

Introduction aux probabilités

M. Yor^{(1),(2)}

27 juin 2005

(1) Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris VI et VII, 4 Place Jussieu - Case 188,
F-75252 Paris Cedex 05

(2) Institut Universitaire de France

C'est avec plaisir que j'ai répondu affirmativement à la demande de Daniel Bennequin de faire une introduction - assez mathématique - aux probabilités, dans le cadre de l'Ecole d'Eté "Mathématiques et Cerveau".

Il m'a semblé raisonnable d'utiliser les 3 heures dont je dispose pour cette introduction de la façon suivante :

1^{ère} heure : Le formalisme (à la Kolmogorov) des probabilités;
Notions d'indépendance, d'indépendance conditionnelle,
d'espérance et de lois conditionnelles.

2^{ème} heure : Les suites de variables aléatoires, et leurs comportements
asymptotiques.

3^{ème} heure : Les processus stochastiques; les grandes hypothèses les
concernant.

Bien entendu, j'essaierai, dans la mesure du possible, d'adapter l'exposé à la demande de l'auditoire; au cours de l'Ecole d'Eté, il serait intéressant de voir, parmi les notions que je présente, celles qui sont utilisées (pas assez?) dans les mathématiques du cerveau...!

1^{ère} heure :

(1.1) Grosso modo, pour un probabiliste, une quantité associée à un phénomène aléatoire n'est pas connue ponctuellement, mais on connaît "sa répartition". De façon plus formelle, une telle quantité, supposée pour des raisons de simplicité à valeurs dans \mathbb{R} , est définie sur un espace mesurable (Ω, \mathcal{F}) , muni d'une probabilité P , et on connaît :

$$P_X(\Gamma) = P(X \in \Gamma), \quad \forall \Gamma \in \mathcal{B}_{\mathbb{R}}$$

P_X est donc une probabilité sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

On dit que X est une variable aléatoire réelle.

Plus généralement, on peut s'intéresser à des variables (X_1, \dots, X_n) à valeurs dans \mathbb{R}^n , et on a alors; pour tous les ensembles boréliens $\Gamma_1, \dots, \Gamma_n$:

$$P_{(X_1, \dots, X_n)}(\Gamma_1 \times \Gamma_2 \times \dots \times \Gamma_n) = P(X_1 \in \Gamma_1; X_2 \in \Gamma_2; \dots; X_n \in \Gamma_n).$$

On a pu dire quelquefois, avec une certaine condescendance, que la théorie des probabilités est un cas particulier de la théorie de la mesure... Mais, le probabiliste s'intéresse à la façon dont une variable "dépend" (au sens probabiliste, ou au sens ordinaire) d'autres variables, ce qui mène à des discussions détaillées de la notion d'indépendance, centrale en théorie des probabilités. Pour comprendre l'importance de cette notion, disons que si X_1, X_2, \dots, X_k sont k variables aléatoires indépendantes, cela correspond à des quantités provenant de phénomènes aléatoires qui n'ont rien à voir entre eux (imaginer des exemples...). De façon mathématique, ceci est défini par la propriété :

$$(1) \quad P_{(X_1, \dots, X_k)} = P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_k}$$

c'est-à-dire que la loi conjointe du vecteur aléatoire (X_1, \dots, X_k) est égale au produit tensoriel des lois "individuelles" dans $(X_i, i \leq k)$. On dit que P_{X_i} est la i^{e} marginale de la loi $P_{(X_1, \dots, X_k)}$.

(1) se traduit aussi par :

$$E[f_1(X_1) \dots f_k(X_k)] = \prod_{j=1}^k E[f_j(X_j)]$$

pour toutes $f_j : \mathbb{R} \rightarrow \mathbb{R}$, bornées, par exemple (on peut prendre : $f_j(x) = 1_{\Gamma_j}(x)$). Ensuite, à partir de là, on peut par exemple, considérer la distribution de $\Phi = \varphi(X_1, \dots, X_k)$, pour n'importe quelle $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$, et on a :

$$\boxed{P_{\Phi} = \varphi(P_{(X_1, \dots, X_k)})}$$

Inversement, si $\underline{X} = (X_1, \dots, X_k)$ est un vecteur aléatoire donné, représentant certaines quantités liées à un phénomène " k -dimensionnel"; il sera souvent intéressant de trouver un autre phénomène, ou plutôt un autre vecteur aléatoire $\underline{Y} = (Y_1, \dots, Y_m)$ dont les composantes sont indépendantes, et tel que l'information engendrée par \underline{X} coïncide avec celle engendrée par \underline{Y} ; autrement dit, pour tout $i \leq k$, il existe $\varphi_i : \mathbb{R}^m \rightarrow \mathbb{R}$ telle que :

$$X_i = \varphi_i(Y_1, \dots, Y_m),$$

et, inversement, pour tout $j \leq m$, il existe $\Psi_j : \mathbb{R}^k \rightarrow \mathbb{R}$ telle que :

$$Y_j = \Psi_j(X_1, \dots, X_k).$$

Tout naturellement, ce type de considération mène à des calculs de lois, qui se ramènent très simplement à des changements de variables dans les intégrales, calculs de Jacobiens, etc...

Bien sûr, les lois en question peuvent être plus ou moins compliquées, et se présenter de différentes manières, i.e :

- $P_X(dx) = \sum_n a_n \delta_{x_n}(dx)$, ⁽¹⁾
i.e: X est discrète: $P(X = x_n) = a_n$, avec: $\sum_n a_n = 1$;
- $P_X(dx) = f(x)dx$, avec $f : \mathbb{R} \rightarrow \mathbb{R}$, telle que: $\int f(x)dx = 1$.

On dit alors que f est une densité de probabilité (et que P_X est absolument continue).

(1.2) Donnons maintenant quelques exemples fondamentaux.

Exemple 1 : (Lois gaussiennes; Vecteurs Gaussiens).

- On dit que G suit la loi de Gauss de moyenne m , et de variance σ^2 si :

$$P_G(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx$$

On montre aisément :

$$E[G] = m; E[(G - E(G))^2] = \sigma^2.$$

- Plus généralement, si G_1, \dots, G_k sont indépendantes, gaussiennes de moyenne m_j et variance σ_j^2 , alors toute combinaison linéaire $\sum a_j G_j$ est encore gaussienne, de moyenne $m = \sum a_j m_j$, et variance $\sigma^2 = \sum a_j^2 \sigma_j^2$.

