Feuille de révision probas

Antoine Roche

16 septembre 2026

1 Disclaimer

Ce document est une feuille de révision destinée aux 2AD pour le cours de probabilités de début d'année. Il est destiné purement à vous aider, notez cependant que ce n'est pas un document "officiel" créé par votre professeur Jaouad Mourtada et qu'il ne fait donc aucune autorité quelle qu'elle soit. En cas de doute sur une affirmation, votre cours et tout le contenu produit par votre professeur sont donc les seules et uniques références.

Table des matières

1 Disclaimer **1**

2 Feuille de révision **2**

2.1 Espaces de probabilités et tribus . . . 2
2.2 Fonction de répartition . . . 2
2.3 Déterminer la loi d'une transformation $Y = g(X)$ . . . 2
2.4 Espérance et moments . . . 3
2.5 Vecteurs aléatoires et indépendance . . . 3
2.5.1 Densités jointes et marginales . . . 3
2.5.2 Somme de variables indépendantes . . . 4
2.5.3 Variables et vecteurs gaussiens . . . 4
2.6 Inégalités et bornes utiles . . . 5
2.7 Limites de variables aléatoires . . . 6
2.7.1 Les modes de convergence . . . 6
2.7.2 Théorème majeur 1 : la loi des grands nombres (LGN) . . . 6
2.7.3 Théorème majeur 2 : le théorème central limite (TCL) . . . 7
2.7.4 Combiner les limites : image continue et Slutsky . . . 8
2.7.5 Méthode générale : convergence en probabilité « à la main » . . . 8
2.7.6 Convergence en loi : les méthodes . . . 9

3 Corrigé de l'exercice 3, TD3 **10**

3.1 Loi de S_n . . . 10
3.2 Convergence en loi vers $\mathcal{P}(\lambda)$ . . . 11

2 Feuille de révision

2.1 Espaces de probabilités et tribus

Définition : tribu, tribu engendrée

Une tribu $\mathcal{F}$ sur Ω contient Ω, est stable par complémentaire et par union dénombrable. Pour $\mathcal{C} \subset \mathcal{P}(\Omega)$, $\sigma(\mathcal{C})$ est la plus petite tribu contenant $\mathcal{C}$ (intersection de toutes les tribus contenant $\mathcal{C}$).

Méthode : montrer que deux tribus engendrées coïncident

Pour montrer $\sigma(\mathcal{B}) = \sigma(\mathcal{C})$, il suffit de montrer

$$\mathcal{B} \subset \sigma(\mathcal{C}) \quad \text{et} \quad \mathcal{C} \subset \sigma(\mathcal{B}).$$

En effet $\mathcal{B} \subset \sigma(\mathcal{C})$ donne $\sigma(\mathcal{B}) \subset \sigma(\mathcal{C})$ (minimalité), et symétriquement. En pratique, on écrit chaque élément de $\mathcal{B}$ comme union/intersection *dénombrable* ou complémentaire d'éléments de $\mathcal{C}$, par exemple

$$[a,b] = \bigcap_{n\geq 1} \left]a - \tfrac{1}{n}, b + \tfrac{1}{n}\right[, \qquad]a,b[= \bigcup_{n\geq 1} \left[a + \tfrac{1}{n}, b - \tfrac{1}{n}\right],$$

$$]-\infty, x] = \bigcap_{n\geq 1}]-\infty, x_n] \quad \text{avec } x_n = \tfrac{\lceil nx \rceil}{n} \in \mathbb{Q},\ x_n \geq x,\ x_n \to x.$$

2.2 Fonction de répartition

Définition

$F_X(x) = \mathbb{P}(X \leq x)$. Elle est croissante, continue à droite, de limites 0 en $-\infty$ et 1 en $+\infty$. Elle **caractérise la loi** de X.

- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a), \quad \mathbb{P}(X = x) = F_X(x) - F_X(x^-)$.
- Si X a une densité f par rapport à Lebesgue, alors $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$.

2.3 Déterminer la loi d'une transformation $Y = g(X)$

On se place dans le cas où l'on connaît la loi de X et où l'on cherche à déterminer la loi de $g(X)$.

Attention : Si une variable aléatoire est définie sur $\mathbb{R}$ (par exemple) alors sa FDR (et son éventuelle densité par rapport à Lebesgue) sont définies aussi sur tout cet ensemble, ainsi, lors de la détermination de ces deux quantités, ne pas oublier de préciser les valeurs que prennent ces fonctions partout même si c'est "évident". Par exemple : Pour $U \sim \mathcal{U}([0,1])$ mentionner que F_U est nulle sur $]-\infty, 0[$ et qu'elle vaut 1 sur $]1, +\infty[$.

