinage de $0$*
2. *$\varphi$ 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.*

**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.*

# 3 Vecteurs aléatoires à densité

**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. $T_1, \dots, T_n$ indépendantes de loi exponentielle de paramètres $\lambda_1, \dots, \lambda_n$.*

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 ;*

**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 ?*

**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.*

**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 !).*

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

# 4 Vecteurs aléatoires à densité (bis)

**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 $f_Z(z) = (\pi(1+z^2))^{-1}$.*
- *Montrer que si $Z$ est une v.a.r. de loi de Cauchy alors il en va de même de $1/Z$.*

**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*

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

**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 $\lambda$. Montrer que $X+Y$ suit la loi Gamma de densité $\lambda^2 x^{-1} e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$ ;*
5. *Établir par récurrence sur $n$ que si $X_1,\ldots,X_n$ sont des v.a.r. indépendantes de loi exponentielle de paramètre $\lambda$ 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 $X_1,\ldots,X_n$ sont des v.a.r. indépendantes avec $X_k \sim \mathcal{N}(m_k,\sigma_k^2)$ pour tout $1 \leq k \leq n$, alors $X_1+\cdots+X_n \sim \mathcal{N}(m_1+\cdots+m_n, \sigma_1^2+\cdots+\sigma_n^2)$.*

**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 $\theta$ sont indépendantes avec $r^2$ de loi exponentielle de paramètre $1/2$ et $\theta$ 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 $n \in \mathbb{N}^*$.*

**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 $\lambda$ 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 $\lambda$, et en déduire un algorithme de simulation de cette loi. Étudier sa complexité.*

**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 $\frac{1}{|D|}\mathbf{1}_D$.*

**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 $[0,1]$.*

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

**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.*

**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,*

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

**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 ?*

**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 $m$*

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

   *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 $B_n$.*

**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 $n \to \infty$ ?*

**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 $m := \mathbb{E}(X_1) = \sum_{k\geq 1} p_k < \infty$ 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 $k \in \mathbb{N}^*$.*

**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 $0 := t_0 < t_1 < \ldots < t_n$, les v.a. $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ 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.*

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

**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 ?*

**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* $\{\pm e_1, \ldots, \pm e_d\}$ *où* $e_1, \ldots, e_d$ *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 ?*

**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* $\mu$ *et de variance* $\sigma^2$*. 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* $\lambda_1, \ldots, \lambda_m$ *les valeurs propres de la matrice symétrique aléatoire* $\frac{1}{n} M M^\top$ *(ce sont des v.a.r. dépendantes à valeurs dans* $\mathbb{R}_+$*). Montrer que*

$$\frac{\lambda_1 + \cdots + \lambda_m}{m} \xrightarrow[mn\to\infty]{p.s.} \sigma^2 + \mu^2.$$

**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.*

**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* $\sum_{n\geq 1} \mu_n([c-\varepsilon, c+\varepsilon]^c) < \infty$
2. *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.r. indépendantes, centrées, et bornée dans* $\mathrm{L}^4$*. 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* $\mathrm{L}^r$ *avec* $r < 4$*.*

# 6 TCL, convergence en loi, fonctions caractéristiques

**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 $\lambda$ ;*
2. *Montrer que si $Y$ suit la loi exponentielle de paramètre $\lambda$ alors $\lfloor Y \rfloor$ et $Y - \lfloor Y \rfloor$ sont indépendantes et $1 + \lfloor Y \rfloor$ suit la loi géométrique de paramètre $e^{-\lambda}$.*

**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*

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

**Exercice 6.3** (Méthode Delta)**.** *Supposons que $A_n(Z_n - B_n) \xrightarrow{\text{loi}} L$ où $(A_n)_{n\geq 1}$ et $(B_n)_{n\geq 1}$ 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 $f'(B) \neq 0$, 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 ?*

**Exercice 6.4** (Convergences de marginales)**.** *Soient $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, $X$, $Y$ 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.*

**Exercice 6.5** (Critère de densité et d'inversion)**.** *Soit $X$ une v.a.r. de loi $\mu$ et de fonction caractéristique $\varphi$ donnée par $\varphi(t) = \mathbb{E}(e^{itX}) = \int e^{itx}\,d\mu(x)$. On se propose d'établir que si $\varphi$ est Lebesgue intégrable alors $\mu$ 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 $\gamma_\sigma$ la densité de $\mathcal{N}(0,\sigma^2)$,*

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

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 $\sigma$ vers $0$.*

**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.*

**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 ?*

**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.*

**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 $\lambda$, 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. ?*

**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 $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} \in \mathbb{R} \cup \{\infty\}$.*

**Exercice 6.11** (Convergence en probabilité impossible dans le TCL)**.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. de carré intégrable, de moyenne nulle et de variance unité. Soit $Z_n := \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$.*

