L'esprit des lois de proba

October 20, 2017

1 Introduction

Commençons par des exemples d'objets aléatoires :

- Le lancé d'un dé ; c'est un entier aléatoire.
- Un appel de la fonction rand() ; c'est un réel aléatoire.
- Le nombre de sardines et la température moyenne de l'eau pour l'océan atlantique, à une date donnée ; c'est un couple aléatoire.
- Le nuage que l'on observera demain dans un coin de ciel. C'est une forme 3D aléatoire.
- La photo d'un carré de pelouse de l'orangerie, photo que j'irais prendre demain. C'est une image aléatoire 2D.

Le mot variable aléatoire (v.a.) désigne la plupart du temps un réel aléatoire. Les deux premiers exemples sont des v.a.

Un objet aléatoire X, est un objet qui n'est pas encore réalisé ou pas encore observé. On ne peut prévoir son aspect que de manière incertaine. On ne peut le décrire qu'au moyen d'une modélisation probabiliste.

Une fois réalisé, ou observé, cet objet n'est plus aléatoire. On parle alors d'une réalisation de X, réalisation que l'on notera x.

Exemple : 3 lancés de dés $X = (X_1, X_2, X_3)$. Ils forment un vecteur aléatoire de $\mathbb{R}^3$. Mais une fois jeté, les trois dès donnent, disons $x = (2, 1, 5)$, qui est un vecteur classique (déterministe). La probabilité de cette réalisation est $\mathbf{P}[X = (2, 1, 5)] = \frac{1}{6^3}$.

La loi d'un objet aléatoire X, c'est la probabilité que X fasse ceci ou cela. Moralement c'est la donnée des $\mathbf{P}[X = x]$ pour tous les x possibles. Cependant, cette définition fonctionne uniquement pour des objets aléatoires discrets, càd qui ne peuvent prendre qu'un nombre fini ou dénombrable de valeurs. Dans le cas général, la loi de X c'est la donnée des $\mathbf{P}[X \in A]$ pour tous les A possibles.

Tout ce cours tourne autour de la notion de loi. Comparé à un cours classique de licence :

- nous mettrons en avant ce qui est utile en modélisation.
- nous mettrons sous le tapis les difficultés théoriques.
- nous emploierons des notations sans doute nouvelles pour vous (Dirac, élément infinitésimal).

2 Préliminaires sur la théorie de la mesure

Si vous voulez plus de détails et de rigueur, après avoir lu cette partie, lisez l'annexe sur la théorie de la mesure.

2.1 Mesure et loi

Une mesure μ sur un ensemble E est une application qui à tout $A \subset E$ associe un réel positif et qui vérifie :

$$A \cap B = \emptyset \;\Rightarrow\; \mu(A \cup B) = \mu(A) + \mu(B) \tag{Additivité}$$

Une loi μ sur un ensemble E est un mesure qui vérifie, en plus de l'additivité :

$$\mu(E) = 1 \tag{Poids 1}$$

Le mot "loi" est synonyme de "mesure de probabilité".

2.2 La notation avec dx

Une fonction sur E se note parfois f ou parfois $f(x)$ où x est un point générique de l'ensemble de E.

Une mesure sur E se note parfois μ ou parfois $\mu(dx)$ où dx est un sous-ensemble infinitésimal de E.

Par exemple, plutôt que de noter $\forall A \; \mathbf{P}[X \in A] = \mu(A)$ on peut noter $\mathbf{P}[X \in dx] = \mu(dx)$. Nous verrons plus loin les avantages de cette notation.

2.3 Intégrale

Considérons une mesure μ sur E. Considérons une fonction $\varphi : E \to \mathbb{R}$. L'intégrale de φ contre μ se note :

$$\int_E \varphi \, \mu \qquad \text{ou bien} \qquad \int_E \varphi(x)\mu(dx)$$

Moralement c'est la "moyenne" de φ pondérée par μ. Imaginons une famille $(dx_i)_i$ de parties de E, telle que: 1/ les dx_i soient très petit, 2/ les dx_i soient disjoints, 3/ $E = \cup_i dx_i$ (vocabulaire : $(dx_i)_i$ est une partition de E). Alors:

$$\int_E \varphi(x)\mu(dx) \simeq \sum_i \varphi(x_i)\mu(dx_i)$$

2.4 Mesures de Lebesgue

Exemple 2.1 La mesure de Lebesgue sur $\mathbb{R}$ donne la longueur des parties de $\mathbb{R}$. En particulier

$$\forall [a, b] \subset \mathbb{R} \qquad Leb_1([a, b]) = b - a$$

Prenons $\varepsilon > 0$ et pour $i \in \mathbb{Z}$ posons $x_i = i\varepsilon$. Notons $dx_i = [x_i, x_{i+1}[$. Ainsi $(dx_i)_{i\in\mathbb{Z}}$ est une partition de $\mathbb{R}$.

$$\int_{\mathbb{R}} \varphi(x)Leb_1(dx) \simeq \sum_i \varphi(x_i)Leb_1(dx_i) = \sum_i \varphi(x_i)\varepsilon$$

Pour que $\simeq$ se transforme en $=$, il faut faire tendre ϵ vers zéro.

Exemple 2.2 la mesure de Lebesgue sur $\mathbb{R}^2$ donne l'aire des parties de $\mathbb{R}^2$. En particulier

$$\forall [a, b] \times [c, d] \subset \mathbb{R} \qquad Leb_2([a, b] \times [c, d]) = (b - a)(d - c)$$

On peut facilement imaginer $(dx_i)_i$ une partition de $\mathbb{R}^2$ composée de petits carrés. Alors :

$$\int_{\mathbb{R}^2} \varphi(x)Leb_2(dx) \simeq \sum_i \varphi(x_i)Aire(dx_i)$$

A partir de ces deux exemples, vous pouvez imaginer comment définir les mesures de Lebesgue sur $\mathbb{R}^n$. Mais attention, les mesures de Lebesgue sont très différentes d'une dimension à l'autre. Par exemple, sur $\mathbb{R}^2$, considérons la droite $D = \{0\} \times \mathbb{R}^+$. On a $Leb_2(D) = 0$ (une demi-droite a une aire nulle). Alors que $Leb_1(\mathbb{R}^+) = +\infty$ (une demi-droite a une longueur infinie).

Notation : La mesure de Lebesgue sur $\mathbb{R}^n$ est si courante qu'on la note couramment dx au lieu de $Leb_n(dx)$. On peut aussi la noter $dx_1 dx_2 ... dx_n$.

2.5 L'indicatrice

Si A est un ensemble, 1_A est la fonction indicatrice de A:

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

DESSIN

Les fonctions indicatrices sont extrêmement simples à intégrer :

$$\int_E 1_A(x)\mu(dx) = \mu(A)$$

Les fonctions étagées sont les fonction de la forme $\varphi = \sum_i w_i 1_{A_i}$. Elles s'intègrent comme ceci :

$$\int_E \varphi(x)\mu(dx) = \sum_i w_i \mu(A_i)$$

Exo : il manque une hypothèse que pour l'égalité ci-dessus soit vraie.

Toute fonction peut-être approchée par des fonctions étagées.

DESSIN

2.6 La Dirac

Soit E un ensemble et $k \in E$. La mesure de Dirac δ_k est une loi concentrée en un seul point. Elle est définie par

$$\forall A \subset E \quad \delta_k(A) = 1_A(k)$$

Ce qui revient aussi à $\delta_k(\{k\}) = 1$ et $\delta_k(E \setminus \{k\}) = 0$

DESSIN

On peut sommer des Dirac. Par exemple, la mesure de comptage sur $\mathbb{Z}$ est $cont = \sum_{i \in \mathbb{Z}} \delta_i$. Si $A \subset \mathbb{Z}$ alors $cont(A) = ...$

DESSIN

Propriété et exo : soient $A, B \subset \mathbb{Z}$ tels que l'on peut passer de l'un à l'autre par une translation, alors $cont(A) = cont(B)$. Quelle autre mesure a cette propriété ?

2.7 Intégrale avec des Dirac

Puisque la mesure de Dirac δ_k met tout son poids en k, elle vérifie donc

$$\int_E \varphi(x)\delta_k(dx) = \varphi(k)$$

Soient $k_i \in E$ et w_i des poids, en utilisant l'équation du dessus et la linéarité de l'intégrale :

$$\int \varphi(x) \textstyle\sum_i w_i \, \delta_{k_i}(dx) = \displaystyle\sum_i w_i \, \varphi(k_i)$$

Nous utiliserons cette formule dans la section sur les lois discrètes.