De manière analogue, ne pas oublier les fonctions indicatrices lors de la détermination d'une éventuelle densité !

Méthode 1 : via la fonction de répartition (FDR)

Calculer $F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(Y \leqslant y)$. Cela suffit à déterminer la loi de Y. En pratique, on réécrit l'événement $(Y \leqslant y)$ en termes de X dont on connaît la FDR. Exemples :

— $Y = U^\alpha$, $U \sim \mathcal{U}([0,1])$: $F_Y(y) = \mathbb{P}(U \leq y^{1/\alpha}) = y^{1/\alpha}$ sur $[0,1]$ nulle pour $y < 0$ et 1 pour $y > 1$. Ici, la FDR est C^1 par morceaux ET continue, on peut donc la dériver pour déduire la densité $f_Y(y) = \frac{1}{\alpha} y^{1/\alpha - 1} \mathbf{1}_{]0,1]}(y)$.

— $Z = -\frac{1}{\lambda} \log U$: Déjà Z est bien définie car $U \neq 0$ p.s. $U \in [0,1]$ donne $Z \geqslant 0$ p.s., pour $z \geqslant 0$, $\mathbb{P}(Z > z) = \mathbb{P}(U < e^{-\lambda z}) = e^{-\lambda z}$, donc $Z \sim \mathcal{E}(\lambda)$.

— $Y = 1/X$, $X \sim \mathcal{E}(1)$: pour $y \leqslant 0, F_Y(y) = 0$, pour $y > 0$, $F_Y(y) = \mathbb{P}(X \geq 1/y) = e^{-1/y}$, donc $f_Y(y) = y^{-2} e^{-1/y} \mathbf{1}(y > 0)$. Par ailleurs, dans ce dernier cas, mentionner aussi que Y est bien définie car $X \neq 0$ p.s.

Bien distinguer les cas où g est croissante ou décroissante (sens des inégalités) et traiter les y hors du support.

Méthode 2 : via une fonction test

Y a pour densité g_Y (par rapport à Lebesgue) si et seulement si, pour toute h mesurable positive (ou bornée),

$$\mathbb{E}[h(Y)] = \int h(y)\, g_Y(y)\, dy.$$

On écrit $\mathbb{E}[h(Y)] = \int h(g(x)) f_X(x)\, dx$ et on fait des changements de variables. Bien déterminer les domaines de départ et d'arrivée et bien justifier que le changement de variable que vous vous apprêtez à faire est un C^1-difféomorphisme.

2.4 Espérance et moments

— Théorème de transfert : sous réserve d'existence $\mathbb{E}[h(X)] = \int h(x) f_X(x)\, dx$ (cas Lebesgue), $= \sum_k h(k) \mathbb{P}(X = k)$ (cas discret).

— $X \in L^p$ ssi $\mathbb{E}[|X|^p] < \infty$. Toujours vérifier l'intégrabilité *avant* de calculer : $f(x) = 3x^{-4} \mathbf{1}(x \geq 1)$ donne $\mathbb{E}[X^2] = \int_1^\infty 3x^{-2} dx = 3 < \infty$ tandis que $Y = 1/X$ avec $X \sim \mathcal{E}(1)$ donne $\mathbb{E}[Y] = \int_0^\infty x^{-1} e^{-x} dx = +\infty$.

2.5 Vecteurs aléatoires et indépendance

Ici, on se place dans le cadre de variables aléatoires dans $\mathbb{R}^d, d \geqslant 1$.

2.5.1 Densités jointes et marginales

Si on prend l'exemple du vecteur aléatoire (X, Y) à valeurs dans $\mathbb{R}^2$ et que l'on suppose que (X, Y) admet pour densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, alors

— Marginales : $f_X(x) = \int_{\mathbb{R}} f(x,y)\, dy$, $f_Y(y) = \int_{\mathbb{R}} f(x,y)\, dx$, ainsi X et Y ont toutes deux des densités par rapport à la mesure de Lebesgue.

— X et Y (à densité pour Lebesgue f_X, f_Y) sont indépendantes ssi $f(x,y) = f_X(x) f_Y(y)$ presque partout.

