J.-P. BENZÉCRI

Dichotomie aléatoire et séparation linéaire

Les cahiers de l'analyse des données, tome 7, nº 4 (1982), p. 485-498

<http://www.numdam.org/item?id=CAD_1982__7_4_485_0>

Les Cahiers de l'Analyse des Données
Vol. VII - 1982 - *n°* 4 - *p.* 485-498

DICHOTOMIE ALÉATOIRE ET SÉPARATION LINÉAIRE
[SEP. ALEAT.]

par J.-P. Benzécri (1)

Dichotomie aléatoire et séparation linéaire

On démontre ici d'après T.M. Cover un résultat de géométrie intégrale cité dans un précédent article consacré à l'analyse discriminante (cf. [*SEP. CORR.*] § 7 *Cahiers* Vol II, n° 4, pp 392 sqq.).

Le nombre des régions délimitées par p cloisons hyperplanes est inférieur ou égal à 2^p, et peut être strictement inférieur à ce nombre si la dimension de l'espace est inférieure à p : par exemple, 3 plans déterminent dans l'espace tridimensionnel $2^3 = 8$ régions, mais 3 droites, non concourantes, déterminent 7 régions dans le plan. De même étant donné dans le plan un ensemble de 4 points non trois à trois alignés, il n'y a que 7 des 8 partitions de cet ensemble en deux classes (dont l'une est éventuellement vide) qu'on puisse réaliser par une cloison rectiligne. Il se trouve que le nombre 7 trouvé dans les deux cas résulte d'une même formule d'analyse combinatoire (cf. *infra* prop. 1, 1", 1').

Au § 1, on redémontre cette formule d'après un mémoire de T. M. Cover (1965 : ce mémoire nous a été signalé par Ch. Masson) ; pour l'analyse combinatoire, Cover lui-même renvoie à Schäfli (1950) mais à ce propos il développe des thèses de reconnaissance des formes, qui si elles sont contestables, sont néanmoins intéressantes. Nous en traiterons au § 2.

1 *Une formule de géométrie combinatoire*

Définition 1 : Soit E un espace vectoriel de dimension d ; A une partie finie de E ; A sera dite en situation générale vectorielle dans E si toute partie B de A, engendre un sous-espace vectoriel L(B) de E dont la dimension est : inf{Card B , d}.

On a les propriétés suivantes, dont, par la suite, nous n'utiliserons que la 3° et la 4°.

---

(1) Professeur de statistique. Université Pierre et Marie Curie.

Propriété 1°. Soit E^N l'espace vectoriel, (de dimension N.d), des suites de N vecteurs de E. A toute suite $s \in E^N$, correspond une partie finie, $\Phi(s) \subset E$, l'ensemble des éléments de la suite. L'ensemble des points $s \in E^N$, tels que $\Phi(s)$ soit un ensemble de N vecteurs (distincts) en situation générale est un ouvert partout dense de E^N ; le complémentaire de cet ouvert est réunion finie de cônes algébriques de E^N ; (chacun de ces cônes est défini par la condition qu'une certaine sous-suite de p vecteurs de rangs donnés de s, engendre un sous-espace de E de dimension strictement inférieure, simultanément, à p et à d). C'est parce qu'il lui correspond un ouvert partout dense que la situation définie ci-dessus, peut être dite générale.

Propriété 2°. Si A est en situation générale dans E, il en est de même de toute partie de A.

Propriété 3°. Soit A en situation générale dans E ; $B \subset A$; l'image de A-B dans le quotient E/L(B) est en situation générale.

Propriété 4°. Si A n'est pas en situation générale dans E, on peut trouver dans A un système B de $p \leq d$ vecteurs engendrant un sous-espace de dimension p-1 ; dans le quotient de E par le sous-espace (droite) engendré par un vecteur non nul b de B, l'image de B-{b}, (donc de A-{b}), n'est pas en situation générale.

Propriété 5°. Soit μ_E^* une mesure de probabilité sur E, N entier > 0; la probabilité qu'une suite s de N vecteurs aléatoires indépendants de loi μ_E^* ait pour image par Φ (cf. 1°) un système de N vecteurs en situation générale est égale à 1 si et seulement si il n'existe pas de sous-espace vectoriel de E de dimension $p = \inf(N,d) - 1$, dont la masse pour μ soit strictement positive.

En effet, d'abord, s'il existe un sous-espace L de dimension p de mesure $m > 0$, il y aura une probabilité $m^{p+1} \neq 0$, que les $p+1$ premiers vecteurs de la suite s soient dans L, ce qui n'est pas une situation générale.

Plus généralement, $\Phi(s)$ n'est pas un système de N vecteurs en situation générale si et seulement si il existe une sous-suite de (q+1) vecteurs de s $(q \leq p)$:

$$s(i_1), s(i_2), \ldots, s(i_{q+1})$$

dont les q premiers vecteurs engendrent un sous-espace vectoriel contenant le dernier : c'est ce que nous appellerons le q-événement $[i_1, \ldots, i_q, i_{q+1}]$. L'ensemble des sous-suites de la suite $[1, \ldots, N]$ étant fini, il nous suffit, pour achever de démontrer 4°, d'établir que si un q-événement à une probabilité non nulle il existe un sous-espace de dimension q de mesure non nulle.