3 Définition de la loi d'un objet aléatoire

3.1 Lien entre "une loi" et "la loi d'un objet aléatoire"

Reprenons une phrase de l'introduction.

«la loi de X c'est la donnée des $\mathbf{P}[X \in A]$ pour tous les A possibles».

Tous les A possible ? Cela signifie tous les $A \subset E$ où E est l'ensemble dans lequel X prend ses valeurs. Il faut toujours bien préciser cet ensemble. Par exemple

- Le lancé d'un dé. $E = \{1, 2, 3, 4, 5, 6\}$
- L'appel de la fonction rand(). $E = [0, 1]$.
- Le nombre de sardines et la température de l'eau pour l'océan atlantique. $E = \mathbb{N} \times \mathbb{R}_+$.
- Le nuage que l'on observera demain dans un coin de ciel. E est l'ensemble des sous-ensembles de $\mathbb{R}^3$.
- La photo d'un carré de pelouse de l'orangerie de 1×1 mètre. $E = [0, 1]^2$.

Exo: Cherchez l'erreur ci-dessus.

Remarque : L'ensemble E dans lequel X prend ses valeurs est aussi appelé espace d'arrivée de X. On n'est pas toujours obligé de définir l'espace d'arrivée le plus petit possible. Par exemple, pour un couple de dés, on pourrait choisir pour E les ensembles $\mathbb{N}^2$ ou $\mathbb{Z}^2$ ou $\mathbb{R}_+^2$ ou $\mathbb{R}^2$. Le plus petit ensemble fermé F telle $\mathbf{P}[X \in F] = 1$ s'appelle le support de la loi de X. Pour un couple de dés, le support est $\{1, 2, 3, 4, 5, 6\}^2$.

Ici nous parlons de "loi d'objet aléatoire". Dans les préliminaires, nous avons parlé de "loi" tout court. Le lien vient maintenant:

Définition 3.1 $X \in E$ a la loi (ou suit la loi) μ quand :

$$\forall A \subset E \qquad \mathbf{P}[X \in A] = \mu(A)$$

Remarquons que l'axiome (Additivité) :

$$A \cap B = \emptyset \ \Rightarrow \ \mu(A \cup B) = \mu(A) + \mu(B) \tag{Additivité}$$

s'interprète alors naturellement :

«Quand A et B sont disjoints, la probabilité que X tombe dans A ou dans B c'est la somme des probabilités. »

ATTENTION : De nombreux étudiants ont tendance à identifier (ou mélanger) les objets aléatoires et leurs lois. Cette erreur est provoquée par un vocabulaire trompeur. Par exemple :

- «Une v.a. gaussienne», est un raccourci de, «une v.a. suivant la loi gaussienne"».
- «Une v.a. exponentielle», est un raccourci de, «une v.a. suivant la loi exponentielle"».
- «Les v.a. X_n convergent en loi vers une gaussienne», est un raccourci de «les lois de X_n convergent vers la loi gaussienne».
- etc.

3.2 Grand Oméga

Dans un cours de probabilité classique on commence par introduire $(\Omega, \mathbf{P})$ où

- Ω est un ensemble qui représente tous les hasards possibles
- $\mathbf{P}$ est une loi (ou mesure de probabilité) sur Ω qui pondère les hasards.

Dès lors, un objet aléatoire est une application $X : \Omega \to E$. Elle fait correspondre à chaque hasard ω une valeur $X(\omega)$.

Dans un cours de modélisation, nous ne définissons précisément ni Ω ni $\mathbf{P}$. Nous ne donnons que les objets aléatoire qui nous intéressent et donnons leurs lois. Un énoncé typique est : « Considérons N étoiles réparties de manière gaussienne dans l'espace tri-dimensionnel». Qu'il faut traduire par «Considérons $X_1, ..., X_N$ des v.a. indépendantes de loi gaussienne sur $\mathbb{R}^3$».

3.3 Formule Générale de l'Espérance (FGE)

Considérons X un réel aléatoire, donc formellement $X : \Omega \to \mathbb{R}$. Son espérance, c'est la moyenne de X sur tous les hasards possibles, pondérée par $\mathbf{P}$:

$$\mathbf{E}[X] = \int_\Omega X(\omega)\mathbf{P}[d\omega] \tag{A}$$

Mais la formule utile est :

$$\mathbf{E}[X] = \int_\mathbb{R} x\,\mathbf{P}[X \in dx] \tag{B}$$

c.à.d. la moyenne des valeurs de X, pondérée par la loi de x.

ATTENTION : l'espérance n'est définie que pour des objets aléatoires à valeurs dans $\mathbb{R}$. Par exemple, cela n'aurait aucun sens de moyenner des objets aléatoires modélisant des nuages. Par contre on pourrait moyenner des volumes de nuage ! Ainsi, si X est un objet aléatoire à valeurs dans un ensemble E, on peut considérer une fonction $\varphi : E \to \mathbb{R}$. Alors $\varphi(X)$ est un réel aléatoire, et son espérance se calcule avec la formule

$$\boxed{\mathbf{E}[\varphi(X)] = \int_E \varphi(x)\mathbf{P}[X \in dx] \qquad (FGE)}$$

Retenez la formule ci-dessus : la Formule Générale de l'Espérance (FGE). Nous allons la décliner dans plusieurs cas particulier (loi discrète, loi à densité).

Si vous voulez comprendre comment les liens entre (A), (B) et (FGE), lisez l'annexe sur le théorème de transfert.

3.4 La variance et l'écart type

Considérons X un réel aléatoire d'espérance m. La variance de X c'est :

$$\mathbf{V}[X] = \mathbf{E}[(X-m)^2] = \int_{\mathbb{R}} (x-m)^2 \; \mathbf{P}[X \in dx]$$

L'écart type de X est la racine carrée de la variance. L'écart type représente l'écart moyen de X autour de son espérance. Ainsi lorsque vous simulez des v.a. d'espérance m et d'écart type σ, vous obtennez des nombres réels x_i qui se répartissent autours de m et tels que les $|x_i - m|$ ont pour ordre de grandeur σ.

Exo. Vous simulez des v.a. exponentielles de paramètre 0.1. Les simulations successives vous donnent

10.13543, 10.26257, 9.9043675, 10.100132, 9.89632...

Votre programme est-il buggé ?

Exo. Montrez que l'on a aussi $\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$.

4 Les lois discrètes

4.1 Définition d'une loi discrète

Une loi μ sur E est dite discrète lorsqu'elle ne charge qu'un nombre fini ou dénombrable de points. Si nous notons $k_i \in E$ ces points, et w_i les poids correspondants, la loi μ peut-être notée

$$\mu(dx) = \sum_i w_i \, \delta_{k_i}(dx)$$

Remarque puisque μ est une loi, les poids w_i doivent vérifier $\sum_i w_i = 1$.

Un objet aléatoire est dit discret lorsqu'il ne peut prendre d'un nombre fini ou dénombrable de valeurs. Bien entendu, la loi d'un objet aléatoire discret est une loi discrète : en notant (k_i) les valeurs que peut prendre X:

$$\mathbf{P}[X \in dx] = \sum_i \mathbf{P}[X = k_i] \, \delta_{k_i}(dx)$$

Exemple : X n'est pas aléatoire : elle prend toujours la valeur k. Sa loi est δ_k.

Exemple : X est le résultat d'un pile ou face. Sa loi c'est $\frac{1}{2}\delta_P + \frac{1}{2}\delta_F$.

Exemple : X prend les valeurs 1 avec probabilité p et 0 avec probabilités $1-p$. La loi de X est appelée loi de Bernouilli de paramètre p.

Exo : Revoyez les lois Binomiales, Géométrique et de Poisson. Il faut les connaitre par coeur. Ecrivez-les avec des Dirac. Par exemple

$$Poisson(\lambda) = e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} \delta_n$$

ou bien sans Dirac : Si X suit une loi de Poisson(λ) alors $\mathbf{P}[X = n] = \frac{\lambda^n}{n!}$.