Exemple (TD2, Ex. 2)

$f(x,y) = x e^{-x(y+1)} \mathbf{1}(x \geq 0) \mathbf{1}(y \geq 0)$. Alors $f_X(x) = e^{-x} \mathbf{1}(x \geqslant 0)$ et $f_Y(y) = \frac{1}{(1+y)^2} \mathbf{1}(y \geqslant 0)$. Comme $f \neq f_X f_Y$ sur un ensemble de mesure non nulle, X et Y ne sont pas indépendantes.

Fubini–Tonelli :

— Tonelli : pour $f \geq 0$ mesurable, on peut intervertir les intégrales sans condition.

— Fubini : idem si $\iint |f| < \infty$.

2.5.2 Somme de variables indépendantes

Méthode : loi de $X+Y$ avec $X \perp\!\!\!\perp Y$

— **Discret** : $\mathbb{P}(X+Y=n) = \sum_k \mathbb{P}(X=k)\mathbb{P}(Y=n-k)$.
Ex. : $\mathcal{P}(\lambda) * \mathcal{P}(\mu) = \mathcal{P}(\lambda+\mu)$ par la formule du binôme.

— **Fonctions caractéristiques** : $\varphi_{X+Y} = \varphi_X \varphi_Y$, puis on reconnaît une loi.

Il y a aussi une formule avec la convolution si on a les densités à disposition que j'ai jugée peu utile ici. A titre indicatif :

Convolution : $f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)\,dx$ (attention aux indicatrices, qui fixent les bornes).

Ex. : $X, X' \sim \mathcal{E}(1)$ i.i.d. : $f_{X+X'}(z) = \int_0^z e^{-x} e^{-(z-x)} dx = z e^{-z} \mathbf{1}(z \geqslant 0)$.

2.5.3 Variables et vecteurs gaussiens

— $X \sim \mathcal{N}(m, \sigma^2)$: densité $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$, $\varphi_X(t) = e^{itm - \sigma^2 t^2/2}$.

— Si $X_1, \dots, X_n$ sont indépendantes, $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$, alors $\sum a_i X_i \sim \mathcal{N}\big(\sum a_i m_i, \sum a_i^2 \sigma_i^2\big)$.

Définition : vecteur gaussien

$(X_1, \dots, X_n)$ est gaussien si toute combinaison linéaire $\sum a_i X_i$ est gaussienne. Sa loi est déterminée par son vecteur moyenne et sa matrice de covariance. Des gaussiennes indépendantes forment un vecteur gaussien, et toute transformation linéaire d'un vecteur gaussien est un vecteur gaussien.

Identités utiles pour les vecteurs Gaussiens :

Soit $X \sim \mathcal{N}(\mu, \Sigma)$ un vecteur Gaussien à valeurs dans $\mathbb{R}^d$. Alors :

— Pour $A \in \mathbb{R}^{d \times d}$, $AX \sim \mathcal{N}(A\mu, A\Sigma A^\top)$ (toujours un vecteur Gaussien de taille d).

— Pour $u \in \mathbb{R}^d$, $\langle u,\, X\rangle \sim \mathcal{N}(\langle u,\, \mu\rangle, \langle u,\, \Sigma u\rangle)$ une variable aléatoire Gaussienne *réelle*.

En particulier, pour $\mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d\times d}$ (semi-définie positive), si $X \sim \mathcal{N}(0, I_d)$ et $Y \sim \mathcal{N}(\mu, \Sigma)$ on a $Y \overset{(\mathrm{d})}{=} \mu + \Sigma^{1/2} X$ (où $\overset{(\mathrm{d})}{=}$ dénote l'égalité en distribution/loi).

Théorème : indépendance dans un vecteur gaussien

Si (U, V) est un vecteur gaussien, alors $U \perp\!\!\!\perp V \iff \mathrm{Cov}(U, V) = 0$.

C'est fondamental, le sens $\implies$ est toujours vrai, le cas très particulier des vecteurs Gaussiens donne la réciproque.

Attention

Il faut que le vecteur soit *conjointement* gaussien : deux variables gaussiennes décorrélées ne sont pas forcément indépendantes. On justifie toujours que (U, V) est l'image linéaire d'un vecteur de gaussiennes indépendantes.

Par exemple, si on prend $X \sim \mathcal{N}(0,1)$ et $\varepsilon \sim \mathrm{Rad}(1/2)$ i.e. $\varepsilon = -1$ avec proba $1/2$ et $\varepsilon = 1$

avec proba $1/2$ avec $\varepsilon \perp\!\!\!\perp X$, on a bien $X, \varepsilon X \sim \mathcal{N}(0,1)$ (de même loi toutes les deux), $\mathrm{Cov}(\varepsilon X, X) = 0$ mais pas $\varepsilon X \perp\!\!\!\perp X$, c'est parce que le vecteur $(X, \varepsilon X)$ n'est **pas** un vecteur Gaussien.

