Séminaire BOURBAKI Janvier 2017
69e année, 2016–2017, no 1124, p. 117 à 134
doi:10.24033/ast.1062

L'INÉGALITÉ DE CORRÉLATION GAUSSIENNE [d'après Thomas Royen]

par **Franck BARTHE**

INTRODUCTION

Les mesures gaussiennes jouent un rôle « central » en théorie des probabilités. Leurs propriétés géométriques et analytiques ont été largement étudiées, et mises en application dans des domaines variés (théorie des processus, statistiques, géométrie des espaces de Banach, théorie des algorithmes, ...). De nombreuses techniques sont maintenant disponibles pour les établir (symétrisation d'Ehrhard, interpolation par le semigroupe d'Ornstein-Uhlenbeck, formules issues du calcul stochastique, transport optimal, entre autres), mais elles semblent plus efficaces pour les propriétés qui font intervenir tous les ensembles mesurables, comme l'isopérimétrie gaussienne. Elles n'ont pas permis, jusqu'à présent, d'établir la célèbre conjecture de corrélation gaussienne, qui fait intervenir des ensembles convexes symétriques. La conjecture est devenue un théorème en 2014, grâce au travail remarquable de Royen :

Théorème 0.1 ([**16**]). — *Soit $n \geq 1$. Soit γ une mesure de probabilité gaussienne sur $\mathbb{R}^n$, de moyenne 0. Pour tous les sous-ensembles $A, B \subset \mathbb{R}^n$ convexes et symétriques par rapport à l'origine, on a*

$$\gamma(A \cap B) \geq \gamma(A)\gamma(B). \tag{1}$$

Royen prouve en fait un résultat pour la famille plus vaste des lois gamma multivariées au sens de Krishnamoorthy et Parthasarathy. Même dans le cas particulier gaussien, sa démonstration utilise de manière astucieuse les propriétés de ces lois (et s'adresse à des lecteurs qui en sont spécialistes). Latała et Matlak [**12**] ont écrit une version auto-contenue et détaillée de l'argument de Royen dans le cas des mesures gaussiennes.

1. FORMULATIONS ÉQUIVALENTES ET BREF HISTORIQUE

En jouant sur les transformations linéaires, on peut donner des formulations équivalentes de la corrélation gaussienne, formellement plus faibles, et ce de deux manières.

Comme toute mesure gaussienne centrée sur $\mathbb{R}^n$ est image linéaire de la mesure gaussienne standard de dimension n,

$$\gamma_n(dx) = e^{-|x|^2/2}\frac{dx}{(2\pi)^{n/2}},$$

le théorème 0.1 revient à

$$\gamma_n(A\cap B) \geq \gamma_n(A)\gamma_n(B)$$

pour tous les convexes symétriques de $\mathbb{R}^n$. Cet énoncé a souvent été formulé en termes fonctionnels, afin d'introduire des techniques analytiques. Toute fonction mesurable $f : \mathbb{R}^n \to \mathbb{R}_+$ peut s'écrire sous la forme

$$f(x) = \int_0^{+\infty} \mathbf{1}_{A_t}(x)\, dt,$$

où les $A_t := \{y \in \mathbb{R}; f(y) \geq t\}$ sont les ensembles de niveaux de f. On dit que f est quasi-concave si ses ensembles de niveau sont tous convexes. On obtient la formulation suivante de la corrélation gaussienne :

Théorème 1.1. — *Soit $n \geq 1$. Soient $f, g : \mathbb{R}^n \to \mathbb{R}_+$ des fonctions paires et quasi-concaves. Alors*

$$\int fg\, d\gamma_n \geq \int f\, d\gamma_n \int g\, d\gamma_n.$$

Dans une autre direction, on peut conserver une mesure gaussienne générale sur $\mathbb{R}^d$ et simplifier les ensembles convexes à traiter. Par approximation, il suffit de considérer des polytopes convexes symétriques, qui sont des intersections de bandes symétriques

$$A = \bigcap_{i=1}^{n_1} \{x \in \mathbb{R}^d; |\langle x, u_i\rangle| \leq t_i\}, \quad B = \bigcap_{i=1+n_1}^{n} \{x \in \mathbb{R}^d; |\langle x, u_i\rangle| \leq t_i\},$$