4.2 La formule de l'espérance et de la variance

En employant la formule générale de l'espérance (FGE) avec une v.a. X discrète de loi $\sum_i \mathbf{P}[X = k_i]\delta_{k_i}$ on trouve

$$\mathbf{E}[\varphi(X)] = \int \varphi(x)\ \mathbf{P}[X \in dx] = \sum_i \varphi(k_i)\ \mathbf{P}[X = k_i]$$

en particulier, si X est un réel aléatoire

$$\mathbf{E}[X] = \int x\ \mathbf{P}[X \in dx] = \sum_i k_i\ \mathbf{P}[X = k_i]$$
$$\mathbf{V}[X] = ...$$

Exo : Révisez comment calculer les espérances et des variances des lois classiques. Ex: quand X suit une loi de Poisson(λ):

$$\mathbf{E}[X] = e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} n = ... = \lambda$$
$$\mathbf{V}[X] = ... = \lambda$$

Question : Si vous simulez des v.a. de Poisson avec un grand paramètre λ et que vous trouvez:

10011, 10093, 9926, 10201, 9876, 10047

Votre programme est-il buggé ?

4.3 Loi uniforme

Soit E un ensemble fini. La loi uniforme sur E est la loi μ définie par :

$$\forall A \subset E \qquad \mu(A) = \frac{\text{nombre de points dans } A}{\text{nombre de points dans } E}$$

Notons n le nombre de points dans E et notons k_i les éléments de E. La loi uniforme sur E se note aussi :

$$\mu = \sum_{k_i \in E} \frac{1}{n}\, \delta_{k_i}$$

Exemple : Soit X le résultat d'un lancé de dé. Il a une loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$. Son espérance vaut ...

En langage courant, lorsque l'on dit «prenons un points X au hasard dans E» et que E est un ensemble fini, il faut comprendre «considérons un point aléatoire $X \in E$ de loi uniforme».

5 Les lois à densité

Attention: lorsque l'on ne précise pas, "loi à densité" signifie "loi à densité par rapport à la mesure de Lebesgue".

5.1 Loi à densité sur $\mathbb{R}$

Une loi μ sur $\mathbb{R}$ est à densité lorsqu'il existe une fonction f positive vérifiant:

$$\forall A \subset E \qquad \mu(A) = \int_{\mathbb{R}} 1_A(x) f(x)\, Leb_1(dx) = \int_{\mathbb{R}} 1_A(x) f(x)\, dx$$

On peut aussi écrire $\mu(dx) = f(x)\, dx$. Moralement μ donne le poids $f(x)\, dx$ à l'élément infinitésimal dx. Donc μ charge davantage les parties où f est grande.

5.2 La gaussienne

Exemple fondamental : la loi gaussienne est définie par :

$$\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

DESSIN

Elle charge beaucoup les parties proches de zéro, et très très peu les parties proches de $+\infty$ ou $-\infty$ (elle est à queue légère).

On dit qu'une v.a. X est gaussienne lorsque sa loi est gaussienne :

$$\mathbf{P}[X \in dx] = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

On a par exemple

$$\mathbf{P}[X \geq 0] = \mathbf{P}\Big[X \in [0, \infty[\Big] = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = \frac{1}{2}$$

5.3 Les loi à densité ne chargent pas les points

Remarque : quand on écrit $\mathbf{P}[X \in dx] = \mu(dx) = f(x)\, dx$, cela illustre bien le fait que : plus $f(x)$ est grand et plus X a de chance de tomber dans dx. Mais attention, cette phrase est une image "à la physicienne" car dx est infinitésimal. Il ne faut pas en déduire que $\mathbf{P}[X = x] = f(x)$. En vérité on a :

$$\mathbf{P}[X = x] = \mathbf{P}[X \in \{x\}] = \int_{\mathbb{R}} 1_{\{x\}}(y) f(y) dy = 0$$

Par ailleurs, il ne faut pas confondre $\int_{\mathbb{R}} f(y)\ 1_{\{x\}}(y)dy$ qui vaut zéro avec $\int_{\mathbb{R}} f(y)\ \delta_x(dy)$ qui vaut $f(x)$. La Dirac a un poids de 1, alors que la mesure $1_{\{x\}}(y)dy$ a un poids de zéro.

Ceci en tête, vous comprendrez pourquoi on ne peut pas définir la loi d'un objet aléatoire par la connaissance des $\mathbf{P}[X = x]$ pour tout x.

5.4 loi à densité sur $\mathbb{R}^n$

Notations : prenons $x = (x_1, x_2) \in \mathbb{R}^2$. L'élément infinitésimal dx correspond au petit carré $dx_1 \times dx_2$. On peut par exemple écrire

$$\mathbf{P}[(X_1, X_2) \in dx] = \mathbf{P}[(X_1, X_2) \in dx_1 \times dx_2] = \mathbf{P}[X_1 \in dx_1 \text{ et } X_2 \in dx_2]$$

Le "et" dans les probabilités sera souvent remplacé par une virgule: $\mathbf{P}[X_1 \in dx_1, X_2 \in dx_2]$

Une loi μ sur $\mathbb{R}^n$ est à densité lorsque qu'il existe une fonction f vérifiant:

$$\forall A \subset E \qquad \mu(A) = \int_{\mathbb{R}^n} 1_A(x) f(x)\, dx = \int_{\mathbb{R}^n} 1_A(x_1, ..., x_n) f(x_1, ..., x_n)\, dx_1...dx_n$$

Exemple de calcul utilisant les densités : Supposons que la loi de $X = (X_1, X_2) \in \mathbb{R}^2$ est donnée par la densité $f(x_1, x_2)$ c.à.d:

$$\mathbf{P}[X_1 \in dx_1, X_2 \in dx_2] = f(x_1, x_2) dx_1 dx_2$$

Calculons la loi de X_1 "tout seul":

$$\begin{aligned}\mathbf{P}[X_1 \in A] &= \mathbf{P}[X_1 \in A, X_2 \in \mathbb{R}] \\ &= \int_{\mathbb{R}^2} 1_A(x_1) 1_{\mathbb{R}}(x_2) f(x_1, x_2) dx_1 dx_2 \\ &= \int 1_A(x_1) \Big(\int_{\mathbb{R}} f(x_1, x_2) dx_2 \Big) dx_1\end{aligned}$$

On en déduit que la loi de X_1 admet la densité suivante

$$x_1 \to \Big(\int_{\mathbb{R}} f(x_1, x_2) dx_2 \Big)$$

5.5 Formule pour l'espérance

Soit $X \in \mathbb{R}^n$ un vecteur aléatoire de densité f. Soit φ est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$. D'après la formule générale de l'espérance (FGE) :

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}^n} \varphi(x) f(x) dx = \int_{\mathbb{R}^n} \varphi(x) \mathbf{P}[X \in dx]$$

En particulier si $n = 1$, on peut calculer l'espérance de X:

$$\begin{aligned}\mathbf{E}[X] &= \int_{\mathbb{R}} x f(x) dx \\ \mathbf{V}[X] &= ...\end{aligned}$$

5.6 Loi uniforme sur une partie de $\mathbb{R}^n$

Soit $E \subset \mathbb{R}^n$. La loi uniforme sur E est la loi

$$\mu(dx) = c_{st} 1_E(x) dx$$

où c_{st} est une constante de normalisation qui vaut forcément $1/Leb_n(E)$. Pour que cela ait du sens il faut que $Leb_n(E)$ soit finie. Par exemple, on ne peut pas définir de loi uniforme sur $\mathbb{R}_+$.

Exemple : Soit X le résultat d'un appel de la fonction rand(). Cette v.a. a une loi uniforme sur $[0, 1]$.

$$\forall A \subset [0,1] \qquad \mathbf{P}[X \in A] = \int_{[0,1]} 1_A dx$$

En particulier pour $[a, b] \subset [0, 1]$ on a $\mathbf{P}\Big[X \in [a, b]\Big] = b - a$.

Exo : Pour $[a, b] \subset \mathbb{R}$, il faut écrire précisément

$$\mathbf{P}[X \in [a, b]] = ...$$

Complétez en utilisant les notations pratiques $c \wedge d = \min(c, d)$ et $c \vee d = \max(c, d)$.

Dans le langage courant, quand on dit : «prenons un point au hasard», cela sous-entend : selon une loi uniforme.

6 Exo

Soit X une v.a. de loi uniforme sur $[0, 1]$. Considérons $Y = 2X - 1$. Elle suit une loi uniforme sur ... et vérifie donc

$$\forall [a, b] \subset [-1, 1] \qquad \mathbf{P}[Y \in [a, b]] = ...$$

Quant à $Z = -Y$, c'est une autre v.a. qui suit la loi ...

Soit X une v.a. de loi uniforme sur $[-2, 2]$. Considérons $Y = X^2$. Quelles sont les valeurs possibles pour Y ? Intuitivement : quelles sont les parties de $\mathbb{R}$ que la loi de Y charge le plus. Nous verrons plus loin comment vérifier notre intuition.