Méthode : montrer qu'un vecteur est un vecteur Gaussien

Plusieurs possibilités, le plus simple est de montrer que le vecteur est lui même une transformation linéaire d'un vecteur Gaussien que l'on connaît. Par exemple supposons donné $(X, Y) \sim \mathcal{N}(0, I_2)$ (vecteur Gaussien standard en dimension 2).
Pour montrer que $(X+Y, X-Y)$ est un vecteur Gaussien il suffit de remarquer

$$\begin{pmatrix} X+Y \\ X-Y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

qui est bien une transformation linéaire d'un vecteur Gaussien. Il est aussi possible d'appliquer la définition (moins pratique en général), s'il est possible de montrer que *toute* combinaison linéaire des éléments du vecteur est Gaussienne alors cela conclut aussi.

Méthode : montrer une indépendance entre gaussiennes

1. Bien montrer que (U, V) est un vecteur Gaussien (fondamental).
2. Calculer $\mathrm{Cov}(U, V)$ et montrer que cette covariance est nulle.

Ex. : $X_\theta = X\cos\theta + Y\sin\theta$: $X_\theta \sim \mathcal{N}(0,1)$, par exemple car

$$X_\theta = \begin{pmatrix} \cos\theta & \sin\theta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = u^\top Z \sim \mathcal{N}(0, \|u\|^2)$$

où $Z \sim \mathcal{N}(0, I_2)$ et $u^\top = (\cos\theta \ \sin\theta)$ est de norme unitaire.
On a donc que le vecteur $(X_\theta \ X_{\theta'})$ est Gaussien pour tout θ, θ' car

$$\begin{pmatrix} X_\theta \\ X'_\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \cos\theta' & \sin\theta' \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Maintenant que l'on a montré que ce vecteur est bien un vecteur Gaussien, il suffit de montrer que X_θ et $X_{\theta'}$ ne sont pas corrélés (covariance nulle) pour montrer qu'ils sont indépendants. Il ne reste qu'à calculer. $\mathrm{Cov}(X_\theta, X_{\theta'}) = \cos(\theta - \theta')$, donc indépendance ssi $\theta' - \theta \in \frac{\pi}{2} + \pi\mathbb{Z}$.

2.6 Inégalités et bornes utiles

Théorèmes : inégalités de concentration élémentaires

— **Markov** : pour $X \geqslant 0$ p.s. intégrable et $a > 0$, $\quad \mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[X]}{a}$.

— **Bienaymé–Tchebychev** : pour $X \in L^2$ et $a > 0$, $\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \dfrac{\mathrm{Var}(X)}{a^2}$.

Note : Markov peut être appliqué à $|X|^p$ pour tout $p > 0$ qui est positive p.s. même si X n'est pas forcément positive p.s., dans ce cas

$$\mathbb{P}(|X| \geqslant a) = \mathbb{P}(|X|^p \geqslant a^p) \leqslant \frac{\mathbb{E}\left[|X|^p\right]}{a^p},$$

quitte à avoir le membre de droite valant $+\infty$ si X n'est pas L^p.

2.7 Limites de variables aléatoires

2.7.1 Les modes de convergence

Définition

Soit (X_n) et X des variables aléatoires réelles.

- **Presque sûre** : $X_n \xrightarrow[n\to\infty]{\text{p.s.}} X$ si $\mathbb{P}\big(\{\omega : X_n(\omega) \to X(\omega)\}\big) = 1$.
- **Dans L^p** : $X_n \xrightarrow[n\to\infty]{L^p} X$ si $\mathbb{E}[|X_n - X|^p] \to 0$.
- **En probabilité** : $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ si pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$.
- **En loi** : $X_n \xrightarrow[n\to\infty]{(\mathcal{L})} X$ si $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ pour toute h continue bornée.

Certaines convergences sont plus fortes que d'autres, sommairement, on peut résumer de la façon suivante.

Théorème : hiérarchie

$$\begin{array}{ccccc} \text{p.s.} & \Longrightarrow & \text{en probabilité} & \Longrightarrow & \text{en loi} \\ & & \Uparrow & & \\ & & L^p & & \end{array}$$

De plus, **si la limite est une constante** c, $X_n \xrightarrow[n\to\infty]{(\mathcal{L})} c \iff X_n \xrightarrow[n\to\infty]{\mathbb{P}} c$.