1. La notation $P_X(dx)$ est très informelle; il serait un peu plus correct d'écrire: $P_X((x, x+dx)) \equiv P(X \in (x, x+dx))$

- Si $m = 0$, et $\sigma^2 = 1$, on dit que G est une variable gaussienne centrée, réduite; on écrit de façon générale :

$$G \sim \mathcal{N}(m, \sigma^2)$$

(le symbole : $X \sim Y$ signifie : X et Y ont même loi).

Exemple 2 : (Variables gamma; variables beta).

- On dit qu'une variable X est distribuée exponentiellement, de paramètre $a > 0$ si : $P_X(dx) = ae^{-ax}dx$.

Considérons G_1 et G_2 deux variables gaussiennes, centrées, réduites, indépendantes; alors : $G_1^2 + G_2^2$ est exponentielle, de paramètre $(1/2)$ (et donc, de moyenne 2). Plus généralement, si G_1, \dots, G_k sont k variables gaussiennes, centrées, réduites, indépendantes, et si l'on considère ⁽²⁾ : $X_k = G_1^2 + \dots + G_k^2$, alors : $X_k \sim 2\gamma_{k/2}$, où γ_r désigne une variable gamma de paramètre $r > 0$, variable dont la loi est donnée par :

$$P_{\gamma_r}(dx) = e^{-x} \frac{x^{r-1} dx}{\Gamma(r)} \quad (x \geq 0).$$

L'identité en loi (évidente, - voir ci-dessus - pour les indices demi-entiers) s'étend en fait à tous $r, r' > 0$:

$$\gamma_r + \gamma_{r'} \stackrel{(loi)}{=} \gamma_{r+r'} \quad (r, r' > 0)$$

où dans le membre de gauche, γ_r et $\gamma_{r'}$ sont indépendantes. Il est alors naturel de présenter les variables $\beta_{a,b}$ ($a, b > 0$), qui satisfont :

$$P(\beta_{a,b} \in dx) = \frac{x^{a-1}(1-x)^{b-1} dx}{B(a,b)} \quad (x \in (0,1)).$$

Considérons à nouveau un couple $(\gamma_r, \gamma_{r'})$ de variables gamma indépendantes, on a alors :

$$(2) \quad \left(\frac{\gamma_r}{\gamma_r + \gamma_{r'}}, \gamma_r + \gamma_{r'} \right) \stackrel{(loi)}{=} (\beta_{r,r'}; \gamma_{r+r'})$$

où, dans le membre de droite, les variables sont également supposées indépendantes.

Illustrons cette décomposition avec à nouveau G_1 et G_2 deux variables gaussiennes centrées, réduites, en écrivant :

$$(3) \quad G_1 + iG_2 = \mathcal{R}e^{i\theta},$$

2. La loi de cette variable est la fameuse loi du chi-2 à k degrés de liberté

d'où :

$$(4) \quad G_1^2 = \mathcal{R}^2(\cos^2(\Theta)).$$

On a choisi en (3) \mathcal{R} et Θ indépendantes; $\mathcal{R}^2 \sim 2\gamma_1$, et Θ uniformément distribuée sur $(0, 2\pi)$.

En (4), on a, à gauche: $G_1^2 \sim 2\gamma_{1/2}$; et, à droite: $\cos^2(\Theta) \sim \beta_{\frac{1}{2}, \frac{1}{2}}$ (pour des raisons évidentes, cette loi est dite loi de l'arc sinus; exercice: exprimer $P(\beta_{\frac{1}{2}, \frac{1}{2}} \leq x)$ à l'aide de la fonction arc sinus).

(1.3) Espérances et lois conditionnelles

- Continuons de considérer l'exemple du couple $(\gamma_r, \gamma_r + \gamma_{r'})$, où γ_r et $\gamma_{r'}$ désignent deux variables gamma indépendantes; à l'évidence, les deux composantes de ce couple ne sont pas indépendantes, et on sait se ramener de façon triviale à un couple indépendant, à savoir $(\gamma_r, \gamma_{r'})$. Mais, l'identité (2) nous explique également, comment connaissant $\gamma_r + \gamma_{r'}$, la loi de γ_r est modifiée, i.e: elle se "ramène" à celle de $\beta_{r, r'}$ (pour que cela soit encore plus parlant, considérer (3) et (4)).
- Essayons de formaliser ceci, et de le comprendre dans un cadre général. On se donne donc un couple (X, Y) de variables aléatoires, et on se propose de discuter comment la loi P_X est modifiée lorsque l'on connaît la valeur de Y ; on écrira: $P(X \in dx | Y = y)$. Pour cela, on a besoin, dans un premier temps, de la notion d'espérance conditionnelle, puis de celle de loi conditionnelle:

- (i) pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$, il existe une fonction $\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}_+$ telle que:

$$\forall g : \mathbb{R} \rightarrow \mathbb{R}_+, \quad E[g(Y)f(X)] = E[g(Y)\tilde{f}(Y)]$$

\tilde{f} est unique, aux ensembles de P_Y - mesure nulle près - et on dit que: $\tilde{f}(Y)$ est l'espérance conditionnelle de $f(X)$ "sachant Y ".

- (ii) A partir de ce résultat d'unicité, on peut montrer qu'il existe en fait une famille de probabilités $\Pi(y; dx)$ sur \mathbb{R} telle que:

$$\forall f : \mathbb{R} \rightarrow \mathbb{R}_+, \quad \tilde{f}(Y) = \int \Pi(Y(w); dx) f(x),$$

et c'est cette famille que l'on appelle la famille des lois conditionnelles de X , sachant Y .

Exercice: Expliciter ces différentes quantités lorsque $X = \gamma_r$, et $Y = \gamma_r + \gamma_{r'}$.

Commentaires : Ce sont là des notions difficiles et fines, sur lesquelles il faut beaucoup réfléchir. Elles donnent un sens mathématique à l'idée que l'aléa (dans notre vie quotidienne, par exemple), est modifié, selon que l'on connaît telle ou telle information; en fait, cet aléa évolue de façon continue avec les informations que nous enregistrons... Relier cela à l'analyse bayésienne.

