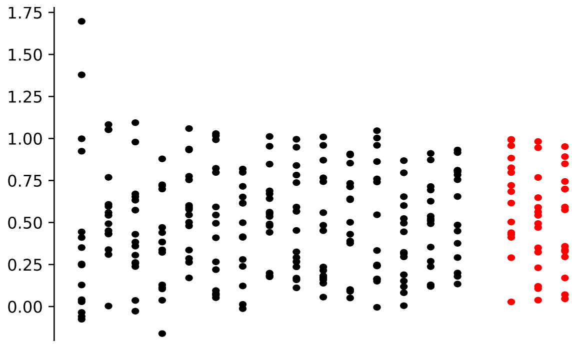
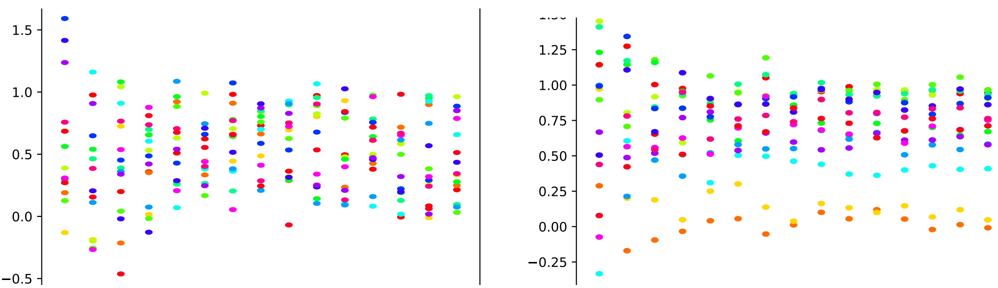
Quelques rappels

1 Sur les convergences probabilistes

Voici une illustration de la convergence en loi. Pour chaque n, j'ai tiré un échantillon de taille 15 d'une certaine variable aléatoire X_n, c'est-à-dire simulé 15 fois X_n de façon indépendante, chaque X_n donnant lieu à un point d'abscisse n. J'ai choisi des X_n qui convergent en loi vers une uniforme sur $[0, 1]$. Pour s'en rendre compte, j'ai simulé en rouge trois échantillons de taille 15 de loi uniforme sur $[0, 1]$. Pour n grand, le nuage de point numéro n « a la même tête » que les nuages de points rouges. En tout cas, il ressemble autant à chacun de ces nuages rouges que ces nuages rouges se ressemblent entre eux. La convergence en loi donne un sens mathématique à cette ressemblance qui frappe l'œil.



La nuance entre convergence en loi et convergence presque sûre se révèle lorsqu'on ne fait plus le dessins ci-dessus en noir et blanc (et rouge) mais en couleurs : désormais, je lance une simulation bleue et dessine tous les X_n correspondant, puis une simulation jaune, etc. On constate que dans la première figure ci-après, pour une couleur donnée, la trajectoire oscille beaucoup tandis que pour la seconde, chaque trajectoire admet une asymptote horizontale : dans le premier cas, on converge en loi mais pas presque sûrement tandis que dans le second, on converge en loi *et* presque sûrement !



La **convergence en loi** dit seulement que, pour n grand, X_n se comporte à peu près comme X. En particulier, elle ne dit pas comment les X_n interagissent entre elles, si elles « s'alignent ou pas du tout ». Autrement dit, pour chaque très grand n, si je regarde le nuage de points au-dessus de n, ça ressemble à ce que ferait X.

La **convergence presque sure** dit que, la suite de réels $(X_n(\omega))$ que j'observerai en pratique converge au sens usuel des nombres réels vers un certain $X(\omega)$

La **convergence en probabilité** dit que, si je fixe *une* grande valeur de n, alors j'aurai une grande probabilité que $X_n(\omega)$ soit proche de $X(\omega)$. Ce n'est pas équivalent à la convergence presque sure car il se pourrait qu'une infinité de fois, à des instants rares mais aléatoires, X_n décide de s'écarter significativement de X.

Dire que X_n converge presque sûrement vers ℓ équivaut à

$$\forall \varepsilon > 0, \quad \mathbf{P}(\forall k \geq n, \, |X_k - \ell| < \varepsilon) \xrightarrow[n\to\infty]{} 1.$$

La convergence en probabilités revient, quant à elle, à

$$\forall \varepsilon > 0, \quad \mathbf{P}(|X_n - \ell| < \varepsilon) \xrightarrow[n\to\infty]{} 1.$$

Voyez-vous la différence ?