En effet si le q-événement $[i_1, \dots, i_q, i_{q+1}]$ a une probabilité non nulle, la probabilité conditionnelle que cet événement se réalise pour des valeurs données $s(i_1), \dots, s(i_q)$ des q premiers vecteurs de la suite s ne sera pas identiquement nulle : or cette probabilité conditionnelle n'est autre que la mesure pour μ du sous-espace vectoriel engendré par les q premiers vecteurs. (Sur E^{q+1}, muni de la probabilité μ^{q+1}, la fonction caractéristique du q-événement a une intégrale partielle par rapport à la dernière variable identiquement nulle, si il n'y a pas de sous-espace de dimension q de mesure non nulle ; donc, sous cette dernière hypothèse, d'après le théorème de Fubini, la probabilité du q-événement est nulle).

La propriété 5° (presque partout probabiliste), comme la propriété 1° (presque partout topologique) fait voir que la situation définie ci-dessus est générale.

Définition 2 : Soit A une partie finie d'un espace vectoriel E, on note :

$$\mathcal{S}(A) = \{B \mid B \subset A; \exists\ l \in E^*,\ \forall\ a \in A: (l(a) \neq 0) \wedge (l(a) > 0 \Leftrightarrow a \in B)\}$$

$\mathcal{S}(A)$ est appelé l'ensemble des parties séparables de A : B est séparable si il existe une forme linéaire homogène l sur E, positive strictement sur B et négative strictement sur A - B ; B est séparé de A - B par le sous-espace vectoriel de codimension 1, d'équation $l(x) = 0$.

Remarque 1 : $\mathcal{S}(A)$ est un nombre pair, car si $B \in \mathcal{S}(A)$, $(A - B) \in \mathcal{S}(A)$: le quotient $\mathcal{S}(A)/2$ est le nombre des partitions de A en deux classes qu'on peut réaliser par une cloison hyperplane passant par l'origine de E.

Nous démontrerons par récurrence la *Proposition 1* : Si A est en situation générale dans E, le cardinal de $\mathcal{S}(A)$ est une fonction déterminée $\mathcal{S}(N;d)$ du cardinal N de A et de la dimension d de E ; Si A n'est pas en situation générale ; le cardinal de $\mathcal{S}(A)$ est strictement inférieur à $\mathcal{S}(N;d)$. Avant de démontrer la proposition 1 et de calculer la fonction $\mathcal{S}(N;d)$ rapprochons du problème étudié ici, des problèmes géométriques analogues.

Remarque 2 Soit A un système fini de N vecteurs dans E^*. On a $B \subset \mathcal{S}(A)$, si et seulement si il existe un cône $\mathcal{C}_B$ ouvert convexe non vide, (composante connexe du complémentaire dans E de la famille des hyperplans $a(x) = 0$, $a \in A$), sur lequel on ait :

$$\forall\ x \in \mathcal{C}_B,\ \forall\ b \in B: b(x) > 0\ ;$$

$$\forall\ x \in \mathcal{C}_B,\ \forall\ a \in A-B: a(x) < 0\ .$$

La fonction $\mathcal{S}(N;d)$ donne donc le nombre de régions délimitées dans E, de dimension d, par un système de N hyperplans (sous-espaces vectoriels de codimension 1) en situation générale.

Après la situation duale, passons en géométrie affine. Nous poserons les définitions suivantes :

Définitions 1' : A , partie finie, sera dite en situation générale affine dans un espace vectoriel E (ou plus généralement un espace affin) de dimension d , si toute partie B de A engendre un sous-espace affin de dimension :

$$\inf(\overline{\overline{B}} - 1,\ d).$$

(Rappelons qu'un sous-espace affin, ou linéaire, d'un espace vectoriel est un translaté d'un sous-espace vectoriel).

Définition 2' : Une partie affinement séparable b de A sera une partie telle qu'il existe une forme linéaire sur E , non nécessairement homogène, qui soit positive sur B , négative sur A - B.

Plongeons E comme un hyperplan ne passant pas par l'origine d'un espace vectoriel E' de dimension d + 1 : A sera en situation générale vectorielle dans E' si et seulement si il est en situation générale affine dans E ; B sera affinement séparable dans E, si et seulement si elle est séparable (vectoriellement) dans E'. On a donc :

Proposition 1' : Soit A un ensemble de N points en situation générale affine dans un espace E de dimension d. Le nombre des parties B de A, affinement séparables (séparables par une cloison hyperplane quelconque) est égal à $\mathcal{S}(N; d+1)$.

On peut encore considérer les régions déterminées par des hyperplans affins.

Définition 1" : Un ensemble fini A d'hyperplans d'un espace E de dimension d , est dit en situation générale affine si les hyperplans éléments d'une partie finie quelconque de A, de cardinal h, ont pour intersection un sous-espace affin de dimension sup(d - h, - 1). (Selon l'usage on pose qu'un point est un sous-espace affin de dimension 0 , et que ∅ est de dimension -1).

En géométrie affine, la dualité entre points et hyperplans n'est pas parfaite : en effet la géométrie affine, n'est autre que la géométrie projective privée d'un hyperplan unique (l'hyperplan à l'infini). C'est pourquoi la proposition 1" ci-dessous diffère de la proposition 1'...