(1.4) Variables conditionnellement indépendantes

- Soient A, B, C trois variables indépendantes; considérons $X = \varphi(A, B)$, et $Y = \Psi(B, C)$, pour deux fonctions φ et Ψ . En général, X et Y ne vont pas être indépendantes, mais, elles le sont toujours conditionnellement à B , ou : "sachant B ".
- On peut formaliser la notion de dépendance, ou d'indépendance conditionnelle, de X et Y , sachant Z , en considérant la loi conditionnelle que nous noterons maintenant $\hat{\Pi}$, du couple (X, Y) sachant Z ; alors, X et Y sont conditionnellement indépendantes sachant Z si on a :
 $P_Z(dz)$ p.s.,

$$\int \hat{\Pi}(z; dx dy) f(x)g(y) = \int \hat{\Pi}(z; dx) f(x) \int \hat{\Pi}(z; dy) g(y)$$

pour toutes fonctions f et g boréliennes, bornées.

(1.5) Outils analytiques

Pour étudier les lois de variables aléatoires à valeurs :

- dans \mathbb{N} , on utilise les fonctions génératrices: $E[a^X] = \sum_n a^n P(X = n)$ (voir en (A.3) un exemple intéressant)
- dans \mathbb{R}_+ , on utilise la transformée de Laplace: $E[e^{-\lambda X}]$ ($\lambda \geq 0$)
- dans \mathbb{R} , on utilise la transformée de Fourier: $E[e^{iuX}]$ ($u \in \mathbb{R}$), également appelée fonction caractéristique de (la loi de).

2^{ème} heure :

(2.1) Nous considérons pendant cette seconde heure une suite de variables aléatoires $X_1, X_2, \dots, X_n, \dots$ indépendantes et équidistribuées (= de même loi). Ceci modélise la répétition "à l'infini" de quantités associées à un phénomène aléatoire, dont l'aléa ne change pas au cours du temps, et qui n'a pas de mémoire. Il s'agit donc là d'une suite certainement très erratique, sujette à de grandes fluctuations, etc... Toutefois, la moyenne dans le temps de ces

quantités⁽³⁾, c'est-à-dire: $\frac{1}{n} \left(\sum_{k=1}^n X_k \right)$ obéit à des comportements universels, lorsque $n \rightarrow \infty$, qui constituent en fait le coeur - et d'une certaine mesure, une justification - de la théorie des probabilités. Outre l'indépendance, et l'équidistribution, on supposera que les X_i ne sont pas trop grandes, c'est-à-dire que: $E(|X_1|) < \infty$, et même $E(X_1^2) < \infty$. On peut alors énoncer 3 théorèmes "universels", c'est-à-dire valables pour toutes ces suites de variables aléatoires.

Théorème 1. (*Loi des grands nombres*)

$$(5) \quad \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{p.s.} E(X_1)$$

Théorème 2. (*Théorème Central limite*)

$$(6) \quad \sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n X_k - E(X_1) \right) \xrightarrow{(loi)} \mathcal{N}(0, \sigma^2) \text{ où } \sigma^2 = E[(X_1 - E(X_1))^2]$$

Théorème 3. (*Théorème des Grandes déviations*)

On suppose ici que $\varphi(t) = E(e^{tX_1})$ existe pour tout t dans un intervalle ouvert I contenant 0. On fixe $s > 0$ tel que: $s > E(X_1)$ et que $t \rightarrow e^{-ts}\varphi(t)$ atteigne son minimum $\alpha(s)$ en un point τ de I . Alors: ⁽⁴⁾

$$(7) \quad \lim_{n \rightarrow \infty} \left(P \left(\frac{1}{n} \sum_{k=1}^n X_k \geq s \right) \right)^{\frac{1}{n}} = \alpha(s)$$

Commentaires sur ces 3 théorèmes :

- (i) Le théorème 1 dit, en première approximation que la moyenne arithmétique des X_k ($1 \leq k \leq n$) est proche de $E(X_1)$. Le théorème 2 et le théorème 3 essaient de préciser cette proximité: en ce qui concerne le théorème 2, c'est une approximation "au second ordre", un peu comme pour les développements limités; en ce qui concerne le théorème 3, on sait, d'après le théorème 1, que, pour tout $\varepsilon > 0$, on a :

$$P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - E(X_1) \right| \geq \varepsilon \right\} \xrightarrow[n \rightarrow \infty]{} 0$$

et on précise à quelle vitesse cette convergence vers 0 a lieu.

3. Penser aux "firing rates" pour les neurones.

4. Dans son livre, Letac (1997) donne une Table des fonctions $\alpha(s)$ calculées explicitement pour de nombreuses lois de X_1 .

- (ii) A partir de la loi des grands nombres, on obtient, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, borélienne, bornée,

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow \infty]{p.s.} E[f(X_1)]$$

Ce résultat sert de fondement à la théorie statistique : en effet, il permet "d'approximer" la loi de X_1 , à partir des lois empiriques :

$$\frac{1}{n} \sum_{k=1}^n \delta_{X_k(w)} \xrightarrow[n \rightarrow \infty]{p.s.} P_{X_1}$$

où δ_x désigne la mesure de Dirac en x .

- (iii) Noter encore que le théorème 3 demande beaucoup plus d'intégrabilité sur X_1 que les théorèmes 1 et 2. En outre, $\alpha(s)$ dépend "beaucoup" de la loi de X_1 , et pas seulement des deux premiers moments.

(2.2) L'énoncé même de ces théorèmes nécessite la définition précise des types de convergence (5) [p.s = presque sure], et (6) "en loi".

Définition :

- On dit qu'une suite de variables aléatoires (Z_n) converge p.s si

$$P \left\{ w : \lim_{n \rightarrow \infty} Z_n(w) \text{ existe} \right\} = 1$$

- On dit qu'une suite de variables aléatoires Z_n converge en loi si la suite des probabilités sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$: $P_{Z_n} \equiv Z_n(P)$ converge étroitement, c'est-à-dire :

$$\boxed{\forall f \in \mathcal{C}_b(\mathbb{R}), E[f(Z_n)] \text{ converge lorsque } n \rightarrow \infty.}$$

Il existe alors une probabilité μ sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que :

$$(8) \quad Z_n(P) \xrightarrow[n \rightarrow \infty]{\text{étroitement}} \mu$$

De façon générale, notons que montrer une convergence p.s. est souvent assez difficile (: c'est un résultat fin), car la preuve met en jeu toute la structure probabiliste de la suite, alors que montrer une convergence en loi ne met en jeu que les lois unidimensionnelles $Z_n(P)$, ce qui n'implique pas que les résultats de convergence en loi soient des trivialisés, bien sûr !