1. *Rappeler la convergence en loi de $(Z_n)_{n\geq 1}$*
2. *En écrivant $Z_{2n} - Z_n$ sous la forme $aZ_n + bZ'_n$ avec $Z'_n$ copie indépendante de $Z_n$, établir que $(Z_{2n} - Z_n)_{n\geq 1}$ converge en loi et identifier la limite.*
3. *En déduire que $(Z_n)_{n\geq 1}$ ne peut pas converger en probabilité.*

**Exercice 6.12** (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 $v \in \mathbb{R}^n$.*

# 7 Vecteurs gaussiens, intervalles de confiance, 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* $n = 10000$*.*

**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% ?*

**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 ?*

**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* $\Sigma$*. Trouver des estimateurs de* $m$ *et de* $\Sigma$ *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* $d \geq 1$*.*

**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* $\lambda$ *et calculer son biais.*

**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* $\theta$*. Construire un intervalle de confiance asymptotique ;*
2. *Montrer que l'estimateur* $M_n = \max(X_1,\ldots,X_n)$ *de* $\theta$ *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* $M_n \overset{\mathbb{P}}{\to} \theta$*. 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* $\theta$*.*

**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* $\mu$ *et* $\nu$ *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* $\mu$ *est une loi sur* $\mathbb{N}$ *alors les propriétés suivantes sont équivalentes :*

*(a)* $(S_n)_{n\geq 1}$ *converge en loi vers* $\mu$ *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*

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

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

**Exercice 8.1** (Théorème de Cochran[1]). *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 $\sigma^{-2}\|Y_k - \mathbf{P}_k m\|^2 \sim \chi^2(d_k)$.*

**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*

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k \qquad \text{et} \qquad \sigma_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \overline{X}_n)^2.$$

*Monter que les variables aléatoires $\overline{X}_n$ et $V_n^2$ sont indépendantes avec*

$$\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \qquad \text{et} \qquad \frac{(n-1)}{\sigma^2}\sigma_n^2 \sim \chi^2(n-1).$$

*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 ?*

**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 ?*

2. *Soit $X_1, \ldots, X_n$ un échantillon d'une loi discrète inconnue $q = (q_1, \ldots, q_k)$ sur $\{1, \ldots, k\}$. On souhaite savoir si cette loi $q$ est égale à une loi discrète de référence $p = (p_1, \ldots, p_k)$ (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 $S_n = S(X_1, \ldots, X_n)$ 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*
   — *$\widehat{q} = (N_1/n, \ldots, N_k/n)$ est l'estimateur empirique de $q$ ;*
   — *$N_i := \mathrm{card}\{1 \leq m \leq n : X_m = i\} = \sum_{m=1}^n \mathbf{1}_{\{X_m = i\}}$ 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 $X_1, \ldots, X_n$, on décide comme suit :*
   — *Si $S_n \in \mathcal{R}_\alpha$, on accepte l'hypothèse $H_0$ ;*

1. Prononcer «Kok-rane»

— *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 $\alpha$ 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 $\alpha$ pour la loi $\chi^2$ ?*
5. *Comment adapter le test du chi-deux au cas des échantillons à densité ? À quoi correspond la terminologie « test non-paramétrique » ?*

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

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

1. *Montrer que pour tout $\omega$, $F_n$ 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 $x \in \mathbb{R}$, $H_n(x) := \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{F^{-1}(U_k)\leq x\}} = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{U_k \leq F(x)\}}$, et que les suites $(\|F_n - F\|_\infty)_{n\geq 1}$ et $(\|H_n - F\|_\infty)_{n\geq 1}$ ont même loi ;*
3. *Montrer que $\|H_n - F\|_\infty \leq \|G_n - G\|_\infty$, 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 $q_{1-\alpha}$.*

**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 :*
$$\sqrt{n}\,\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{loi}} K_1 \quad \textit{et} \quad \sqrt{n}\,\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{loi}} K_2,$$
*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*
$$F_{K_1}(]-\infty,u]) = 1 - e^{-2u^2} \quad \textit{et} \quad F_{K_2}(]-\infty,u]) = 1 + 2\sum_{k=1}^\infty (-1)^k e^{-2k^2u^2}.$$
*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 $H_1$.*

# 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 $\alpha$ et de variance $\sigma^2$. 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 $\beta$, et de variance $\tau^2$. 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 $\alpha, \beta, a, b$, 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 $n$ ?*

**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 $p_x = \frac{x}{N}(1-u) + \left(1 - \frac{x}{N}\right)v$,*
$$P_{x,y} = \mathbb{P}(X_{t+1} = y | X_t = x, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = \binom{N}{y} p_x^y (1-p_x)^{N-y};$$
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 $x_0$) ;*

**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 distributées de loi exponentielle de paramètre $\lambda$. 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 $\lambda t$, 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 $p \geq 1$,*

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