Proposition 1" : Soit A un système de N hyperplans (affins) en situation générale affine dans E , de dimension d. Le complémentaire de la réunion dans E des N cloisons, éléments de A , est formé de composantes ouverts connexes non vides dont le nombre est :

$$1/2 \,.\, \mathcal{S}(N + 1\ ;\ d + 1)$$

En effet, plongeons E dans E' comme un hyperplan ne passant pas par l'origine ; et soit E_0 le sous-espace vectoriel parallèle à E . Les N sous-espaces vectoriels de codimension 1 de E', définis chacun par l'origine de E' et une des cloisons de E, forment avec E_0 un système de N + 1 hyperplans en situation vectorielle générale dans E'. Ces sous-espaces divisent donc E' en $\mathcal{S}(N + 1,\ d + 1)$ régions coniques. Parmi ces régions, $\mathcal{S}(N + 1,\ d + 1)/2$ sont situées du même côté de E_0 que E (il y a, en effet, symétrie par rapport à l'origine de E') ; elles correspondent biunivoquement aux composantes déterminées dans E par le système des cloisons.

On va maintenant démontrer la proposition 1 par réccurrence sur N. Auparavant on peut traiter deux cas particuliers.

Lemme 1 : La proposition 1 est vraie pour tout couple (1,d) (d ≥ 1), et l'on a $\mathcal{S}(1,d) = 2$. (Quand A comprend un seul vecteur non nul), $\mathcal{S}(A) = \{A, \emptyset\}$; *cf. Déf.2 p.487*).

Lemme 2 : La proposition 1 est vraie de tout couple (N,1) (N ⩾ 1) et l'on a $\mathcal{S}(N;1) = 2$ (Quand E est une droite, $\mathcal{S}(A)$ comprend deux parties, dont l'une est éventuellement vide, qui sont les intersections de A avec les deux demi-droites que détermine sur E l'origine).

Lemme 3 : Soit N un entier (≥ 1) tel que la proposition 1 soit vraie de tout couple (N,d) (d ≥ 1). Alors la proposition 1 est vraie de tout couple (N + 1, d) (d ≥ 1) et l'on a, quel que soit d ≥ 2 :

$$\mathcal{S}(N+1\,;d) = \mathcal{S}(N\,;d) + \mathcal{S}(N\,;d-1).$$

Soit A un système de N + 1 vecteurs en situation vectorielle générale dans E de dimension d ≥ 2. Posons $A = A_0 \cup \{a\}$, où $\operatorname{Card} A_0 = N$. On va déterminer $\mathcal{S}(A)$ à partir de $\mathcal{S}(A_0)$. D'abord si $B_0 = \mathcal{S}(A_0)$, il est clair que l'une au moins des deux parties B_0 et $B = B_0 \cup \{a\}$ appartient à $\mathcal{S}(A)$. De même si $B \in \mathcal{S}(A)$, $B \cap A_0$ appartient à $\mathcal{S}(A_0)$. Le cardinal de $\mathcal{S}(A)$ sera donc égal au cardinal $\mathcal{S}(N,d)$ de $\mathcal{S}(A_0)$, augmenté du cardinal de l'ensemble $\mathcal{S}_a(A_0)$.

$$\mathcal{S}_a(A_0) = \{B_0 \mid B_0 \subset A_0\,;\ B_0 \in \mathcal{S}(A)\,;\ B_0 \cup \{a\} \in \mathcal{S}(A)\}$$

Le lemme 3 et donc la proposition 1 seront établis quand on aura montré que le cardinal de $\mathcal{S}_a(A_0)$ est $\mathcal{S}(N;d-1)$. Cela résultera du lemme 4.

Lemme 4 : Reprenons les notations du lemme 3. $\mathcal{S}_a(A_0)$ est l'ensemble des parties $B_0 \subset A_0$, telles qu'il existe une forme linéaire $l \in E^*$, positive sur B_0, nulle sur a, négative sur $A - B_0 - \{a\} = A - B$.

En effet, s'il existe une telle forme l, on pourra en la modifiant arbitrairement peu par une forme εu (ε réel, $u \in E^*$, $u(a) > 0$), séparer soit B soit B_0. Réciproquement, si B et B_0 appartiennent à $\mathcal{S}(A)$, il existe u (resp. u_0) $\in E^*$ positive sur B (resp. B_0) et négative sur $A - B$ (resp. $A - B_0$) : la forme linéaire combinaison à coefficients positifs de u et u_0.

$$-\,u_0(a)\,u + u(a)\,u_0$$

sera une forme b convenable.

Maintenant notons E' le quotient de E par le sous-espace (droite) engendré par a ; soit A'_0 l'image de A_0 dans ce quotient. Il résulte du lemme 4 que $\mathcal{S}_a(A_0)$ n'est autre que l'ensemble $\mathcal{S}(A'_0)$ des parties de A'_0 séparables dans E' ; d'après déf. 1, 3°, A'_0 est en situation générale dans E' ; d'après l'hypothèse de récurrence $\mathcal{S}(N,d-1)$ est donc bien le cardinal de $\mathcal{S}_a(A_0)$.

Pour achever de démontrer la proposition 1, il reste à considérer le cas d'un système A de vecteurs qui ne sont pas en situation générale. Alors, (si A ne contient pas le vecteur nul, cas trivial) on peut trouver dans A un vecteur a tel que, (nous reprenons les notations ci-dessus), A'_0 ne soit pas en situation générale dans E', d'où résultera selon l'hypothèse de récurrence que $\mathcal{S}_0(A_0)$ a un cardinal strictement inférieur à $\mathcal{S}(N;d-1)$. D'où pour $\mathcal{S}(A)$ un cardinal strictement inférieur à $\mathcal{S}(N+1\ ;d)$.

