

SÉMINAIRE DE PROBABILITÉS (STRASBOURG)

PIERRE CARTIER

Sur certaines variables aléatoires associées au réarrangement croissant d'un échantillon

Séminaire de probabilités (Strasbourg), tome 4 (1970), p. 28-36

http://www.numdam.org/item?id=SPS_1970__4__28_0

© Springer-Verlag, Berlin Heidelberg New York, 1970, tous droits réservés.

L'accès aux archives du séminaire de probabilités (Strasbourg) (<http://portail.mathdoc.fr/SemProba/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE

Laboratoire Associé au C.N.R.S.

Rue René Descartes

STRASBOURG

1968-69

- Séminaire de Probabilités -

SUR CERTAINES VARIABLES ALÉATOIRES

ASSOCIÉES AU RÉARRANGEMENT CROISSANT D'UN ÉCHANTILLON

par P. CARTIER

I. RÉARRANGEMENTS CROISSANTS.

On suppose donnés un espace probabilisé (Ω, \mathcal{F}, P) et une suite finie de variables aléatoires réelles X_1, \dots, X_n (avec $n \geq 2$), en dépendance symétrique. Pour toute suite a_1, \dots, a_n de nombres réels, on note $\bar{a}_1, \dots, \bar{a}_n$ son réarrangement croissant : c'est la suite croissante (au sens large) qui ne diffère de la suite a_1, \dots, a_n que par l'ordre des termes. Le réarrangement croissant $\bar{X}_1, \dots, \bar{X}_n$ est alors la suite de variables aléatoires telle que $\bar{X}_1(\omega), \dots, \bar{X}_n(\omega)$ soit le réarrangement croissant de $X_1(\omega), \dots, X_n(\omega)$ pour tout $\omega \in \Omega$.

THÉOREME 1.

Soit b une fonction borélienne réelle sur \mathbb{R}^2 , telle que $b(x, y) = 0$ pour $x \geq y$. Posons

$$U = \sum_{i=1}^n b(\bar{X}_i, X_i) \quad \text{et} \quad V = \sum_{i=1}^{n-1} b(X_i, X_{i+1}) \quad .$$

Les variables aléatoires U et V ont même loi.

Comme on sait, il suffit de prouver que U et V ont même fonction caractéristique. Nous fixerons dans la suite un nombre réel t , et nous établirons l'égalité $E[e^{itU}] = E[e^{itV}]$. On note c la fonction borélienne complexe e^{itb} sur \mathbb{R}^2 et l'on pose

$$(1) \quad f(x_1, \dots, x_n) = \prod_{i=1}^n c(\bar{x}_i, x_i)$$

pour x_1, \dots, x_n réels. Alors f est une fonction borélienne complexe bornée sur \mathbb{R}^n et l'on a $e^{itU} = f(X_1, \dots, X_n)$. De manière analogue, on définit une fonction borélienne complexe bornée g sur \mathbb{R}^n par

$$(2) \quad g(x_1, \dots, x_n) = \prod_{i=1}^{n-1} c(x_i, x_{i+1}) \quad ;$$

il est clair qu'on a $e^{itV} = g(X_1, \dots, X_n)$.

On notera $[n]$ l'ensemble des entiers de 1 à n et \mathfrak{S}_n l'ensemble des permutations de $[n]$. Pour toute permutation $\sigma \in \mathfrak{S}_n$, le vecteur aléatoire $X_{\sigma(1)}, \dots, X_{\sigma(n)}$ a par hypothèse la même loi que X_1, \dots, X_n ; on a donc

$$(3) \quad E[f(X_1, \dots, X_n)] = E[f(X_{\sigma(1)}, \dots, X_{\sigma(n)})] \quad ,$$

et par sommation sur les $n!$ permutations $\sigma \in \mathfrak{S}_n$, on en déduit

$$(4) \quad E[e^{itU}] = \frac{1}{n!} E\left[\sum_{\sigma \in \mathfrak{S}_n} f(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \right] \quad .$$

On a évidemment une formule analogue où U et f sont remplacés respectivement par V et g . Pour établir l'égalité $E[e^{itU}] = E[e^{itV}]$, il suffit donc d'établir l'identité

$$(5) \quad \sum_{\sigma \in \mathfrak{S}_n} f(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = \sum_{\sigma \in \mathfrak{S}_n} g(x_{\sigma(1)}, \dots, x_{\sigma(n)}) \quad ,$$

où x_1, \dots, x_n sont des nombres réels arbitraires.

