Prépa. Agrég. écrit d'Analyse, Annexe n° 8.

Dénombrabilité en Analyse et Probabilités

Une suite ne peut remplir aucun intervalle, d'après Cantor

L'argument qui suit a été publié par G. Cantor vers 1875 ; on peut imaginer que bien d'autres mathématiciens auraient pu le faire avant, si on leur avait posé la question, mais Cantor est l'un des tous premiers à s'être intéressé à ces problèmes de *cardinalité*. A la même époque, il montre que l'ensemble des nombres algébriques est dénombrable ; la conjonction des deux arguments donne une raison très simple pour l'existence de nombres transcendants, existence précédemment montrée directement par Liouville.

Considérons une suite quelconque (w_p) de nombres réels et un intervalle $[a, b]$, $a < b$; on va montrer qu'il existe au moins un élément z de l'intervalle qui ne fait pas partie de la suite.

Désignons par m_1 le plus petit indice m tel que $a < w_m < b$; s'il n'en existe aucun, notre affirmation sera montrée en prenant n'importe quel z de $]a, b[$; de même désignons par n_1 le plus petit indice n tel que $a < w_n < b$ et $w_n \neq w_{m_1}$; s'il n'en existe aucun, notre affirmation est montrée comme avant, en prenant pour z un point de $]a, w_{m_1}[$ ou bien un point de $]w_{m_1}, b[$; si m_1 et n_1 existent, on note que $n_1 > m_1$ et on pose $x_1 = \min(w_{m_1}, w_{n_1})$ et $y_1 = \max(w_{m_1}, w_{n_1})$; on a $a < x_1 < y_1 < b$.

Désignons ensuite par m_2 le plus petit indice m tel que $x_1 < w_m < y_1$; s'il n'en existe aucun, notre affirmation sera montrée en prenant n'importe quel z de $]x_1, y_1[$; de même désignons par n_2 le plus petit indice n tel que $x_1 < w_n < y_1$ et $w_n \neq w_{m_2}$; s'il n'en existe aucun, notre affirmation est montrée comme avant ; sinon, on note que $n_2 > m_2 > n_1$ et on pose $x_2 = \min(w_{m_2}, w_{n_2})$, $y_2 = \max(w_{m_2}, w_{n_2})$; on a $a < x_1 < x_2 < y_2 < y_1 < b$.

On continue ainsi par récurrence, choisissant m_k, n_k plus grands que $n_{k-1} > m_{k-1}$ de façon que $x_k = \min(w_{m_k}, w_{n_k})$ et $y_k = \max(w_{m_k}, w_{n_k})$ soient “les premiers” éléments de la suite (w_p) à vérifier $x_{k-1} < x_k < y_k < y_{k-1}$. On pose aussi $x_0 = a$ et $y_0 = b$.

La suite des intervalles emboîtés $[x_k, y_k]$ contient au moins un point z, et on a $x_k < z < y_k$ pour tout entier k ; il n'est pas possible que z soit l'un des (w_p), par exemple $z = w_{p_0}$; en effet, la suite des indices n_k tend vers l'infini puisque $n_k < m_{k+1} < n_{k+1}$ pour tout entier $k \geq 1$; soit $k_0 = k$ le plus petit indice tel que $p_0 \leq n_k$; si on avait $p_0 \leq m_k$ et $z = w_{p_0}$, alors z serait le premier point x de la suite (w_p) vérifiant $x_{k-1} < x < y_{k-1}$, ce qui n'est pas le cas puisque ce point est x_k ou bien y_k, tous deux différents de z ; si on avait $m_k < p_0 \leq n_k$, alors z serait le premier point y de la suite (w_p) vérifiant $x_{k-1} < y < y_{k-1}$ et $y \neq w_{m_k}$, ce qui n'est pas le cas à nouveau, puisque ce point est encore y_k ou x_k.

Exercices élémentaires

1. Montrer que toute partie X de $\mathbb{R}^2$, munie de la distance d induite par $\mathbb{R}^2$, est un espace séparable.