7 Autres type de loi

En apprenant les probabilités en licence, on peut avoir l'impressions que toutes les v.a. ont des lois à densité ou des lois discrète. C'est loin d'être le cas, même si l'on se restreint aux exemples concrets.

7.1 Loi portée par un sous-ensemble fin

Considérons le carré $E = [0,1]^2$. Prenons X un point pris au hasard sur périmètre P de E. La loi de X n'est pas discrète (car le périmètre P a un nombre infini de point). Mais la loi de X n'est pas non plus à densité. Raisonnons par l'absurde. Si $\mathbf{P}[X \in dx] = f(x)dx$ alors

$$1 = \mathbf{P}[X \in P] = \int_{[0,1]^2} 1_P(x) f(x) dx$$

Or, P ayant une aire nulle, l'intégrale ci-dessus vaut zéro.

De manière plus générale, si X est à valeurs dans $\mathbb{R}^n$, mais prend ses valeurs que des ensembles de dimension strictement inférieur à n, alors X ne suit pas une loi à densité.

Exo : Dessinez les supports des lois de couple $(X, Y) \in \mathbb{R}^2$ suivants. Indiquez s'ils ont une loi à une densité, une loi discrète, ou ni l'un ni l'autre.

- X est une loi uniforme sur $[-2, 2]$ et $Y = X^2$.
- X est le nombre de sardines et Y la température moyenne de l'océan atlantique à une date donnée.
- X est le premier résultat de la fonction rand(). Y est le second résultat de la fonction rand().
- X et Y sont deux lancés de dé.

7.2 Une loi continue sur une surface

Le tirage d'un point "au hasard" sur la terre, peut-être modélisé par une v.a. X de loi uniforme sur la sphère $\mathbb{S}$ de circonférence 40 075 km. Ce qui signifie que pour tout $A \subset \mathbb{S}$, $\mathbf{P}[X \in A]$ vaut la surface de A divisé par surface totale de $\mathbb{S}$.

Question : Quelle est la probabilité que X tombe exactement sur l'équateur ?

DESSIN

Notons $\mu(dx)$ la loi uniforme sur le globe, décrite précédemment. La position d'un homme tiré au hasard sur terre aura comme loi $\nu(dx) = f(x)\mu(dx)$ où f désigne la densité de population sur terre. Par exemple f est nulle en mer, et admet des pics en chaque ville.

Vocabulaire. Comme précédemment, considérons une mesure ν qui s'écrive $\nu(dx) = f(x)\mu(dx)$. On dit que f est la densité de ν relativement à μ. On décrit même $f(x) = \frac{\nu(dx)}{\mu(dx)}$. Mais rappelons, que, si l'on ne précise pas, les densités sont toujours relatives à la mesure de Lebesgue.

7.3 Loi mélangeant discret et densité

Imaginez le jeu suivant : vous tirez un pile ou face.

- Si pile tombe, alors vous lancez un dé et vous gagnez le chiffre du dé en euros.
- Si face tombe, vous lancez la fonction rand(), et vous gagnez 10 fois le résultat en euros.

DESSIN

La loi du gain X est alors

$$\mathbf{P}[X \in dx] = \frac{1}{2}\sum_{i=1}^{6}\frac{1}{6}\delta_i(dx) \ + \ \frac{1}{2} * 10 * 1_{[0,1]}dx$$

L'espérance d'un fonction φ du gain est donc :

$$\mathbf{E}[\varphi(X)] = \frac{1}{2}\sum_{i=1}^{6}\frac{1}{6}\varphi(i) \ + \ \frac{1}{2} * 10 * \int_{[0,1]}\varphi(x)dx$$

Exo : corrigez l'erreur bête dans les 2 équations ci-dessus.

7.4 Lois étranges

Ce dernier exemple n'est pas du tout concret.

Il existe des lois étranges, même sur $\mathbb{R}$. Connaissez-vous l'escalier de Cantor ? C'est une fonction croissante F qui est continue mais très irrégulière.

DESSIN

On peut définir une mesure sur $\mathbb{R}$ par $\forall]a,b] \subset \mathbb{R}$ $\mu([a,b]) = F(b) - F(a)$. Cette loi ne peux s'écrire ni comme une somme de Dirac, ni comme une loi à densité, ni même comme un mélange des deux.

7.5 Lois sur des gros espaces.

Considérons E l'ensemble des fonctions de $\mathbb{R}^+$ dans $\mathbb{R}^n$. Un élément $x \in E$ représente souvent la trajectoire d'une quantité dans le temps. Donc si $X = (X_t)_{t\in\mathbb{R}_+}$ est un objet aléatoire à valeurs dans E, X représente une trajectoire aléatoire. Pour décrire la loi d'un objet si compliqué, nous nous contenterons de décrire la loi des vecteurs $(X_{t_1}, ..., X_{t_n})$ pour toutes les familles finies d'indice $(t_1, ..., t_n)$. Nous en reparlerons.

Remarque : L'ensemble E ci-dessus est de dimension infinie. Dans ce cas, on ne préfère pas définir de mesure de Lebesgue. Donc on ne dira pas que les lois des trajectoires aléatoires sont a densité (ou alors ce sera une densité par rapport à une mesure qu'il faut préciser).

8 Caractérisation des lois

8.1 Caractérisation avec des fonctions tests

Une fonction test φ sur E est une fonction telle que $\int_E \varphi\mu$ est toujours bien définie pour toute loi μ. Par exemple, sur $\mathbb{R}$, on ne choisira pas $x \to \frac{1}{x}$ comme fonction test. Comme ensemble Φ de fonction test on peut prendre

- l'ensemble des indicatrices.
- l'ensemble de toutes les fonctions bornées.
- l'ensemble des fonctions positives.
- l'ensemble des fonctions continues bornées (cette dernière famille nous sera utile pour la convergence de loi).

Je conseille de choisir par défaut les fonctions positives. Puisque Φ est suffisamment "riche", la loi d'un objet aléatoire $X \in E$ est "caractérisée" par les quantités:

$$\Big(\mathbf{E}[\varphi(X)] : \varphi \in \Phi\Big)$$

Le mot "caractérisé" signifie que :

$$\forall \varphi \in \Phi : \mathbf{E}[\varphi(X)] = \int \varphi(x)\mu(dx) \qquad \Leftrightarrow \qquad \mathbf{P}[X \in dx] = \mu(dx)$$

Variante pour 2 objets aléatoires :

$$\forall \varphi \in \Phi : \mathbf{E}[\varphi(X_1)] = \mathbf{E}[\varphi(X_2)] \qquad \Leftrightarrow \qquad \mathbf{P}[X_1 \in dx] = \mathbf{P}[X_2 \in dx]$$

8.2 Exo : Le carré d'une v.a. uniforme, suite

Considérons une v.a. X de loi uniforme sur $[-2, 2]$. Considérons $Y = X^2$. Intuitivement, nous avons vu que Y chargeait beaucoup les voisinages de zéro. Calculons explicitement la loi de Y. Prenons une fonction test positive φ.

$$\begin{aligned}\mathbf{E}[\varphi(Y)] = \mathbf{E}[\varphi(X^2)] &= \int_{[-2,2]} \varphi(x^2)\mathbf{P}[X \in dx] \\ &= \int_{[-2,2]} \varphi(x^2)\frac{1}{4}dx \\ &= \frac{1}{2}\int_{[0,2]} \varphi(x^2)dx\end{aligned}$$

Changement de variable : $x^2 \to y, x \to \sqrt{y}, dx \to \frac{dy}{2\sqrt{y}}, x \in [0, 2] \;\Leftrightarrow\; y \in [0, 4]$

$$= \frac{1}{2}\int_{[0,4]} \varphi(y)\,\frac{1}{2\sqrt{y}}\,dy$$

Nous venons d'établir que

$$\mathbf{E}[\varphi(Y)] = \int_{\mathbb{R}} \varphi(y)1_{[0,4]}\,\frac{1}{4\sqrt{y}}\,dy$$

Puisque l'égalité ci-dessus est vraie pour n'importe quelle fonction test φ, nous avons :

$$\mathbf{P}[Y \in dy] = 1_{[0,4]}\,\frac{1}{4\sqrt{y}}\,dy$$

DESSIN

8.3 Caractérisation sur $\mathbb{R}$.

Soit X un réel aléatoire (une v.a.). La loi de X est caractérisée par l'un ou l'autre des objets ci-dessous:

- Sa fonction de répartition, c.à.d. la fonction $x \mapsto \mathbf{P}[X \leq x]$.
- Ses probabilités d'appartenir aux intervalles, c.à.d. par la famille $(\mathbf{P}[X \in I] : I$ intervalle de $\mathbb{R})$.
- Sa fonction caractéristique, c.à.d. la fonction $u \mapsto \mathbf{E}[e^{iuX}]$.

