

Recherches

Décomposition de lois, Fonctions Caractéristiques, et Caractérisation.

Probability Laws Decomposition, Characteristic Function and Characterization of Laws

par Serge PROSPERI

THOMSON CSF, Division RCC
160, bd de Valmy F-92704 Colombes cedex

Résumé

Après avoir présenté les résultats nécessaires d'analyse complexe et les propriétés de base des fonctions caractéristiques, on introduit les notions de stabilité et de décomposabilité, qui sont utilisées pour caractériser certaines classes de lois, en particulier les lois de Gauss et de Poisson.

Comme application de ces notions, on cite l'étude des problèmes de convergence de lois ainsi que les problèmes d'identification, et l'exemple significatif que constitue l'Analyse en Composantes Indépendantes.

Mots clés : Analyse Complexe, Caractérisations, Convergence, Décompositions, Fonctions Caractéristiques, Fonctions Entières, Identification, Lois de probabilité.

Abstract

After a short presentation of complex analysis results, we give an insight into the basic properties of characteristic functions. We then introduce the stability and decomposability properties, which are used to characterize several classes of probability laws, among which the normal and Poisson laws.

As first applications, one can consider the limit theorems and the identification problems, with the lightening example of Independent Component Analysis.

Key words : Complex Analysis, Characteristic Functions, Characterization, Convergence, Decomposition, Identification, Probability laws.

1. Introduction

L'objet de cet article est de présenter dans une forme cohérente certains résultats relatifs à la décomposition des lois et aux sommes finies de variables aléatoires indépendantes. Outre leur intérêt dans la recherche de lois limites, ces résultats voient se développer actuellement de nouvelles applications en traitement du signal, notamment dans l'utilisation grandissante des statistiques d'ordre supérieur, et les traitements en environnement non gaussien.

Il est bien connu que les fonctions caractéristiques permettent de caractériser l'indépendance des variables aléatoires. Il existe de fait une relation surprenante au premier abord entre la nature des lois et l'indépendance des variables, où les fonctions caractéristiques jouent un rôle central. Nous aurons l'occasion au cours de cet exposé de revenir sur certaines caractéristiques très particulières de la loi normale, qui fait l'objet entre autres des théorèmes de Cramer et Darmois que nous verrons au paragraphe 3.

Les fonctions caractéristiques, introduites en 1937 par Paul Lévy [1], constituent depuis lors un des outils les plus féconds en théorie

des probabilités. Leur usage concerne aussi bien les problèmes de convergence que l'étude des combinaisons finies de variables qui nous intéresse ici. En fait, elles ne sont autres que les transformées de Fourier des lois associées. Il s'agit là d'un outil familier en traitement du signal, qui prend un intérêt particulier dans le cas où la fonction de départ est réelle positive, spécialement dans le cadre de fonctions de la variable complexe qui sera développé ici.

Le plan suivi est le suivant :

- 1) Notations
- 2) Fonctions caractéristiques, propriétés élémentaires et utilisation de la variable complexe
- 3) Décomposition des lois de Probabilité et caractérisations
- 4) Applications. Cas de l'Analyse en Composantes Indépendantes.
- 5) Annexes
 - A1. Rappels de topologie générale
 - A2. Rappels d'analyse complexe et résultats sur les fonctions entières

A3. Démonstration des théorèmes sur la stabilité des décompositions

Dans un souci de clarté de l'exposé, on insistera sur les principales propriétés élémentaires des fonctions caractéristiques. Les résultats concernant la caractérisation des lois sont désormais classiques mais restent peu connus par les non spécialistes et sont rarement cités dans les ouvrages généraux de Probabilités. Ainsi, on présente au paragraphe 3 différentes caractérisations des lois de Gauss et de Poisson. Ces résultats ne prétendent pas à l'exhaustivité, le dénominateur commun étant la référence aux décompositions de lois. Le lecteur intéressé pourra, pour des propriétés plus élémentaires des lois classiques, se référer par exemple à [7]. Pour ce qui est des démonstrations, nous ne renvoyons que dans un nombre limité de cas aux références, les autres étant exposées dans le texte de l'article. Cela est cohérent avec notre souhait d'un traitement autonome. On notera que certaines diffèrent de façon sensible de celles présentées dans la littérature. C'est que ces dernières nous ont paru trop liées à la construction d'un ouvrage et de ce fait difficiles à extraire de leur contexte, ou de nature trop technique. Nous avons essayé, à travers l'exposition de ces résultats, et de leur enchaînement, de souligner les mécanismes mis en jeu. En outre, le paragraphe 4 constitue autant une tentative d'éclairage nouveau des sujets présentés qu'une application des résultats contenus dans l'article.

C'est également ce même souci qui justifie les rappels faits en annexe, concernant les notions de base d'analyse complexe et de topologie.

Le théorème de Hadamard qui, on le verra intervient de manière centrale, permet d'étudier l'ordre des fonctions analytiques, et de fournir des critères pour la normalité des lois (théorèmes de Cramer et Marcinkiewicz). En fait, les corollaires 1 à 3 du théorème sont suffisants pour établir la plupart des résultats.

La métrique de Lévy sera introduite au paragraphe 2 pour établir les propriétés de stabilité des décompositions. Ces résultats sont indépendants de ceux concernant les caractérisations, mais ils nous ont semblé entrer naturellement dans le cadre de l'exposé, car ils permettent de compléter la distinction entre le cas des sommes finies et infinies de variables aléatoires d'une part, et d'autre part interviennent dans certaines généralisations des théorèmes limites, esquissées au paragraphe 4. La démonstration du théorème 3.7, qui est le principal résultat de stabilité nécessite l'emploi de plusieurs lemmes, mais nous la proposons toutefois en annexe car elle est significative de la nature des raisonnements qui entrent en jeu dans l'étude de ce type de problèmes.

Pour finir, on présente une application prometteuse dans différents domaines du traitement du signal, aussi bien pour des problèmes d'identification que de déconvolution de signaux. Le point de départ de la méthode est un prolongement remarquable du théorème de Darmais, appliqué à la théorie des systèmes linéaires.

2. Notations

On désigne par \mathbf{R} le corps des nombres réels, par \mathbf{C} le corps des nombres complexes, et \mathcal{Q} celui des nombres rationnels. On conviendra de noter z^* le nombre complexe conjugué de z . On utilisera la notation $\log(x)$ pour désigner le logarithme népérien de x , et $Arg(z)$ désigne un représentant modulo 2π de l'argument de z . On notera $Re(z)$ et $Im(z)$ les parties réelle et imaginaire de z .

Pour la convergence, l'écriture $f_n \rightarrow f$ signifiera que la suite f_n converge vers f quand n tend vers l'infini. Si f_n est une suite de fonctions, il s'agira de la convergence simple sauf mention contraire.

Pour les développements en série, on utilise la notation de Landau, dans laquelle $o(x)$ désigne une fonction négligeable devant x ($\frac{o(x)}{x} \rightarrow 0$), et $O(x)$ une fonction comparable à x .

On notera également $f_n \sim g_n$ (f_n équivalent à g_n) si la relation $\frac{f_n}{g_n} \rightarrow 1$ est satisfaite. Une fonction p fois continuellement dérivable sera dite de classe C^p .

Dans tout le texte, on notera les fonctions caractéristiques en minuscule (f, g) et les lois en majuscule (F, G). La densité associée à F , si elle existe, sera simplement notée dF . Etant donnée une variable aléatoire X , on notera f_X et Ψ_X les première et seconde caractéristiques de X .

On conviendra enfin de noter C_n^p le nombre de combinaisons de p éléments parmi n , et δ_{ij} l'indice de Kronecker.

3. Fonctions caractéristiques

On suppose connue la théorie élémentaire des probabilités et la notion de loi, terme que l'on emploiera également par abus de langage pour désigner les fonctions de répartition.

Nous limiterons l'exposé au cas des variables aléatoires réelles. L'extension au cas de vecteurs aléatoires ne présente pas de difficultés, mais alourdirait inutilement l'exposé; en outre, il s'agira d'une extension immédiate en ce qui concerne les propriétés des lois marginales.

On rappelle la caractérisation des fonctions de répartition (fonctions croissantes de \mathbf{R} dans $[0, 1]$, continues à droite, et telles que $F(-\infty) = 0$ et $F(+\infty) = 1$). En outre, une fonction de répartition, comme toute fonction croissante de \mathbf{R} , a une limite à gauche en tout point.

On remarque en premier lieu que l'ensemble des points de discontinuité de F est au plus dénombrable, ce qui se vérifie en considérant la suite croissante d'ensembles

$$D_n = \{x \in D, d(x) \geq 1/n\}, \text{ avec } d(x) = F(x) - F(x^-)$$

L'ensemble D_n a au plus n éléments, donc $Card(D_n) \leq n$, et $Card(D) \leq card(\mathbf{N})$.

On dit que la loi est discrète si F est constante, sauf sur un ensemble au plus dénombrable (les « atomes » de la loi de probabilité, où elle se trouve concentrée), et est continue si elle n'a pas d'atomes. Plus généralement, les atomes sont les points de discontinuité de F .

Toute loi est mélange d'une loi continue et d'une loi discrète (formule de décomposition de Lebesgue) :

$$F = p_1 F_1 + p_2 F_2, \text{ avec } F_1 \text{ discrète, } F_2 \text{ continue, et } p_1 \geq 0, p_2 \geq 0, p_1 + p_2 = 1$$

Une loi sera absolument continue si elle est dérivable par rapport à la mesure de Lebesgue sur \mathbf{R} , ou simplement si F est dérivable.

Pour toute loi F , on définit sa fonction caractéristique f comme sa transformée de Fourier-Stieltjes

$$f(x) = \int e^{itx} dF(t) \quad (1)$$

qui est toujours définie sur \mathbf{R} . Si f ne s'annule pas, on définit la seconde caractéristique Ψ comme le logarithme de f (ou plus exactement comme sa détermination principale contenant 0).