Transformons le premier membre de (5). Pour toute permutation σ , le réarrangement croissant de $x_{\sigma(1)}, \dots, x_{\sigma(n)}$ coïncide évidemment avec le réarrangement croissant $\bar{x}_1, \dots, \bar{x}_n$ de x_1, \dots, x_n ; de plus, il existe une permutation ρ telle que $\bar{x}_i = x_{\rho(i)}$ pour $1 \leq i \leq n$. Le premier membre de (5) est donc égal à $\sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^n c(x_{\rho(i)}, x_{\sigma(i)})$; en faisant le changement d'indices de sommation $i = \rho^{-1}(j)$, $\sigma = \tau\rho$, on obtient $\sum_{\sigma \in \mathfrak{S}_n} \prod_{j=1}^n c(x_j, x_{\tau(j)})$. La formule (5) peut donc s'écrire sous la forme

$$(6) \quad \sum_{\tau \in \mathfrak{S}_n} \prod_{i=1}^n c(x_i, x_{\tau(i)}) = \sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^{n-1} c(x_{\sigma(i)}, x_{\sigma(i+1)}) .$$

Nous allons maintenant démontrer la formule (6) par un argument combinatoire. Soit $\tau \in \mathfrak{S}_n$, et soient C_1, \dots, C_p les cycles de τ ; notons i_k le plus grand élément du cycle C_k ; on peut supposer que l'on a $i_1 > \dots > i_p$, ce qui détermine sans ambiguïté la numérotation des cycles C_1, \dots, C_p . Si m_k est le nombre d'éléments de C_k , les éléments de C_k sont de la forme $\tau^j(i_k)$ pour $1 \leq j \leq m_k$, et l'on a $\tau^{m_k}(i_k) = i_k$. Considérons maintenant la suite

$$(S) \quad \tau(i_1), \tau^2(i_1), \dots, \tau^{m_1}(i_1), \tau(i_2), \dots, \tau^{m_2}(i_2), \dots, \tau(i_p), \dots, \tau^{m_p}(i_p) ;$$

on l'obtient en énumérant successivement les éléments des cycles C_1, \dots, C_p de sorte que le plus grand élément de chaque cycle se trouve le dernier du cycle.

Les éléments de la suite (S) sont distincts; il existe donc une permutation $\sigma \in \mathfrak{S}_n$ bien définie par le fait que la suite (S) soit identique à la suite $\sigma(1), \dots, \sigma(n)$. Nous allons maintenant montrer que la correspondance entre τ et σ est bijective; comme \mathfrak{S}_n est un ensemble fini, il suffit de montrer que τ est déterminée par la permutation associée σ . Il revient au

même de dire que la suite (S) détermine sans ambiguïté les entiers m_1, \dots, m_p , car on pourra alors reconstituer les différents cycles de τ à partir de la suite (S). Appelons S_1, \dots, S_p les portions successives de (S) correspondant aux cycles C_1, \dots, C_p de τ , de sorte que S_k a m_k éléments. Le dernier élément de S_k est i_k , et tout élément de S_k distinct de i_k est majoré par i_k , qui se trouve à sa droite dans (S). Par ailleurs, i_k majore i_{k+1}, \dots, i_p et a fortiori les éléments des diverses suites S_{k+1}, \dots, S_p . Appelons dominants les éléments de la suite (S) qui majorent tous les termes de (S) situés à leur droite ; ce qui précède montre que les termes dominants de (S) sont i_1, \dots, i_p , qui sont de rangs respectifs $m_1, m_1 + m_2, \dots, m_1 + \dots + m_p$. On a montré que les entiers m_1, \dots, m_p sont bien déterminés par (S).