Remarque : Ces caractérisations, sont aussi des caractérisations à l'aide de fonctions tests. Par exemple, en utilisant

$$\mathbf{P}[X \leq x] = \mathbf{E}[1_{\{X \leq x\}}] = \mathbf{E}\Big[1_{]-\infty,x]}(X)\Big]$$

Nous voyons que, connaitre la fonction de répartition, revient à connaitre la famille : $\big(\mathbf{E}[\varphi(X)] : \varphi \in \Phi\big)$ avec comme ensemble de fonctions tests

$$\Phi = \{1_{]-\infty,x]} : x \in \mathbb{R}\}$$

8.4 Les deux cas particuliers

Enfonçons des portes ouvertes:

Si X est une v.a. ayant une loi à densité, alors sa loi est caractérisée par cette densité.

Si X est une v.a. ayant une loi discrète, alors sa loi est caractérisée par la liste des points (k_i) et des poids (w_i).

8.5 Exo : D'un caractérisation à l'autre

Soit X une v.a.

(a). Calculez les $\mathbf{P}\Big[X \in]a,b]\Big]$ à partir des $\mathbf{P}[X \leq x]$.

(b). Calculez les $\mathbf{P}[X \in I]$, pour I intervalle ouvert, fermé, semi-ouvert, à partir des $\mathbf{P}\Big[X \in]a,b]\Big]$.

(c). Calculez les $\mathbf{P}\Big[X \in [a,b]\Big]$ à partir des $(\mathbf{E}[\varphi(X)] : \varphi :$ fonction continue bornée).

Indices :

(a). Faites une soustraction ensembliste.

(b). Faites des passages à la limite en utilisant des égalités comme $[0,1] = \cap_n [0, 1 + \frac{1}{n}[$

(c). N'essayez pas de formaliser, un bon dessin suffira.

8.6 Vecteur aléatoire

Soit $X = (X_1, X_2)$ un vecteur aléatoire de $\mathbb{R}^2$. Les lois de X_1 et de X_2 sont appelées : les lois marginales de X ; ce sont des lois sur $\mathbb{R}$ alors que la loi de X est une loi sur $\mathbb{R}^2$. Ces deux lois marginales ne caractérisent pas la loi de X. Autrement dit, la donnée des $\mathbf{P}[X_1 \in A]$ et $\mathbf{P}[X_2 \in B]$ pour tout A, B parties de $\mathbb{R}$ ne suffit pas pour calculer les $\mathbf{P}[X \in C]$ pour C partie de $\mathbb{R}^2$.

Exemple: Prenons X_1, X_2 deux tirages successifs de la fonction rand(). Considérons $X = (X_1, X_2)$ et $Y = (X_1, X_1)$. Ces deux vecteurs aléatoires ont les mêmes lois marginales, mais n'ont pas du tout la même loi. Par exemple, si Δ désigne la diagonale du carré $[0, 1] \times [0, 1]$ on a

$$\mathbf{P}[X \in \Delta] = 0$$
$$\mathbf{P}[Y \in \Delta] = 1$$

DESSIN

Par contre la loi de $X \in \mathbb{R}^2$ est caractérisée par la donnée de tous les $\mathbf{P}[X \in A \times B]$ pour tout A, B parties de $\mathbb{R}$ (ou peut même se contenter des intervalles de $\mathbb{R}$). C'est assez logique, toute partie C de $\mathbb{R}^2$ peut-être approchée aussi finement que l'on veut par des unions de petits rectangles $A_i \times B_i$ disjoints. Et donc

$$C \simeq \bigcup_i A_i \times B_i \;\Rightarrow\; \mathbf{P}[X \in C] \simeq \sum_i \mathbf{P}[X \in A_i \times B_i]$$

DESSIN

9 Indépendances

9.1 Indépendance de deux objets aléatoires

Deux objets aléatoires X_1 et X_2 à valeurs dans E_1 et E_2 sont indépendants lorsque

$$\forall A \subset E_1 \; \forall B \subset E_2 \qquad \mathbf{P}[X_1 \in A, X_2 \in B] = \mathbf{P}[X_1 \in A]\, \mathbf{P}[X_2 \in B]$$

Remarque: on en déduit que quand X_1 et X_2 sont indépendants, la proba $\mathbf{P}[(X_1, X_2) \in A \times B]$ est déterminée par $\mathbf{P}[X_1 \in A]$ et $\mathbf{P}[X_2 \in B]$. Donc les lois marginales caractérisent la loi du couple.

Quand X_1 et X_2 sont discrets. L'indépendance est équivalente à

$$\forall a \in E_1, \forall b \in E_2 \qquad \mathbf{P}[X_1 = a, X_2 = b] = \mathbf{P}[X_1 = a]\, \mathbf{P}[X_2 = b]$$

9.2 Exemple discret

Exo : Soient $X_1 \in \{a, b\}$ et $X_2 \in \{c, d\}$ des objets aléatoires. Supposons que leur loi jointe (=la loi du couple) est décrite par le tableau suivant:

	a	b
c	$\frac{1}{6}$	$\frac{2}{6}$
d	$\frac{1}{6}$	$\frac{2}{6}$

Par exemple $\mathbf{P}[X_1 = a, X_2 = c] = \frac{1}{6}$. Montrez que X_1 et X_2 sont indépendants. Indice : commencez par calculer les lois marginales.

9.3 Caractérisation avec des fonctions testes (splitting)

X_1 et X_2 sont indépendants si et seulement si pour toutes fonctions testes $\varphi : E_1 \to \mathbb{R}$ et $\psi : E_2 \to \mathbb{R}$ on a

$$\mathbf{E}[\varphi(X_1)\psi(X_2)] = \mathbf{E}[\varphi(X_1)]\mathbf{E}[\psi(X_2)]$$

(on peut "splitter" l'espérance)

Exo : En supposant le "splitting" ci-dessous, montrez l'indépendance.

Exo : En supposant l'indépendance, montrer que le splitting est vrai, dans le cas particulier des fonctions testes φ et ψ étagées (de la forme $\sum_i c_i 1_{A_i}$). Le cas général s'en déduit par un passage à la limite.

Attention: Dans le cas de v.a., l'égalité $\mathbf{E}[X_1 X_2] = \mathbf{E}[X_1]\mathbf{E}[X_2]$ n'implique pas l'indépendance.

9.4 Indépendance d'une famille finie

Les éléments d'un vecteur aléatoire $(X_1, ..., X_n)$ sont indépendants lorsque ... (imaginez la définition. Imaginez la caractérisation avec le splitting)

En général, les indépendances des couples (X, Y), (Y, Z) et (X, Z) , n'impliquent pas l'indépendance du triplet (X, Y, Z).

Exemple : considérons X et Y deux v.a. indépendantes de loi uniforme sur $\{0, 1, ..., 9\}$. Notons $\dot{+}$ la sommation modulo 10, par exemple $2\dot{+}9 = 1$. Notons $Z = X\dot{+}Y$. On a $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Z$, $Y \perp\!\!\!\perp Z$ mais (X, Y, Z) n'est pas indépendant, car Z est une fonction de (X, Y).