Exemples :

1/ La fonction caractéristique d'une loi normale, de densité

$$\frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \text{ est } f(x) = e^{i\mu x - \sigma^2 x^2/2}.$$

2/ La fonction caractéristique d'une loi discrète, $\sum p_k u(x - a_k)$, où u est la fonction échelon, est $\sum p_k e^{ia_k x}$. En particulier, dans le cas d'une loi de Poisson, pour laquelle les points de discontinuité a_k sont les entiers positifs, et les probabilités sont $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, la fonction caractéristique est

$$f(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{ikx} = e^{\lambda(e^{ix} - 1)}$$

On cite des propriétés élémentaires des fonctions caractéristiques, puis les résultats sur la convergence faible.

Pour toute fonction caractéristique f , on peut établir sans difficulté les relations suivantes

- i) $f(0) = 1$
- ii) $|f| \leq 1$
- iii) f est uniformément continue
- iv) $f(-x) = f(x)^*$
- v) Si f est la fonction caractéristique de la variable X , celle de

$$Y = aX + b \text{ est } g(x) = e^{ibx} f(ax)$$

En outre, si une fonction caractéristique est de la forme $f = e^P$, où P est un polynôme de degré 2, c'est la fonction caractéristique

d'une loi gaussienne. En effet, si $P = R + iQ$, d'après iv) R est pair et Q impair, et d'après i) $P(0) = 0$, donc $P(x) = ax^2 + ibx$, et comme $R \leq 0$, $a = -\frac{1}{2\sigma^2} \leq 0$.

Théorème 2.1 : toute fonction caractéristique absolument continue est négligeable à l'infini.

Démonstration

$f(x) = \int e^{itx} dF(t)$, et lorsque x tend vers l'infini les variations de F sont lentes devant celles de l'exponentielle complexe. Plus précisément, dF est Riemann-intégrable dans tout compact, et par suite limite (par valeurs supérieures) de fonctions continues, qui vérifient la propriété, qui sera vraie d'après le théorème de convergence de Lebesgue. En effet, si $g = dF$ est continue, soit $c > 0$ arbitraire, il existe un nombre positif y tel que

$$\int_{-y}^y g(x) dx > 1 - \frac{c}{2}, \text{ et en posant } a_n = -y + 2ny/N,$$

$$\begin{aligned} \int_{-y}^y e^{iux} g(x) dx &= \sum_{n=0}^N \int_{a_n}^{a_{n+1}} e^{iux} g(x) dx \\ &= \sum_{n=0}^N \int_{a_n}^{a_{n+1}} e^{iux} (g(a_n) + [g(x) - g(a_n)]) dx \end{aligned}$$

g étant uniformément continue sur le compact $[-y, y]$, pour tout nombre réel positif c_1 il existe n_0 tel que pour $n \geq n_0$,

$$|x - t| \leq 2y/n \Rightarrow |f(x) - f(t)| \leq c_1;$$

en outre, g est bornée sur $[-y, y]$ ($|g(x)| \leq M$).

Par suite, en posant $N = n_0$,

$$\begin{aligned} \left| \int_{-y}^y e^{iux} g(x) dx \right| &\leq 2y c_1 + \sum_{n=0}^N \frac{|e^{i\frac{u}{N}} - 1|}{u} |g(a_n)| \\ &\leq 2y c_1 + 2M(N+1)/|u| \end{aligned}$$

On choisit alors $c_1 = \frac{c}{8y}$, et donc pour $|u| \geq \frac{c}{8M(N+1)}$

$$\left| \int e^{iux} g(x) dx \right| < c \quad \square$$

Le produit de fonctions caractéristiques est une fonction caractéristique, associée au produit de convolution des lois.

Étant données deux lois F_1 et F_2 associées à f_1 et f_2 , on définit le produit de convolution $F = F_1 * F_2$ par

$$F(x) = \int F_1(x-u) dF_2(u) \quad (2)$$

on en déduit

$$\begin{aligned} f(t) &= \int e^{-itx} dF(x) = \iint e^{-itx} dF_1(x-u) dF_2(u) \\ &= \int \int e^{-it(x-u)} dF_1(x-u) e^{-itu} dF_2(u) \end{aligned}$$

et donc

$$f(t) = f_1(t) f_2(t)$$

Théorème 2.2 : f désignant une fonction caractéristique, et F la loi associée, les trois énoncés suivants sont équivalents :

1) les moments m_j de F existent jusqu'à l'ordre n pour n pair, et $n - 1$ pour n impair.

2) f est n fois dérivable sur \mathbf{R} .

3) f a un développement limité à l'ordre n en 0.

En outre, le coefficient de x^j dans le développement de f est alors $f^{(j)}(0) = i^j \frac{m_j}{j!}$ (cf. [4] pour la démonstration).

Corollaire : Si $f(t) = 1 + e(t)$, où $e(t)$ est négligeable devant t^2 au voisinage de 0, $f = 1$.

Démonstration

C'est une conséquence directe du théorème 2.1, car la moyenne et la variance doivent être nulles, ce qui implique que la loi associée est dégénérée en 0, d'où $f = 1$ \square .

En particulier, une fonction caractéristique ne peut être de la forme e^g , où g est une fonction entière (ou un polynôme) de valuation (degré du terme de plus bas degré) strictement plus grande que deux, ce qui est également une conséquence du théorème de Marcinkiewicz.

Formule d'inversion : Pour tout couple (a, b) de nombres réels

$$F(b) - F(a) = \lim_{x \rightarrow \infty} \frac{1}{2\pi} \int_{-x}^x \frac{e^{iya} - e^{iyb}}{iy} f(y) dy \quad (3)$$

Cette relation n'est autre que la transformée de Fourier inverse pour l'intégrale de Stieltjes. Elle pourrait d'ailleurs s'obtenir en intégrant la formule inverse sur les densités, donc pour l'intégrale de Lebesgue, dans le cas où f est absolument continue.

En effet, en désignant par δ la distribution de Dirac,

$$dF(y) = \int_u \delta(u - y) dF(u) = \int_u \int_x \frac{1}{2\pi} e^{ix(u-y)} dx dF(u)$$

Et le problème réside alors dans le changement de l'ordre d'intégration. En fait, en langage des distributions, si TF désigne la transformation de Fourier

$$dF(y) = \langle \delta_y, dF \rangle = \langle TF \frac{1}{2\pi} e^{-ixy}, dF \rangle = \frac{1}{2\pi} \langle e^{-ixy}, TF dF \rangle,$$

ce qui peut encore s'écrire

$$dF(y) = \frac{1}{2\pi} \int e^{-ixy} f(x) dx \quad (4)$$

PROPRIÉTÉS DES LOIS DISCRÈTES

Une loi discrète est dite régulière si ses points de discontinuité sont équidistants.

Proposition 2.1 : f est la fonction caractéristique d'une loi discrète régulière si et seulement si il existe une valeur x_0 non nulle telle que $|f(x_0)| = 1$.

Démonstration

En effet, si F est discrète régulière, $f(x) = \sum p_k e^{ia_k x}$ avec $a_k = a_0 + kd$, et donc $|f(\frac{2\pi}{d})| = 1$.

Réciproquement, si $|f(x_0)| = 1$, $f(x_0) = e^{ix_0 \xi_0} = \int e^{itx_0} dF(t)$ par suite

$$e^{ix_0 \xi_0} \int [1 - e^{ix_0(t-\xi_0)}] dF(t) = 0$$

et donc

$$\int [1 - \cos(x_0(t - \xi_0))] dF(t) = 0$$

La fonction sous l'intégrale est positive, et la loi est discrète, avec des points de discontinuité vérifiant

$$t_k = \xi_0 + \frac{k2\pi}{x_0} \quad \square$$

Corollaire : Si $|f| \neq 1$, il existe au plus une famille dénombrable de points vérifiant

$$|f(x)| = 1$$

En effet, d'après la démonstration précédente, si deux tels points x_0 et x_1 (non nuls) existent, et si la loi a au moins deux points de discontinuité, il existe des entiers k_1, k_2, l_1, l_2 vérifiant $k_1/x_0 - k_2/x_0 = l_1/x_1 - l_2/x_1$, et le rapport x_1/x_0 sera un nombre rationnel. \square

CONVERGENCE EN LOI ET THÉORÈMES DE HELLY [3].

Une suite de mesures de probabilité F_n converge faiblement vers une loi F si la convergence a lieu en tout point de continuité de F . Une suite de variables aléatoires converge en loi si les lois associées convergent faiblement.

Le *premier théorème de Helly* affirme que toute suite de fonctions croissantes uniformément bornées admet une sous suite convergente sur \mathbf{R} par densité.

On indexe alors $\mathcal{Q} = (r_i)_{i \in \mathbf{R}}$, et on construit une séquence de sous suites imbriquées F_{in} qui convergent aux points r_i . Selon le principe diagonal classique, on choisit alors la suite F_{nn} qui converge alors en tout point de \mathcal{Q} vers une fonction F , qui est alors prolongée à \mathbf{R} , en posant $F(x) = \text{Sup}\{F(r), r \in \mathcal{Q} \text{ et } r < x\}$. F est bien croissante et bornée sur \mathcal{Q} par passage à la limite, et par extension sur \mathbf{R} . \square

Dans le cas de lois, il faudra vérifier que la limite est continue à droite.