Un théorème de Paul Lévy affirme que (8) équivaut à la convergence simple des fonctions caractéristiques, i.e :

$$(9) \quad \forall \xi \in \mathbb{R}, E[\exp(i\xi Z_n)] \rightarrow \int \mu(dx) \exp(ix\xi)$$

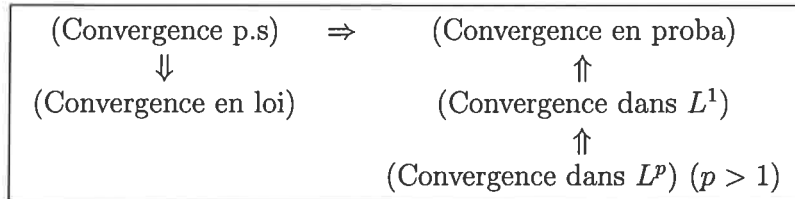
Il existe encore d'autres types de convergence, par exemple la convergence dans L^1 et plus généralement dans L^p : ($p \geq 1$)

$$(10) \quad E[|Z_n - Z|^p] \xrightarrow{(n \rightarrow \infty)} 0,$$

ou encore en probabilité :

$$(11) \quad \forall \varepsilon > 0, P(|Z_n - Z| \geq \varepsilon) \xrightarrow{(n \rightarrow \infty)} 0$$

Notons les implications suivantes :



(2.3) Ebauche de démonstration des théorèmes 1, 2, 3

- (i) En ce qui concerne le Théorème 1, plutôt que de chercher à démontrer le résultat dans toute sa finesse, nous montrons la convergence dans L^2 , sous l'hypothèse : $E(X_1^2) < \infty$.

On a :

$$(12) \quad E \left[\left(\frac{1}{n} \sum_{i=1}^n X_i - E(X_1) \right)^2 \right] = \frac{1}{n} E((X_1 - E(X_1))^2)$$

résultat que l'on obtient très simplement en développant le carré, et en utilisant :

$$0 = E[(X_i - E(X_1))(X_j - E(X_1))] \quad \text{pour } i \neq j.$$

Remarque :

Notons déjà, en amorce de la preuve du Théorème 2, qu'il peut sembler plausible, à partir de (12), que la suite des lois de :

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - E(X_1) \right)$$

converge, lorsque $n \rightarrow \infty$.

(ii) En utilisant le théorème de Paul Lévy, il s'agit de montrer que :

$$(13) \quad E \left[\exp \left(i\xi\sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n X_k - E(X_1) \right) \right) \right] \xrightarrow{n \rightarrow \infty} \exp \left(-\frac{\xi^2 \sigma^2}{2} \right)$$

Or, le terme de gauche est égal à :

$$\left[E \exp \left(i \frac{\xi}{\sqrt{n}} X_1 \right) \right]^n \exp(-i\xi\sqrt{n}E(X_1))$$

et, à l'aide du développement de Taylor à l'ordre 2 de la fonction caractéristique de X_1 , on obtient la convergence vers le membre de droite de (13).

(2.4) En guise d'illustration, voici maintenant une application des théorèmes 1 et 2 : un résultat souvent attribué (à tort, semble-t-il) à H. Poincaré est le suivant : si σ_n désigne la loi uniforme sur la sphère de rayon \sqrt{n} dans \mathbb{R}^n , et si : $\pi_k : \mathbb{R}^n \rightarrow \mathbb{R}^k$ est la projection canonique : $(x_1, \dots, x_n) \rightarrow (x_1, \dots, x_k)$ ($k \leq n$) alors : $\pi_k(\sigma_n) \xrightarrow[n \rightarrow \infty]{\text{étroitement}} G_k$, où G_k désigne la loi gaussienne standard sur \mathbb{R}^k .

Ce résultat assez spectaculaire (: on "voit" la loi de Gauss apparaître comme limite des lois uniformes sur les sphères) peut s'expliquer très simplement à l'aide de la loi des grands nombres (alors que la présence de \sqrt{n} pourrait faire penser a priori que c'est le théorème central limite qui est en jeu...).

En effet, considérons $(N_1, N_2, \dots, N_n, \dots)$ une suite de variables gaussiennes, indépendantes, centrées, réduites, et notons

$$\begin{aligned} \underline{N}_n &= (N_1, \dots, N_n) && , \text{ puis encore :} \\ &= |\underline{N}_n| \Theta_n && = \sqrt{\frac{\sum_{i=1}^n N_i^2}{n}} (\sqrt{n} \Theta_n) \\ &&& \equiv \rho_n (\sqrt{n} \Theta_n) \end{aligned}$$

Par symétrie, $(\sqrt{n} \Theta_n)$ a pour loi $(5) \sigma_n$; d'autre part, d'après la loi des grands nombres, $\frac{1}{n} \sum_{i=1}^n N_i^2$ converge p.s vers 1; ainsi, $\rho_n \xrightarrow{p.s} 1$. On écrit ensuite :

$$\underline{N}_k = \pi_k(\underline{N}_n) = \rho_n \pi_k(\sqrt{n} \Theta_n).$$

D'où :

$$\pi_k(\sqrt{n} \Theta_n) = \frac{1}{\rho_n} \underline{N}_k \xrightarrow[n \rightarrow \infty]{p.s} \underline{N}_k$$

On peut affiner ce résultat à l'aide du théorème central limite.

$$(14) \quad \begin{aligned} \sqrt{n}(\pi_k(\sqrt{n} \Theta_n) - \underline{N}_k) &= \sqrt{n} \left(\frac{1}{\rho_n} - 1 \right) \underline{N}_k \\ &= \sqrt{n} \frac{(1 - \rho_n^2)}{\rho_n(1 + \rho_n)} \underline{N}_k \end{aligned}$$

Or,

$$\rho_n(1 + \rho_n) \xrightarrow[n \rightarrow \infty]{p.s.} 2,$$

et, d'autre part, d'après le théorème central limite :

$$\sqrt{n}(1 - \rho_n^2) \xrightarrow[n \rightarrow \infty]{(loi)} \mathcal{N}(0,3) \stackrel{(loi)}{=} \sqrt{3}\mathcal{N}(0,1)$$

On déduit ensuite de (14) que :

$$(15) \quad \sqrt{n}(\pi_k(\sqrt{n}\Theta_n) - \underline{N}_k) \xrightarrow[n \rightarrow \infty]{(loi)} \left(\frac{\sqrt{3}}{2}\right) \mathcal{N}(0,1) \cdot \underline{N}_k$$

où dans le terme de droite de (15) les variables $\mathcal{N}(0,1)$ et \underline{N}_k sont indépendantes.