Démo : Remarquons que pour $a \in \{0, 1, ..., 9\}$ on a $\sum_{y=0}^{9} \psi(y\dot{+}a) = \sum_{y=0}^{9} \psi(y)$ (grâce au modulo). Ainsi

$$\begin{aligned}
\mathbf{E}[\varphi(X)\psi(Z)] &= \mathbf{E}[\varphi(X)\psi(X\dot{+}Y)] \\
&= \frac{1}{10}\frac{1}{10}\sum_{x=0}^{9}\sum_{y=0}^{9}\varphi(x)\psi(x\dot{+}y) \\
&= \frac{1}{10}\frac{1}{10}\sum_{x=0}^{9}\varphi(x)\sum_{y=0}^{9}\psi(x\dot{+}y) \\
&= \frac{1}{10}\frac{1}{10}\sum_{x=0}^{9}\varphi(x)\sum_{y=0}^{9}\psi(y)
\end{aligned}$$

En particulier en remplaçant φ par la fonction constante 1, on voit que Z suit aussi une loi uniforme sur $\{0, 1, ..., 9\}$. Et donc l'égalité ci-dessus peut s'écrire $\mathbf{E}[\varphi(X)\psi(Z)] = \mathbf{E}[\varphi(X)]\,\mathbf{E}[\psi(Z)]$. Donc $X \perp\!\!\!\perp Z$. Par symétrie $Y \perp\!\!\!\perp Z$. Quant à l'indépendance entre X et Y, elle est donnée par l'énoncé. □

Par contre. L'indépendance de (X, Y) avec Z et l'indépendance de X avec Y implique l'indépendance du triplet (X, Y, Z). Démo:

$$\begin{aligned}
\mathbf{P}[X \in dx, Y \in dy, Z \in dz] &= \mathbf{P}[(X, Y) \in dx \times dy, Z \in dz] \\
&= \mathbf{P}[(X, Y) \in dx \times dy]\,\mathbf{P}[Z \in dz] \\
&= \mathbf{P}[X \in dx]\mathbf{P}[Y \in dy]\,\mathbf{P}[Z \in dz]
\end{aligned}$$

Exo : Refaites cette preuve en utilisant le splitting.

Exo : Vérifiez que la réciproque est vraie : si le triplet est indépendant alors ...

9.5 Exemples à densité

Exo : considérons (X, Y, Z) un vecteur aléatoire de $\mathbb{R}^3$ dont la loi est donnée par

$$\mathbf{P}[(X, Y, Z) \in C] = \int_{\mathbb{R}^n} 1_C(x, y, z) f(x) g(y) h(z) dx dy dz$$

Vérifiez leur indépendance.

Remarque : Il existe une réciproque à cela : si (X, Y, Z) admet une densité $F(x, y, z)$ alors (X, Y, Z) sont indépendants si et seulement si F se factorise en 3 fonctions d'arguments respectifs x, y, z.

Exemple: Considérons un point $(X, Y) \in \mathbb{R}^2$ qui suit une loi uniforme sur le disque unité. Est-ce que X et Y sont indépendants ?

DESSIN

Notons (R, Θ) les coordonnées polaires du point (X, Y). Calculons leurs lois :

$$\begin{aligned} \mathbf{E}[\varphi(R, \Theta)] &= \frac{1}{\pi} \int_{\mathbb{R}^2} \varphi(\sqrt{x^2 + y^2}, \operatorname{arctan}_2(x, y)) 1_{\{x^2+y^2<1\}} dx dy \\ &= \frac{1}{\pi} \int_{\mathbb{R}_+} \int_{[0,2\pi]} \varphi(r, \theta)) 1_{\{r<1\}} r \, dr d\theta \end{aligned}$$

Alors ? ...

9.6 Exemple mixte

Dans $\mathbb{R}^2$. Considérons le segment S_1 reliant $(0, -1)$ à $(1, 0)$. Let segment S_2 reliant $(-1, 0)$ à $(0, 1)$. Considérons P un point uniforme sur $S_1 \cup S_2$.

Considérons (K, L) les coordonnées de P dans la base (v_1, v_2) avec $v_1 = (1, 1)$ et $v_2 = (-1, 1)$.

DESSIN

Quelle est la loi de (K, L) ? Pas besoin de calcul, votre intuition suffira.

K et L sont indépendants. K suit une loi uniforme sur $[-\frac{1}{2}, \frac{1}{2}]$ et L suit $\frac{1}{2}\delta_{\frac{1}{2}} + \frac{1}{2}\delta_{-\frac{1}{2}}$.

9.7 Indépendance d'une famille infinie

Une famille infinie $(X_t)_{t\in\mathbb{T}}$ est indépendante indépendante lorsque pour n et tout $(t_1,...,t_n) \in \mathbb{T}^n$, le vecteur $(X_{t_1},...,X_{t_n})$ est indépendant. Cette définition est compatible avec la caractérisation des lois des familles finie (cf. plus haut).

On dit qu'une suite de $X_1,...,X_n,...$ est i.i.d (indépendant et identiquement distribuée), lorsqu'elle est indépendante et que tous les X_i ont même loi. Typiquement, lorsqu'on observe un phénomène naturel plusieurs fois, dans des conditions semblables, mais à des instants espacés, les observations sont i.i.d. Si ces observations sont des réels, on peut les moyenner, et l'on obtient alors ...

Théoreme 9.1 (Loi Forte des Grands Nombres) . Soient $X_1,X_2,...$ une suite de v.a. i.i.d. Alors

$$\frac{1}{n}(X_1+...+X_n)\to \mathbf{E}[X_1] \qquad (LFGN)$$

Les observations ne sont pas toujours réelles ; ex: les (X_i) sont des nuages. Dans ce cas on peut moyenner le volume $\varphi(X_i)$ de ces nuages :

Soient $X_1,X_2,...$ une suite i.i.d. d'objets aléatoires à valeurs dans E. Soit $\varphi : E\to R$ alors

$$\frac{1}{n}\Big(\varphi(X_1)+...+\varphi(X_n)\Big)\to \mathbf{E}[\varphi(X_1)]$$

Remarque : Comme d'habitude, nous mettons quelques difficultés sous le tapis : Certaines lois μ ont des queues si lourdes, qu'on ne peut pas définir correctement $\mathbf{E}[X_1]=\int x\mu(dx)$. Ex : la loi de Cauchy $\mu(dx)=\frac{1}{\pi}\frac{1}{1+x^2}\,dx$. Dans ce cas là, il ne faut pas espérer avoir (LFGN).

9.8 Modélisation

En général, les indépendances sont données dans les énoncées, soit de manière explicite : «considérons X et Y deux v.a. indépendantes ..», soit de manière implicite : « considérons X le nombre de poules dans le poulailler et Y la note de Simon à son devoir de math». Si dans l'énoncé il n'est pas précisé que Simon joue avec les poules au lieu de réviser ses maths, alors on peut raisonnablement supposer que X et Y sont indépendants.

Exo : Y-a-t-il indépendance entre les v.a. décrites.

- L'indice de masse corporelle d'une maman babouin et le nombre de ses enfants.
- La température à New York le 7 novembre 2000 et la température à Strasbourg le 1 janvier 1900.
- Le nombre de sardines et le nombre de homards dans l'océan Atlantique.
- Les notes de Robert à ses examens et les notes de sa petite amie.

Remarque : il n'y a pas toujours de réponse claire à ce genre de question. Cela dépend beaucoup du degré de précision désiré. Les modélisations simples supposent de nombreuses indépendances, ce qui facilite les calculs de lois. Les modélisations compliquées supposent de nombreuses corrélations ce qui demande l'estimation de plus de paramètres.

9.9 Indépendance et dépendance fonctionnelle

Cette partie met le doigts sur une confusion très répandue. Beaucoup de gens pensent à tord que :

«X et Ysont indépendants»
est synonyme de :
«X et Y ne sont reliées par aucune fonction».

C'est vrai dans un sens : si $X = f(Y)$ ou bien $Y = g(X)$ alors X et Y ne sont pas indépendants (à moins que l'un des deux ne soit constant).

Par contre la réciproque est fausse. Construisons deux v.a. X et Y qui ne sont pas indépendantes, et pour lesquelles il n'existe aucune fonction f vérifiant $X = f(Y)$, ni de fonction g vérifiant $Y = g(X)$. Pour cela considérons X à valeurs dans $\{a, b\}$ et Y à valeurs dans $\{c, d\}$ dont les probabilités d'apparitions sont données par:

	a	b
c	$\frac{1}{8}$	$\frac{1}{8}$
d	$\frac{4}{8}$	$\frac{2}{8}$

Démontrez que X et Y satisfont les propriétés souhaitées. Aide :

- Vérifiez que X et Y ne sont pas indépendantes en calculant les lois marginales.
- si $X = f(Y)$ alors le tableau de la loi doit comporter au moins deux zéros ...
- si $Y = g(X)$ alors le tableau de la loi doit comporter au moins deux zéros ...