Pour que la suite F_n converge en loi, il est nécessaire et suffisant que les fonctions caractéristiques associées convergent simplement vers une fonction continue en 0 - il s'agit alors d'une fonction caractéristique et la convergence est uniforme sur tout compact (Théorème de Paul Lévy) - ou que pour toute fonction continue bornée g , on ait (deuxième théorème de Helly)

$$\int g(x) dF_n(x) \text{ converge vers } \int g(x) dF(x)$$

Dans la suite, la métrique associée à la norme sup pour les fonctions caractéristiques est notée ρ . Dans l'étude de la convergence des lois, on utilise également la métrique de Lévy définie par :

$$L(G, H) = \inf\{h \geq 0, \text{ pour tout } x\} \\ H(x-h) - h \leq G(x) \leq H(x+h) + h \quad (5)$$

Il est clair qu'elle est dominée par ρ . Pour tout couple de lois F et G ,

$$L(F, G) \leq \rho(F, G) \quad (6)$$

Si G est absolument continue, M étant la borne supérieure de dG ,

$$\rho(F, G) \leq (1 + M)L(F, G) \quad (7)$$

Dans ce cas, les théorèmes de convergence établis pour l'une des deux métriques seront valables pour l'autre. La convergence suivant ces métriques est évidemment plus forte que pour la convergence faible. Certains auteurs [5] emploient pourtant le terme de convergence faible pour la métrique de Lévy. Il s'agit en fait d'une convergence uniforme faible. Dans la représentation géométrique des fonctions F et G , la métrique de Lévy définit une distance non pas selon l'axe vertical, mais selon la seconde diagonale (avec un coefficient $\frac{1}{\sqrt{2}}$ car la définition (2.5) considère la projection sur l'axe horizontal).

Fonctions caractéristiques de la variable complexe

Jusqu'à présent, on a considéré le seul cas réel. On peut définir une fonction caractéristique sur \mathcal{C} , par prolongement analytique. Si on s'intéresse seulement au cas de variables aléatoires réelles, on crée ainsi un outil puissant d'investigation. Mais pour ce qui concerne les décompositions de lois, qui font intervenir la somme des variables, toutes les identités pourront se décomposer suivant les parties réelles et imaginaires. Dans ce qui suit, on suppose donc que les variables aléatoires sont à valeurs réelles.

On note pour commencer une propriété élémentaire mais d'un grand intérêt pratique. C'est en effet cette propriété des fonctions caractéristiques qui va nous permettre, en l'appliquant à un polynôme, d'établir le théorème de Marcinkievicz.

Proposition 2.3 : Pour tout nombre complexe z tel que $f(z)$ est définie, $|f(z)| \leq f(i \operatorname{Im}(z))$.

En effet, si on pose :

$$z = x - iy, f(z) = \int e^{itz} dF(t) = \int e^{itx} e^{-ty} dF(t)$$

On peut en tirer une conséquence qui sera également utile dans la suite.

Proposition 2.4 : Le maximum de $|f|$ dans un disque $|z| \leq r$ est atteint en l'un des points $-ir, ir$.

4. Décomposition des lois

L'étude de la décomposition des fonctions caractéristiques est liée à celle de la somme des variables aléatoires indépendantes. En effet, la loi d'une somme de variables indépendantes est le produit de

convolution des termes de la somme, et sa fonction caractéristique est le produit des fonctions caractéristiques. L'étude des convergences des sommes de variables indépendantes se fait, depuis les travaux de Paul Lévy, à partir de leurs fonctions caractéristiques. Notre propos est ici, en langage des variables aléatoires, d'étudier le lien entre la loi d'une somme finie et les lois de ses termes. La loi normale joue un rôle central dans tous les domaines du traitement du signal. Chacun sait que la somme de variables aléatoires gaussiennes est gaussienne. Bien que cela s'exprime en des termes tout aussi simples, le fait qu'une somme finie de variables indépendantes ne peut être gaussienne que si tous ses termes le sont est moins souvent cité, et ne peut être établi qu'avec l'introduction des fonctions caractéristiques analytiques complexes. Ce résultat, connu sous le nom de théorème de Cramer, s'énoncera alors : une fonction caractéristique gaussienne ne peut avoir pour composantes que des fonctions caractéristiques gaussiennes, ou encore la classe des fonctions caractéristiques des lois normales est indivisible.

Une fonction caractéristique f est décomposable si elle s'écrit sous la forme $f = f_1 f_2$, où f_1 et f_2 sont des fonctions caractéristiques (composantes, ou facteurs de f), ou de façon équivalente si la loi associée est décomposable ($F = F1 * F2$).

On dit que f est infiniment divisible si, pour tout entier n , f est décomposable en un produit de n fonctions caractéristiques identiques ($f = f_n^n$).

Une fonction caractéristique peut être indécomposable, et dans le cas contraire, ses facteurs peuvent eux-mêmes être décomposables ou non. Une variable aléatoire dont la fonction caractéristique est indécomposable ne peut être la somme de variables indépendantes. On peut citer l'exemple des variables uniformes sur un compact. On peut montrer (théorème de Khintchine) que toute fonction caractéristique est le produit de fonctions indécomposables et de fonctions sans composantes indécomposables, lesquelles sont indéfiniment décomposables [5].

Stabilité des lois

On se propose d'étudier les relations existant entre certaines fonctions caractéristiques et leurs facteurs, en particulier quelles sont les classes de lois dont les facteurs appartiennent à la même classe.

On dit que deux lois sont équivalentes, et on note $F \sim G$ si elles sont déduites l'une de l'autre par translation, donc si leurs fonctions caractéristiques vérifient $f(x) = e^{iax}g(x)$ pour une constante a . Deux lois auront le même type si elles sont équivalentes à un facteur d'échelle près. Des variables ont le même type si elles se déduisent par transformation affine.

On dira enfin qu'une classe de lois est stable si toute somme finie d'éléments de la classe est du type d'un autre élément de la classe. En général, une classe contiendra tous les éléments d'un même type. Dans certains cas, elle peut être réduite à un seul type.

Les lois de Gauss et de Poisson sont stables, de même que les lois dont la fonction caractéristique est analytique. Nous nous intéressons dans ce qui suit au problème réciproque, c'est-à-dire

aux cas où les facteurs de la décomposition appartiennent à la classe de leur produit. On dira alors que la classe est invariante, ou indivisible.

On peut immédiatement remarquer que la classe des lois discrètes est invariante. Soit par ailleurs F_1 une loi continue et $F_2 = \sum p_k u(x - a_k)$ une loi discrète. $F(x) = (F_1 * F_2)(x) = \sum p_k F_1(x - a_k)$.

F est un mélange de lois continues, et est donc continue.

On commence par établir un résultat analogue pour la classe des fonctions caractéristiques analytiques, puis on établit les théorèmes de Marcinkiewicz et de Darmois. On montre ensuite que les lois de Gauss et de Poisson forment des classes indivisibles.

Décomposition des fonctions caractéristiques entières

Si f est une fonction caractéristique entière, son développement en série à l'origine aura pour coefficients

$$c_j = i^j \frac{m_j}{j!}$$

les m_j étant les moments de la loi associée (d'après le théorème 2.2).

Théorème 3.1 : Si f est une fonction caractéristique analytique décomposable du type $f = f_1 f_2$, ses facteurs sont également des fonctions caractéristiques analytiques sur le même domaine.

Démonstration

Soit $F = F_1 * F_2$ la loi d'une fonction caractéristique analytique dans la bande $-a < Im(z) < b$, et soit v un nombre réel compris entre $-a$ et b , on a

$$\infty > I = \int e^{vx} dF(x) = \int e^{vx} \int dF_2(x-t)$$

D'après le théorème de Fubini-Tonnelli sur l'inversion de l'ordre d'intégration, le second membre de l'égalité s'écrit aussi

$$\int \int e^{v(x-t)} dF_2(x-t) \cdot e^{vt} dF_1(t)$$

ou encore

$$\int e^{vt} dF_1(t) \cdot \int e^{vx} dF_2(x) (= I_1 \cdot I_2)$$

Les deux intégrales I_1 et I_2 sont finies. f_1 et f_2 sont donc uniformément bornées pour $|z| < v$, de même que les fonctions $\int i z e^{izt} dF_j(t)$, et par suite elles sont dérivables sous l'intégrale. D'après le théorème de Cauchy, ce sont des fonctions caractéristiques analytiques dans la bande $-a < Im(z) < b$ □

En outre, si f est entière, on aura alors d'après la proposition 2.3, $M(f, r)$ désignant le maximum de $|f|$ sur la boule de rayon r ,

$$M(f_1, r) \cdot M(f_2, r) \leq M(f, r)$$

Corollaire 3.1 : L'ordre d'un facteur ne peut donc excéder l'ordre de f .

Lemme 3.1 : Si

1. la fonction caractéristique f ne s'annule pas
2. il existe un nombre $\eta > 0$ tel que $\int e^{\eta^2 x^2} dF(x) < \infty$

la loi F est normale.

Cela montre qu'il ne peut y avoir de loi vérifiant 1. et dont la fonction caractéristique (ou la densité) converge plus vite à l'infini que la normale. L'hypothèse 1 implique entre autres que F a un spectre à bande infinie.

Cette propriété est liée à la concentration de la loi normale autour de sa moyenne, ce qui peut expliquer le rôle très particulier joué dans les problèmes de convergence, où la convolution des lois (ou le produit des fonctions caractéristiques) tend à en augmenter la concentration.

Démonstration

D'après la proposition 2.4,

$$|f(x)| \leq \int e^{-uIm(x)} dF(u) \leq e^{\frac{\eta^{-2}|x|^2}{4}} \cdot \int e^{\eta^2 u^2} dF(u)$$

car

$$-uIm(x) \leq \frac{\eta^{-2}Im(x)^2}{4} + \eta^2 u^2$$

Par suite, la dernière intégrale étant convergente, l'ordre de f est au plus égal à 2 par le corollaire 1 du théorème de décomposition de Hadamard, et donc la loi associée est normale. □

On peut également comparer la situation du lemme 3.1 au comportement des fonctions caractéristiques au voisinage de 0 (proposition 2.1).

Théorème 3.2 (Marcinkiewicz) : Si une fonction caractéristique s'écrit $f = e^P$, où P est un polynôme, alors soit P est de degré deux et la loi associée est Normale, soit P est de degré 1 et la loi est dégénérée.

On peut l'établir soit en revenant à la loi associée, et en se servant du lemme précédent, soit en utilisant directement la fonction caractéristique, ce que nous faisons.