Remarquons que la fonction caractéristique (sur \mathbb{R}^k) de $\mathcal{N}(0,1) \cdot \underline{N}_k$ est donnée par : (on note $x \cdot y$ le produit scalaire euclidien dans \mathbb{R}^k)

$$\begin{aligned} E[\exp i\mathcal{N}(\lambda \cdot \underline{N}_k)] &= E \left[\exp \left(-\frac{1}{2}(\lambda \cdot \underline{N}_k)^2 \right) \right] \\ &= E \left[\exp \left(-\frac{|\lambda|^2}{2} N^2 \right) \right] = (1 + |\lambda|^2)^{-\frac{1}{2}} \quad (\lambda \in \mathbb{R}^k). \end{aligned}$$

Mais, on peut également calculer directement la loi de $\mathcal{N}\underline{N}_k$ sur \mathbb{R}^k ; en effet, pour toute $f : \mathbb{R}^k \rightarrow \mathbb{R}$ Borélienne, on a :

$$\begin{aligned} E[f(\mathcal{N}\underline{N}_k)] &= \frac{1}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} dx e^{-\frac{|x|^2}{2}} E[f(\mathcal{N}x)] \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} dy E \left[\frac{1}{|\mathcal{N}|^{\frac{k}{2}}} e^{-\frac{|y|^2}{2\mathcal{N}^2}} \right] f(y) \end{aligned}$$

La densité de la loi de $(\mathcal{N}\underline{N}_k)$ est donc :

$$\begin{aligned} \delta_k(y) &= \frac{1}{(2\pi)^{\frac{k}{2}}} E \left[\frac{1}{|\mathcal{N}|^{\frac{k}{2}}} \exp \left(-\frac{|y|^2}{2\mathcal{N}^2} \right) \right] \\ &= \frac{2}{(2\pi)^{\frac{k+1}{2}}} \int_0^\infty d\nu \frac{e^{-\frac{\nu^2}{2}}}{\nu^{\frac{k}{2}}} e^{-\frac{|y|^2}{2\nu^2}} \end{aligned}$$

Il n'est pas difficile de montrer :

$$E[e^{-\frac{\lambda}{\mathcal{N}^2}}] = e^{-\sqrt{2\lambda}} \quad (\lambda \geq 0)$$

En conséquence, on a, pour p entier :

$$E\left[\frac{1}{\mathcal{N}^{2p}} e^{-\frac{\lambda}{\mathcal{N}^2}}\right] = (-1)^p \frac{d^p}{d\lambda^p} (e^{-\sqrt{2\lambda}})$$

et donc :

$$\delta_{2p}(y) = \frac{1}{(2\pi)^p} (-1)^p \frac{d^p}{d\lambda^p} (e^{-\sqrt{2\lambda}}) \Big|_{\lambda=\frac{|y|^2}{2}}$$

Voir quelques autres calculs en appendice.

(2.5) La notion de martingale

- A une suite $X_1, X_2, \dots, X_n, \dots$ de variables aléatoires, on peut associer la suite croissante des tribus $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$, $n \in \mathbb{N}$ engendrée par la suite (X_i) . On dit que (M_n) suite adaptée à (\mathcal{F}_n) est une $((\mathcal{F}_n), P)$ martingale si :

$$E[M_{n+1} | \mathcal{F}_n] = M_n$$

Cette notion, toute simple, de "jeu équitable", est extrêmement féconde.

- Donnons tout de suite quelques exemples.

Exemple 1 :

$$\sum_{i=1}^n Y_i, \text{ avec les } (Y_i) \text{ indépendantes et centrées.}$$

Exemple 2 :

$$\frac{\exp\{\lambda(X_1 + \dots + X_n)\}}{(\Phi(\lambda))^n}$$

où $\Phi(\lambda) = E[\exp(\lambda X_1)]$, et les (X_i) sont indépendantes et équidistribuées.

- La principale raison pour laquelle cette notion est féconde est le nombre important d'applications du théorème d'arrêt de Doob, qui s'énonce comme suit : si $T : (\Omega, \mathcal{F}) \rightarrow \bar{\mathbb{N}} \equiv \mathbb{N} \cup \{\infty\}$ est un temps d'arrêt, c'est-à-dire qu'il satisfait : $(T \leq n) \in \mathcal{F}_n$, pour tout n , et si la suite $(M_n)_{n \in \mathbb{N}}$ est uniformément intégrable, alors : $E[M_T] = E[M_n]$, pour tout n . Autrement dit, la quantité $E[M_T]$ ne dépend pas de T , temps d'arrêt. Pour une application en temps continu, voir (A.2).

3^{ème} heure :

(3.1) Les considérations précédentes qui portaient sur les suites de variables aléatoires correspondent à un échantillonnage du temps en unités, i.e: la se-

conde, ou la minute, ou le jour, etc... Les techniques et besoins actuels nécessitent bien sûr de pouvoir traiter les phénomènes, aléatoires ou autres, en temps réel; il est donc souhaitable de pouvoir travailler avec des familles $(X_t, t \in \mathbb{R}_+)$ de variables aléatoires, indexées par $t \in \mathbb{R}_+$.

Il faut souligner que, historiquement, ceci a posé des problèmes théoriques formidables, car de façon générale, chaque variable X_t est définie à un ensemble négligeable Λ_t près; ainsi, il y a donc une famille non dénombrable $(\Lambda_t, t \in \mathbb{R}_+)$ de tels ensembles négligeables "gênants", et il faut faire en sorte que l'on puisse se ramener à un seul ensemble négligeable...