Les lemmes 1, 2, 3 permettent de calculer la fonction $\mathcal{S}$ pour tout couple (N,d) d'entiers supérieurs ou égaux à 1 : par un procédé tout analogue au calcul du triangle de Pascal, on trouve le tableau 1 ; afin que l'équation du lemme 3 soit vraie pour tout (N,d) où N≥2, mais d de signe quelconque, on a complété le tableau par des zéros , on a fait de même dans le tableau 2 pour le coefficient binomial $\binom{N}{d} = C_N^d = N!/(d!\times(N-d)!)$. Ce parallélisme fournit une expression de $\mathcal{S}(N;d)$, par une méthode qui rappelle l'intégration des équations différentielles linéaires.

N ↑

0	0	0	2	12	32	52	62	64
0	0	0	2	10	22	30	32	32
0	0	0	2	8	14	16	16	16
0	0	0	2	6	8	8	8	8
0	0	0	2	4	4	4	4	4
0	0	0	2	2	2	2	2	2

→ d

Tableau 1 : La fonction $\mathcal{S}(N;d)$; $(N \geq 1)$

N ↑

0	0	1	5	10	10	5	1	0
0	0	1	4	6	4	1	0	0
0	0	1	3	3	1	0	0	0
0	0	1	2	1	0	0	0	0
0	0	1	1	0	0	0	0	0
0	0	1	0	0	0	0	0	0

→ d

Tableau 2 : La fonction $\binom{N}{d} = \frac{N!}{d!(N-d)!}$; $(N \geq 0)$

On a le lemme suivant :

Lemme 5 : Soit Φ une fonction de deux variables entières N et d définies pour $N \geq N_0$ et d quelconque dans Z (entier algébrique). Supposons que pour $N \geq N_0$, $d \in Z$, Φ satisfait à l'équation du lemme 3 :

$$\Phi(N+1,\, d) = \Phi(N,d) + \Phi(N,d-1) \;;$$

Alors quels que soient $N \geq N_0$, $d \in Z$, Φ est égale à la fonction ψ :

$$\psi(N,d) = \Sigma\{\Phi(N_0\,, d_0) \cdot \begin{pmatrix} N - N_0 \\ d - d_0 \end{pmatrix} \mid d_0 \in Z\}$$

Remarquons d'abord que la série donnant Ψ est sommable car le coefficient binomial est non nul si et seulement si d_0 est dans l'intervalle fini où est satisfaite l'inégalité :

$$0 \leq d - d_0 \leq N - N_0$$

De plus la fonction Ψ, somme de translatés de la fonction $\binom{N}{d}$, satisfait comme celle-ci à l'équation, (invariante par translation), du lemme 3.

Enfin Ψ et Φ coïncident pour $N = N_0$, $d \in Z$. Or l'équation du lemme 3 permet de déterminer une fonction pour $N \geq N_0$ si on connaît ses valeurs pour $N = N_0$; d'où le lemme.

Des lemmes 1, 3, 5 on déduit l'expression suivante de $\mathcal{S}$:

$$\mathcal{S}(N;d) = \Sigma\{\mathcal{S}(1;d_0) \cdot \begin{pmatrix} N-1 \\ d_0 - d \end{pmatrix} \mid d_0 \in Z\}$$

ou, en éliminant les termes sûrement nuls :

$$\mathcal{S}(N;d) = 2\,\Sigma\,\{\begin{pmatrix} N-1 \\ d' \end{pmatrix} \mid d' = 0, \ldots, d-1\}$$

En particulier, comme on peut le vérifier sur le tableau 1, on a pour $d \geq N$, $\mathcal{S}(N;d) = 2^N$.

2 *Applications à la reconnaissance des formes*

Nous énumérerons d'abord les applications citées par Cover.

2.1 Théorème de Wendel : Soit E espace vectoriel de dimension d ; $N \geq 1$; μ une loi de probabilité sur E^N, telle que :

a) la mesure μ sur E^N soit invariante par le groupe G des 2^N transformations ε de la forme, $s \rightarrow \varepsilon s$:

$$(s(1), \ldots, s(i), \ldots, s(N)) \rightarrow (\varepsilon_1 s(1), \ldots, \varepsilon_i s(i), \ldots, \varepsilon_N s(N))$$

où chacun des ε_i vaut 1 ou -1.

b) l'ouvert des suites de N vecteurs en situation vectorielle générale est de mesure 1 pour μ (une suite tirée au hasard suivant la loi μ est en situation générale avec une probabilité 1).

Alors la probabilité qu'une suite tirée suivant μ soit contenue dans un demi-espace de E (limité à un hyperplan passant par l'origine) est

$$(1/2)^N \mathcal{S}(N;d)$$

La démonstration utilise le lemme suivant :

Lemme 6 : Soit s une suite en situation générale, $s \in E^N$; parmi les 2^n transformés de s par G, il y en a $\mathcal{S}(N,d)$ qui sont contenus dans un demi-espace.

En effet soit $\varepsilon \in G$; εs est dans un demi-espace si et seulement si est séparable dans E (cf. définition 2) la partie de la suite s (plus exactement, cf. définition 1, 5°, l'ensemble associé à la suite) formée des vecteurs s(i) tels que $\varepsilon_i > 0$.