Indication. On suppose X non vide. Soit (x_m) une suite dense dans $\mathbb{R}^2$, (r_n) une énumération des rationnels > 0, et z_0 un point fixé de X ; pour tout couple (m, n) on pose $z_{m,n} = z_0$ si la boule $\mathrm{B}(x_m, r_n)$ ne rencontre pas X, sinon on prend pour $z_{m,n}$ un point de l'intersection. Montrer que l'ensemble dénombrable $(z_{m,n})$ est dense dans X.

2. Montrer que la tribu borélienne de $\mathbb{R}^2$ coïncide avec le produit tensoriel de la tribu borélienne de $\mathbb{R}$ avec elle-même.

Désintégration d'une probabilité sur $\mathbb{R}^2$

Désignons par μ une probabilité sur $\mathbb{R}^2$ (c'est à dire une probabilité sur la tribu borélienne de $\mathbb{R}^2$) et par ν la probabilité sur $\mathbb{R}$ qui est l'image de μ par la première projection $p_1 : \mathbb{R}^2 \to \mathbb{R}$, définie par $p_1(x,y) = x$. On a donc

$$\int_{\mathbb{R}^2} h(x)\, d\mu(x,y) = \int_{\mathbb{R}} h(x)\, d\nu(x)$$

pour toute fonction borélienne positive h sur $\mathbb{R}$.

Théorème. *Il existe une application $x \in \mathbb{R} \to \mu_x$, où chaque μ_x est une probabilité sur $\mathbb{R}$, telle que*

– *pour tout borélien A de $\mathbb{R}$, l'application $x \to \mu_x(\mathrm{A})$ est borélienne*
– *pour toute fonction borélienne positive φ sur $\mathbb{R}^2$, on a*

$$\int_{\mathbb{R}^2} \varphi(x,y)\, d\mu(x,y) = \int_{\mathbb{R}} d\nu(x) \Big(\int_{\mathbb{R}} \varphi(x,y)\, d\mu_x(y) \Big).$$

On a de plus le résultat d'unicité suivant : si $(\tilde{\mu}_x)$ est une autre famille de probabilités sur $\mathbb{R}$ qui vérifie les deux propriétés précédentes, alors $\tilde{\mu}_x = \mu_x$ pour ν-presque tout $x \in \mathbb{R}$.

Exemples évidents.

1. Si γ_1, γ_2 sont deux probabilités sur $\mathbb{R}$, si $\mu = \gamma_1 \otimes \gamma_2$, alors $\nu = \gamma_1$ et on a $\mu_x = \gamma_2$ pour γ_1-presque tout x ; dans ce cas la formule ci-dessus est simplement le théorème de Fubini pour la mesure produit.

2. Si γ est une probabilité sur $\mathbb{R}$, et si on désigne par μ l'image de γ par l'application $x \to (x,x)$ qui envoie $\mathbb{R}$ sur la première diagonale de $\mathbb{R}^2$, on obtient encore $\nu = \gamma$ comme dans l'exemple précédent, mais maintenant on aura $\mu_x = \delta_x$ (la masse de Dirac au point x) pour γ-presque tout x.

Idée de preuve du théorème. On va trouver les μ_x grâce à leur fonction de répartition. Si $\varphi(x,y) = h(x)\, \mathbf{1}_{y \le b}$, on devra avoir

$$\int_{\mathbb{R}} d\nu(x)\, h(x) \mu_x(]-\infty, b]) = \int_{\mathbb{R}^2} h(x)\, \mathbf{1}_{y \le b}\, d\mu(x,y).$$

On va donc chercher, pour chaque b fixé, une fonction $x \to g_b(x)$ qui vérifie la propriété que doit avoir $x \to \mu_x(]-\infty, b])$, c'est à dire

$$\int_{\mathbb{R}} h(x)\, g_b(x)\, d\nu(x) = \int_{\mathbb{R}^2} h(x)\, \mathbf{1}_{y \le b}\, d\mu(x,y)$$

pour toute h borélienne bornée. Le point essentiel qui rattache ce développement aux applications de la dénombrabilité est le suivant : pour déterminer la fonction de répartition $b \to \mu_x((-\infty, b])$, il suffit de la déterminer lorsque $b = q$ est rationnel. On va donc chercher une famille dénombrable de fonctions $(g_q)_{q \in \mathbb{Q}}$ de la variable x, qui reconstitueront, pour x fixé et q variant, la fonction de répartition de la probabilité μ_x voulue.