2 Limsup et liminf

Si vous avez besoin de réviser les notions de limsup et liminf, je vous invite à regarder cette vidéo.

J'ajoute à cela l'explication suivante. Si $u = (u_n)$ désigne une suite de réels, on peut noter $E(u)$ l'ensemble de toutes les limites possibles pour des suites extraites de (u_n), où l'on autorise les limites $+\infty$ et $-\infty$. Ainsi, si une certaine suite extraite de u admet une limite ℓ, alors ℓ appartiendra à $E(u)$, tandis que si la suite extraite considérée n'admet pas de limite, elle n'ajoutera aucun élément à $E(u)$. Il se trouve que l'ensemble $E(u)$ n'est jamais vide et qu'il admet toujours un maximum et un minimum dans $[-\infty, +\infty]$. La limsup de la suite u n'est autre que le maximum de $E(u)$. La liminf est son minimum.

3 Lemme de Borel–Cantelli

Commençons par énoncer ce lemme.

Lemme de Borel–Cantelli
Soit $(A_n)_{n\geq 0}$ une suite d'événements. On note $B = \bigcap_n \bigcup_{k\geq n} A_k$.
1. Si $\sum_{n=0}^{\infty} \mathbf{P}(A_n) < \infty$, alors B est de probabilité 0.
2. Si $\sum_{n=0}^{\infty} \mathbf{P}(A_n) = \infty$ et si on suppose de plus que les événements A_n sont <u>indépendants</u>, alors B est de probabilité 1.

Pour comprendre ce lemme, il convient de comprendre la signification de l'événement B. Il signifie « il y a une infinité de valeurs de k pour lesquelles l'événement A_k a lieu ». En effet, on a

$$B = \{\omega \in \Omega \,:\, \forall n,\, \exists k \geq n,\, \omega \in A_k\}\,.$$

Cet événement est souvent noté $\limsup A_n$. À titre informatif, la raison derrière cela est que $\mathbf{1}_B = \limsup \mathbf{1}_{A_n}$ — mais c'est quelque chose d'assez anecdotique.

L'énoncé 1 indique que B est de probabiltié 0. Cela signifie que son complémentaire est de probabilité 1. Or le complémentaire de B est l'événement $\bigcup_n \bigcap_{k\geq n} A_k^c$, qui peut se reformuler « il existe un rang n à partir duquel aucun des événements A_k n'a lieu ».

Subtilité fondamentale : Le lemme de Borel–Cantelli ne dit *pas du tout* que, dans l'énoncé 1, il existe un entier n tel que, presque sûrement, les événements A_k n'ont pas lieu à partir du rang n. Et l'énoncé 2 ne dit *pas du tout* qu'il existe une partie infinie $S \subset \mathbf{N}$ telle que, presque sûrement, pour tout $k \in S$, l'événement A_k ait lieu. Non, le lemme de Borel-Cantelli version 1 dit que, pour presque tout ω, il existe un rang n qui dépend de ω tel que les événements A_k n'aient pas lieu à partir du rang n. De même, le lemme de Borel-Cantelli version 2 dit que, pour presque tout ω, il existe une partie infinie $S \subset \mathbf{N}$ qui dépend de ω telle que les événements A_k aient lieu pour tout $k \in S$. Autrement dit, le second lemme de Borel–Cantelli dit que l'ensemble aléatoire $\{k \in \mathbf{N} \,:\, \omega \in A_k\}$ est infini presque sûrement.

Utilité du lemme : Si vous avez des estimées sur des probabilités d'événements, alors vous pouvez obtenir des informations presque sures sur des comportements asymptotiques. Typiquement, si vous souhaitez montrer que, presque sûrement, pour tout n assez grand, les choses se passent comme vous voulez, vous pouvez poser A_n l'événement « les choses ne se passent pas comme je veux » et voir si la version 1 du lemme de Borel–Cantelli s'applique.

Voici comment comprendre la premier lemme de Borel–Cantelli. Pour la démonstration du second, je vous réfère à la démonstration présentée en appendice B4 du polycopié de cours.