pour n entier, des $t_i \geq 0$ et des vecteurs unitaires $u_i \in \mathbb{R}^d$. Si Y est un vecteur aléatoire sur $\mathbb{R}^d$ de loi gaussienne γ, l'inégalité $\gamma(A\cap B) \geq \gamma(A)\gamma(B)$ revient à

$$\mathbb{P}\big(\forall i, |\langle Y, u_i\rangle| \leq t_i\big) \geq \mathbb{P}\big(\forall i \leq n_1, |\langle Y, u_i\rangle| \leq t_i\big)\, \mathbb{P}\big(\forall i > n_1, |\langle Y, u_i\rangle| \leq t_i\big),$$

où l'on a sous-entendu la condition $i \in \{1, \ldots, n\}$. Si $t_i \neq 0$, on définit $X_i = \langle Y, u_i\rangle / t_i$. Si $t_i = 0$, comme $\langle Y, u_i\rangle$ est une variables aléatoire gaussienne centrée, $\mathbb{P}\big(|\langle Y, u_i\rangle| = 0\big)$ peut seulement valoir 0 ou 1. Dans le premier cas l'inégalité de corrélation est évidente, dans le second on peut enlever la condition toujours vraie. On obtient ainsi un vecteur gaussien X. En récrivant l'inégalité précédente en fonction de X on arrive à une autre formulation équivalente du théorème 0.1 :

Théorème 1.2. — *Soient $n \geq n_1 \geq 1$ des entiers et $X = (X_1, \ldots, X_n)$ un vecteur gaussien centré (i.e., d'espérance nulle) à valeurs dans $\mathbb{R}^n$. Alors*

$$\mathbb{P}\Big(\max_{1\leq i\leq n} |X_i| \leq 1\Big) \geq \mathbb{P}\Big(\max_{1\leq i\leq n_1} |X_i| \leq 1\Big)\, \mathbb{P}\Big(\max_{n_1<i\leq n} |X_i| \leq 1\Big).$$

D'après [**2**], l'histoire de ce problème remonte à un article de Dunnett et Sobel datant de 1955 [**3**]. Elle a été marquée par les travaux indépendants de Khatri [**8**] et Šidák [**21**] en 1967, qui établissent le théorème 1.2 lorsque $n_1 = 1$, ce qui revient à (1) lorsque l'un des convexes est une bande symétrique de la forme $\{x;\ |\langle x, u\rangle| \leq 1\}$. En 1977, Pitt [**13**] a démontré l'inégalité de corrélation (1) en dimension 2, voir [**1**] pour une extension. Une étape importante vers la confirmation de la conjecture est franchie dans l'article de Schechtman, Schlumprecht et Zinn [**18**], qui établit (1) pour les couples d'ellipsoïdes, et pour les ensembles contenus dans la boule centrée à l'origine et de rayon $c\sqrt{n}$ pour une certaine constante c. Plus récemment, des progrès substantiels ont été effectués par Hargé : il a établi (1) lorsqu'un seul des convexes est un ellipsoïde [**6**], puis il a montré une inégalité de corrélation négative entre une fonction paire convexe et une fonction paire log-concave, ce qui est un cas limite du théorème 1.1 lorsque l'une des fonctions est une perturbation d'une constante [**7**]. On pourrait citer bien d'autres travaux, dont certains ont exploré l'affaiblissement des hypothèses (mesures non gaussiennes, ensembles non convexes, notions de centrage plus faibles que la symétrie par rapport à 0) ou ont proposé des versions « à constante près » de l'inégalité de corrélation, comme [**20**] qui présente des applications aux probabilités de petites boules pour les mouvements browniens fractionnaires. Citons enfin l'article [**11**] qui propose un panorama des inégalités géométriques gaussiennes (isopérimétrie, inégalités de type Brunn-Minkowski, mesures des ensembles dilatés notamment).