Démonstration

Soit n le degré de P , on peut écrire $P(z) = a_n z^n + o(z^n)$ avec $a_n \neq 0$.

Soit $z = r e^{i\theta}$, alors

$$P(z) = |a_n| r^n e^{i(n\theta + Arg a_n)} + o(r^n)$$

On suppose que θ vérifie $Arg(a_n) + n\theta = 0[2\pi]$. D'après la proposition 2.3, $Re P(z) \leq |P(iIm(z))|$ donc

$$Re P(z) = |a_n| r^n + o(r^n) \leq |a_n| r^n \sin \theta^n + o(r^n)$$

donc $\sin \theta^n \geq 1 - o(1)$

Lorsqu'on fait tendre r vers l'infini, on obtient $\sin \theta = \pm 1$ et par suite $\theta = \frac{\pi}{2}$ ou $\theta = \frac{3\pi}{2}$.

Si on considère l'ensemble $\Theta = \{\theta, Arg(a_n) + n\theta = 0[2\pi]\}$, il est clair que Θ a n éléments.

Donc $n \leq 2$, et les considérations de symétrie permettent de conclure. \square

Remarque : Sous cette forme, on peut dégager un résultat plus général, car on a utilisé uniquement le fait que $f = goP$, où g est une fonction croissante sur \mathcal{R} , et vérifie $|g(z)| \leq g(|z|)$ à l'infini ce qui montre que l'on peut prendre aussi pour g un polynôme dont le coefficient directeur est positif, ou une composition d'exponentielles (comme dans [4]). Les démonstrations classiques du théorème de Marcinkiewicz sont calculatoires, et beaucoup plus longues de celle présentée ici. On peut observer l'importance de la proposition 2.3, d'apparence sybilline, mais qui fournit une caractérisation analytique des fonctions caractéristiques. Le résultat précédent se généralise au cas où g est une fonction entière quelconque, en utilisant cette fois le théorème de Hadamard dans sa forme générale.

Théorème 3.3 (Darmois) : Si deux combinaisons linéaires de variables aléatoires complexes indépendantes de la forme $z_1 = \sum a_i x_i$ et $z_2 = \sum b_i x_i$ sont elles-mêmes indépendantes, toutes les composantes x_i telles que $a_i b_i \neq 0$ sont soit gaussiennes soit dégénérées.

On peut l'énoncer sous une forme légèrement différente, en posant

$$z_1 = \sum a_i x_i + u \text{ et } z_2 = \sum b_i x_i + v,$$

où u et v sont indépendantes des x_i .

Lemme 3.2 : Soient (f_1, \dots, f_n) n fonctions continues dans un ouvert U de \mathcal{C} , et $(a_i, b_i, 1 \leq i \leq n, 2n$ nombres complexes vérifiant pour tout couple (u, v) d'éléments d'un ouvert $U_1 \subset U$,

$$g(u, v) = \sum_{i=1}^n f_i(w_i) = 0 \text{ (en posant } w_i = a_i u + b_i v) \quad (8)$$

Si les vecteurs $t_i = (a_i, b_i)$ sont tous non proportionnels, alors les fonctions f_i sont des polynômes de degré $p \leq n - 2$.

Démonstration

La démonstration générale se fait en utilisant les différences finies, mais elle est rigoureusement identique à celle que nous proposons ici, et qui utilise les dérivées successives des f_i . On suppose donc les fonctions f_i $n - 1$ fois dérivables.

On va montrer que pour tout indice j , $f_j^{n-1} = 0$. On peut sans restriction supposer $j = 1$. Il est équivalent que les vecteurs t_i d'une part, et w_i d'autre part soient non proportionnel. (t_{n-1}, t_n) forme un système libre, ce qui signifie que w_{n-1} et w_n sont deux variables indépendantes.

On peut donc en prenant pour nouvelles variables $u_1 = w_{n-1}$ et $v_1 = w_n$ et en dérivant par rapport à u_1 annuler le n -ième terme de la somme (8). Les w_i s'expriment alors en fonction des nouvelles variables (les nouvelles combinaisons étant encore non proportionnelles, puisqu'il s'agit d'une propriété des w_i). En répétant l'opération, on annulera successivement tous les termes jusqu'à l'ordre 1, et on obtient $A_1 f_1^{n-1}(u_{n-1}) = 0$, $A_1 (\neq 0)$ étant le produit des coefficients des w_k dans w_1 , d'où la conclusion. \square

Démonstration

Elle découle directement du lemme, avec l'écriture des fonctions caractéristiques associées aux variables. En effet, en notant f_z la f.c. associée à la variable Z , et Ψ_Z sa seconde caractéristique, c'est-à-dire la détermination principale de son logarithme, nulle en 0,

$$f_{Z_1, Z_2}(u, v) = f_{Z_1}(u) f_{Z_2}(v)$$

soit encore

$$\sum_{i=1}^n \Psi_{x_i}(a_i u + b_i v) = \sum_{i=1}^n \Psi_{x_i}(a_i u) + \sum_{i=1}^n \Psi_{x_i}(b_i v) \quad (9)$$

Les secondes caractéristiques des lois x_1 sont donc des polynômes, et par application du théorème de Marcinkiewicz ces lois sont gaussiennes. \square

Théorème 3.4 (Cramer) : Si une fonction caractéristique décomposable est associée à une loi de Gauss, chacune des composantes est une loi de Gauss.

(Si une somme finie de variables aléatoires indépendantes est Gaussienne, chacune des composantes est Gaussienne).

Démonstration

La conclusion est immédiate, car les composantes sont des fonctions entières, d'ordre inférieur ou égal à deux d'après le corollaire 3.1, et par suite (corollaire 1.2) les secondes caractéristiques sont des polynômes de degré deux, ce qui n'est possible que pour des lois de Gauss. \square

Pour le cas des lois de Poisson, on a une propriété analogue.

Théorème 3.5 (Raïkov) : Si une loi de Poisson de paramètre λ est décomposable, chaque facteur est également une loi de Poisson, et la somme des paramètres est égale à λ .

Démonstration

On suppose $f = f_1 f_2$, avec

$$f(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{ikx} = e^{\lambda(e^{ix} - 1)}$$

D'après la proposition 2.1, f_1 et f_2 sont discrètes uniformes, car si $|f(x_0)| = 1$, on aura également $|f_j(x_0)| = 1$, $j = 1, 2$. En outre, d'après le Théorème 3.1, f_1 et f_2 sont entières et ne s'annulent pas.

Par suite, $f_j(x) = \sum p_{jk} e^{ikx}$, et en posant, $y = e^{ix}$,

$$f_j(x) = g_j(y) = \sum p_{jk} y^k$$

comme $f(x) = g(y) = g_1(y)g_2(y)$, $e^{-\lambda} \frac{\lambda^k}{k!}$ est la convolution des suites p_{j1}, \dots, p_{jk} , $j = 1, 2$ formées de nombres positifs, et par suite, $p_{1k} \leq p_{20}^{-1} e^{-\lambda} \frac{\lambda^k}{k!}$

g_1 est donc une fonction entière sur \mathcal{C} , qui vérifie

$$p_{20} M(r, g_1) \leq M(r, g),$$

et l'ordre de g_1 est majoré par celui de g , donc d'après les corollaires 2 et 3 du théorème de Hadamard, $g_1(y) = e^{\lambda_1(y-y_1)}$, et de même $g_2(y) = e^{\lambda_2(y-y_2)}$.

De $g = g_1 g_2$, on déduit que $y_1 = y_2 = 1$ et $\lambda_1 + \lambda_2 = \lambda$ \square

Théorème 3.6 (Linnik) : Si un produit est un mélange de lois normales et de lois de Poisson, chacun des facteurs est de cette forme.

La démonstration est dérivée de la précédente, mais plus technique, et on renvoie à [5].

On peut généraliser ces théorèmes (Linnik), en considérant une généralisation de la notion de décomposition au cas où les composantes ne sont pas des fonctions caractéristiques, mais des puissances non entières de fonctions caractéristiques :

$$f(x) = f_1(x)^{\alpha_1} \dots f_n(x)^{\alpha_n}, \text{ avec } \alpha_i > 0.$$

On montre alors que si le produit est une fonction entière, les composantes sont encore des fonctions entières. La démonstration du théorème 3.1, fondée sur une majoration se transpose en effet directement.

On généralise alors le théorème de Cramer : si la loi du produit est normale, les composantes sont des fonctions entières d'ordre inférieur ou égal à deux, et ce sont donc des fonctions caractéristiques de lois gaussiennes, (cela montre entre autre que les exposants α_i sont égaux à 1). Il suffit en outre, d'après le principe de prolongement, que l'égalité soit vérifiée pour une infinité de points.

Stabilité des décompositions

Un théorème de décomposition vérifié pour une classe K de lois sera dit stable pour des métriques ρ_1 et ρ_2 données si un faible écart à la loi produit conduit également à de faibles écarts pour les facteurs, plus précisément si étant données une loi F de K et les décompositions

$$F_n = \prod_{j=1}^N F_{n_j}$$

$$\begin{aligned} F \in K \text{ et } \varepsilon_n = \rho_1(F_n, F) \rightarrow 0 \\ \Rightarrow \delta_n = \max_j \inf_{G \in K} \rho_2(F_{n_j}, G) \rightarrow 0 \end{aligned} \quad (10)$$

Les notations introduites ici seront reprises dans la suite. Le théorème suivant contribue à justifier l'importance de la métrique de Lévy.

Théorème 3.7 : Tout théorème de décomposition est stable pour la métrique de Lévy.

La démonstration est présentée en annexe.

Il existe des résultats plus forts concernant le cas des lois normales.