Soient τ et σ comme précédemment. On fait l'hypothèse $x_1 \leq \dots \leq x_n$. Le produit $\prod_{i=1}^n c(x_i, x_{\tau(i)})$ est le produit des p contributions F_1, \dots, F_p correspondant aux p cycles C_1, \dots, C_p , et l'on a évidemment

$$(7) \quad F_k = \prod_{j=1}^{k-1} c(x_{\tau^j(i_k)}, x_{\tau^{j+1}(i_k)}) \times c(x_{i_k}, x_{\tau(i_k)}) \quad .$$

Or, on a $i_k \geq \tau(i_k)$ d'où $x_{i_k} \geq x_{\tau(i_k)}$, et l'hypothèse faite sur b entraîne $c(x, y) = 1$ lorsque $x \geq y$ (le cas d'égalité $i_k = \tau(i_k)$ se produit dans le cas des cycles à un élément, et ce cas seulement). Le dernier facteur dans F_k est donc égal à 1. Par ailleurs, la définition de σ à partir de la suite (S) entraîne

$$(8) \quad \prod_{i=1}^{n-1} c(x_{\sigma(i)}, x_{\sigma(i+1)}) = F_1 \dots F_p c(x_{i_1}, x_{\sigma(i_2)}) \dots c(x_{i_{p-1}}, x_{\sigma(i_p)}) \quad .$$

Pour $k = 1, \dots, p-1$, on a $i_k > i_{k+1} \geq \sigma(i_{k+1})$, d'où $x_{i_k} \geq x_{\sigma(i_{k+1})}$ et

finalement $c(x_{i_k}, x_{\sigma(i_{k+1})}) = 1$. On a donc

$$(9) \quad \prod_{i=1}^n c(x_{i_k}, x_{\tau(i)}) = F_1 \dots F_p = \prod_{i=1}^{n-1} c(x_{\sigma(i)}, x_{\sigma(i+1)}) .$$

Sous l'hypothèse faite $x_1 \leq x_2 \leq \dots \leq x_n$, les termes des deux sommes dans (6) se correspondent donc un à un, d'où l'égalité cherchée. Comme les deux membres de (6) sont évidemment des fonctions symétriques des n variables réelles x_1, \dots, x_n , ceci suffit à entraîner l'égalité dans le cas général. Nous avons vu précédemment comment déduire le théorème 1 de l'égalité (5) , donc de (6) .

2. UNE IDENTITÉ EXPONENTIELLE SUR LES SÉRIES GÉNÉRATRICES.

On suppose maintenant donnée une suite infinie $(X_n)_{n \geq 1}$ de variables aléatoires réelles mutuellement indépendantes et de même loi ; on se donne aussi une fonction borélienne réelle b sur \mathbb{R}^2 , mais on ne fait plus l'hypothèse que l'on a $b(x, y) = 0$ pour $x \geq y$. On définit alors deux suites de variables aléatoires réelles $(Y_n)_{n \geq 1}$ et $(Z_n)_{n \geq 1}$ par les formules

$$(10) \quad Y_n = \sum_{i=1}^n b(\bar{X}_i^{(n)}, X_i)$$

$$(11) \quad Z_n = b(X_1, X_2) + \dots + b(X_{n-1}, X_n) + b(X_n, X_1)$$

où $\bar{X}_1^{(n)}, \dots, \bar{X}_n^{(n)}$ désigne le réarrangement croissant de la suite finie

X_1, \dots, X_n . La formule (11) n'a de sens que pour $n \geq 2$; on posera par convention $Z_1 = 0$. Par ailleurs, on note F_n la fonction caractéristique de Y_n et G_n celle de Z_n .

THÉOREME 2.

Pour tout nombre réel t et tout nombre complexe u avec $|u| < 1$,

on a

$$(12) \quad 1 + \sum_{n \geq 1} u^n F_n(t) = \exp \sum_{k \geq 1} \frac{u^k}{k} G_k(t) \quad .$$

On fixe t et u comme dans l'énoncé, et l'on note c la fonction borélienne complexe bornée e^{itb} sur \mathbb{R}^2 . Pour tout entier $n \geq 1$, on définit une fonction borélienne complexe bornée f_n sur \mathbb{R}^n et une fonction complexe h_n sur \mathfrak{S}_n par les formules

$$(13) \quad f_n(x_1, \dots, x_n) = \prod_{i=1}^n c(\bar{x}_i, x_i)$$

$$(14) \quad h_n(\sigma) = E \left[\prod_{i=1}^n c(X_i, X_{\sigma(i)}) \right] \quad .$$