2. PRÉLIMINAIRES

2.1. Notations

Même si ce n'est pas indispensable, il est commode de munir $\mathbb{R}^n$ du produit scalaire canonique $\langle\cdot,\cdot\rangle$. La norme euclidienne est notée $|\cdot|$. On notera $\mathbb{R}_+^n := [0, +\infty[^n$. Pour $m \in \mathbb{N}^*$, on note $[m]$ l'ensemble $\{1, \ldots, m\}$. Si S est un ensemble fini $|S|$ représente son cardinal. Si $x \in \mathbb{R}^n$ et $S \subset [n]$, on pose $x_S = (x_i)_{i\in S}$. On note $M_{m,n}$ l'ensemble des matrices réelles à m lignes et n colonnes. Si $A \in M_{m,n}$ et $S \subset [m]$ et $T \subset [n]$, $A_{S,T}$ désigne la sous-matrice $(a_{i,j})_{i\in S, j\in T}$. Lorsque $A \in M_n := M_{n,n}$ est une matrice carrée, on note (encore !) $|A|$ pour son déterminant. De plus, si $S \subset [n]$, on note simplement A_S la sous-matrice carrée $A_{S,S} = (a_{i,j})_{i,j\in S}$. Si $x \in \mathbb{R}^n$, $\mathrm{diag}(x) = \mathrm{diag}(x_1, \ldots, x_n)$ est la matrice diagonale associée. On note A^* la transposée de la matrice A. Enfin $A > 0$ signifie que A est une matrice symétrique définie positive. On note $A \geq 0$ si elle est seulement (semi-définie) positive ; $\sqrt{A}$ désigne alors son unique racine carrée positive. Enfin la matrice identité de taille n est notée I_n.

2.2. Mesures gaussiennes

Les mesures de probabilités gaussiennes sur $\mathbb{R}$ sont notées $\mathscr{N}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+$. Lorsque $\sigma > 0$, ces mesures sont à densité

$$\mathscr{N}(m, \sigma^2)(dt) = e^{-\frac{(t-m)^2}{2\sigma^2}} \frac{dt}{\sigma\sqrt{2\pi}}.$$

De plus $\mathscr{N}(m, 0)$ est la masse de Dirac en m. Une mesure sur $\mathbb{R}^n$ est dite gaussienne si ses images par toutes les formes linéaires sont des mesures gaussiennes sur $\mathbb{R}$. Elles sont caractérisées par leur espérance $m \in \mathbb{R}^n$ et leur matrice de covariance $C \geq 0$, $C \in M_n$. On les note logiquement $\mathscr{N}_n(m, C)$. Si $G_1, \ldots, G_n$ sont des variables aléatoires réelles indépendantes et de loi $\mathscr{N}(0, 1)$, alors le vecteur $G = (G_1, \ldots, G_n)$ suit la loi $\mathscr{N}_n(0, I_n)$ que nous avons déjà notée γ_n. Pour $m \in \mathbb{R}^n$ et $C \in M_n$ positive, le vecteur $m + \sqrt{C}G$ suit la loi $\mathscr{N}_n(m, C)$. Lorsque $C > 0$ cette loi est à densité par rapport à la mesure de Lebesgue :

$$\mathscr{N}_n(m, C)(dx) = e^{-\frac{1}{2}\left\langle x-m, C^{-1}(x-m)\right\rangle} \frac{dx}{(2\pi)^{n/2}|C|^{1/2}}.$$

L'inégalité de corrélation gaussienne fait intervenir des vecteurs centrés, i.e., $m = 0$.

2.3. Quelques transformées de Laplace utiles

Dans sa preuve, Royen ne travaille avec les lois de probabilités que par l'intermédiaire de leur transformée de Laplace et tire habilement parti de leur forme simple. Comme l'énoncé que l'on cherche à prouver (le théorème 1.2) ne fait intervenir que les variables positives $|X_i|$, on va définir, pour une mesure μ sur $\mathbb{R}_+^n$, sa transformée de Laplace par

$$L_\mu(\lambda) = \int_{\mathbb{R}_+^n} e^{-\langle \lambda, x\rangle} \mu(dx), \quad \lambda \in \mathbb{R}_+^n.$$

Ainsi elle prend toujours des valeurs finies pour les mesures de probabilité. On étendra la définition aux mesures signées. Si X est un vecteur à valeurs dans $\mathbb{R}_+^n$, on définit sa transformée de Laplace comme celle de sa loi $\mathbb{P}_X$, ce qui revient à $L_X(\lambda) = \mathbb{E}e^{-\langle \lambda, X\rangle}$. Il est connu que dès que la transformée de Laplace d'une mesure (finie ou signée) est finie sur un ouvert, la donnée de ses valeurs sur l'ouvert caractérise la mesure. Commençons par calculer la transformée des « carrés » de vecteurs gaussiens centrés.