Démonstration. Pour tout n, l'événement B est inclus dans $\bigcup_{k\geq n} A_k$. On a donc

$$\mathbf{P}(B) \leq \mathbf{P}\left(\bigcup_{k\geq n} A_k\right) \leq \sum_{k=n}^{\infty} \mathbf{P}(A_k),$$

où la seconde inégalité vaut par sous-σ-additivité. Le membre de droite tend vers 0 quand n tend vers l'infini, en tant que reste d'une série convergente. Donc $\mathbf{P}(B) \leq 0$. Puisque $\mathbf{P}(B)$ est clairement positif, en tant que probabilité, on a $\mathbf{P}(B) = 0$. $\square$

Autre démonstration. Puisque chaque variable aléatoire $\mathbf{1}_{A_n}$ est positive, on a la première égalité de la chaîne d'égalités suivantes

$$\mathbb{E}\left(\sum_{n=0}^{\infty}\mathbf{1}_{A_n}\right)=\sum_{n=0}^{\infty}\mathbb{E}\left(\mathbf{1}_{A_n}\right)=\sum_{n=0}^{\infty}\mathbf{P}\left(A_n\right)<\infty,$$

où la finitude est vraie par hypothèse. La variable aléatoire $\sum_{n=0}^{\infty}\mathbf{1}_{A_n}$ compte combien des événements A_k sont satisfaits. On vient de voir que cette variable aléatoire était d'espérance finie. Cela qui implique en particulier qu'elle est finie presque sûrement, ce qu'on souhaitait démontrer. □

Remarque. L'ensemble des parties infines de $\mathbf{N}$ n'est pas dénombrable. Cela explique qu'il n'y ait aucune contradiction dans la situation suivante : je tire à pile ou face une infinité de faces avec une pièce équilibrée. Il est intuitivement clair, et confirmé par la version 2 du lemme de Borel-Cantelli, que presque sûrement, il y aura une infinité de piles. Pourtant, quelle que soit la partie infinie $S \subset \mathbf{N}$ que je me donne, l'événement « tous les lancers de S donnent pile » est de probabilité nulle. Cela est quelque peu paradoxal car on vient d'écrire un événement de probabilité 1 sous la forme $\bigcup_S C_S$, où l'union porte sur toutes les parties infinies $S \subset \mathbf{N}$ et où C_S désigne l'événement de la phrase précédente... qui est de probabilité nulle ! Cela serait contradictoire si l'ensemble des parties infinies de $\mathbf{N}$ était dénombrable, par sous-σ-additivité. Mais nous sommes sauvés : l'ensemble des parties infinies de $\mathbf{N}$ n'est pas dénombrable ! Cela peut se voir à la main mais on peut aussi considérer qu'on vient de le démontrer de façon probabiliste, en raisonnant par l'absurde : si cet ensemble était dénombrable, on aurait obtenu une contradiction.

4 Sur l'espérance et ses propriétés

Si on se donne une variable aléatoire à valeurs dans $[0,\infty]$, il existe, sans hypothèse supplémentaire, une façon de définir son espérance. Cela donne un résultat que je vais noter $\mathbb{E}_{\geq 0}(X)$ et qui appartient à $[0,\infty]$. Même si X est à valeurs dans $[0,\infty[$, le résultat $\mathbb{E}_{\geq 0}(X)$ appartient a priori à $[0,\infty]$: il peut être infini.

Si on travaille avec une variable aléatoire réelle X, alors $|X|$ est une variable aléatoire positive. On peut donc parler de $\mathbb{E}_{\geq 0}(|X|)$. Si $\mathbb{E}_{\geq 0}(|X|)<\infty$, on dit que X est *intégrable*. Dans le cas où la variable aléatoire réelle X est intégrable, il existe une façon de définir son espérance, que je noterai $\mathbb{E}_{\mathbf{R}}(X)$ et qui appartient à $\mathbf{R}$.

Chacune de ces deux notions d'espérance admet une propriété de linéarité. Commençons par le cas positif. Dans les cours d'intégration et de probabilités, on convient que $0\times\infty=0$, tout en gardant à l'esprit que la multiplication n'est alors pas une opération continue sur $[0,\infty]^2$: $1/n$ tend vers 0, n tend vers l'infini et pourtant $1=(1/n)\times n$ ne converge pas vers $0\times\infty=0$.

Si X et Y sont des variables aléatoires à valeurs dans $[0,\infty]$ et si a et b sont des éléments de $[0,\infty]$, alors $aX+bY$ est aussi une variable aléatoire à valeurs dans $[0,\infty]$.