Attention

La convergence en loi ne porte que sur les lois : elle ne dit rien de $X_n - X$, et on ne peut pas additionner ou multiplier deux convergences en loi sans hypothèse supplémentaire (voir Slutsky pour ce que l'on peut dire dans certains cas). Les réciproques de la hiérarchie sont fausses sans hypothèses additionnelles.

2.7.2 Théorème majeur 1 : la loi des grands nombres (LGN)

Théorème : loi forte des grands nombres

Si $(W_i)_{i\geqslant 1}$ sont **i.i.d.** et **intégrables** (dans L^1), alors

$$\frac{1}{n}\sum_{i=1}^{n} W_i \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}[W_1].$$

Version faible : le même énoncé donne que la convergence a aussi lieu en probabilité car la convergence presque sûre est plus forte.

Méthode : quand et comment appliquer la LGN

On pense à la LGN dès qu'on voit une quantité de la forme « **somme de n termes divisée par n** » et qu'il est demandé une limite p.s./proba. Check-list :

1. **Réécrire** la quantité sous la forme $\frac{1}{n}\sum_{i=1}^{n} W_i$, en identifiant les W_i (souvent une fonction d'un bloc de variables).
2. **Indépendance** : les W_i sont-elles indépendantes ? C'est souvent direct, parfois cela

peut être des fonctions de blocs *disjoints* de variables indépendantes (cf TD3 exo 2).

3. **Même loi** : Bien vérifier que les W_i ont la même loi (souvent donné dans l'énoncé lui même).
4. **Intégrabilité** : $\mathbb{E}[|W_1|] < \infty$.
5. **(si nécessaire) Calculer** $\mathbb{E}[W_1]$: c'est la limite.

Exemple (TD3, Ex. 2.1)

$Y_n = \sum_{i=1}^n (X_{2i-1} - X_{2i})^2$. On pose $W_i = (X_{2i-1} - X_{2i})^2$: les blocs $(X_1, X_2), (X_3, X_4), \dots$ sont disjoints donc les W_i sont i.i.d. ; $X \in L^2$ donc $W_1 \in L^1$; et $\mathbb{E}[W_1] = \mathrm{Var}(X_1 - X_2) = 2\sigma^2$. Donc $Y_n/n \xrightarrow[n\to\infty]{\text{p.s.}} 2\sigma^2$.

Attention : quand la LGN ne s'applique pas directement

— Variables **non identiquement distribuées** (ex. TD3, Ex. 4 : la loi de X_k dépend de k).

— Variables non intégrables.

— Les termes W_i ne sont pas indépendants (blocs qui se chevauchent, par ex. $(X_i - X_{i+1})^2$).

Dans ces cas, on revient à la méthode générale (section 2.7.5) ou on calcule directement une loi.

2.7.3 Théorème majeur 2 : le théorème central limite (TCL)

Théorème central limite

Si (W_i) sont **i.i.d.**, de carré intégrable (dans L^2), de moyenne μ et variance $\sigma^2 > 0$, alors

$$\sqrt{n}\,\frac{\bar{W}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{(\mathcal{L})} \mathcal{N}(0,1), \qquad \text{soit} \qquad \frac{\sum_{i=1}^n W_i - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{(\mathcal{L})} \mathcal{N}(0,\sigma^2).$$

Méthode : quand appliquer le TCL

On pense au TCL quand on voit une **somme centrée divisée par** $\sqrt{n}$ (ou une moyenne centrée multipliée par $\sqrt{n}$) et quand on cherche une limite en distribution. Même check-list que pour la LGN, en remplaçant l'intégrabilité par $\mathbb{E}[W_1^2] < \infty$, puis :

1. vérifier le centrage ;
2. calculer la variance $\sigma^2 = \mathrm{Var}(W_1)$: c'est la variance de la loi limite.

Utilisation : $\mathbb{P}\big(a \leqslant \frac{\sum W_i - n\mu}{\sigma\sqrt{n}} \leqslant b\big) \to \Phi(b) - \Phi(a)$ car Φ est continue.

Exemple (TD3, Ex. 2.2)

$Z_n = \sum_{i=1}^n (X_{2i-1} - X_{2i})$: les $W_i = X_{2i-1} - X_{2i}$ sont i.i.d., centrés, de variance $2\sigma^2$. Donc $Z_n/\sqrt{n} \xrightarrow[n\to\infty]{(\mathcal{L})} \mathcal{N}(0, 2\sigma^2)$.

Attention

Le TCL donne une convergence **en loi seulement** : $Z_n/\sqrt{n}$ ne converge ni p.s. ni en probabilité.