Théorème 3.8 (Sapogov [5]) : Soit ϕ la loi normale réduite, a_j et σ_j les deux premiers moments de F_j , et en posant

$$\phi_j = \phi\left(\frac{x - a_j}{\sigma_j}\right), \text{ et } \varepsilon = \rho(F, \phi) \quad (11)$$

alors

$$\rho(F_j, \phi_j) < \frac{C}{\sigma^3 \sqrt{-\ln(\varepsilon)}} \quad (12)$$

Théorème 3.9 : Avec les notations précédentes, et en considérant la métrique de Lévy,

$$L(F_j, \phi_j) < C (-\ln(\varepsilon))^{\frac{1}{8}} \quad \varepsilon = L(F, \phi) \quad (13)$$

L'utilité pratique de ce théorème est néanmoins discutable, étant donnée la très faible vitesse de convergence de la fonction dans le second membre.

Il existe des théorèmes analogues concernant les lois de Poisson.

On peut noter pour finir une autre propriété très remarquable de la loi de Gauss.

Théorème 3.10 (Réciproque forte du Théorème de Cramer) : Soit f une fonction caractéristique décomposable. Si tous ses facteurs sont du type de f , c'est-à-dire si tout facteur g de f s'écrit $g(x) = e^{i\mu x} f(ax)$, f est la fonction caractéristique d'une loi gaussienne.

5. Applications

UNE GÉNÉRALISATION DES THÉORÈMES CLASSIQUES DE CONVERGENCE.

On peut noter que dans des conditions très générales les lois stables d'une part et infiniment divisibles d'autre part sont les lois de sommes infinies de variables indépendantes (problème de la limite centrale).

En outre, l'étude de la stabilité fournit un premier résultat, avec le théorème 3.7.

On a donc une relation asymptotique entre les éléments de la série étudiée et les composantes de la limite F .

On peut alors généraliser les théorèmes concernant la convergence vers les lois de Gauss et de Poisson, en utilisant la convergence des éléments des séries étudiées.

Le théorème le plus classique est celui concernant des variables aléatoires indépendantes et de même loi, dont la somme converge en loi vers la normale, avec le facteur de normalisation $\sigma\sqrt{n}$. Le problème général de la limite centrale considère des sommes de variables indépendantes quelconques, avec des conditions de régularité des loi, et des coefficients de normalisation appropriés pour assurer la convergence. En particulier la condition de Lindeberg [3]. Ce qui permet alors d'assurer la convergence est la négligeabilité des termes élevés de la série normalisée (par exemple, pour le théorème classique, avec une décroissance en $\frac{1}{\sqrt{n}}$). L'utilisation des décompositions de lois permet de rendre non nécessaire cette condition, en la remplaçant par la convergence des éléments de la série vers des lois normales (la convergence vers 0 en est un cas particulier, dégénéré). En fait, cela peut aussi bien s'appliquer à la convergence vers les lois de Poisson.

Il est remarquable que si on considère les termes des séries normalisées (problème de la limite centrale) le comportement des sommes finies est radicalement opposé de celui des sommes infinies. Si on considère à nouveau le cas des lois gaussiennes, qui sont limites de sommes de variables indépendantes quelconques, d'après le théorème de Cramer elles ne peuvent être la somme d'un nombre fini de variables indépendantes non gaussiennes. Qui plus est, les théorèmes 3.8 et 3.9 fournissent une minoration de l'écart entre les sommes finies et une limite éventuelle. On peut voir ces majorations comme étant duales des majorations de l'écart des lois des sommes partielles.

En particulier (inégalité de Berry-Essen [3]), si F_n est la loi de la somme de n variables indépendantes de même loi F , et ρ_3 désigne le moment absolu d'ordre 3 de F ,

$$\rho(F_n, \phi) \leq 3 \frac{\rho_3}{\sigma^3 \sqrt{n}} \quad (14)$$

Il ne s'agit pas de la meilleure borne. On peut aller plus avant dans la majoration, et montrer par exemple que ¹

$$\rho(F_n, \phi) \leq 1, 2 \frac{\rho_3}{\sigma^3 \sqrt{n}} + 6 \left(\frac{\rho_3}{\sigma^3 \sqrt{n}} \right)^2 \quad (15)$$

Du théorème 3.8, on peut déduire

$$\rho(F_n, \phi) > e^{-\frac{\sqrt{n}}{c^2 \sigma^3 \rho(F, \phi)}} \quad (16)$$

L'écart des sommes partielles à la loi normale est donc minoré, et on obtient un nombre minimal d'échantillons pour obtenir une erreur donnée dans le théorème de la limite centrale.

En particulier, l'utilisation d'une transformée de Fourier discrète n'a pas, comme on le voit parfois affirmer, pour effet de rendre les signaux rigoureusement gaussiens (c'est le cas seulement pour des processus aléatoires blancs). Si on veut étudier cette question, on est amené à observer la convergence des sommes d'échantillons des signaux temporels, qui sont faiblement corrélés si on limite l'échantillonnage du signal. L'argument invoqué est alors un théorème de limite centrale (généralisé au cas de variables non indépendantes). La vitesse de convergence est faible, en particulier dans le cas de lois multimodales, ce d'autant plus que les échantillons sont fortement corrélés, et donc que les signaux sont suréchantillonnés (la convergence en loi est en $\frac{1}{\sqrt{n}}$ pour des échantillons indépendants de longueur n).

LES PROBLÈMES D'IDENTIFICATION AVEUGLE ET L'ANALYSE EN COMPOSANTES INDÉPENDANTES.

On considère un système linéaire, représenté par la relation $y = Ax$, où x désigne le vecteur des entrées, supposées indépendantes, et y la sortie du système. A étant une matrice carrée d'ordre m . Si on suppose que le vecteur y est gaussien, et la matrice A inversible, alors $x = A^{-1}y$ est également gaussien,

par linéarité. Mais d'après le théorème de Cramer, il est suffisant que les colonnes et les lignes de A soient non nulles (si des lignes de A sont nulles, il y aura en réalité un nombre de sorties n plus petit que le nombre m d'entrées).

On voit alors que si la sortie d'un système linéaire quelconque est gaussienne, l'entrée sera également gaussienne. En outre, si on considère le cas d'un processus aléatoire blanc, qui appliqué à un filtre linéaire orthogonal fournit des sorties gaussiennes, non seulement le processus d'entrée est un processus gaussien, mais la sortie du filtre est également un processus gaussien, car les échantillons sont alors non corrélés.

Dans ce qui suit, on s'intéresse à la propriété d'indépendance des sorties. D'après le théorème de Darmois, si la matrice A a tous ses éléments non nuls, et si les sorties du système sont indépendantes, alors les entrées seront nécessairement gaussiennes. On commence par établir une conséquence du théorème de Darmois, puis on montre comment utiliser ce résultat pour un problème pratique de traitement du signal.

Lemme 4.1 : soient x, u, v trois variables indépendantes. Si les variables $z_1 = ax + u$ et $z_2 = bx + v$ sont elles-mêmes indépendantes et x non dégénérée, alors $a = 0$ ou $b = 0$.

En effet, si $ab \neq 0$, on peut supposer $a = b = 1$. ψ_x désignant la seconde caractéristique de x , pour tout couple (α, β) , d'après l'hypothèse $\psi_{z_1, z_2}(\alpha, \beta) = \psi_{z_1}(\alpha) + \psi_{z_2}(\beta)$, et donc $\psi_x(\alpha + \beta) = \psi_x(\alpha) + \psi_x(\beta)$. On peut comme le lemme 3.2 supposer que ψ_x est deux fois dérivable. On déduit alors que $\psi_x'' = 0$, et par suite x est dégénérée. \square

Théorème 4.1 : Soit x un vecteur aléatoire à n composantes indépendantes et non dégénérées, dont l'une au plus est Gaussienne. Soit C une matrice rectangulaire quelconque à m lignes et n colonnes, et $z = Cx$. On suppose qu'aucune ligne ni colonne de C n'est nulle (ce qui reviendrait soit à éliminer une composante de x , soit à prendre une composante de z nulle). Les trois propriétés suivantes sont équivalentes.

- i) Les composantes de z sont indépendantes deux à deux
- ii) Les composantes de z sont indépendantes dans leur ensemble
- iii) Il existe une sous matrice carrée L d'ordre m de C dont tous les éléments non nuls sont sur des colonnes distinctes (L est de la forme ΛP , P étant une permutation et Λ une matrice diagonale).

En particulier, C ne peut avoir strictement plus de lignes que de colonnes ($m \leq n$).

Si $m = n$, la matrice C s'écrit elle-même comme le produit d'une matrice de permutation et d'une matrice diagonale : $C = \Lambda_1 P_1$.

Démonstration

Supposons qu'il existe dans C deux éléments non nuls situés sur une même colonne, C_{i1} et C_{j1} , on considère

$$z_i = \sum_{k=1}^n C_{ik} x_k, z_j = \sum_{k=1}^n C_{jk} x_k \quad (17)$$

1. Résultat non publié.

Par hypothèse, $\forall k \neq 1, C_{ik}C_{jk} = 0$ (sinon x_k est gaussienne d'après Darmois) et donc

$$z_i = C_{i1}x_1 + u, z_j = C_{j1}x_1 + v \quad (18)$$

x_1, u, v sont indépendantes, ce qui contredit le lemme.

Dans C , il existe donc un élément non nul, au plus, par colonne, et par suite un exactement d'après l'hypothèse. Il existe également par hypothèse un élément non nul par ligne, d'où $m \leq n$, et la conclusion. \square

L'Analyse en Composantes Indépendantes, telle qu'elle est présentée par Pierre Comon dans [6], repose sur cette propriété. On considère un problème d'identification, où on cherche à estimer l'entrée d'un système à partir de l'observation de sa sortie.

Étant donné un système linéaire $y = Ax$ dont les entrées sont supposées indépendantes et non gaussiennes, si on détermine un filtre $z = Cy$ dont les sorties sont elles-mêmes indépendantes, alors les vecteurs z et x seront identiques (à une permutation et un facteur d'échelle près).