Notons i_1 le plongement isométrique de $\mathrm{L}_2(\mathbb{R}, \nu)$ dans $\mathrm{L}_2(\mathbb{R}^2, \mu)$ obtenu par la formule $i_1(h)(x,y) = h(x)$, c'est à dire que nous considérons tout simplement une fonction d'une seule variable h comme une fonction de deux variables $i_1(h)$ qui ne dépend que de la première variable ! L'opérateur adjoint $\mathrm{P} = i_1^*$, défini de $\mathrm{L}_2(\mathbb{R}^2, \mu)$ dans $\mathrm{L}_2(\mathbb{R}, \nu)$ est caractérisé par

$$\forall h \in \mathrm{L}_2(\nu), \quad \langle \mathrm{P}f, h\rangle_{\mathrm{L}_2(\nu)} = \langle f, i_1 h\rangle_{\mathrm{L}_2(\mu)}$$

c'est à dire que la fonction $g = \mathrm{P}f$ d'une seule variable est caractérisée par le fait que

$$\int_{\mathbb{R}} h(x)g(x)\, d\nu(x) = \int_{\mathbb{R}^2} h(x)f(x,y)\, d\mu(x,y)$$

pour toute $h \in \mathrm{L}_2(\mathbb{R}, \nu)$. On vérifie facilement les propriétés suivantes.

(1). L'application P est linéaire et continue en norme L_2.
(2). On a $\mathrm{P}(k(x)f(x,y)) = k(x)(\mathrm{P}f)(x)$ pour toute k borélienne bornée.
(3). Si $f \geq 0$, alors $\mathrm{P}f \geq 0$ ν-presque partout.
(4). Et finalement, $\mathrm{P}(\mathbf{1}) = \mathbf{1}$.

Pour chaque rationnel q posons $f_q(x,y) = \mathbf{1}_{y \leq q}$ et désignons par g_q une fonction borélienne sur $\mathbb{R}$, définie partout, qui soit un ν-représentant de $\mathrm{P}f_q$; pour chaque couple $q < r$ de rationnels, on a $g_q \leq g_r$ ν-presque partout d'après (3), puisque $f_q \leq f_r$, donc

$$\mathrm{N}_1 = \bigcup \{x \in \mathbb{R} : g_q(x) > g_r(x),\, q < r,\, q, r \in \mathbb{Q}\}$$

est ν-négligeable, comme réunion dénombrable d'ensembles négligeables. De même puisque $0 \leq f_q \leq 1$ on a $0 \leq g_q \leq 1$, donc

$$\mathrm{N}_2 = \bigcup \{x \in \mathbb{R} : g_q(x) < 0,\, q \in \mathbb{Q}\}$$

et

$$\mathrm{N}_3 = \bigcup \{x \in \mathbb{R} : g_q(x) > 1,\, q \in \mathbb{Q}\}$$

sont ν-négligeables. Posons $\mathrm{N} = \mathrm{N}_1 \cup \mathrm{N}_2 \cup \mathrm{N}_3$. Pour chaque réel $x \notin \mathrm{N}$, la fonction $q \in \mathbb{Q} \to g_q(x)$ est croissante et comprise entre 0 et 1 ; pour $x \in \mathrm{N}$, on posera par exemple $g_q(x) = 0$ si $q < 0$ et $g_q(x) = 1$ si $q \geq 0$; ce changement sur l'ensemble négligeable N ne modifie pas le fait qui nous intéresse, à savoir que g_q est toujours un représentant de $\mathrm{P}f_q$. On pose maintenant pour tous t, x réels