Alors, tous les membres de l'égalité suivante ont un sens et l'égalité est valide

$$\mathbb{E}_{\geq 0}(aX + bY) = a\mathbb{E}_{\geq 0}(X) + b\mathbb{E}_{\geq 0}(Y).$$

Remarque. C'est bien la positivité de a, b, X et Y qui importe ici. Si on suppose seulement la positivité de $aX + bY$, ça ne fonctionne pas. Par exemple, on pourrait prendre X positive d'espérance infinie, $Y = X$, $a = 1$ et $b = -1$. On aurait alors $aX + bY = X - X = 0$ qui est bien positive, d'espérance nulle. Pourtant, la formule n'aurait aucun sens, prétendant que $0 = \infty - \infty$.

Quittons le cas positif pour traiter le cas des variables aléatoires réelles intégrables. Si X et Y sont des variables aléatoires réelles *intégrables* et si a et b sont des réels, alors $aX + bY$ est aussi une variable aléatoire réelle intégrable. Alors, tous les membres de l'égalité suivante ont un sens et l'égalité est valide

$$\mathbb{E}_{\mathbf{R}}(aX + bY) = a\mathbb{E}_{\mathbf{R}}(X) + b\mathbb{E}_{\mathbf{R}}(Y).$$

Remarque. C'est bien le caractère intégrable de X et Y qui importe ici. Si on suppose seulement l'intégrabilité de $aX + bY$, ça ne fonctionne pas. Par exemple, on pourrait prendre X positive d'espérance infinie, $Y = X$, $a = 1$ et $b = -1$. On aurait alors $aX + bY = X - X = 0$ qui est bien intégrable. Pourtant, la formule n'aurait aucun sens, prétendant que $0 = \infty - \infty$.

En dehors de ce rappel, les notations $\mathbb{E}_{\geq 0}(X)$ et $\mathbb{E}_{\mathbf{R}}(X)$ ne sont jamais employées : on utilise toujours la même notation $\mathbb{E}(X)$. Pourquoi ? La raison est la suivante :

— si une variable aléatoire X est à valeurs dans $[0, \infty]$ et vérifie $\mathbb{E}_{\geq 0}(X) = \infty$, alors il n'y a pas ambiguïté : la notation $\mathbb{E}(X)$ ne peut que se référer $\mathbb{E}_{\geq 0}(X)$ car pour parler de $\mathbb{E}_{\mathbf{R}}(X)$, il faudrait que X soit intégrable, ce qui n'est pas le cas ici ;

— si une variable aléatoire X est à valeurs dans $[0, \infty]$ et vérifie $\mathbb{E}_{\geq 0}(X) < \infty$, alors X est intégrable et $\mathbb{E}(X)$ peut aussi bien référer à $\mathbb{E}_{\geq 0}(X)$ qu'à $\mathbb{E}_{\mathbf{R}}(X)\dots$ ce qui ne pose pas problème car il se trouve que ces deux quantités ont le bon goût d'être égales ;

— si X est une variable aléatoire intégrable qui prend des valeurs strictement négatives, alors seule $\mathbb{E}_{\mathbf{R}}(X)$ a un sens, si bien que $\mathbb{E}(X)$ ne peut se rapporter qu'à cette quantité.

Je rappelle que la linéarité de l'espérance ne fait appel à *aucune* hypothèse d'indépendance. Cela en fait un outil puissant, puisqu'il s'applique à moindres frais. Où l'indépendance intervient-elle ? Quand on calcule l'espérance de *produits* !

Soient $X_1, \dots, X_n$ des variables aléatoires positives (resp. réelles intégrables). On suppose que ces variables aléatoires forment une famille *indépendante*. Alors le produit $X_1 \dots X_n$ est une variable aléatoire positive (resp. réelle intégrable) et son espérance vaut $\mathbb{E}(X_1) \dots \mathbb{E}(X_n)$.

Remarque. En général, un produit de variables aléatoires intégrables peut parfaitement ne pas être intégrable. Par exemple, considérer $X \times X$ pour une variable aléatoire qui est dans $\mathbf{L}^1$ mais pas dans $\mathbf{L}^2$. Ainsi, dans le paragraphe précédent, l'hypothèse d'indépendance ne sert pas seulement à donner la formule pour $\mathbb{E}(X_1 \dots X_n)$, elle sert déjà avant, pour affirmer que $X_1 \dots X_n$ est intégrable (dans le cas où les X_i sont intégrables et forment une famille indépendante).