Les composantes z_i de z ne seront jamais rigoureusement indépendantes, mais on cherche un filtre qui les rende le moins dépendantes possible. Pour cela, on minimise une fonction de contraste, qui caractérise le degré d'indépendance des z_i . Le problème considéré est dit d'identification aveugle, dans la mesure où on n'a aucune information a priori sur la matrice de transfert A . En fait, on ne cherche pas non plus à l'estimer.

L'étape suivante est la recherche de « bonnes » fonctions de contraste. Il faudra en effet pouvoir mettre en œuvre des algorithmes de minimisation ayant une vitesse de convergence suffisante. Il faut par ailleurs pouvoir estimer la fonction de contraste, qui sera une fonction de la loi des signaux, a priori inconnue.

Si on considère le cas de l'information mutuelle, définie pour un vecteur aléatoire x de composantes (x_1, \dots, x_n) dont la loi est absolument continue par

$$I(p_x) = \int p_x(u) \log \frac{p_x(u)}{\prod p_{x_i}(u_i)} du \quad (19)$$

I est minimale si la loi produit est égale au produit des lois, c'est-à-dire si les composantes de x sont indépendantes. Si on reprend les notations précédentes, où z est le vecteur dont on cherche à rendre les composantes indépendantes, les lois de z et des z_i sont inconnues. On peut en fait estimer $I(p_z)$ à partir des cumulants de z (les coefficients du développement en série de la seconde caractéristique) qui peuvent être estimés directement à partir d'un échantillon.

On utilise pour ce faire un développement de la loi par rapport à la loi normale. On montre en effet que

$$I(p_z) = J(p_z) - \sum J(p_{z_i}) \quad (20)$$

où

$$J(p_z) = \int p_z(u) \log \frac{p_z(u)}{\phi_z(u)} du$$

est la divergence de Kullback entre la loi p_z et la loi normale de mêmes moyenne et variance, et le rapport $\frac{p_z(u)}{\phi_z(u)}$ se développe en fonction des cumulants de z .

En définissant $\psi(p_{z'}) = -I(p_{Cz'})$, où y' est la variable normalisée associée à y , le facteur de normalisation étant la racine carrée de la matrice de covariance de y , on peut montrer que ψ est un contraste dans le sens suivant

1. ψ est invariant par changement d'échelle ($\psi(p_{\Lambda z'}) = \psi(p_{z'})$, pour toute matrice diagonale régulière Λ)
2. Si z est à composantes indépendantes, $\psi(p_{Az'}) \leq \psi(p_{z'})$, pour toute matrice régulière A . La normalisation du contraste se réduit alors, au premier ordre, à

$$\sum 4K_{iii}^2 + K_{iiii}^2 + 7K_{iii}^4 - 6K_{iii}^2 K_{iiii} \quad (21)$$

où K_{i_1, \dots, i_n} désigne le cumulants des variables z'_{i_1} à z'_{i_n} et K_{iii} vérifie

$$K_{iii} = \sum_{pqr} C_{ip} C_{iq} C_{ir} \Gamma_{pqr} \quad (22)$$

Γ_{pqr} étant le cumulants des variables y'_p, y'_q et y'_r

Le contraste ψ est bien une fonction de la matrice C cherchée, sous la forme d'un polynôme des coefficients de C .

Dans [6], le contraste ψ est remplacé par

$$\psi_1(C) = \sum K_{iii}^2 \quad (23)$$

qui est encore un contraste discriminant, dont l'expression nécessite une moindre charge de calculs.

L'algorithme utilisé pour maximiser ψ_1 est du même type que l'algorithme de Jacobi, utilisé dans la diagonalisation des matrices réelles symétriques. Il consiste à prendre pour matrice C une matrice orthogonale, obtenue en accumulant les rotations de Givens maximisant la fonction de contraste pour tous les couples de composantes de z , pris comme axes de rotation.

BIBLIOGRAPHIE

- [1] Paul LÉVY, « Théorie de l'addition des variables aléatoires », Gauthiers-Villars, Paris, 1937.
- [2] Georges VALIRON, « Théorie des fonctions », Gauthiers-Villars, Paris, 1960.
- [3] W.J. FELLER, « An introduction to Probability Theory and its applications », Wiley, New-York, 1966.
- [4] G. LUKACS, « characteristic Functions », Griffin, 1960.
- [5] J.V. LINNIK, I.V. OSTROVSKI, « Décomposition of random variables and vectors », AMS, Vol 68, Providence, 1977.
- [6] Pierre COMON, « Independent Component Analysis », *International Signal Processing Workshop on High-order Statistics*, Chamrousse, 1991, publié dans « Higher Order Statistics », J.L. LACOURME editor, Elsevier, 1992.
- [7] N.L. JOHNSON, S. KOTZ, « Distributions in statistics », Wiley (tomes 1 à 4).

Remerciements

Je tiens à remercier Pierre Comon, qui à travers de longues discussions concernant les problèmes d'identification et l'Analyse en Composantes Indépendantes est également à l'origine de cet article, ainsi que Michel Grojnowski pour sa relecture attentive, tout particulièrement celle de l'ensemble des démonstrations.

6. Annexes

A.1. RAPPEL DE NOTIONS ÉLÉMENTAIRES DE TOPOLOGIE

Soit un espace topologique X sur lequel on a défini l'ensemble O des ouverts (stable par réunions et par intersections finies, et contenant l'ensemble vide), les complémentaires des ouverts étant les fermés. Une *base* d'ouverts ou base de la topologie est une famille d'éléments de O génératrice pour la réunion d'ensembles.

Les ouverts de \mathcal{R} sont les réunions dénombrables d'intervalles ouverts. La topologie naturelle de \mathcal{C} est la topologie produit associée à \mathcal{R} , engendrée par les produits d'intervalles ouverts. Une partie E de X est *connexe* si les seules parties à la fois ouvertes et fermées de E sont E lui-même et l'ensemble vide. On définit les composantes connexes comme les classes pour la relation d'équivalence : $A \approx B$ si $\{A, B\} \subset F$, F connexe. Comme $\{A\}$ est connexe, la relation est bien réflexive, et les composantes connexes forment une partition de E . E sera *connexe par arcs* si deux points A et B peuvent être joints par un chemin, ce qui est une condition plus forte. On définit de même les composantes connexes par arcs. Tout chemin est par définition connexe par arcs. On peut noter que très souvent, si la connexité par arcs paraît plus naturelle, il est bien plus aisé de démontrer directement la connexité, à l'aide d'arguments élémentaires.

On peut définir une troisième notion de connexité, la connexité simple. Une partie E de X est *simplement connexe* si tout lacet γ (chemin dont les extrémités coïncident) de E est homotope à un point, c'est-à-dire si on peut « déformer » γ continument pour le réduire à un seul point de E . Cela signifie l'absence de « trous ». Cette propriété est plus forte que les précédentes. Si on se limite aux lacets formés d'un seul point, on retrouve la connexité par arcs.

Une partie E de X est *quasi compacte* si pour toute famille d'ouverts $O_i, i \in I$, recouvrant ($E \subset \cup O_i$) on peut en extraire une sous famille finie vérifiant la même inclusion (axiome de Borel Lebesgue). Par suite, de toute suite de E on peut extraire une sous suite convergente (axiome de Bolzano Weierstrass). Les deux propriétés sont équivalentes si on peut passer d'une famille finie d'ouverts à une famille dénombrable, donc si la topologie a une base dénombrable d'ouverts. C'est le cas en particulier si X est un espace métrique.

Une partie *compacte* est une partie quasi compacte *séparée* (au sens de Hausdorff), c'est-à-dire dont deux points distincts appartiennent à des ouverts disjoints. Si toute suite a une sous suite qui converge dans X , on parle de compacité relative.

A.2. RÉSULTATS D'ANALYSE COMPLEXE

On revient brièvement sur les propriétés élémentaires des fonctions analytiques et les théorèmes de Cauchy avant de donner un aperçu de la théorie des fonctions entières.

On rappelle qu'une série entière $\sum_{n=0}^{\infty} c_n(z-a)^n$ converge normalement dans son disque de convergence, c'est-à-dire pour $|z-a| < R$, R étant le rayon de convergence de la série.

Une fonction complexe définie sur un ouvert D est analytique si elle est développable en série en tout point de D , le rayon de convergence étant au moins égal à la distance du point à la frontière de D .

Principe des zéros isolés. Si f est une fonction analytique non identiquement nulle pour $|z-a| < R$, alors, à partir de son développement au point a , elle peut s'écrire $f(z) = (z-a)^n g(z)$, où $g(a) \neq 0$ et g est elle-même analytique, donc continue, et il existe $r > 0$ tel que g , et par suite f , ne peut s'annuler pour $0 < |z-a| < r$. a est donc un zéro « isolé » de f . On peut énoncer comme corollaire que si f s'annule dans tout voisinage de a , f est nulle à l'intérieur d'un de ces voisinages.

On en déduit le *Principe du Prolongement Analytique* : Si l'ensemble des zéros d'une fonction f analytique sur un ouvert connexe U a un point d'accumulation a , f est nulle dans U . Elle est en effet nulle dans un voisinage de a , et il suffit alors de considérer la composante connexe S de l'ensemble E des zéros de f , contenant a . Tout d'abord, S est ouvert dans U , car S contient un voisinage de a , et si $z (\neq a)$ est un élément de S , z n'est pas un zéro isolé, sinon il formerait une composante connexe de E , et par suite S contient un voisinage de z .

Mais S est également fermé (comme composante connexe de E , qui est fermé), et de par la connexité de U , $S = U = E$.

Par suite, si deux fonctions f et g , analytiques dans U , coïncident dans un ouvert $V \subset U$, elles sont identiques.

On peut établir de manière simple deux résultats importants relatifs aux bornes des fonctions analytiques complexes : le théorème de Liouville et le principe du maximum.

Pour commencer, on remarque que si f a un développement en série du type

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n, |z-a| < R \quad (\text{A.1})$$

alors pour tout $r, 0 < r < R$, la fonction définie par

$$g(\theta) = f(a + re^{i\theta}) = \sum_{n=0}^{\infty} c_n r^n e^{in\theta} \quad (\text{A.2})$$

existe, la série étant normalement convergente en θ . Par suite, g est intégrable et les $c_n r^n$ sont les coefficients de sa série de Fourier.