DESSIN

10 Loi conditionnelle

10.1 Probabilité "désintégrée"

Etant donnée une probabilité $\mathbf{P}[\cdot]$ et un objet aléatoire $X \in E$, il existe une (presque)-unique famille de probabilité $(\mathbf{P}[\cdot/X = x] : x \in E)$ qui vérifie les axiomes suivants:

$$\mathbf{P}[X = x/X = x] = 1 \qquad (Concentration)$$

$$\mathbf{P}[\cdot] = \int_E \mathbf{P}[\cdot/X = x]\mathbf{P}[X \in dx] \qquad (Recollement)$$

Les mesures de probabilités $\mathbf{P}[\cdot/X = x]$ s'appellent des mesures de probabilités conditionnelles. Mais il y a une autre appellation très parlante : La famille $(\mathbf{P}[\cdot/X = x] : x \in E)$ est une désintégration de $\mathbf{P}[\cdot]$ selon X. En d'autres termes : la probabilité conditionnelle est un découpage de $\mathbf{P}$ selon les lignes de niveau de X. La formule (Recollement) permet de recoller les morceaux.

Ensuite on note $\mathbf{E}[\cdot/X = x]$ l'espérance calculée avec $\mathbf{P}[\cdot/X = x]$. De l'axiome de recollement on déduit :

$$\mathbf{E}[\cdot] = \int_E \mathbf{E}[\cdot/X = x]\mathbf{P}[X \in dx]$$

Exemple : Considérons un objet aléatoire discret X. On a alors :

$$\mathbf{P}[\cdot/X = x] = \frac{\mathbf{P}[\cdot \cap \{X = x\}]}{\mathbf{P}[X = x]}$$

Exo : Vérifiez les axiomes (Concentration) et (Recollement)

10.2 Loi conditionnelle

Considérons une désintégration ($\mathbf{P}[\cdot/X = x] : x \in E$). Considérons $Y \in F$ un autre objet aléatoire. Considérons $\mathbf{P}[Y \in dy/X = x]$, la loi de Y sous $\mathbf{P}[\cdot/X = x]$.

propriété :

$$\mathbf{P}[X \in dx, Y \in dy] = \mathbf{P}[Y \in dy/X = x]\,\mathbf{P}[X \in dx] \qquad (\textit{Jointe} \leftrightarrow \textit{Conditionnelle})$$

Démo : prenons une fonction test φ:

$$\begin{aligned}
\mathbf{E}[\varphi(X,Y)] &= \int \mathbf{E}[\varphi(X,Y)/X = x]\,\mathbf{P}[X \in dx] && \text{(Recollement)}\\
&= \int \mathbf{E}[\varphi(x,Y)/X = x]\,\mathbf{P}[X \in dx] && \text{(Concentration)}\\
&= \int\int \varphi(x,y)\mathbf{P}[Y \in dy/X = x]\,\mathbf{P}[X \in dx] && \text{(FGE)}
\end{aligned}$$

Puisque c'est vrai sur toutes les fonctions tests, la propriété est vraie. □

Exemple : Considérons un couple aléatoire (X,Y) admettant une densité $f_{X,Y}(x,y)$. Notons f_X la densité de X seul. On peut alors définir

$$\mathbf{P}[Y \in dy/X = x] = \frac{f_{X,Y}(x,y)}{f_X(x)}dy$$

Vérifiez que c'est bien une mesure de probabilité en dy (c.à.d. qu'en intégrant sur les y on trouve 1). Vérifiez la propriété (Jointe $\leftrightarrow$Conditionnelle)

10.3 Exemple de modélisation à deux couches

Quand on modélise un phénomène aléatoire en plusieurs couches de hasard, on utilise des conditionnements :

Considérons N le nombre de sardines dans l'atlantique, qui suit une loi de Poisson de paramètre Λ, où Λ est la température de l'eau, qui suit une loi exponentielle de paramètre 1. Essayons de déterminer la loi du couple (N,Λ).

Exo : Il y a plus d'étapes que nécessaire dans le calcul ci-dessous. Mais c'est pour vous permettre d'écrire pour chaque égalité, si l'on utilise les formules (Recollement), (Concentration), (FGE), ou bien une donnée de l'énoncé.

$$\begin{aligned}
\mathbf{E}[\varphi(\Lambda,N)] &= \int \mathbf{E}[\varphi(\Lambda,N)/\Lambda = \lambda]\mathbf{P}[\Lambda \in d\lambda]\\
&= \int \mathbf{E}[\varphi(\lambda,N)/\Lambda = \lambda]\mathbf{P}[\Lambda \in d\lambda]\\
&= \int \mathbf{E}[\varphi(\lambda,N)/\Lambda = \lambda]e^{-\lambda}1_{\{\lambda>0\}}\,d\lambda\\
&= \int\int \varphi(\lambda,y)\mathbf{P}[N \in dy/\Lambda = \lambda]e^{-\lambda}1_{\{\lambda>0\}}\,d\lambda\\
&= \int\int \varphi(\lambda,y)\Big(\sum_n e^{-\lambda}\frac{\lambda^n}{n!}\delta_n(dy)\ e^{-\lambda}1_{\{\lambda>0\}}d\lambda\Big)
\end{aligned}$$

Donc la loi de (N,Λ) c'est tout ce qui apparait dans la grande parenthèse : c'est une loi mixte : un mélange de Dirac $\delta_n(dy)$ et de Lebesgue $d\lambda$. On peut supprimer une intégrale grâce aux Dirac :

$$\mathbf{E}[\varphi(\Lambda,N)] = \int \sum_n \varphi(\lambda,n)\frac{\lambda^n}{n!}e^{-2\lambda}1_{\{\lambda>0\}}\,d\lambda$$

Déterminons maintenant la loi de N seul:

$$\begin{aligned}
\mathbf{E}[\varphi(N)] &= \int \sum_n \varphi(n) \frac{\lambda^n}{n!} e^{-2\lambda} 1_{\{\lambda>0\}}\, d\lambda \\
&= \sum_n \varphi(n) \frac{1}{n!} \int \lambda^n e^{-2\lambda} 1_{\{\lambda>0\}}\, d\lambda \\
&= \sum_n \varphi(n) \frac{1}{n!} \int \frac{\lambda^n}{2^n} e^{-\lambda} 1_{\{\lambda>0\}}\, \frac{1}{2} d\lambda \\
&= \sum_n \varphi(n) \frac{1}{2} \frac{1}{2^n} = \int \varphi(x) \sum_n \frac{1}{2} \frac{1}{2^n} \delta_n(dx)
\end{aligned}$$

Dingue : nous constatons que N suit une loi géométrique de paramètre $\frac{1}{2}$.

10.4 Conditionnement en situation d'indépendance

L'indépendance entre X et Y signifie que la probabilité que Y fasse ceci ou cela n'est pas influencée par les valeurs que prend X (et on peut inverser X et Y dans cette phrase).

Mathématiquement :

$$X \perp\!\!\!\perp Y \qquad \Leftrightarrow \qquad \forall x\ \mathbf{P}[Y \in dy / X = x] = \mathbf{P}[Y \in dy]$$

Conséquence : Quand $X \in E$ et Y sont indépendants, on a :

$$\mathbf{E}[\varphi(X, Y)] = \int_E \mathbf{E}[\varphi(x, Y)] \mathbf{P}[X \in dx]$$

Mais sans l'indépendance il faut rajouter ...

10.5 Enchainement de conditionnement

Considérons un triplet aléatoire (X, Y, Z). On a :

$$\mathbf{P}[Z \in dz, Y \in dy, X \in dx] = \mathbf{P}\Big[Z \in dz \Big/ Y = y, X = x\Big]\ \mathbf{P}[Y \in dy \Big/ X = x]\ \mathbf{P}[X \in dx]$$

Maintenant vous êtes capables de calculer la loi du nombre de sardines dans le modèle suivant :

Z est le nombre de sardines dans l'atlantique, qui suit une loi de Poisson de paramètre Y, où Y est la température de l'eau, qui suit une loi exponentielle de paramètre X, qui suit une loi uniforme sur $[0, 1]$.

10.6 Indépendance conditionnelle

Un exemple suffit à comprendre cette notion :

«Considérons N le nombre de sardines, et M le nombre de homards, et Λ la température moyenne pour l'océan Atlantique. Supposons que N suive une loi de Poisson de paramètre Λ. Que M suive une loi de Poisson de paramètre Λ^2. Supposons que Λ suive une loi exponentielle de paramétre 1. »

Dans cette énoncé M et N ne sont pas indépendants puisqu'ils sont tous deux influencés par la température de l'eau (plus elle est chaude et plus ils y a de sardines et de homards). Par contre, à température fixée $\Lambda = \lambda$, les sardines et les homards ne se faisant pas concurrence, les v.a. M et N sont indépendance. Donc plus précisément :

$$\mathbf{P}[M = m, N = n / \Lambda = \lambda] = \mathbf{P}[M = m / \Lambda = \lambda]\, \mathbf{P}[N = n / \Lambda = \lambda]$$

Ainsi, l'hypothèse implicite de l'énoncé, c'est que M et N sont indépendants conditionnellement à Λ.