L'indépendance intervient aussi quand on calcule la variance d'une somme. Si on a des variables aléatoires $X_1, \dots, X_n$ dans $\mathbf{L}^2$, leur somme est encore dans $\mathbf{L}^2$. Ainsi, les variances $V(X_1), \dots, V(X_n)$ et $V(X_1 + \cdots + X_n)$ sont toutes bien définies. La dernière variance est-elle la somme des précédentes ? Pas toujours ! Prenez X une variable aléatoire de variance non-nulle et posez $X_1 = X_2 = X$. Alors $X_1 + X_2 = 2X$ donc la variance de X_1+X_2 vaut $2^2V(X) = 4V(X)$, ce qui ne vaut pas du tout $V(X_1)+V(X_2) = 2V(X)$. Mais si on suppose en outre que les variables aléatoires X_i forment une famille indépendante, alors il devient bien vrai que la variance de la somme est la somme des variances. Il suffit en fait de supposer que pour tous i et j distincts, la covariance de X_i et X_j est nulle, ce qui est une hypothèse beaucoup plus faible que l'indépendance.

5 Mesures produits, indépendance et Fubini

5.1 Produits d'espaces mesurés

On sait définir le produit d'ensembles : si $E_1, \dots, E_d$ désignent des ensembles, alors on note $E_1 \times \cdots \times E_d$ l'ensemble des n-uplets $(x_1, \dots, x_d)$ où, pour tout i, l'élément x_i appartient à E_i. De façon analogue, peut-on définir un produit d'espaces mesurés ?

La réponse est affirmative, sous une petite hypothèse technique : on doit considérer des espaces σ-finis. Un espace mesuré $(E, \mathscr{E}, \mu)$ est dit *σ-fini* s'il existe une suite (A_n) d'éléments de $\mathscr{E}$ telle que $E = \bigcup_n A_n$ et, pour tout n, on ait $\mu(A_n) < \infty$. Bref, un espace mesuré est σ-fini si l'espace peut s'écrire comme union dénombrable de blocs de mesure finie.

Exemples fondamentaux :

1. Tout espace mesuré vérifiant $\mu(E) < \infty$ est σ-fini, en posant $A_n = E$.
2. Cas particulier de l'exemple précédent : tout espace probabilisé est σ-fini.
3. L'espace $\mathbf{R}^d$ muni de la mesure de Lebesgue est σ-fini.

Voici un quatrième exemple pouvant être sauté en première lecture. Si $(E, \mathscr{E}, \mu)$ est un espace σ-fini et $f : E \to [0, \infty[$ est mesurable, on peut définir une mesure ν par $\nu(A) = \int_A f \, \mathrm{d}\mu$. La mesure ν est alors automatiquement σ-finie. En effet, prenons (A_n) témoignant de la σ-finitude de μ. Pour tous m et n, on peut poser $B_{m,n} = A_n \cap \{x : f(x) \leq m\}$, qui vérifie $\nu(B_{m,n}) \leq m\mu(A_n) < \infty$. L'ensemble E s'écrit donc comme union indexée par $\mathbf{N}^2$ de blocs de ν-mesure finie. On conclut en se rappelant que $\mathbf{N}^2$ est dénombrable — une bijection explicite de $\mathbf{N}^2$ vers $\mathbf{N}$ est donnée par $(m,n) \mapsto 2^m(2n+1)-1$. Trouverez-vous une bijection explicite de $\mathbf{N}^3$ vers $\mathbf{N}$?

Soient $(E_1, \mathscr{E}_1, \mu_1), \dots, (E_d, \mathscr{E}_d, \mu_d)$ des espaces mesurés σ-finis. On cherche à définir un espace mesuré $(E', \mathscr{E}', \mu')$ qui mériterait de s'appeler le produit de ces espaces mesurés.

Pour E', c'est simple : on pose $E' = E_1 \times \cdots \times E_d$. Quelle tribu mettre sur cet espace ? Il existe deux approches naturelles et il se trouve qu'elles définissent bien la même tribu. Vérifier que les deux approches donnent la même tribu est un exercice de difficulté raisonnable.

Approche 1 : La moindre des choses serait que pour tous $A_1 \in \mathscr{E}_1, \dots, A_d \in \mathscr{E}_d$, l'ensemble $A_1 \times \cdots \times A_d$ appartienne à $\mathscr{E}'$. N'ayant pas d'autre exigence nous venant à l'esprit, on prendra la plus petite tribu sur E' contenant ces ensembles. Autrement dit, on pose

$$\mathscr{E}' = \sigma(\{A_1 \times \cdots \times A_d \,:\, A_1 \in \mathscr{E}_1, \dots, A_d \in \mathscr{E}_d\}).$$