$$\tilde{g}_t(x) = \lim_{q \searrow t} g_q(x).$$

Par construction, cette fonction $t \to \tilde{g}_t(x)$ est une fonction croissante continue à droite sur $\mathbb{R}$, comprise entre 0 et 1, et pour chaque t fixé $x \to \tilde{g}_t(x)$ est borélienne comme limite simple de fonctions boréliennes. Lorsque t est rationnel, on vérifie par convergence dominée lorsque $q \searrow t$, appliquée à $f_q \to f_t$ et $g_q \to \tilde{g}_t$ que la fonction $x \to \tilde{g}_t(x)$ vérifie aussi les propriétés qui caractérisent $\mathrm{P}f_t$, à savoir

$$\int_{\mathbb{R}} h(x)\tilde{g}_t(x)\, d\nu(x) = \int_{\mathbb{R}^2} h(x)\mathbf{1}_{y \leq t}\, d\mu(x,y)$$

pour toute $h \in \mathrm{L}_2(\mathbb{R}, \nu)$. On en conclut que $g_q = \tilde{g}_q$ ν-presque partout pour tout rationnel q, et quitte à faire une nouvelle modification à un ensemble négligeable près on pourra supposer que $\tilde{g}_q = g_q$ partout, et on notera simplement $g_t(x)$, pour tout t réel.

La fonction $t \in \mathbb{R} \to g_t(x)$ admet aussi des limites $g_{-\infty}(x)$ et $g_\infty(x)$ aux deux infinis ; ces deux limites sont des fonctions boréliennes de la variable x. On vérifie comme ci-dessus que ces limites sont des représentants de $\mathrm{P}(0)$ et $\mathrm{P}(\mathbf{1})$ respectivement, donc $g_{-\infty} = 0$ et $g_\infty = 1$ ν-presque partout. En enlevant encore un négligeable on trouve finalement un borélien ν-négligeable $\mathrm{N}' \subset \mathbb{R}$ tel que : pour tout $x \notin \mathrm{N}'$, la fonction $t \in \mathbb{R} \to g_t(x)$ est une fonction croissante continue à droite, dont les limites en $\pm\infty$ sont égales à 0 et 1 ; c'est donc la fonction de répartition d'une probabilité μ_x sur $\mathbb{R}$. Pour les autres valeurs de x on posera par exemple $\mu_x = \delta_0$, si on veut.

Par construction $x \to \mu_x(\mathrm{A})$ est borélienne lorsque $\mathrm{A} = (-\infty, q]$ avec q rationnel, puisque c'est la fonction g_q de la construction précédente. La classe $\mathcal{C}$ formée par tous ces intervalles est stable par intersection finie, et la classe $\mathcal{M}$ des boréliens $\mathrm{A} \subset \mathbb{R}$ tels que $x \to \mu_x(\mathrm{A})$ soit borélienne est monotone et stable par différence, elle contient la classe $\mathcal{C}$, donc elle contient aussi la tribu engendrée par $\mathcal{C}$ (lemme des classes monotones), c'est à dire la tribu borélienne de $\mathbb{R}$. On a ainsi montré la première affirmation du théorème.

On a aussi par construction

$$\int_{\mathbb{R}^2} \mathbf{1}_\mathrm{A}(x)\mathbf{1}_{y\le q}\, d\mu(x,y) = \int_\mathbb{R} 1_\mathrm{A}(x)\Big(\int_\mathbb{R} \mathbf{1}_{y\le q}\, d\mu_x(y)\Big)\, d\nu(x)$$

puisque

$$\int_\mathbb{R} \mathbf{1}_{y\le q}\, d\mu_x(y) = \mu_x((-\infty,q]) = g_q(x),$$

et on passe au cas général par classes monotones à nouveau, en considérant la classe $\mathcal{M}$ des boréliens B de $\mathbb{R}^2$ tels que

$$\int_{\mathbb{R}^2} \mathbf{1}_\mathrm{B}(x,y)\, d\mu(x,y) = \int_\mathbb{R}\Big(\int_\mathbb{R} \mathbf{1}_\mathrm{B}(x,y)\, d\mu_x(y)\Big)\, d\nu(x)$$

qui contient donc les produits $\mathrm{A} \times (-\infty, q]$ par la formule précédente. On passe ensuite aux fonctions étagées, puis mesurables par les techniques standard.