Lemme 2.1. — *Soit X un vecteur gaussien de loi $\mathscr{N}_n(0, C)$, et $Z = \left(\frac{X_1^2}{2}, \ldots, \frac{X_n^2}{2}\right)$. Alors pour tout $\lambda \in \mathbb{R}_+^n$,*

$$L_Z(\lambda) = |I_n + \Lambda C|^{-1/2},$$

où $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Si $C > 0$, la loi de Z est à densité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+^n$.

Preuve. — Soit G de loi $\mathscr{N}_n(0, I_n)$. On utilise que X a même loi que $\sqrt{C}\,G$

$$\begin{aligned}
\mathbb{E}\exp\left(-\sum_{i\le n}\lambda_i\frac{X_i^2}{2}\right) &= \mathbb{E}\exp\left(-\frac{1}{2}\langle \Lambda X, X\rangle\right) = \mathbb{E}\exp\left(-\frac{1}{2}\langle \Lambda\sqrt{C}G, \sqrt{C}G\rangle\right)\\
&= \int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\langle\sqrt{C}\Lambda\sqrt{C}x, x\rangle - \frac{1}{2}\langle x, x\rangle\right)\frac{dx}{(2\pi)^{n/2}}\\
&= \int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}\langle Mx, x\rangle\right)\frac{dx}{(2\pi)^{n/2}},
\end{aligned}$$

où $M := I_n + \sqrt{C}\Lambda\sqrt{C} > 0$. En effectuant le changement de variables $y = \sqrt{M}x$, on obtient

$$L_Z(\lambda) = \left|\sqrt{M}\right|^{-1} = \left|I_n + \sqrt{C}\Lambda\sqrt{C}\right|^{-1/2} = |I_n + \Lambda C|^{-1/2},$$

en utilisant en dernier lieu que pour des matrices carrées A, B de même taille, AB et BA ont même polynôme caractéristique.

Si $C > 0$, la loi de X est à densité. On en déduit la densité de Z en découpant $\mathbb{R}^n$ en 2^n orthants. Sur chacun d'eux, les coordonnées sont de signe constant, et le changement de variables $x_i^2/2 = z_i$ est bijectif. On obtient que

$$\mathbb{P}_Z(dz) = \mathbf{1}_{]0,+\infty[^n}(z)\frac{|C|^{-1/2}(4\pi)^{-n/2}}{\prod_{i\le n}\sqrt{z_i}}\sum_{\varepsilon\in\{-1,1\}^n}\exp\left(-\langle C^{-1}(\varepsilon\cdot\sqrt{z}), (\varepsilon\cdot\sqrt{z})\rangle\right)\,dz,$$

où $\varepsilon\cdot\sqrt{z} = (\varepsilon_i\sqrt{z_i})_{i=1}^n$. □

Nous aurons besoin d'identifier d'autres transformées de Laplace qui sont liées à la précédente.

Lemme 2.2. — *Soit $k \in \mathbb{N}^*$. Soit $C \ge 0$ une matrice de taille n. Alors il existe une mesure de probabilité μ sur $\mathbb{R}_+^n$ telle que pour tout $\lambda \in \mathbb{R}_+^n$,*

$$L_\mu(\lambda) = |I_n + \Lambda C|^{-k/2},$$

où $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Si $C > 0$ alors μ est à densité sur $\mathbb{R}_+^n$, $\mu(dz) = h(z)dz$. De plus, lorsque $k \ge 3$, on a pour tout ensemble $S \subset [n]$ les propriétés suivantes

1. *La dérivée « diagonale » $\dfrac{\partial^{|S|}}{\partial x_S}h$ existe sur $]0, +\infty[^n$ et appartient à $L^1(\mathbb{R}_+^n)$.*
2. *Si $i \in [n] \setminus S$, alors $\displaystyle\lim_{x_i\to 0^+}\frac{\partial^{|S|}}{\partial x_S}h(x) = 0$.*