Approche 2 (bonus) : La moindre des choses serait que les projections sur chaque facteur soient mesurables. Autrement dit, on veut mettre sur E' une tribu telle que, pour tout i, la fonction $\pi_i : E_1 \times \cdots \times E_d \to E_i$ définie par $\pi_i : x \mapsto x_i$ soit mesurable. N'ayant pas d'autre exigence nous venant à l'esprit, on prendra la plus petite tribu sur E' rendant toutes les fonctions $\pi_1, \dots, \pi_d$ mesurables. Une telle tribu existe bien : il suffit de prendre l'intersection de toutes les tribus rendant toutes les fonctions $\pi_1, \dots, \pi_d$ mesurables. Considérant une intersection de tribus, on pourrait s'inquiéter que la famille de tribus considérée soit vide : tel n'est pas le cas car la tribu formée par toutes les parties de E' rend toutes les fonctions $\pi_1, \dots, \pi_d$ mesurables (pourquoi ?).

La tribu $\mathscr{E}'$ définie par l'une quelconque de ces approches est notée $\mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_d$.

Il nous reste à définir la mesure μ' sur $(E', \mathscr{E}')$. La construire requiert un travail conséquent dans un cours d'intégration. On ne va donc pas la reconstruire dans ce rappel mais la caractériser par un théorème-définition, qu'on ne démontrera pas ici.

Théorème définissant la mesure produit. *Il existe une unique mesure μ' sur $(E', \mathscr{E}')$ vérifiant, pour tous $A_1 \in \mathscr{E}_1, \dots, A_d \in \mathscr{E}_d$, l'égalité*

$$\mu'(A_1 \times \cdots \times A_d) = \mu_1(A_1) \times \cdots \times \mu_d(A_d).$$

La mesure μ' fournie par le théorème précédent est celle qu'on choisit pour définir le produit de nos espaces mesurés σ-finis. Elle s'appelle la *mesure produit* et on la note $\mu_1 \otimes \cdots \otimes \mu_d$. Le symbole $\otimes$ se prononce « tens' », ce qui vient du terme d'algèbre « produit tensoriel ».

5.2 Lien avec l'indépendance

Mesure produit et indépendance faisant toutes deux intervenir des formules de produit, il serait naturel que ces notions soient liées d'une certaine façon. Tel est bien le cas.

Soient $(E_1, \mathscr{E}), \dots, (E_d, \mathscr{E})$ des espaces mesurables. Soient $X_1, \dots, X_d$ des variables aléatoires, à valeurs respectivement dans $(E_1, \mathscr{E}), \dots, (E_d, \mathscr{E})$. Tout d'abord, je prétends que $(X_1, \dots, X_d)$ est une variable aléatoire à valeurs dans le $(E', \mathscr{E}')$ de la section précédente. J'en propose une démonstration, qui peut être sautée en première lecture.

Démonstration. Tout d'abord, quand on parle de la fonction $(X_1, \dots, X_d)$, de quoi parle-t-on ? Implicitement, il y a un espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ sur lequel toutes les X_i sont définies. La fonction $(X_1, \dots, X_d)$ est alors définie sur cet espace probabilisé, à valeurs dans E' et on pose $(X_1, \dots, X_d)(\omega) = (X_1(\omega), \dots, X_d(\omega))$. Il s'agit maintenant de montrer que cette fonction est bien mesurable.

Vérifions que, pour tous $A_1 \in \mathscr{E}_1, \dots, A_d \in \mathscr{E}_d$, l'ensemble

$$\{\omega \in \Omega \,:\, (X_1(\omega), \dots, X_d(\omega)) \in A_1 \times \cdots \times A_d\}$$

appartient à la tribu $\mathscr{F}$. Cet ensemble n'est autre que $\{X_1 \in A_1\} \cap \cdots \cap \{X_d \in A_d\}$. Les X_i étant des variables aléatoires, il s'agit donc d'une intersection finie d'événéments, qui est donc à son tour un événement. Je rappelle que « événement » signifie « élément de $\mathscr{F}$ ».

D'accord, mais pourquoi vérifier cela suffit-il ? Soit Y une fonction définie sur $(\Omega, \mathscr{F})$ et à valeurs dans un espace mesurable $(E, \mathscr{E})$. On se donne une collection $\mathscr{C}$ de parties de E qui vérifie $\mathscr{E} = \sigma(\mathscr{C})$. Pourquoi suffit-il de vérifier $\forall A \in \mathscr{C}$, $\{Y \in A\} \in \mathscr{F}$ pour affirmer que Y est mesurable ?