Considérons maintenant sur $E^N \times G$ la mesure Π produit de μ par la probabilité uniforme sur G (chaque ε a une masse $1/2^N$). D'après l'hypothèse a) l'image de Π par l'application

$$(s,\varepsilon) \rightarrow \varepsilon s$$

de $E^N \times G$ dans E^N n'est autre que μ. La probabilité cherchée est donc égale à la mesure pour Π de l'ensemble des couples (s,ε) tels que εs soit dans un hémisphère. Intégrons la fonction caractéristique de cet ensemble suivant le théorème de Fubuni : l'intégration partielle par rapport à ε donne une fonction de s qui d'après le lemme 6 vaut $(1/2)^N \mathcal{S}(N;d)$ en tout s en situation générale. D'après l'hypothèse b, la fonction a cette valeur μ presque partout ; d'où le théorème.

Le théorème de Wendel s'applique en particulier si μ est la puissance $N^{\text{ème}}$ d'une mesure de probabilité symétrique μ_E sur E, satisfaisant à la condition (fort peu restrictive) donnée en définition 1, 5°; par exemple si μ_E est la densité uniforme sur la sphère de centre l'origine et de rayon s (pour une métrique euclidienne mise sur E).

2.2 *Dichotomie aléatoire* : Considérons plus généralement une loi μ_E sur l'espace vectoriel E de dimension d, μ_E satisfaisant à la condition de la définition 1, 5° (pour qu'un échantillon soit presque sûrement en situation générale). A partir de la mesure μ, Cover définit ainsi les dichotomies aléatoires : on tire au sort suivant la loi μ_E une suite infinie dénombrable s, (s = s(1), s(2),...), de points indépendants ; et à chaque point s(i) de la suite, on attribue, indépendamment une étiquette de classe qui peut être équiprobablement 0 ou 1 (que la probabilité d'appartenir à l'une ou l'autre classe soit égale à 1/2 est essentiel pour la suite). Ainsi une dichotomie aléatoire associée à la loi μ_E est un couple d'une suite $s \in E^N$ (N ensemble des entiers strictement positifs) et d'une partie B de N (e.g. la première des deux classes) ; B peut encore être considéré comme une suite de chiffres binaires b_i ($b_i = 1$ si s(i) est étiqueté de première classe, 0 sinon) et donc comme le développement binaire d'un nombre réel, 0, b_1 b_2 ,..., qui appartient au segment (0,1). On écrira donc :

$$(s,B) \in E^N \times (0,1)$$

La suite s a pour loi de probabilité la loi μ puissance dénombrable de μ_E. Sous l'hypothèse faite d'équiprobabilité de l'appartenance d'un point à chacune des deux classes, B est distribué uniformément sur (0,1) : i.e. la loi de B est la mesure de Lebesgue λ , du segment (0,1). Et la dichotomie aléatoire (s,B) est distribuée suivant la loi produit $\mu \times \lambda$. Les probabilités que nous désirons calculer sont d'ailleurs celles d'événements dont la réalisation ne dépend que de la dichotomie tronquée (s_P , B_P), où P est un entier positif:

$$s_P = \{s(1), \ldots, s(P)\} \quad ; \; B_P = \{b_1, \ldots, b_P\} \in (0,1)^P \; ;$$

la loi du couple (s_P , B_P) est le produit de la puissance P-ème de μ_E par la puissance P-ème de la mesure uniforme (masse 1/2 pour chaque point) sur $\{0,1\}$.

Une dichotomie aléatoire (S,B) est dite séparable jusqu'au rang P au moins, si la sous-suite

$$\{s(i) \mid i \in B \; ; \; 1 \leq i \leq P\}$$

est une partie de s_P. La probabilité qu'une dichotomie soit séparable jusqu'au rang P est :

$$(1/2)^P \; \mathcal{J}'(P \; ; d) \; ;$$

la démonstration est analogue à celle donnée en 2.1 : en bref, s_P étant donné il y a $\mathcal{J}(P;d)$ des B_P sur les 2^P possibles, qui définissent des dichotomies séparables ; le fait que les $B_P^{\bullet}$ aient tous même probabilité, (cf. *supra* étiquetage équiprobable des éléments de la suite) joue ici un rôle essentiel.

Une dichotomie séparable jusqu'au rang P , mais non jusqu'au rang P + 1 sera dite de rang P. La probabilité qu'une dichotomie soit de rang P est :

$$(1/2)^P \mathcal{J}(P;d) - (1/2)^{P+1} \mathcal{J}(P+1;d) = (1/2)^P \binom{P-1}{d-1} .$$

Comme l'espérance mathématique du rang P de la dichotomie aléatoire (s,B) (distribuée comme on l'a dit ci-dessus suivant la loi $\mu \times \lambda$) est $\cdot$ 2 d ; Cover propose de dire que 2 d est la capacité de séparation associée à une représentation d-dimensionnelle des objets à ranger en deux classes. Quel que soit l'intérêt des recherches géométriques de Cover, il est clair que cette définition est peu réaliste, car elle ne tient auncun compte du fait qu'une dichotomie à laquelle on soit fondé à s'intéresser, est tout autre qu'aléatoire : Arkadiew et Bravermann marquent bien ce point quand ils parlent de "régularité du système des domaines" (cf. [*DIALOGUE FORMES*] *in Cah.* Vol VI n° 2 pp 157-174).