On a donc, d'après la relation de Parseval :

$$\sum_{n=0}^{\infty} |c_n|^2 r^{2n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(a + re^{i\theta})|^2 d\theta \quad (\text{A.3})$$

On en déduit le *Théorème de Liouville*. Ce dernier affirme que toute fonction analytique sur \mathcal{C} , bornée, est nécessairement constante. En effet, soit M un majorant de $|f|$, alors pour tout $r > 0$ et pour tout entier n , $|c_n|^2 r^{2n} \leq M^2$. Ceci implique que $c_n = 0$ pour $n \neq 0$, et donc que f est constante.

Principe du Maximum : ce principe précise que le module d'une fonction analytique non constante ne peut avoir de maximum relatif dans un ouvert D .

Soit f une fonction analytique complexe telle que $|f|$ a un maximum en un point a de D , et on suppose que pour $|z - a| < r$, $|f(z)| \leq |f(a)|$. Soit alors le développement de f au point a , $f(z) = \sum c_n(z - a)^n$.

D'après (A.3)

$$\sum |c_n|^2 r^{2n} = |c_0|^2 + |c_1|^2 r^2 + \dots \leq |f(a)|^2 = |c_0|^2.$$

Tous les coefficients d'indices strictement positifs de la suite c_n sont donc nuls. Par suite, $f(z) = c_0$ sur la boule $B(a, r)$, et par prolongement analytique $f = c_0$ sur D .

La théorie de Cauchy joue un rôle important, non seulement par les formules de calcul explicites qu'elle fournit, en particulier pour le calcul d'intégrales, mais aussi car elle permet de déduire l'un des résultats les plus profonds sur les fonctions analytiques complexes, que toute fonction dérivable sur un ouvert de \mathcal{C} est analytique. Cette propriété est à comparer avec la situation dans \mathbf{R} , où une fonction indéfiniment dérivable n'est pas toujours analytique, avec par exemple $f(x) = e^{-1/x}$ pour $x > 0$, et 0 ailleurs.

Pour mieux comprendre la situation, on peut la comparer avec celle de \mathbf{R}^2 , où d'après le théorème de Schwartz la C^1 différentiabilité est équivalente à l'existence et la continuité des dérivées suivant chaque variable alors que dans \mathcal{C} cela suppose l'existence de dérivées vectorielles suivant toutes les directions. On peut également remarquer qu'on ne peut définir dans \mathcal{C} de fonction du type de la fonction ci-dessus (par exemple avec un processus de révolution) sinon en remplaçant $e^{-1/x}$ par $e^{-1/|z-a|}$ qui n'est plus dérivable.

On rappelle qu'un chemin γ est une application continue, à variations bornées (c'est-à-dire $\sum |\gamma(t_i + 1) - \gamma(t_i)| < M$ pour toute famille finie t_1, \dots, t_n), d'un segment T de \mathbf{R} dans \mathcal{C} . Son image $\bar{\gamma}$ est un contour de \mathcal{C} . Il est dit fermé si les images des extrémités du segment coïncident.

On définit l'intégrale d'une fonction complexe f sur un contour $\bar{\gamma}$ (ou simplement γ) comme l'intégrale de Riemann-Stieljes :

$$\int_{\gamma} f(z) dz = \int_T (f \circ \gamma)(t) d\gamma(t) \quad (\text{A.4})$$

Cette intégrale généralise l'intégrale de Riemann au cas où le pas d'intégration est déterminé par les écarts de γ . Elle existe lorsque γ est à variations bornées, ou ce qui est équivalent si elle définit une courbe rectifiable (sur laquelle on peut définir une longueur). C'est le cas en particulier si γ est continument dérivable par morceaux.

Il est aisé d'établir qu'elle vérifie les propriétés de linéarité et de décomposabilité, et, lorsque γ est dérivable, qu'elle coïncide avec une intégrale de Riemann

$$\int (f \circ \gamma)(t) d\gamma(t) = \int (f \circ \gamma)(t) \gamma'(t) dt \quad (\text{A.5})$$

Théorème de Cauchy : si f est dérivable sur un ouvert D simplement connexe de \mathcal{C} , alors pour tout contour fermé γ contenu dans D ,

$$\int_{\gamma} f(z) dz = 0$$

Ce théorème peut également s'énoncer : si f est dérivable dans un ouvert D , et si γ_1 et γ_2 sont deux contours fermés homotopes contenus dans D , alors

$$\int_{\gamma_1} f(z) dz = \int_{\gamma_2} f(z) dz$$

En appliquant le théorème à la fonction $g(z) = \frac{f(z) - f(z_0)}{z - z_0}$ qui est dérivable pour $z \neq z_0$ on déduit la formule intégrale de Cauchy : si z_0 n'est pas dans $\bar{\gamma}$,

$$\int_{\gamma} \frac{f(z)}{z - z_0} dz = f(z_0) \cdot \int_{\gamma} \frac{1}{z - z_0} dz \quad (\text{A.6})$$

Car g est continue, donc bornée au voisinage de z_0 , et on considère le chemin $\gamma_n(t) = z_0 + 1/n e^{i2\pi t}$ qui est homotope à γ .

En outre, si z_0 est un point intérieur à γ ,

$$\int_{\gamma} \frac{f(z)}{z - z_0} dz = 2i\pi f(z_0)$$

d'où on peut déduire l'existence et la valeur des dérivées successives de f

$$\int_{\gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz = \frac{2i\pi}{n!} f^{(n)}(z_0) \quad (\text{A.7})$$

La formule de Cauchy permet de définir une fonction dérivable à partir uniquement de ses valeurs sur un contour, ce qui est proche du principe de prolongement. Mais également d'après (5) qu'une fonction dérivable est indéfiniment dérivable. En fait, la formule intégrale permet également de montrer l'existence de développements en série entière, d'où on déduit l'analyticité.

Fonctions entières et le théorème de décomposition de Hadamard.

On définit l'ordre d'une fonction entière (analytique sur \mathcal{C}) par la quantité τ , éventuellement infinie :

$$\tau = \limsup_{r \rightarrow \infty} \frac{\log \log M(r)}{\log(r)} \quad (\text{A.8})$$

avec

$$M(r) = \sup_{|z| \leq r} |f(z)| = \sup_{|z|=r} |f(z)|.$$

L'ordre permet de décrire le comportement asymptotique des fonctions entières. Le Théorème de Hadamard joue un rôle fondamental dans la théorie des décompositions de lois et permet de déduire des propriétés sur la forme des composantes des fonctions caractéristiques (paragraphe 3).

Théorème de Hadamard : si f est entière d'ordre ρ ,

$$f(z) = e^{Q(z)} \cdot z^m \cdot g(z),$$

où Q est un polynôme de degré au plus ρ et g (produit canonique de Weierstrass) est soit un polynôme (si f a un nombre fini de zéros), soit s'écrit sous forme d'un produit infini

$$g(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z}{a_n}\right) e^{c_n} \text{ avec } c_n = \frac{z}{a_n} + \dots + \frac{z^p}{pa_n^p} \quad (\text{A.9})$$

les a_n étant les zéros non nuls de f .

On ne propose pas ici de démonstration de ce théorème, que l'on pourra trouver par exemple dans [2]. Les trois propriétés énoncées ci-dessous seront suffisantes pour établir les résultats de cet exposé comme les théorèmes de Cramer et de Raïkov (le théorème est nécessaire à la démonstration de la forme la plus générale du théorème de Marcinkiewicz). Elles sont des conséquences du théorème de Hadamard, mais on peut en proposer une démonstration directe.

Corollaire 1 : si une fonction entière f vérifie :

$$f(z) = O(z^m), \quad |z| \rightarrow \infty,$$

alors f est un polynôme de degré au plus égal à m .

Pour $n = 0$, on retrouve le théorème de Liouville pour les fonctions entières.

Démonstration

Il suffit d'appliquer la relation (1) pour r au voisinage de l'infini :

$$\sum_{n=0}^{\infty} |c_n|^2 r^{2n} \leq M r^{2m}$$

Les coefficients d'ordre strictement supérieur à m sont nuls, et par suite f est un polynôme de degré inférieur ou égal à m . \square

Corollaire 2 : si f est une fonction entière d'ordre inférieur ou égal à n , alors f est un polynôme de degré au plus n .

Il s'agit d'une conséquence directe du corollaire 1.

Corollaire 3 : si une fonction entière f n'a pas de zéros, elle s'écrit $f = e^g$ où g est elle-même entière. Il suffit de considérer la fonction f'/f qui est entière, et admet donc une primitive g qui vérifie alors $f = e^g$ sur \mathbf{R} , et donc sur \mathcal{C} par prolongement analytique.

A.3. DÉMONSTRATIONS

Démonstration du théorème 3.7

On commence par établir une condition de compacité pour une suite de lois, puis on exprime la métrique de Lévy pour les lois convoluées.

Lemme 3.3 : Pour que toute sous suite de lois F_n admette une sous suite convergeant faiblement vers une loi F , c'est-à-dire pour que l'ensemble $\{F_n\}$ soit relativement compact, il faut et il suffit que pour tout $\varepsilon > 0$, il existe $a = a(\varepsilon)$ tel que pour tout entier n ,

$$F_n(-a) < \varepsilon \quad \text{et} \quad F_n(a) > 1 - \varepsilon$$

ou encore que $1 - F_n(a) + F_n(-a) < \varepsilon \quad (\text{A.10})$

Démonstration

Pour montrer que la condition est suffisante, on construit une suite de sous suites successives $F_{n_k, p}$, qui coïncident pour les p premières valeurs, et convergent simplement sur l'intervalle compact $[-a(\frac{1}{p}), a(\frac{1}{p})]$. Il est aisé de vérifier que la suite $F_{n_p, p}$ converge faiblement vers une loi F .