Les calculs faits page 2, en particulier la formule (4) et la transformation du premier membre de (5), établissent la relation

$$(15) \quad F_n(t) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} h_n(\sigma) \quad .$$

Notons γ_n la permutation circulaire $1 \rightarrow 2 \rightarrow 3 \rightarrow \dots \rightarrow n-1 \rightarrow n \rightarrow 1$; on a

$$(16) \quad e^{itZ_n} = c(X_1, X_2) \dots c(X_{n-1}, X_n) c(X_n, X_1) = \prod_{i=1}^n c(X_i, X_{\gamma_n(i)}) \quad ,$$

d'où

$$(17) \quad G_n(t) = h_n(\gamma_n) \quad .$$

La suite de la démonstration va résulter d'une analyse des propriétés des fonctions h_n .

a) On a $h_n(\sigma \tau \sigma^{-1}) = h_n(\tau)$ quels que soient σ et τ dans \mathfrak{S}_n :

cela résulte du calcul suivant

$$\begin{aligned} h_n(\sigma\tau\sigma^{-1}) &= E\left[\prod_{i=1}^n c(X_i, X_{\sigma\tau\sigma^{-1}(i)})\right] = E\left[\prod_{j=1}^n c(X_{\sigma(j)}, X_{\sigma\tau(j)})\right] \\ &= E\left[\prod_{j=1}^n c(X'_j, X'_{\tau(j)})\right] = E\left[\prod_{j=1}^n c(X_j, X_{\tau(j)})\right] = h_n(\tau) . \end{aligned}$$

On passe de la deuxième expression à la troisième par le changement d'indice $i = \sigma(j)$. On a posé $X'_j = X_{\sigma(j)}$; la quatrième égalité résulte alors de ce que les vecteurs aléatoires X_1, \dots, X_n et X'_1, \dots, X'_n ont même loi, puisque les variables aléatoires X_1, \dots, X_n sont indépendantes et de même loi.

b) On a $h_{m+n}(\rho \oplus \sigma) = h_m(\rho) \cdot h_n(\sigma)$ pour ρ dans \mathfrak{S}_m et σ dans \mathfrak{S}_n : la permutation $\rho \oplus \sigma$ appartient à \mathfrak{S}_{m+n} ; elle est définie comme la permutation τ telle que $\tau(i) = \rho(i)$ pour $1 \leq i \leq m$ et $\tau(m+j) = m + \sigma(j)$ pour $1 \leq j \leq n$. On a

$$\begin{aligned} (18) \quad h_{m+n}(\rho \oplus \sigma) &= E\left[\prod_{i=1}^m c(X_i, X_{\rho(i)}) \times \prod_{j=1}^n c(X_{m+j}, X_{m+\sigma(j)})\right] \\ &= E\left[\prod_{i=1}^m c(X_i, X_{\rho(i)})\right] \times E\left[\prod_{j=1}^n c(X_{m+j}, X_{m+\sigma(j)})\right] \end{aligned}$$

puisque les vecteurs aléatoires X_1, \dots, X_m et X_{m+1}, \dots, X_{m+n} sont indépendants. Mais comme le vecteur aléatoire X_{m+1}, \dots, X_{m+n} a même loi que X_1, \dots, X_n , le produit précédent est égal à $h_m(\rho) \cdot h_n(\sigma)$.

Ces points étant acquis, fixons l'entier $n \geq 1$. Si p_1, \dots, p_n sont des entiers positifs tels que $1 \cdot p_1 + 2 \cdot p_2 + \dots + n \cdot p_n = n$, il existe dans \mathfrak{S}_n un nombre égal à $n! / \prod_{k=1}^n (k^{p_k} \cdot p_k!)$ de permutations dont la décomposition en cycles comprenne p_1 cycles de longueur 1, ..., p_n cycles de longueur n . Si σ est l'une de ces permutations, il existe alors une permutation τ dans \mathfrak{S}_n telle que

$$\tau \sigma \tau^{-1} = \underbrace{\gamma_1 \oplus \dots \oplus \gamma_1}_{p_1} \oplus \underbrace{\gamma_2 \oplus \dots \oplus \gamma_2}_{p_2} \oplus \dots \oplus \underbrace{\gamma_n \oplus \dots \oplus \gamma_n}_{p_n} ;$$