La démonstration du résultat d'unicité est laissée au lecteur, s'il en reste un.

Probabilités conditionnelles

On se donne un couple (X,Y) de v.a. réelles et on désigne sa loi jointe par μ. La loi de X est l'image ν de μ par l'application p_1. Les μ_x sont les *lois conditionnelles de* Y *sachant que* $\mathrm{X} = x$. Pour toute fonction borélienne bornée $f(x,y)$, on obtient un représentant de l'espérance conditionnelle de la variable aléatoire $\mathrm{Z} = f(\mathrm{X},\mathrm{Y})$ sur la tribu $\mathcal{A} = \sigma(\mathrm{X})$ engendrée par X, qui se met sous la forme

$$\mathrm{E}\big(f(\mathrm{X},\mathrm{Y}) \,|\, \mathcal{A}\big) = g(\mathrm{X}),$$

où

$$g(x) = \int_\mathbb{R} f(x,y)\, d\mu_x(y).$$

Du point de vue probabiliste, les deux "exemples évidents" donnés après le théorème correspondent aux deux cas (extrêmes) suivants : dans le cas **1**, on a deux v.a. indépendantes X,Y de lois γ_1 et γ_2, et l'indépendance dit que l'information $\mathrm{X} = x$ ne change pas notre connaissance de la loi de Y, la loi conditionnelle μ_x est simplement la loi γ_2 de Y ; le cas **2** est celui où X est de loi γ et où $\mathrm{Y} = \mathrm{X}$: dans ce cas évidemment, la loi de Y sachant que $\mathrm{X} = x$ est δ_x !

Exercices tordus, avec de la dénombrabilité

1. On désigne par Ω le premier ordinal non dénombrable et on suppose que $(\mathrm{U}_\alpha)_{\alpha<\Omega}$ est une famille croissante d'ouverts de $\mathbb{R}^2$, indexée par les ordinaux $\alpha < \Omega$ (c'est à dire indexée par la famille des ordinaux dénombrables). Montrer qu'il existe un ordinal $\alpha_0 < \Omega$ tel que $\mathrm{U}_\beta = \mathrm{U}_{\alpha_0}$ pour tout $\beta \ge \alpha_0$ (la famille est stationnaire).

Indication. Soit $(\mathrm{V}_m)_{m\in\mathbb{N}}$ une base de la topologie de $\mathbb{R}^2$. Pour chaque m fixé, considérer la famille des ordinaux α tels que $\mathrm{V}_m \subset \mathrm{U}_\alpha$; si elle n'est pas vide, elle possède un plus petit élément β_m. Montrer que $\alpha_0 = \sup_m \beta_m$ convient.

2. Montrer que tout espace **topologique** compact séparable K a au plus la puissance $2^{\mathbf{c}}$ (celle de $\mathcal{P}(\mathcal{P}(\mathbb{N}))$; la notation $\mathbf{c}$ désigne ici la puissance du continu, celle de $\mathcal{P}(\mathbb{N})$).

Indication. Soit $(x_n)_{n\in\mathbb{N}}$ une suite dense dans K ; pour chaque $\mathrm{M} \subset \mathbb{N}$ désignons par $\mathrm{K_M}$ l'adhérence dans K de l'ensemble $(x_m)_{m\in\mathrm{M}}$, puis pour chaque $\mathrm{A} \subset \mathcal{P}(\mathbb{N})$ désignons par $\mathrm{K^A}$ l'intersection $\bigcap_{\mathrm{M}\in\mathrm{A}} \mathrm{K_M}$. Enfin désignons par $\mathcal{F}$ le sous-ensemble de $\mathcal{P}(\mathcal{P}(\mathbb{N}))$ formé des A tels que $\mathrm{K^A}$ soit un singleton, et posons $\mathrm{K^A} = \{\varphi(\mathrm{A})\}$ pour tout $\mathrm{A} \in \mathcal{F}$. Montrer que φ est surjective de $\mathcal{F}$ sur K. Conclure.

Remarque : si K est compact **métrisable** il a au plus la puissance du continu.