Pour établir la réciproque, il suffit de revenir à la définition de la métrique de Lévy (5).

Soit $\varepsilon > 0$, il existe un entier n_0 tel que pour tout x et pour $n \geq n_0$,

$$F_n(x) \leq F(x + \frac{\varepsilon}{2}) + \frac{\varepsilon}{2}$$

on choisit $a = a(\varepsilon)$ tel que

$$F_n(-a) \leq \varepsilon \text{ pour } n < n_0 \text{ et } F(-a - \frac{\varepsilon}{2}) < \frac{\varepsilon}{2} \quad \square$$

Lemme 3.4 : Pour toutes lois P_1, P_2, Q_1, Q_2 :

$$L(P_1 * P_2, Q_1 * Q_2) \leq L(P_1, Q_1) + L(P_2, Q_2) \quad (\text{A.11})$$

Démonstration

Il s'agit d'une conséquence directe de la définition du produit de convolution (2)

Notons l_1 et l_2 les termes du second membre de (A.11).

Il faut montrer que

$$(Q_1 * Q_2)(x) \leq (P_1 * P_2)(x + l_1 + l_2) + l_1 + l_2 \quad (\text{A.12})$$

$$\begin{aligned} (Q_1 * Q_2)(x) &= \int Q_2(x - u) dQ_1(u) \\ &\leq \int [P_2(x + l_2 - u) + l_2] dQ_1(u) \\ &\leq l_2 + \int P_2(x + l_2 - u) dQ_1(u) \\ &\leq l_2 + \int Q_1(x + l_2 - u) dP_2(u) \\ &\leq l_2 + \int [P_1(x + l_2 + l_1 - u) + l_1] dP_2(u) \\ &\leq l_1 + l_2 + (P_1 * P_2)(x + l_1 + l_2) \end{aligned}$$

On montre de la même façon que

$$(Q_1 * Q_2)(x) \geq (P_1 * P_2)(x - l_1 - l_2) - l_1 - l_2 \quad \square \quad (\text{A.13})$$

Lemme 3.5 : Si $F = F_1 * F_2$, et si $\lambda_1, \lambda_2, y_1, y_2$ sont respectivement les médianes de F_1 et de F_2 et les quartiles de F , alors $y_1 \leq \lambda_1 + \lambda_2 \leq y_2$

Démonstration

Par définition,

$$F(y_1) \leq \frac{1}{4} \leq F(y_1 + 0) \text{ et } F(y_2) \leq \frac{3}{4} \leq F(y_2 + 0)$$

et de même,

$$F_1(\lambda_1) \leq \frac{1}{2} \leq F(\lambda_1 + 0) \text{ et } F_1(\lambda_2) \leq \frac{1}{2} \leq F(\lambda_2 + 0)$$

$$\begin{aligned} F(\lambda_1 + \lambda_2) &= (F_1 * F_2)(\lambda_1 + \lambda_2) \\ &= \int F_2(\lambda_1 + \lambda_2 - t) dF_1(t) \\ &= \int F_2[\lambda_2 + (\lambda_1 - t)] dF_1(t) \\ &= \int_{t \geq \lambda_1} F_2[\lambda_2 + (\lambda_1 - t)] dF_1(t) \\ &\quad + \int_{t < \lambda_1} F_1[\lambda_2 + (\lambda_1 - t)] dF_1(t) \\ &\geq \frac{1}{2} \int_{t \geq \lambda_1} dF_1(t) \geq \frac{1}{4} \geq F(y_1) \end{aligned}$$

Donc $y_1 \leq \lambda_1 + \lambda_2$, et de la même façon, $y_2 \leq \lambda_1 + \lambda_2 \quad \square$

Lemme 3.6 : Si $\varepsilon_n = L(F_n, F) \rightarrow 0$, les suites F_{n_j} ont des sous suites convergeant vers des composantes de F .

Démonstration

Il suffit de considérer le cas de deux composantes, la généralisation étant immédiate.

On a les décompositions : $F_n = F_{n_1} * F_{n_2}$, ou en langage des variables aléatoires $Z_n = X_n + Y_n$, avec X_n et Y_n indépendantes, et on considère μ_1 et μ_2 leurs médianes.

On peut supposer l'une des variables (X_n) centrée par rapport à sa médiane, en changeant X_n en $X_n - \mu_1$ et Y_n en $Y_n + \mu_1$.

On montre que les sous suites F_{n_1} et F_{n_2} sont relativement compactes, en utilisant le lemme 3.3.

$$\begin{aligned} P(Y_n \geq c) &\leq 2P(X_n \geq 0, Y_n \geq c) \leq 2P(Z_n \geq c) \\ \text{et} \quad P(Y_n < -c) &\leq 2P(Z_n < -c) \end{aligned}$$

soit $\varepsilon > 0$, par hypothèse, il existe c tel que pour tout n ,

$$F_n(-c) < \frac{\varepsilon}{2} \text{ et } F_n(c) > 1 - \frac{\varepsilon}{2}$$

car pour $h > 0$ fixé, il existe n_0 ,

$$F(x - h) - h < F_n(x) < F(x + h) + h, \quad n \geq n_0$$

Soit C tel que $F(-C) < \frac{\varepsilon}{4}$,

on prend $h = \frac{\varepsilon}{4}$ et $c = \text{Max}(c_n, c + h)$, où $F(-c_n) < \frac{\varepsilon}{4}$

par suite, $P(Y_n \geq c) \leq \varepsilon$, et $P(Y_n < -c) < \varepsilon$ et F_{n_1} est compacte et a donc une sous suite $F_{n_k,1}$ convergeant vers une loi F_1 (si $X_{n_k,1} - \mu_1$ converge vers X_1 , $X_{n_k,1}$ converge vers $X_1 - \mu_1$).

Pour la même raison, F_{n_2} est également compacte, et on peut prendre une sous suite commune d'entiers n_k pour laquelle les deux séquences convergent simultanément.

On a alors $F = F_1 * F_2$, d'après le lemme 3.4 \square

Démonstration du théorème

Soit K_F la classe des composantes de F . Si on suppose que la décomposition n'est pas stable, il existe une sous suite $F_{n_k,j}$ vérifiant $L(F_{n_k,j}, K_F) \geq \varepsilon$, pour $\varepsilon > 0$.

Mais la suite F_{n_k} vérifie alors les hypothèses du lemme 3.6, et par suite $F_{n_k,j}$ a une sous suite convergeant vers une composante de F , qui est élément de K_F ce qui contredit l'hypothèse. \square

Démonstration du théorème 3.10 :

On montre tout d'abord que f est infiniment divisible avec des facteurs du même type.

Soit $f = f_1 f_2$, avec $f_i(x) = f(a_i x)$, $f(x) = f(a_1 x) f(a_2 x)$ et en itérant la décomposition,

$$f(x) = e^{i\mu x} \prod_{p=0}^n f(a_1^p a_2^{n-p} x)^{C_n^p} \quad (\text{A.14})$$

Pour tout entier n , C_n^p désignant les combinaisons de p éléments parmi n .

Par suite, pour tout entier n , $e^{i\mu_1 x} f(a_1 a_2^{n-1} x)^n = f(bx)$, donc $f(x) = f(b_n x)^n e^{i\mu_n x}$

(ce qui induit également que la loi est stable).

En particulier, on a

$$|f(x)| = |f(c_1 x)|^2 = |f(c_2^2 x)|^4 = |f(c_3^3 x)|^8 = \dots$$

Si f a un moment d'ordre deux, on aura au voisinage de 0, d'après le théorème 2.1,

$$|f(x)| = 1 + \alpha x^2 + o(x^2) \quad (\text{A.15})$$

On a aussi

$$|f(x)| = (1 + \alpha c_k^{2k} x^2)^{2^k} + o(x^2) = 1 + 2^k \alpha c_k^{2k} x^2 + o(x^2)$$

donc $c_k = \frac{1}{\sqrt{2}}$

On aura alors

$$e^{-i\mu_k x} f(x) = f(2^{\frac{-k}{2}} x) 2^k = (1 + h(2^{\frac{-k}{2}} x))^{2^k} \sim e^{-\alpha x^2}$$

et donc

$$f(x) = e^{-i\mu x - \alpha x^2} \quad \square$$

L'existence de la variance est une hypothèse restrictive, non nécessaire. On peut en effet établir le théorème à partir de l'expression générale de f pour des lois stables, dont on pourra trouver la démonstration dans [3] :

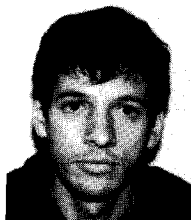
$$\log f(x) = i\mu x - b|x|^\alpha \left[1 + ic \frac{x}{|x|} \operatorname{tg}\left(\frac{\pi}{2}\alpha\right) \right] \quad (\text{A.16})$$

avec $0 < \alpha < 2$, $b \geq 0$, et $|c| \leq 1$. La loi est gaussienne pour $\alpha = 2$, et si $\alpha \neq 2$,

$$\log f(x) = \log f_1(x) + \log f_2(x)$$

avec $b = b_1 + b_2$ et $bc = b_1c_1 + b_2c_2$, f_1 et f_2 étant de types distincts.

L'AUTEUR



L'auteur est diplômé de l'École Nationale Supérieure des Télécommunications, et Agrégé de Mathématiques. Après une première expérience de développement, il a rejoint en Décembre 1987 le service Traitement du Signal et de l'Information de THOMSON SINTRA activités sous-marines, pour des travaux et recherches principalement dans les domaines des traitements d'antennes, du filtrage et de la poursuite de cibles, ainsi qu'en théorie des probabilités. Il exerce depuis janvier 1993 une fonction d'ingénieur d'études système au sein de la division RCC de THOMSON CSF, dans le cadre d'un projet Européen de Télécommunications Spatiales, avec la responsabilité technique des aspects traitements d'antennes et calculs de performances.

Manuscrit reçu le 14 octobre 1992.