d'après les propriétés a) et b), on a alors $h_n(\sigma) = \prod_{k=1}^n h_k(\gamma_k)^{p_k}$. On en déduit

$$(19) \quad \frac{1}{n!} \sum_{\sigma \in S_n} h_n(\sigma) = \sum_{1p_1 + \dots + np_n = n} \prod_{k=1}^n \frac{1}{p_k!} \left(\frac{h_k(\gamma_k)}{k} \right)^{p_k},$$

et d'après (15) et (17), on a donc

$$(20) \quad F_n(t) = \sum_{1p_1 + \dots + np_n = n} \prod_{k=1}^n \frac{1}{p_k!} \left(\frac{G_k(t)}{k} \right)^{p_k}.$$

Par sommation sur n , on déduit de (20) la relation

$$(21) \quad 1 + \sum_{n \geq 1} u^n F_n(t) = \sum u^{\sum k p_k} \prod_{k \geq 1} \left(\frac{G_k(t)}{k} \right)^{p_k},$$

où la sommation est étendue à toutes les suites (p_1, p_2, \dots) d'entiers positifs, dont un nombre fini seulement sont non nuls. D'après la formule de développement d'un produit infini de séries, le second membre de (21) est égal à

$$\prod_{k \geq 1} \sum_{p \geq 0} \frac{1}{p!} \left(\frac{u^k G_k(t)}{k} \right)^p = \prod_{k \geq 1} \exp \frac{u^k}{k} G_k(t) = \exp \sum_{k \geq 1} \frac{u^k}{k} G_k(t) ;$$

les séries et les produits infinis écrits convergent absolument en vertu des inégalités $|u| < 1$ et $|G_k(t)| \leq 1$, ce qui justifie les calculs formels ci-dessus. On a prouvé le théorème 2.

3. COMPLÉMENTS.

Nous conservons les notations du n° 2, et nous définissons les variables aléatoires T_n par $T_1 = 0$ et $T_n = \sum_{i=1}^{n-1} b(X_i, X_{i+1})$ pour $n \geq 2$. On note H_n la fonction caractéristique de T_n .

THÉOREME 3.

Sous l'hypothèse $b(x, y) = 0$ pour $x \geq y$, on a la relation

$$(22) \quad 1 + \sum_{n \geq 1} u^n H_n(t) = \exp \sum_{k \geq 1} \frac{u^k}{k} G_k(t)$$

lorsque $|u| < 1$, pour tout nombre réel t .

Le théorème 1 montre que Y_n et T_n ont même loi, d'où $F_n = H_n$ (on utilise ici l'hypothèse sur b). Il suffit alors d'appliquer le théorème 2.

4. INDICATIONS BIBLIOGRAPHIQUES.

Les théorèmes 1 et 3 sont dus à Foata. Dans sa thèse (Publ. Inst. Stat. Univ. Paris, 14 (1965), p. 81-241), il les a déduit de résultats combinatoires assez délicats. Ces derniers ont été simplifiés et généralisés dans un travail commun de Foata et moi-même (Lecture Notes in Mathematics, n° 85, Springer (1969)); nous en avons déduit les théorèmes 1 à 3. De ces théorèmes combinatoires, il ne reste ici que la construction de la bijection $\sigma \mapsto \tau$ de \mathfrak{S}_n dans lui-même étudiée à la page 3. Pour toute permutation $\sigma \in \mathfrak{S}_n$, soit $v(\sigma)$ le nombre des entiers $i \in [n]$ tels que $i < \sigma(i)$, et $\xi(\sigma)$ le nombre des entiers $i \in [n-1]$ tels que $\sigma(i) < \sigma(i+1)$; avec les notations de la page 3, on a $v(\sigma) = \xi(\tau)$, ce qui s'établit immédiatement par les considérations de l'endroit cité. On en déduit qu'il existe, pour tout entier positif m , autant de permutations σ avec $v(\sigma) = m$ que de permutations σ avec $\xi(\sigma) = m$. C'est là un théorème de P. Mac Mahon (voir son livre : Combinatory Analysis, Chelsea, New-York (1960), à la page 186 du volume I).