2.3 <u>Apprentissage d'une dichotomie</u> : Tirons au hasard comme des événements indépendants une suite de N+1 vecteurs de E (loi μ) et une partie B de $\{1,\ldots, N\}$ (l'ensemble des 2^N parties est muni de la loi uniforme). Restreignons-nous au cas où est séparable la partie de la suite des N premiers vecteurs, formée de ceux de ces vecteurs dont le rang est dans B ; en bref nous dirons que B est séparable dans la suite de longueur N. Le (N + 1)-ème vecteur sera dit ambigu si B et B∪{N}, sont tous deux séparables dans la suite des longueurs N + 1. La probabilité de cet événement est (cf. lemme 3, 4) :

$$\mathcal{J}(N , d - 1)/ \mathcal{J}(N , d) ;$$

Selon Cover cela correspond au cas où un apprentissage ayant été fait sur la suite des N premiers vecteurs par une machine qui a placé une cloison de séparation (pour séparer B), la classe à laquelle cette machine affectera le (N + 1)-ème point dépend de la cloison particulière mise en place, et non pas seulement de la suite des N + 1 vecteurs et de la partie B.

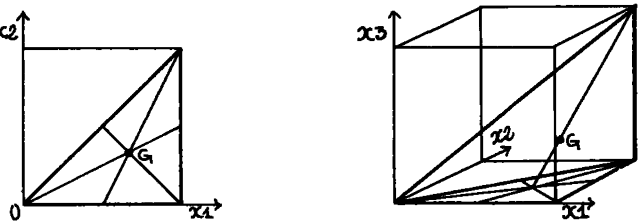
Cependant, la probabilité qu'un objet soit bien classé après un apprentissage portant sur N objets, dépend manifestement de la forme des classes et de leur disposition relative. A ce propos examinons un problème.

Problème 1 : Soit μ une loi de probabilité sur un espace affin F ; quelle est la distribution de la variable aléatoire "étendue d'un échantillon d'effectif N" (i.e. intégrale de la loi μ étendue à l'enveloppe convexe d'un échantillon de N points indépendants tirés suivant cette loi)?

Il s'agit, en bref, de déterminer avec quelle précision un échantillon d'effectif N nous donne le champ de variation d'un point aléatoire... Sur ce problème, l'étude bibliographique de Moran (1966 ; cette étude nous a été signalée par Ch. Masson) donne deux références : Renyi &Sulanka (1964) et Efron (1965). Dans l'article de Renyi & Sulanka ; il s'agit exclusivement d'une loi μ dont la densité est constante sur une partie convexe compacte K du plan, et nulle en dehors de K. Si K a pour frontière une ligne dont la courbure est bornée inférieurement et supérieurement, l'espérance mathématique de l'étendue d'un échantillon d'effectif N est asymptotiquement égale à $1-(s(K)/(N^{2/3}))$ où s(K) est une constante caractéristique du convexe ; si K est un carré, on a la formule, bien différente, $1 - (8LN/3N)$. Efron traite d'une mesure μ quelconque sur le plan ou dans l'espace F de dimension 3 ; et il donne dans ce cas des formules complètes. Mais il ne nous paraît pas simple de déterminer, d' après les intégrales d'Efron, comment tend asymptotiquement vers 1 l'espérance mathématique de l'étendue d'un échantillon dont l'effectif N tend vers l'infini. (en particulier : pour quelles distributions l'étendue tend-elle le moins vite vers 1 ?)

Le cas particulier où F est une droite est très simple : si μ a une densité, on peut, par un changement de paramètre monotone, se ramener au cas où cette densité est uniforme sur un segment et nulle ailleurs (en effet, en dimension 1, l'enveloppe convexe d'un échantillon de N points répartis uniformément sur (0,1) : l'espérance mathématique de cette étendue est $1-(2/(N + 1))$.

[Rappelons brièvement comment on procède dans l'étude de la loi uniforme sur (0,1) (cf. P. Jouve, inédit 1966). Soit $(a1,..., aN)$ un échantillon de N nombres obtenus par tirages successifs indépendants suivant cette loi ; convenons de désigner ces nombres par $(x1,...,xN)$ de telle sorte que dans leur ensemble les xi soient les mêmes nombres que les ai, mais les xi étant rangés dans l'ordre décroissant : $xN < xN-1 < \cdots < x2 < x1$. Par exemple si N = 3, x1 est le plus grand des ai, x3 le plus petit et x2 le médian. Alors le point de coordonnée $(x1,..., xN)$ est distribué de façon uniforme dans un simplexe, qu'il nous suffira de dessiner dans deux cas : N = 2 (triangle) et N = 3 (tétraèdre). Il est clair que le centre de gravité du simplexe a pour coordonnées les valeurs moyennes de $x1, \ldots, xN$; et, en particulier la différence entre sa 1-ère et sa N-ème coordonnée n'est autre que l'espérance mathématique de l'étendue, qui vaut donc bien $1-(2/(N+1))$].



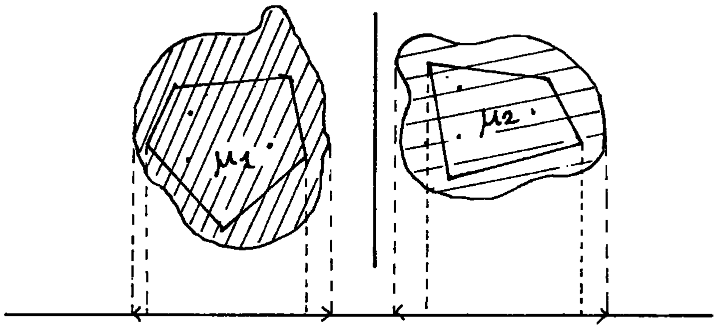
A titre d'exercice considérons encore une mesure plane dont le support est formé de p rayons d'un même cercle, chaque rayon portant une masse $(1/p)$ répartie uniformément sur sa longueur ; approximativement, sur N points d'un échantillon, chaque rayon en portera N/p ; et la partie laissée libre à l'extrémité d'un rayon vaudra, en moyenne, environ $\frac{p}{N}$ fois la masse du rayon ; d'où pour l'étendue d'un échantillon d'effectif N (N grand) en moyenne, environ, $1-(p/N)$.