La raison est la suivante. Posons $\mathscr{G} = \{A \subset E \,:\, \{Y \in A\} \in \mathscr{F}\}$. Je vous laisse le soin de vérifier (exercice formateur) que $\mathscr{G}$ définit bien une tribu sur E. Comme cette tribu contient $\mathscr{C}$, on en déduit que $\mathscr{E} = \sigma(\mathscr{C}) \subset \mathscr{G}$, ce qui est une reformulation de ce qu'on cherchait à établir. □

Nous sommes alors en mesure de formuler le lien entre indépendance et mesure produit. Soient $\mu_1, \dots, \mu_d$ des mesures de probabilités sur les $(E_1, \mathscr{E}_1), \dots, (E_d, \mathscr{E}_d)$. On a alors l'équivalence entre les deux énoncés suivants :

1. la loi de $(X_1, \dots, X_d)$ est $\mu_1 \otimes \cdots \otimes \mu_d$,
2. les variables aléatoires $X_1, \dots, X_n$ sont indépendantes et, pour chaque i, la variable aléatoire X_i est de loi μ_i.

Démonstration. Commençons par $2 \implies 1$. Supposons 2 et notons μ la loi de $(X_1, \dots, X_d)$. Il s'agit d'une mesure de probabilité sur $(E', \mathscr{E}')$. Dire que les variables aléatoires X_i sont indépendantes, c'est dire que pour tous $A_1 \in \mathscr{E}_1, \dots, A_d \in \mathscr{E}_d$, on a

$$\mathbf{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbf{P}(X_1 \in A_1) \times \cdots \times \mathbf{P}(X_d \in A_d).$$

Le membre de gauche peut se ré-écrire $\mathbf{P}((X_1, \dots, X_d) \in A_1 \times \cdots \times A_d)$, qui à son tour peut se ré-écrire, par définition de μ, sous la forme $\mu(A_1 \times \cdots \times A_d)$. Le membre de droite de l'égalité se ré-écrit, quant à lui, sous la forme $\mu_1(A_1) \times \dots \mu_d(A_d)$, puisque X_i est de loi μ_i. On a donc montré que pour tous choix de A_i, on a $\mu(A_1 \times \cdots \times A_d) = \mu_1(A_1) \times \cdots \times \mu_d(A_d)$. C'est exactement la propriété qui caractérise la mesure produit $\mu_1 \otimes \cdots \otimes \mu_d$.

Traitons maintenant $1 \implies 2$. On suppose 1. De manière analogue à ce qu'on vient de voir, en écrivant la définition de ce qu'est la loi de $(X_1, \ldots, X_d)$ ainsi que la définition de mesure produit, on obtient pour tous $A_1 \in \mathscr{E}_1, \ldots, A_d \in \mathscr{E}_d$ l'égalité

$$\mathbf{P}(X_1 \in A_1, \ldots, X_d \in A_d) = \mu_1(A_1) \times \cdots \times \mu_d(A_d).$$

On aura gagné dès lors qu'on aura montré que, pour tout i, la loi de X_i est μ_i. Or cela s'observe en prenant A_j quelconque pour $j = i$ et égal à E_j pour $j \neq i$. Cela utilise le fait que les mesures μ_j ont, dans notre énoncé, été supposées de probabilité. □

5.3 Fubini, ou intégrer sur des espaces produits

La théorie de la mesure sert en majeure partie à définir des intégrales : on a notre mesure produit — comment intégrer par rapport à elle ? C'est à cette question que répondent les théorèmes de Fubini. Ils permettent de ramener le calcul d'intégrales par rapport à $\mu_1 \otimes \cdots \otimes \mu_d$ à des intégrales successives par rapport aux mesures μ_i.

Ces théorèmes sont soigneusement énoncés dans le polycopié, aussi je vous réfère aux théorèmes 3.4 et 3.5 de l'appendice A du poly de cours. Si vous les appelez théorème de Fubini positif et théorème de Fubini intégrable au lieu de Fubini–Tonelli et Fubini–Lebesgue, on vous comprendra.

Dans le théorème 3.5, la notation $\mathscr{L}^1(E, \mathscr{E}, \mu)$ désigne l'ensemble des fonctions intégrables définies sur $(E, \mathscr{E}, \mu)$ et à valeurs réelles. La condition d'intégrabilité inclut implicitement la condition de mesurabilité !

Je précise que dans ces théorèmes, l'hypothèse de mesurabilité de f est bien relativement à la tribu produit $\mathscr{A} \otimes \mathscr{B}$.