11 Enfonçons le clou sur la loi

En guise de conclusion, voici une série de faits, pour vous aider à ne jamais confondre la notion d'objet aléatoire et la notion de loi.

- Soient $X_1, X_2, X_3, ...$ les appels successifs de la fonction rand(). Toutes ces variables ont la même loi $1_{[0,1]}dx$, mais elles sont toutes <u>différentes</u> ; et même mieux, elles sont indépendante. D'ailleurs

$$\mathbf{P}[X_i = X_j] = \int_{[0,1]^2} 1_{x=y}\, dxdy = 0$$

  Les $1 - X_i$ sont d'autres v.a. de même loi. Si on considère les $-\log(X_i)$ ils ont une autre loi (nous verrons qu'elles suivent des lois exponentielles).

- Considérons un algorithme qui fait appel plusieurs fois à la fonction rand(). Par exemple :
  - Si rand() $<$ 1/2 alors return floor(rand()*6)+1
  - Sinon return 10*rand()

  (Chaque rand() ci-dessus correspond à un nouveau tirage). En lançant deux fois cet algo on obtient deux v.a. indépendantes, mais de même loi. Exo: écrivez cette loi. Mais oui, vous savez le faire.

- Une variation sur le l'item précédent : Soient deux vecteurs aléatoires (X_1, Y_2, Z_2) et (X_2, Y_2, Z_2) sont indépendants et de même loi. Alors $f(X_1, Y_1, Z_1)$ et $f(X_2, Y_2, Z_2)$ sont indépendants et de même loi.

- Nous avons déjà signalé que: lorsque l'on observe un phénomène naturel aléatoire, à des instants éloignés, mais dans des conditions identiques, les observations sont i.i.d. Supposons par exemple que notre observation X_i est la températures à Strasbourg le 14 juillet de l'année $1900 + i$. Quelle est la loi de X_1 ? a priori elle est donnée par notre modélisation. Mais comment construit-t-on le modèle ? En général, on a une petite idée de ce que devrait être la loi de X_1 : mais pour ne pas être trop arbitraire, on se donne une famille de modèle possible, une famille paramétrée. Par ex: X_1 suit une loi Gaussienne de moyenne m (c'est le paramètre que l'on ne fixe pas) :

$$\mathbf{P}[X_1 \in dx] = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-m)^2}\, dx$$

  Ensuite, on cale notre modèle en estimant le paramètre. Dans notre exemple on estime m en moyennant un grand nombre de X_i :

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  La loi des grands nombres (ou le bon sens) nous indique que $\widehat{m} \approx m$. Une fois le modèle calé, on peut le faire parler, en calculant des quantités comme par exemple l'entropie $\mathbf{E}[X_1 \log(X_1)]$, qui nous permet de faire des prévisions météorologiques. Voilà, nous venons de faire un petit pas vers les statistiques.

12 ANNEXE : le théorème de transfert

Plaçons nous dans un cadre abstrait. Considérons une application $\theta : E \to F$.

Notation : Si $B \subset F$ on note $\{\theta \in B\} = \{x \in E : \theta(x) \in B\} \subset E$.

Considérons une loi μ sur E. Définissons la mesure μ_θ sur F par

$$\forall B \subset F \quad \mu_\theta(B) = \mu(\{\theta \in B\})$$

(une autre bonne notation pour $\mu_\theta(dy)$ c'est $\mu(\theta \in dy)$).

Théoreme 12.1 (de transfert)

$$\int_E \varphi(\theta(x))\ \mu(dx) = \int_F \varphi(y)\mu_\theta(dy)$$

Démonstration. Supposons que $\varphi = 1_B$ pour $B \subset F$. Dans ce cas, l'égalité du théorème est simplement la définition de μ_θ. Supposons que φ est une fonction étagée. Dans ce cas, l'égalité du théorème se déduit par linéarité. Le cas général se déduit par passage à la limite, car toute fonction φ est limite de fonctions étagées. $\square$

Considérons les applications :

$$\Omega \xrightarrow{X} E \xrightarrow{\varphi} \mathbb{R}$$

On ajoute une mesure proba $\mathbf{P}[d\omega]$ sur Ω. Ce qui nous donne $\mathbf{P}_X$ une mesure de proba sur E, et même $\mathbf{P}_{\varphi(X)}$ une mesure de proba sur $\mathbb{R}$:

$$(\Omega, \mathbf{P}) \xrightarrow{X} (E, \mathbf{P}_X[dx]) \xrightarrow{\varphi} (\mathbb{R}, \mathbf{P}_{\varphi(X)})$$

Le théorème de transfert nous donne alors

$$\int_\Omega \varphi(X(\omega))\, \mathbf{P}[d\omega] = \int_E \varphi(x)\, \mathbf{P}_X[dx] = \int_\mathbb{R} y\, \mathbf{P}_{\varphi(X)}[dy]$$

Idem avec des notations plus probabilistes (donc mieux) :

$$\mathbf{E}[\varphi(X)] = \int_E \varphi(x)\, \mathbf{P}[X \in dx] = \int_\mathbb{R} y\, \mathbf{P}[\varphi(X) \in dy]$$

Exo : Considérez la situation suivante :

$$\Omega \xrightarrow{X} E \xrightarrow{\varphi} F \xrightarrow{\psi} \mathbb{R}$$

Considérez une mesure de probabilité $\mathbf{P}$ sur Ω. Transportez-la sur les autres ensembles. Ecrivez plein d'égalités d'intégrales.

13 ANNEXE : un peu plus de théorie de la mesure

Nous avons dit que, pour être une mesure, il suffisait à μ de vérifier l'additivité :

$$A \subset B = \emptyset \;\Rightarrow\; \mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B] \qquad (Additivite)$$

Mais pour faire plus de théorie, le bon axiome, c'est la σ-additivité : pour toute famille (A_i) de parties de E disjointes :

$$\mu[\cup_i A_i] = \sum_i \mu[A_i] \qquad (\sigma Additivite)$$

Cette axiome est équivalent à : pour toute famille (B_i) de parties, avec $B_i \subset B_{i+1}$, on a

$$\mu[\cup_i B_i] = \lim_i \mu[B_i] \qquad (continuite \uparrow)$$

Exo : démontrez cette équivalence. Indice : trouvez comment passer d'une suite de parties disjointes à une suite de parties croissantes. Et vise-versa.

L'axiome ($continuite \uparrow$) peut se réécrire avec des espérances: Si (φ_i) est une suite croissante de fonctions indicatrices, alors

$$\int \lim_i \varphi_i\, \mu = \lim_i \int \varphi_i\, \mu$$

(oui, c'est la même chose car si $\varphi_i = 1_{B_i}$ alors $\lim_i \varphi_i = 1_{\cup_i B_i}$).

On se convaint facilement que cette dernière propriété, se généralise : Si (φ_i) est une suite croissante de fonctions positives, alors

$$\int \lim_i \varphi_i \, \mu = \lim_i \int \varphi_i \, \mu$$

(on a remplacé "indicatrice" par "positive")

Question : peut-on enlever le mot "croissante" dans la propriété ci-dessous. La réponse est non. Pour le voir, considérons $\varphi_i = 1_{[i,\infty[}$, qui forme une suite décroissante. On a

$$\int \lim_i \varphi_i \mu = \int 0 \, \mu = 0 < \lim_i \int \varphi_i \mu = \lim_i +\infty = +\infty$$

Ainsi, on ne peut pas toujours inverser les limites et les intégrales. Mais pour se faire, il suffit d'éviter les problèmes d'intégrale infinie :

Théoreme 13.1 (Convergence dominée) soit φ_i une suite de fonctions sur E. Soit μ une mesure sur E. Supposons qu'il existe une fonction $f \geq 0$ telle que $\forall i |\varphi_i| < f$ et tel que $\int f \, \mu < \infty$ (f domine les φ_i). Dans ce cas

$$\int \lim_i \varphi_i \, \mu = \lim_i \int \varphi_i \, \mu$$

Ce théorème est très utile en analyse et en probabilités. Selon le niveau d'application des modélisations, il peut être utile ou pas. Mais dans tous les cas, un tel théorème n'est possible qu'en supposant la (σAdditivité) des mesures. La simple (Additivité) ne suffit pas.