Remarque : Plus rare mais si jamais, le TCL peut servir dans certains cas à montrer une convergence en probabilité vers 0 d'une certaine suite de variables aléatoire, en effet, si l'on a grâce au TCL un résultat de la forme :

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{(\mathcal{L})} \mathcal{N}(0, \sigma^2),$$

on sait donc par Slutsky que pour tout $\delta > 0$, puisque $\frac{1}{n^\delta} \xrightarrow[n\to\infty]{} 0$, que

$$\frac{\sqrt{n}}{n^\delta}(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{(\mathcal{L})} 0.$$

Puisque 0 est une constante, cela donne la convergence en proba vers 0.

2.7.4 Combiner les limites : image continue et Slutsky

Théorème de l'application continue

Si g est continue et $X_n \to X$ p.s. (resp. en probabilité, resp. en loi), alors $g(X_n) \to g(X)$ p.s. (resp. en probabilité, resp. en loi). (Il suffit que g soit continue en tout point d'un ensemble de mesure 1 pour la loi de X.)

Théorème de Slutsky

Si $X_n \xrightarrow[n\to\infty]{(\mathcal{L})} X$ et $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$ où c **est une constante**, alors

$$X_n + Y_n \xrightarrow[n\to\infty]{(\mathcal{L})} X + c, \qquad X_n Y_n \xrightarrow[n\to\infty]{(\mathcal{L})} cX, \qquad X_n / Y_n \xrightarrow[n\to\infty]{(\mathcal{L})} X/c \quad (\text{si } c \neq 0).$$

Plus généralement $(X_n, Y_n) \xrightarrow[n\to\infty]{(\mathcal{L})} (X, c)$, et on applique l'application continue.

Méthode : LGN + TCL + Slutsky

1. Découper l'expression en un facteur qui converge en loi (TCL) et un facteur qui converge vers une *constante* (LGN).
2. Convertir la convergence p.s. de la LGN en convergence en probabilité (p.s. $\Rightarrow$ proba).
3. Appliquer Slutsky, puis identifier la loi limite (par ex. $a\mathcal{N}(0, s^2) = \mathcal{N}(0, a^2 s^2)$).

Ex. (TD3, Ex. 2.3) :

$$\frac{\sqrt{n}Z_n}{Y_n} = \frac{Z_n/\sqrt{n}}{Y_n/n} \xrightarrow[n\to\infty]{(\mathcal{L})} \frac{1}{2\sigma^2}\mathcal{N}(0, 2\sigma^2) = \mathcal{N}\Big(0, \frac{1}{2\sigma^2}\Big).$$

Attention

Bien comprendre que Slutsky est faux si Y_n converge en loi vers une variable *non constante* : on ne peut pas conclure sur $X_n + Y_n$ sans connaître la loi jointe.

2.7.5 Méthode générale : convergence en probabilité « à la main »

Quand la LGN ne s'applique pas, on revient à la définition.

Méthode : introduire ε et borner la probabilité de déviation

Pour montrer $X_n \xrightarrow[n\to\infty]{\mathbb{P}} c$:

1. Fixer $\varepsilon > 0$.
2. Borner $\mathbb{P}(|X_n - c| > \varepsilon)$ par une quantité qui tend vers 0, avec :
   - Markov : $\mathbb{P}(|X_n - c| > \varepsilon) \leq \mathbb{E}[|X_n - c|^p]/\varepsilon^p$;
   - Tchebychev si l'on connaît espérance et variance ;
   - un calcul exact de loi si c'est possible.
3. Conclure en faisant $n \to \infty$ à ε fixé.

Exemple (TD3, Ex. 4) : variables indépendantes non identiquement distribuées

$X_k = \sqrt{k+1}$ avec probabilité $1/\sqrt{k+1}$, 0 sinon. Alors $\mathbb{E}[X_k] = 1$, $\mathrm{Var}(X_k) = \sqrt{k+1} - 1$, donc

$$\mathbb{E}[S_n] = n, \qquad \mathrm{Var}(S_n) = \sum_{k=1}^{n} (\sqrt{k+1} - 1) \leq n\sqrt{n+1}, \quad \text{(indépendance)}.$$

Par Tchebychev : $\mathbb{P}\big(|S_n/n - 1| > \varepsilon\big) \leq \dfrac{\mathrm{Var}(S_n)}{n^2\varepsilon^2} \leq \dfrac{\sqrt{n+1}}{n\varepsilon^2} \to 0$. Donc $S_n/n \xrightarrow[n\to\infty]{\mathbb{P}} 1$.