Nous allons chercher à donner un exemple naturel en partant là encore d'une suite de variables équidistribuées

$$X_1, \dots, X_n, \text{ avec } E(X_1) = 0, E(X_1^2) = \sigma^2 < \infty.$$

On a alors le théorème de Donsker :

$$(16) \quad \left(\frac{1}{\sqrt{n}} \left(\sum_{k=1}^{[nt]} X_k \right), t \geq 0 \right) \xrightarrow[n \rightarrow \infty]{(d.f)} (B_t, t \geq 0)$$

[(d.f) signifie : convergence en loi, au sens des marginales de rang fini.]

où, à gauche $[x]$ désigne la partie entière de $x \geq 0$, et à droite, $(B_t, t \geq 0)$ désigne le mouvement brownien, c'est-à-dire une famille de variables aléatoires telle que :

$(B_t, t \geq 0)$ est à accroissements indépendants, homogènes dans le temps et gaussiens : si $t_1 < t_2 < \dots < t_k$,

(i) $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}}$, sont des variables indépendantes,

(ii) $B_{t_k} - B_{t_{k-1}} \stackrel{(loi)}{=} B_{(t_k - t_{k-1})}$,

(iii) $B_s \sim \mathcal{N}(0, \sigma^2 s)$.

Il est facile de voir que le théorème (16) est une extension en temps continu du théorème central limite : en effet, si l'on considère : $t_1 < t_2 < \dots < t_k$, les variables :

$$\sum_{j=1}^{[nt_1]} X_j, \sum_{j=[nt_1]+1}^{[nt_2]} X_j, \dots, \sum_{j=[nt_{k-1}]+1}^{[nt_k]} X_j$$

sont indépendantes, et on peut appliquer le théorème central limite à chacune de ces sommes normalisées par \sqrt{n} . Il est bien clair que l'on va ainsi obtenir des lois limites pour chacune de ces composantes, qui sont :

$$\mathcal{N}(0, \sigma^2 t_1), \dots, \mathcal{N}(0, \sigma^2 (t_k - t_{k-1})),$$

et donc une loi limite sur \mathbb{R}^k , qui est la loi d'un vecteur gaussien centré, dont nous venons de décrire les lois unidimensionnelles.

En fait, le résultat (16) peut être affiné de plusieurs manières, par exemple, en montrant l'existence d'un processus

(iv) $(\tilde{B}_t(w); t \geq 0)$ à trajectoires presque sûrement continues tel que :

$$\forall t, P(B_t = \tilde{B}_t) = 1.$$

C'est un tel processus (\tilde{B}_t) qui, outre la propriété de continuité, possède également les propriétés (i), (ii), (iii) ci-dessus que l'on appelle mouvement brownien (de variance σ^2 , disons; le mouvement brownien standard correspond à $\sigma^2 = 1$). C'est Norbert Wiener qui en 1923, a montré l'existence d'une mesure de probabilité, notée maintenant traditionnellement W , sur l'espace canonique des fonctions continues telle que relativement à W , le processus $X_t(w) \equiv w(t)$ obtenu tout simplement en considérant les valeurs dans le temps de la fonction continue générique w satisfasse les propriétés (i), (ii), (iii) et (iv) ci-dessus.

Le mouvement brownien dont nous venons de parler, est certainement le plus célèbre parmi les processus $(\mathcal{L}_t, t \geq 0)$ à accroissements indépendants, et homogènes dans le temps. Depuis un peu plus d'une dizaine d'années, on appelle ces processus processus de Lévy, car c'est à Lévy que l'on doit leur description en toute généralité. Celle-ci peut être présentée par la célèbre formule de Lévy-Khintchine qui exprime la fonction caractéristique de \mathcal{L}_t , à l'instant t :

$$(17) \quad E[\exp(i\xi\mathcal{L}_t)] = \exp(-t\Psi(\xi)),$$

où :

$$\Psi(\xi) = im\xi + \sigma^2 \frac{\xi^2}{2} + \int \nu(dy)(e^{i\xi y} - 1 - i\xi y 1_{(|y| \leq 1)})$$

Comme on le voit, cette fonction ξ - dite souvent fonction ou exposant, de Lévy - fait apparaître

$$3 \text{ paramètres } \begin{cases} m, & \text{appelé dérive;} \\ \sigma^2, & \text{coefficient de diffusion;} \\ \nu(dy), & \text{mesure de Lévy; c'est une mesure } \geq 0 \\ & \text{qui satisfait : } \int \nu(dy)(y^2 \wedge 1) < \infty. \end{cases}$$

A partir de la formule (17), on peut bien entendu présenter la fonction caractéristique conjointe de $(\mathcal{L}_{t_1}, \mathcal{L}_{t_2} - \mathcal{L}_{t_1}, \dots, \mathcal{L}_{t_k} - \mathcal{L}_{t_{k-1}})$ puisque ces accroissements sont indépendants.

Expliquons un peu maintenant la signification de $m, \sigma^2, \nu(dy)$. On peut déjà écrire : $\mathcal{L}_t = mt + \sigma B_t + \tilde{\mathcal{L}}_t$, où (B_t) est un mouvement brownien standard,

et $(\widehat{\mathcal{L}}_t)$ n'a plus pour paramètres que $(0,0,\nu)$, et est indépendant de (B_t) .

Parmi de tels processus $(\widehat{\mathcal{L}}_t)$, sans terme de dérive, ni de partie brownienne, le plus simple d'entre eux est certainement le processus de Poisson de paramètre λ disons; on notera ce processus $(N_t, t \geq 0)$; ses trajectoires sont faciles à décrire, car elles se présentent sous forme de marches d'escalier; plus précisément :

$$N_t = \sum_{n=1}^{\infty} 1_{(T_n \leq t)},$$

où les variables $(T_n, 1 \leq n < \infty)$, qui sont les instants de saut de (N_t) , peuvent être définies au moyen de: $T_n = \varepsilon_1 + \dots + \varepsilon_n$, avec (ε_k) suite des variables exponentielles, indépendantes de paramètre λ . En particulier, pour $\lambda = 1$, T_n est une variable γ_n .

En fait, assez souvent, en travaillant avec une certaine variable aléatoire X_t , indexée par $t > 0$, on s'aperçoit que sa fonction caractéristique s'écrit sous la forme: $\exp(-t\Psi(\xi))$.