Précisons maintenant le rapport entre le problème 1, et la discrimination des formes. Soient d'abord deux formes, représentées par deux mesures de probabilité μ_1, μ_2 dont les supports sont des intervalles non empiétants de la droite. Tirons au sort N objets de chaque forme : on aura ainsi sur la droite deux échantillons entre lesquels est un intervalle, (où tombera la cloison) qui contient une masse m_1 de la loi μ_1 et une masse m_2 de la loi de μ_2 : on vient de voir que :

$$E(m_1) = E(m_2) = 1/(N+1).$$

En un certain sens, $1/(N+1)$ est un ordre de grandeur de la probabilité de l'erreur, après un apprentissage fait sur deux échantillons d'effectif N. Mais il s'agit plutôt d'une borne supérieure : si entre les supports de μ_1 et de μ_2 se trouve un large vide ; la probabilité que la cloison tombe dans ce vide sera élevée, même si les supports de μ_1 et μ_2 sont mal connus d'après les échantillons présentés durant l'apprentissage. Pratiquement les objets sont décrits non par un point d'une droite, mais par un point d'un espace F de dimension élevée ; or le problème 1 n'est quelque peu étudié qu'en dimension 2 ou 3. Mais dans le cas d'une dichotomie (cas où il n'y a que deux formes à séparer par un hyperplan), on peut admettre que la direction de la cloison étant à peu près déterminée, on retrouve un problème unidimensionnel, relatif aux projections sur la direction perpendiculaire à la cloison (cf. Figure).

Et, à propos des cloisons hyperplanes, se pose un nouveau problème de géométrie statistique, que nous nous bornerons à énoncer :



Problème 1' : Soit μ une loi de probabilité sur un espace affin F ; ε, un échantillon d'effectif N de cette loi ; appelons "étendue de l'onglet maximum relatif à ε ", le maximum de l'intégrale de μ étendue à un demi-espace ne contenant pas de point de ε. Quelle est (en fonction de N) la distribution de la variable aléatoire "étendue de l'onglet maximum relatif à un échantillon d'effectif N".

(L'énoncé de ce problème nous a été suggéré par J. Robert).

2.4 *Points extrémaux :* Soit A une partie finie d'un espace vectoriel E : un point a est dit point extrémal de A (ou de l'enveloppe convexe de A) si a n'appartient pas à l'enveloppe convexe de $A-\{a\}$; il revient au même de dire que $\{a\}$ et $A-\{a\}$ sont affinement séparables. Plus généralement étant donnée une partie séparable $B \subset A$, on va définir l'ensemble des points extrémaux de B et de $C = A - B$ relativement à la dichotomie (B,C) : si $B = A$, (donc $C \neq \emptyset$), on retrouvera les points extrémaux de A. Dans le travail de Cover, il s'agit de séparation vectorielle (cf. déf. 2) ; mais nous préférons traiter ici de la séparation affine (cf. déf. 2'), nous bornant à signaler quelques difficultés de détail, propres à la séparation vectorielle. Il sera commode de noter $\mathcal{S}_A(A)$ l'ensemble des parties affinement séparables de A ; et E^*_A l'espace vectoriel des fonctions linéaires, non nécessairement homogènes sur E.

Soit $A \subset E$, A fini ; $B \in \mathcal{S}_A(A)$; $C = A - B$. Notons

$$B_1 = \{b \mid b \in B \ ; \ C \cup \{b\} \in \mathcal{S}_A(A)\} \ ;$$

$$C_1 = \{c \mid c \in C \ ; \ B \cup \{c\} \in \mathcal{S}_A(A)\} \ ;$$

nous dirons $B_1 \cup C_1$ est l'ensemble des points extrémaux relativement à la dichotomie affinement séparable (B,C) si et seulement si elle sépare B_1 de C_1. Plus précisément définissons un ensemble P de paires de parties :

$$P=\{(B',C') \mid B'\subset B; C'\subset C; \forall\ l\in E^*_A : (lB>0 \wedge lC<0) \Leftrightarrow (lB'>0 \wedge lC'<0)\} \ ,$$

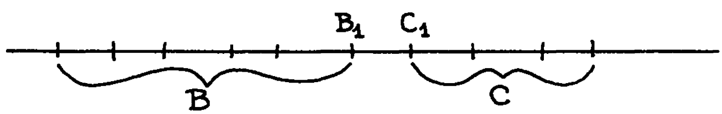
(où $lB > 0$) signifie : l est strictement positif sur B). On va montrer que la paire (B_1, C_1) est la paire minimale de P en ce sens qu'on a :

$$P = \{(B', C') \mid B_1 \subset B' \subset B \ ; \ C_1 \subset C' \subset C\}.$$

D'abord, il est clair que toute paire $(B', C') \in P$, est telle que $B_1 \subset B'$, $C_1 \subset C'$: car s'il existait une paire $(B',C') \in P$, et un $b \in B_1$

tels que $b \notin B'$, il existerait, d'après la définition de B_1, une forme linéaire (non homogène) l, $(l \in E^*_A)$, qui soit strictement positive sur $B - \{b\}$ (donc sur B'), et strictement négative sur C (donc sur C') et aussi en b ; mais alors (B', C') ne serait pas dans P.