Règles de calcul. Si $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ et $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y$, alors $X_n + Y_n \xrightarrow[n\to\infty]{\mathbb{P}} X + Y$ et $X_n Y_n \xrightarrow[n\to\infty]{\mathbb{P}} XY$.

2.7.6 Convergence en loi : les méthodes

On voit ici les méthodes générales pour montrer une convergence en loi (sans forcément passer par le TCL donc).

Théorème : caractérisations de la convergence en loi

Les assertions suivantes sont équivalentes :

1. $X_n \xrightarrow[n\to\infty]{(\mathcal{L})} X$;
2. $F_{X_n}(x) \to F_X(x)$ **en tout point x où F_X est continue** ;
3. $\varphi_{X_n}(t) \to \varphi_X(t)$ pour tout $t \in \mathbb{R}$ (théorème de Lévy).

Méthode A : limite des fonctions de répartition

1. Calculer $F_{X_n}(x)$ explicitement (typique pour un max, un min, une transformation simple), lorsqu'on a indépendance par exemple, le max plus petit qu'un seuil t est une probabilité d'intersection. Pour le minimum, c'est 1− une proba d'intersection.
2. Fixer x et faire tendre $n \to \infty$ (on peut supposer n assez grand pour que x soit dans le support de X_n).
3. Identifier la limite comme une fonction de répartition F : vérifier les limites en $\pm\infty$ (sinon il y a « perte de masse »).
4. La convergence n'est requise qu'aux points de continuité de F.

Exemple (TD3, Ex. 5) : maximum d'exponentielles et loi de Gumbel

$X_i \sim \mathcal{E}(1)$ i.i.d. Pour $x \in \mathbb{R}$ fixé et n assez grand pour que $\log n + x \geq 0$:

$$\mathbb{P}(M_n - \log n \leq x) = \left(1 - e^{-(\log n + x)}\right)^n = \left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow[n\to\infty]{} \exp(-e^{-x}) = G(x).$$

G est continue sur $\mathbb{R}$, donc $M_n - \log n \xrightarrow[n\to\infty]{(\mathcal{L})}$ Gumbel, de densité $g(x) = e^{-x}\exp(-e^{-x})$.

Méthode B : variables entières, limite des probabilités ponctuelles

Voire la solution ci-jointe Section 3. On regarde la limite de la fonction de masse à k fixé.

Méthode C : fonctions caractéristiques (Lévy)

Peut être utile pour les sommes de variables indépendantes (exemple de l'exercice 3 TD3) : $\varphi_{S_n} = \prod \varphi_{X_i}$. Ex. : $\varphi_{S_n}(t) = \left(1 + \frac{\lambda(e^{it}-1)}{n}\right)^n \to e^{\lambda(e^{it}-1)} = \varphi_{\mathcal{P}(\lambda)}(t)$.
C'est aussi la méthode de preuve du TCL ($\varphi(t/\sqrt{n})^n$ avec un développement à l'ordre 2).

Il existe d'autres conditions (Scheffé) (condition suffisante pour la convergence en loi si on a convergence presque partout des densités etc...). Non central ici.

3 Corrigé de l'exercice 3, TD3

Remarque préliminaire. Pour que la loi $\mathcal{B}(\lambda/n)$ ait un sens, il faut $\lambda/n \in [0,1]$, c'est-à-dire $n \geqslant \lambda$. On peut donc supposer dans toute la suite que soit $n \geqslant \lambda$ (ce qui est sans conséquence pour la question 2, où $n \to \infty$), soit $\lambda \in [0,1]$ ou encore tout simplement, si on n'impose rien, définir la variable aléatoire de Bernoulli de paramètre λ/n comme étant égale à 1 presque sûrement si $\lambda/n \geqslant 1$. On note $p_n = \lambda/n$.

3.1 Loi de S_n

Les variables $X_{1,n}, \ldots, X_{n,n}$ sont à valeurs dans $\{0,1\}$, donc S_n est à valeurs dans $\{0,1,\ldots,n\}$. En particulier, $\mathbb{P}(S_n = k) = 0$ pour tout entier $k > n$.