Ces théorèmes se généralisent à d facteurs au lieu de 2. Cela peut se déduire du cas $d = 2$, par récurrence. Prenons d espaces mesurés σ-finis $(E_1, \mathscr{E}_1, \mu_1), \ldots, (E_d, \mathscr{E}_d, \mu_d)$.

1. Soit $f : E_1 \times \cdots \times E_d \to [0, \infty]$ une fonction $\mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_d$-mesurable. Alors les intégrales intervenant dans la formule suivante ont un sens et l'égalité est valide :

$$\int_{E_1 \times \cdots \times E_d} f \, \mathrm{d}\mu_1 \otimes \ldots \mu_d = \int_{E_1} \cdots \int_{E_d} f(x_1, \ldots, x_d) \, \mathrm{d}\mu_1(x_1) \ldots \mathrm{d}\mu_d(x_d).$$

2. Soit $f : E_1 \times \cdots \times E_d \to \mathbf{R}$ une fonction $\mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_d$-mesurable et intégrable. Alors les intégrales intervenant dans la formule suivante ont un sens et l'égalité est valide :

$$\int_{E_1 \times \cdots \times E_d} f \, \mathrm{d}\mu_1 \otimes \ldots \mu_d = \int_{E_1} \cdots \int_{E_d} f(x_1, \ldots, x_d) \, \mathrm{d}\mu_1(x_1) \ldots \mathrm{d}\mu_d(x_d).$$

De plus, dans chacune de ces situations, les d opérations d'intégration effectuées dans le membre de droite peuvent être effectuées dans n'importe quel ordre : cela donnera à chaque fois le même résultat, à savoir le membre de gauche $\int f \,\mathrm{d}\mu_1 \otimes \dots \mu_d$.

Comment vérifier l'hypothèse d'intégrabilité dans l'item 2 ci-dessus ? Il s'agit de montrer que $\int |f| \,\mathrm{d}\mu_1 \otimes \cdots \otimes \ \mu_d$ est finie. Pour cela, il est commode d'appliquer l'item 1 à la fonction positive $|f|$, afin de ramener ce calcul à des intégrales sur les $(E_i, \mathscr{E}_i, \mu_i)$ plutôt que sur un espace produit.

Le dernier point à gérer consiste à vérifier la mesurabilité par rapport à une tribu $\mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_d$ sans prendre peur. Le cas qui nous occupera principalement cette année est lorsque chaque E_i est égal à $\mathbf{R}$. Je note $\mathscr{B}(\mathbf{R}^d)$ la tribu borélienne associée à l'espace topologique $\mathbf{R}^d$. Si on prend le produit des tribus boréliennes sur $\mathbf{R}$, retombe-t-on bien sur la tribu borélienne de l'espace topologique produit $\mathbf{R}^d$? La réponse est oui : tout va bien, on a bien $\mathscr{B}(\mathbf{R}) \otimes \cdots \otimes \mathscr{B}(\mathbf{R}) = \mathscr{B}(\mathbf{R}^d)$. En particulier, si une fonction $f : \mathbf{R}^d \to \mathbf{R}$ est continue, alors elle est $\mathscr{B}(\mathbf{R}^d)$-mesurable donc $\mathscr{B}(\mathbf{R}) \otimes \cdots \otimes \mathscr{B}(\mathbf{R})$-mesurable.

Plus généralement, il n'y a qu'une seule tribu naturelle sur $\mathbf{R}^d$: peu importe dans quel ordre on regroupe les facteurs ; peu importe également, à chaque étape, si on prend la tribu borélienne d'un produit d'espaces topologiques ou un produit de tribus. Par exemple, pour $d = 3$, les six tribus suivantes sont égales :

$$\mathscr{B}(\mathbf{R}) \otimes \mathscr{B}(\mathbf{R}) \otimes \mathscr{B}(\mathbf{R}) = (\mathscr{B}(\mathbf{R}) \otimes \mathscr{B}(\mathbf{R})) \otimes \mathscr{B}(\mathbf{R}) = \mathscr{B}(\mathbf{R}) \otimes (\mathscr{B}(\mathbf{R}) \otimes \mathscr{B}(\mathbf{R})) = \dots$$

$$\cdots = \mathscr{B}(\mathbf{R}^2) \otimes \mathscr{B}(\mathbf{R}) = \mathscr{B}(\mathbf{R}) \otimes \mathscr{B}(\mathbf{R}^2) = \mathscr{B}(\mathbf{R}^3).$$