En conséquence :

(i) $X_t \stackrel{(loi)}{=} X_{t/n}^{(1)} + X_{t/n}^{(2)} + \dots + X_{t/n}^{(n)}$, où, à droite les variables $(X_{t/n}^{(i)}; 1 \leq i \leq n)$ sont indépendantes et équidistribuées; on dit alors que X_t est indéfiniment divisible

(ii) pour tout t , X_t a donc même loi que la valeur en l'instant t du processus de Lévy \mathcal{L}_t , de fonction de Lévy $\Psi(\xi)$.

(3.2) Les processus de Lévy sont des exemples très particuliers de processus de Markov, notion que je discute maintenant.

On dit qu'un processus $(X_t, t \geq 0)$ est de Markov, ou est Markovien, si pour tout $t > 0$, le passé de X , jusqu'en t , et son futur, après l'instant t , sont indépendants conditionnellement à son état présent X_t , ce qui signifie que pour tous $s_1 < s_2 < \dots < s_k < t < t_1 < t_2 < \dots < t_p$ les variables $(X_{s_1}, X_{s_2}, \dots, X_{s_k})$ et $(X_{t_1}, \dots, X_{t_p})$ sont indépendantes conditionnellement à X_t .

On peut aussi étudier cette propriété de façon asymétrique en considérant

$$E[f(X_{t_1}) | X_{s_1}, X_{s_2}, \dots, X_{s_k}, X_t],$$

et en montrant que cette quantité ne dépend que de X_t (et bien sûr, t_1 , et f), autrement dit qu'elle est égale à :

$$E[f(X_{t_1}) | X_t].$$

Si, en outre, le résultat ne dépend que de $(t_1 - t)$ (et f bien sûr), le processus de Markov est dit homogène dans le temps.

Toutes ces propriétés sont satisfaites par les processus de Lévy, qui satisfont, en outre :

$$E[f(\mathcal{L}_{t_1})|\mathcal{L}_t = x] = E[f(x + \mathcal{L}_{t_1-t})]$$

On voit donc ici que la dépendance par rapport à la position présente x , s'exprime (membre de droite) au moyen de la seule addition de x . Pour un processus de Markov général, la dépendance en x peut être beaucoup plus compliquée.

(3.3) De même que Paul Lévy a pu décrire l'ensemble de tous les processus de Lévy à l'aide de trois paramètres, on peut se proposer de décrire, disons, l'ensemble de tous les processus de Markov homogènes dans le temps. La tâche est beaucoup plus ardue, mais pas impossible, moyennant certaines hypothèses de régularité sur le processus (X_t) .

On parvient en fait souvent à caractériser le processus (X_t) à partir de propriétés de martingales du style : pour $f \in \mathcal{C}_b^2$ (ou même seulement \mathcal{C}_b^∞),

$$\left(C_t^f \stackrel{\text{def}}{=} f(X_t) - f(X_0) - \int_0^t Af(X_s)ds, \quad t \geq 0 \right)$$

est une P_x martingale ⁽⁶⁾, pour toute valeur de départ x . L'opérateur A est, la plupart du temps, un opérateur intégro-différentiel.

$$Af(x) = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial f}{\partial x_i}(x) + \int n(x; dy)(f(y) - f(x) - \nabla f(x) \cdot (y - x))$$

(3.4) Revenons à notre discussion relative au résultat de Poincaré que nous allons démontrer cette fois en utilisant des arguments de processus. Considérons une suite indépendante de mouvements browniens $\overline{B^{(1)}}, \overline{B^{(2)}}, \dots, \overline{B^{(n)}}, \dots$

unidimensionnels, et définissons : $R_n^2(t) = \sum_{i=1}^n (B_t^{(i)})^2$.

En s'appuyant toujours sur la loi des grands nombres, on montre facilement que :

$$\left(\frac{1}{n} R_n^2(t), t \leq T \right) \xrightarrow[n \rightarrow \infty]{p.s.} (t; t \leq T)$$

et ceci pour la convergence uniforme sur $[0, T]$, pour tout $T > 0$. On déduit ensuite de ce résultat que, en notant :

$$H_n = \inf \left\{ t : \frac{1}{n} R_n^2(t) = 1 \right\}, \text{ alors : } H_n \xrightarrow[n \rightarrow \infty]{(P)} 1$$

Ce résultat redonne bien le lemme de Poincaré, puisque, si l'on note :

$$\underline{B}_n(t) = (B_t^{(1)}, \dots, B_t^{(n)}),$$

on a :

$$\underline{B}_n(H_n) \sim \sigma_n, \text{ et donc : } \underline{B}_k(H_n) \sim \pi_k(\sigma_n),$$

et finalement :

$$\pi_k(\sigma_n) \xrightarrow[\text{(étroitement)}]{n \rightarrow \infty} \underline{B}_k(1)$$

Conclusion :

J'ai essayé pendant ces 3 heures, de travailler dans des cadres probabilistes de plus en plus riches. Ceci traduit le fait que l'on peut toujours considérer qu'un phénomène aléatoire est "une partie" d'un phénomène aléatoire qui "englobe" le premier... On peut apprendre sur le "petit" phénomène à partir du "grand", et inversement intuire des résultats sur le "grand" à partir de certains résultats (plus élémentaires) connus sur le "petit". Pour de nombreux exemples d'applications, on pourra consulter :

- L. Chaumont, M. Yor : Exercices in Probability, CUP (2003).

Un modèle pour l'élaboration de ce livre a été :

- G. Letac : Intégration et Probabilités; Analyse de Fourier. Masson (2^e édition), 1997. Cf. Exercice 410, p. 96.

Ce livre d'exercices est à rapprocher de :

- P.S. Toulouse : Thèmes de Probabilités et Statistiques; Dunod (1999)

que j'ai beaucoup utilisé dans mes cours de préparation à l'Agrégation de Mathématiques - Option Probabilités et Statistiques.

Mais, il y a bien sûr beaucoup d'autres livres d'introduction aux probabilités; citons par exemple :

- J. Pitman : Probability; Springer (1985).
- D. Williams : Weighing the Odds, CUP (2003);

où l'auteur essaie de réconcilier Probabilités et Statistique.

A lire également, le remarquable :

- D. Williams : Probability with Martingales; CUP (1991).