Soit $k \in \{0,\ldots,n\}$. L'événement $\{S_n = k\}$ est réalisé si et seulement si exactement k des variables $X_{i,n}$ valent 1, les autres valant 0. On a donc la réunion disjointe

$$\{S_n = k\} = \bigsqcup_{\substack{I \subset \{1,\ldots,n\} \\ |I| = k}} \Big(\bigcap_{i \in I} \{X_{i,n} = 1\} \cap \bigcap_{i \notin I} \{X_{i,n} = 0\} \Big).$$

Par indépendance des $X_{i,n}$, pour tout I de cardinal k,

$$\mathbb{P}\Big(\bigcap_{i \in I} \{X_{i,n} = 1\} \cap \bigcap_{i \notin I} \{X_{i,n} = 0\} \Big) = \prod_{i \in I} \mathbb{P}(X_{i,n} = 1) \prod_{i \notin I} \mathbb{P}(X_{i,n} = 0) = p_n^k (1 - p_n)^{n-k}.$$

Comme il y a $\binom{n}{k}$ parties I de cardinal k, on obtient

$$\boxed{\mathbb{P}(S_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad \text{pour } 0 \leqslant k \leqslant n, \qquad \mathbb{P}(S_n = k) = 0 \quad \text{pour } k > n.}$$

Autrement dit, S_n suit la loi binomiale $\mathcal{B}(n, \lambda/n)$.

Seconde méthode : Plus simplement, on peut aussi remarquer qu'à n fixé, les $(X_{i,n})_{i\in\{1,\dots,n\}}$ sont des Bernoulli indépendantes de paramètre λ/n de telle sorte que $\sum_{i=1}^n X_{i,n}$ est binomiale de paramètres $(n, \lambda/n)$.

3.2 Convergence en loi vers $\mathcal{P}(\lambda)$

Convergence des probabilités ponctuelles. Fixons $k \in \mathbb{N}$. Pour $n > \max(k, \lambda)$, on écrit

$$\mathbb{P}(S_n = k) = \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

k est fixé et on regarde chaque terme lorsque $n \to \infty$:

— $\dfrac{n(n-1)\cdots(n-k+1)}{n^k} = \dfrac{n}{n} \times \dfrac{n-1}{n} \times \dfrac{n-2}{n} \times \cdots \times \dfrac{n-k}{n} = \displaystyle\prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \xrightarrow[n\to\infty]{} 1$, comme produit d'un nombre fixe k de termes tendant vers 1 ;

— $\left(1 - \dfrac{\lambda}{n}\right)^{-k} \xrightarrow[n\to\infty]{} 1$, pour la même raison ;

— pour $n > \lambda$ (ok car on regarde bien la limite), $\left(1 - \dfrac{\lambda}{n}\right)^n = \exp\left(n \ln\left(1 - \dfrac{\lambda}{n}\right)\right)$, et comme $\ln(1-x) = -x + o(x)$ quand $x \to 0$,

$$n \ln\left(1 - \frac{\lambda}{n}\right) = n\left(-\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right) = -\lambda + o(1),$$

d'où, par continuité de l'exponentielle, $\left(1 - \dfrac{\lambda}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-\lambda}$ (c'est aussi une limite classique).

Par conséquent, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(S_n = k) \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}(S = k).$$

Si le théorème de Scheffé est dans le cours, cela suffit pour conclure, car il énonce que si $(X_n)_n$ est une suite de vecteurs aléatoires dans $\mathbb{R}^d$ et que si μ est une mesure σ-finie sur $\mathbb{R}^d$ muni de la tribu borélienne, alors si $f_n \xrightarrow[n\to\infty]{} f, \mu$-p.p. où f est bien une densité (positive qui intègre à 1) alors $(X_n)_n$ converge en distribution vers une variable aléatoire de loi limite de densité f par rapport à μ. Sinon :

Convergence des fonctions de répartition. Notons F_n et F les fonctions de répartition respectives de S_n et de S. Soit $x \in \mathbb{R}$.

— Si $x < 0$: comme S_n et S sont à valeurs dans $\mathbb{N}$, $F_n(x) = 0 = F(x)$ pour tout n.

— Si $x \geqslant 0$: comme S_n et S sont à valeurs dans $\mathbb{N}$,

$$F_n(x) = \mathbb{P}(S_n \leqslant x) = \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(S_n = k) \xrightarrow[n\to\infty]{} \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(S = k) = F(x),$$

la convergence découlant de l'étape 1, puisqu'il s'agit d'une somme d'un nombre *fini* (et indépendant de n) de termes.

Ainsi $F_n(x) \to F(x)$ pour tout $x \in \mathbb{R}$, a fortiori en tout point de continuité de F. Par caractérisation de la convergence en loi avec les FDR, on conclut que

$$S_n \xrightarrow[n\to\infty]{(\mathcal{L})} S, \qquad S \sim \mathcal{P}(\lambda).$$