Reste à montrer que $(B_1, C_1) \in P$. Il est commode de procéder par récurrence sur la dimension d de E. Si $d = 1$, et que B et C sont non vides séparables, B_1 et C_1 comprennent chacun un seul point ; (si $C = \emptyset$, $B \neq \emptyset$, B comprend deux points!) ; la situation est triviale



comme on peut le voir sur la figure.

Supposons maintenant la démonstration faite jusqu'au rang $d - 1$; pour passer au rang d nous utiliserons seulement ce corollaire très simple de l'hypothèse de récurrence : soit B, C affinement séparables dans E de dimension $d - 1$; l'ensemble $B_1 \cup C_1$, (des points extrémaux relativement à la dichotomie (B, C), est non vide. Nous procéderons par l'absurde.

Supposons qu'il existe $\ell_1 \in E^*_A$, telle que :

$$\ell_1 . B_1 > 0 \quad ; \quad \ell_1 . C_1 < 0 \; ;$$

mais que, cependant ℓ_1 soit négative en certains points de B, ou positive en certains points de C, ou nulle en certains points de $A = B \cup C$: nous dirons qu'en ces points ℓ_1 a des anomalies relativement à la dichotomie (B, C). Nous allons modifier ℓ_1 afin qu'il n'ait plus qu'une seule anomalie, un zéro.

Soit $l \in E^*_A$, telle que $lB > 0$, $lC < 0$: on peut trouver $k \geq 0$ tel que $\ell'_1 = \ell_1 + k$ n'ait d'autre anomalie que d'être nulle en un ou plusieurs points de $A - (B_1 \cup C_1)$. On peut supposer que la forme ℓ'_1 n'est pas identiquement nulle (ℓ'_1 ne pourrait être nulle que si, $B_1 \cup C_1$ étant vide, ℓ était le produit de ℓ_1 par une constante négative ; mais même alors, on pourrait modifier légèrement ℓ_1). L'hyperplan E' où s'annule ℓ_1, contient les points anormaux : ces points se répartissent en deux ensembles $B' \subset B - B_1$, et $C' \subset C - C_1$; et l'on a par hypothèse $b' \cup C' \neq \emptyset$. D'après l'hypothèse de récurrence (et les lemmes 3 et 4) on peut trouver une forme ℓ' linéaire non homogène sur E', positive ou nulle sur B', négative ou nulle sur C', et nulle en un seul point de $B' \cup C'$; ℓ' se prolonge en une forme linéaire $\ell'' \in E^*_A$; pour ε suffisamment petit positif la forme $\ell''_1 = \ell'_1 + \varepsilon''$ n'a qu'une seule anomalie ; un zéro, (le point de $B' \cup C'$ où ℓ' est nulle). Mais cela est absurde, car, (cf. lemmes 3 et 4), une forme qui n'a qu'une seule anomalie doit l'avoir sur $B_1 \cup C_1$. Nous avons ainsi montré que $(B_1, C_1) \in P$.

Un mot des difficultés propres à la séparation vectorielle. On doit supposer : dimension de $E \geq 2$, et A en situation générale. Quand E est une droite, B_1 et C_1 (définis comme ci-dessus, mais avec $\mathcal{S}(A)$ au lieu de $\mathcal{S}_A(A)$) sont vides. Quand A n'est pas en situation générale, $B_1 \cup C_1$ peut encore être vide : c'est le cas si $C = \emptyset$ et $B = A$ se compose des quatre points du plan : (1,0), (2,0), (0,1), (0,2).

Enonçons maintenant une remarque due à Cover (en la transposant toutefois de la séparation vectorielle affine). Soit A, de cardinal N, en situation générale dans E de dimension d ; chaque point $a \in A$ appartient à l'ensemble extrémal de $2\,\mathcal{S}(N-1\,;d)$ dichotomies séparables de A (cf. lemme 4 et propriété 1') : l'effectif moyen des ensembles extrémaux des $\mathcal{S}(N\,;d+1)$ dichotomies affinement séparables sur A est donc :

$$\frac{2N\,\mathcal{S}(N-1\,;d)}{\mathcal{S}(N\,;d)}$$

Pour terminer, signalons un problème.

<u>Problème 2</u> : Soit A de cardinal N en situation générale affine dans E, de dimension d. Supposons connu $\mathcal{S}_A(A) \subset \mathfrak{P}(A)$. Peut-on, pour d donné et N assez grand retrouver presque surement la figure A avec une précision arbitrairement grande (mais à une transformation projective près) ?

C'est un problème de rigidité analogue à celui qu'après R.N. Shepard nous avons étudié dans l'analyse des proximités. Ici, la connaissance de toutes les dichotomies séparables par des cloisons hyperplanes nous donne approximativement les relations d'alignement entre points de A ; mais, il faut noter que les transformations projectives, qui respectent l'alignement, ne respectent pas les distances, ni même, (au sens usuel du terme), la forme d'une figure (en particulier il existe des convexes dont les transformés projectifs approchent arbitrairement un convexe quelconque !) ; "rigidité projective" n'est donc pas rigidité au sens usuel du terme !