Là, le leitmotiv est : "Chercher la martingale", qui, bien choisie, doit permettre de résoudre tel ou tel problème de probabilités.

Appendice : Autour de la loi de $T \stackrel{(loi)}{=} \frac{1}{\mathcal{N}^2}$

(A.1) Nous montrons ici que l'on peut faire le calcul du terme de gauche de la formule (cf. fin de 2^{ème} heure) :

$$\gamma \stackrel{def}{=} E \left[\frac{1}{\mathcal{N}^{2p}} \exp \left(\frac{\lambda}{\mathcal{N}^2} \right) \right] = (-1)^p \frac{d^p}{d\lambda^p} \left(e^{-\sqrt{2\lambda}} \right)$$

par intégration.

En effet, on a :

$$\gamma = \frac{1}{\Gamma(p)} \int_0^\infty d\mu \mu^{p-1} E \left[\exp \left(-\mu \mathcal{N}^2 - \frac{\lambda}{\mathcal{N}^2} \right) \right]$$

(on a utilisé l'astuce de calcul : $\frac{1}{r^p} = \frac{1}{\Gamma(p)} \int_0^\infty d\mu \mu^{p-1} e^{-\mu r}$ ($r > 0$))

Or, on a :

$$\begin{aligned} E \left[\exp \left(-\mu \mathcal{N}^2 - \frac{\lambda}{\mathcal{N}^2} \right) \right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty dx e^{-\frac{x^2}{2}(1+2\mu) - \frac{\lambda}{x^2}} \\ &= \frac{1}{\sqrt{2\pi(1+\mu)}} \int_{-\infty}^\infty dy e^{-\frac{y^2}{2} - \frac{\lambda(1+2\mu)}{y^2}} \\ &= \frac{1}{\sqrt{1+2\mu}} e^{-\sqrt{2\lambda(1+2\mu)}} \end{aligned}$$

où l'on a utilisé : $E \left[\exp \left(-\frac{\alpha}{\mathcal{N}^2} \right) \right] = \exp(-\sqrt{2\alpha})$ ($\alpha \geq 0$)

On a donc :

$$\gamma = \frac{1}{\Gamma(p)} \int_0^\infty d\mu \mu^{p-1} \frac{e^{-\sqrt{2\lambda(1+2\mu)}}}{\sqrt{1+2\mu}} = \frac{1}{2^p \Gamma(p)} \int_1^\infty \frac{dx}{\sqrt{x}} (x-1)^{p-1} e^{-\sqrt{2\lambda x}}$$

On a ainsi obtenu ici une formule intégrale pour γ , que l'on peut bien sûr encore transformer, à l'aide de : $x = y^2$, etc...

(A.2) La variable $T \stackrel{(loi)}{=} \frac{1}{\mathcal{N}^2}$ intervient très souvent dans l'étude du mouvement brownien, car on a :

$$T \stackrel{(loi)}{=} \sigma \stackrel{def}{=} \inf \{ t : B_t = 1 \},$$

comme l'application du théorème d'arrêt de Doob à la martingale :

$$\exp \left\{ \lambda B_t - \frac{\lambda^2}{2} t \right\}$$

le montre : cette martingale, qui est une version continue de celles de l'Exemple 2 (fin de 2^{ème} heure), arrêtée en σ , est bornée, et donc :

$$1 = E \left[\exp \left\{ \lambda B_\sigma - \frac{\lambda^2}{2} \sigma \right\} \right] = (\exp \lambda) E \left[\exp \left(-\frac{\lambda^2}{2} \sigma \right) \right],$$

d'où le résultat.

(A.3) Une famille de variables à valeurs entières

On va décrire ici les lois des variables ($N_\alpha^{(u)}$) à valeurs dans \mathbb{N} , dont la fonction génératrice est :

$$(18) \quad E \left[a^{N_\alpha^{(u)}} \right] = \left(\frac{1-u}{1-ua} \right)^\alpha \quad (0 < a < 1)$$

On a, pour tous $\alpha, \beta, u > 0$: $N_\alpha^{(u)} + \tilde{N}_\beta^{(u)} \stackrel{(loi)}{=} N_{\alpha+\beta}^{(u)}$, où, dans le membre de gauche, $N_\alpha^{(u)}$ et $\tilde{N}_\beta^{(u)}$ sont indépendantes. En écrivant, le membre de droite de (18) sous forme exponentielle, on a :

$$\begin{aligned} & \exp \alpha (\log(1-u) - \log(1-ue^{-\lambda})) \\ &= \exp -\alpha \int \nu(dx) (1 - e^{-\lambda x}), \end{aligned}$$

soit $\frac{ue^{-\lambda}}{1-ue^{-\lambda}} = \sum_{n=1}^{\infty} u^n e^{-\lambda n}$, d'où : $\nu(dx) = \sum_{n=1}^{\infty} u^n \delta_n(dx)$. Et on trouve finalement que ($N_\alpha^{(u)}, \alpha \geq 0$) est un processus de Poisson composé c'est-à-dire que l'on peut écrire :

$$(19) \quad \left(N_\alpha^{(u)}, \alpha \geq 0 \stackrel{(loi)}{=} \sum_{i=0}^{N_\alpha^{(\lambda)}} X_i, \alpha \geq 0 \right)$$

où, dans le membre de droite, les variables (X_i) sont indépendantes, et équidistribuées, et indépendantes du processus de Poisson $N_\alpha^{(\lambda)}$ de paramètre λ . Il s'agit de trouver μ , la loi commune des (X_i) et le paramètre λ ; ceux-ci sont obtenus de la façon suivante :

$$(20) \quad \nu(dx) = \lambda \mu(dx)$$

Or, d'après la formule ci-dessus, on a :

$$\lambda = \sum_{n=1}^{\infty} u^n = \frac{u}{1-u}$$

et donc :

$$\mu(dx) = \left(\frac{1-u}{u} \right) \sum_{n=1}^{\infty} u^n \delta_n(dx),$$

c'est-à-dire

$$P(X_i = n) = \frac{(1-u)}{u} u^n \quad (n \geq 1),$$

et donc $(X_i - 1)$ est une variable géométrique de paramètre u .

Noter que, parmi les processus de Lévy, les processus de Poisson composés constituent le sous ensemble dont les mesures de Lévy sont finies, et qui n'ont ni terme brownien, ni terme de dérive.