

VECTEURS GAUSSIENS

PRÉPARATION À L'AGRÉGATION EXTERNE DE MATHÉMATIQUES DE L'UNIVERSITÉ RENNES 1¹

ANNÉE 2014-2015

1. DEFINITIONS ET PREMIERES PROPRIETES

Définition On dit que X est un vecteur gaussien de \mathbb{R}^n si il existe $m \in \mathbb{R}^n$ et $\Sigma \in \mathcal{M}_n(\mathbb{R})$ symétrique semi-définie positive tels que la fonction caractéristique φ_X de X s'écrit :

$$\varphi_X(u) := \mathbb{E}e^{i\langle u, X \rangle} = \exp\left(i\langle u, m \rangle - \frac{1}{2}u^\top \Sigma u\right), \forall u \in \mathbb{R}^n.$$

Dans ce cas, on note $X \sim \mathcal{N}_n(m, \Sigma)$.

Remarques

- (a) Cette définition inclut notamment le cas où X suit une loi de Dirac en m (cas $\Sigma = 0$);
- (b) Si $X \sim \mathcal{N}_n(m, \Sigma)$, alors $\mathbb{E}(X) = m$ et $\mathbb{V}(X) = \Sigma$. En particulier, un vecteur gaussien admet un moment d'ordre 2 et, de manière plus générale, des moments de tous ordres.
- (c) Tout sous-vecteur d'un vecteur gaussien est gaussien, et en particulier, les composantes d'un vecteur gaussien sont des v.a.r. gaussiennes. Par contre, il se peut que X_1, \dots, X_n soient des v.a. réelles gaussiennes sans pour autant que le vecteur $(X_1, \dots, X_n)^\top$ soit gaussien : en effet, soient $X_1 \sim \mathcal{N}_1(0, 1)$ et $X_2 = \varepsilon X_1$, où $\varepsilon \perp X_1$, et de loi définie par $\mathbb{P}(\varepsilon = \pm 1) = 1/2$. Alors, $X_2 \sim \mathcal{N}_1(0, 1)$ mais la fonction caractéristique du vecteur $(X_1, X_2)^\top$ n'est pas une fonction caractéristique gaussienne.

Souvent, un moyen simple de montrer qu'un vecteur aléatoire est gaussien est d'utiliser la proposition suivante, qui constitue une définition alternative efficace d'un vecteur gaussien :

Proposition 1.1 Un vecteur aléatoire est gaussien si, et seulement si, toute combinaison linéaire de ses composantes est une v.a. réelle gaussienne.

Si X est un vecteur aléatoire de carré intégrable, alors $X - \mathbb{E}(X) \in \text{Im}(\mathbb{V}(X))$ p.s. Par suite, X ne possède pas de densité si $\det \mathbb{V}(X) = 0$. Dans le cas où X est un vecteur gaussien, la condition d'inversibilité de sa matrice de variance suffit à établir l'existence d'une densité, comme le montre le résultat ci-dessous.

Théorème 1.1 Soit $X \sim \mathcal{N}_n(m, \Sigma)$ avec $\det \Sigma \neq 0$. Alors, X admet une densité qui est

$$\frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-m)^\top \Sigma^{-1}(x-m)\right), x \in \mathbb{R}^n.$$

D'après la proposition 1.1, les lois gaussiennes sont stables par transformation affine. La forme très particulière de la loi image, décrite dans le résultat suivant sur lequel on pourrait être tenté -à tort- de jeter un coup d'oeil distrait, est d'utilité constante dans la manipulation des vecteurs gaussiens.

Proposition 1.2 Si $A \in \mathcal{M}_{k,n}(\mathbb{R})$, $b \in \mathbb{R}^k$ et $X \sim \mathcal{N}_n(m, \Sigma)$, on a :

$$AX + b \sim \mathcal{N}_k(Am + b, A\Sigma A^\top).$$

Lorsque Σ est réelle, symétrique et semi-définie positive, on peut construire à l'aide du théorème de Schur une matrice $\Sigma^{1/2}$ vérifiant $\Sigma^{1/2} \Sigma^{1/2} = \Sigma$. Cette matrice est inversible, d'inverse $\Sigma^{-1/2}$, lorsque Σ est définie positive.

Proposition 1.3 Soient $m \in \mathbb{R}^n$ et $\Sigma \in \mathcal{M}_n(\mathbb{R})$ symétrique semi-définie positive.

- (i) Si $\det \Sigma \neq 0$ et $X \sim \mathcal{N}_n(m, \Sigma)$, alors $\Sigma^{-1/2}(X - m) \sim \mathcal{N}_n(0, \text{Id})$;
- (ii) Si $X \sim \mathcal{N}_n(0, \text{Id})$, alors $m + \Sigma^{1/2}X \sim \mathcal{N}_n(m, \Sigma)$.

1. Benoît Cadre - ENS Rennes

Lorsque 2 v.a.r. sont indépendantes, leur covariance est nulle. En revanche, la réciproque est fautive, sauf dans le cas des vecteurs gaussiens :

Théorème 1.2 Soit X un vecteur gaussien. Les composantes de X sont des v.a.r. indépendantes si, et seulement si la matrice de variance de X est diagonale.

La preuve de ce résultat, sur laquelle nous ne nous arrêterons pas, est une illustration de l'intérêt de la fonction caractéristique, et donc de la transformée de Fourier (cf [OUVRARD]). A ce titre, elle peut être insérée dans une leçon d'analyse et probabilités portant sur ce thème.

Remarque Le fait que les composantes du vecteur aléatoire X soient des v.a.r. gaussiennes n'est pas suffisant pour établir cette équivalence (reprendre l'exemple du (c) de la remarque ci-dessus).

APPLICATION : SIMULATION DE VECTEURS GAUSSIENS. Il est facile de simuler un vecteur gaussien de matrice de variance diagonale, car les composantes sont alors des v.a.r. indépendantes de lois gaussiennes. Supposons donc que l'on ait à simuler une réalisation d'un vecteur gaussien $X \sim \mathcal{N}_n(m, \Sigma)$. Il existe une matrice orthogonale P telle que $\Sigma = P\Delta P^\top$ (\star), où $\Delta = \text{diag}(\lambda_1, \dots, \lambda_n)$. Si $Z = P^\top(X - m)$, alors $Z \sim \mathcal{N}_n(0, \Delta)$. Il suffit donc de simuler les n v.a.r. indépendantes Z_1, \dots, Z_n de lois $\mathcal{N}_1(0, \lambda_1), \dots, \mathcal{N}_1(0, \lambda_n)$ constituant le vecteur gaussien Z pour obtenir une réalisation de X , car $X = m + PZ$. L'algorithme de simulation de X est :

1. Calculer la décomposition (\star) pour Σ . NB : en pratique, Scilab fournit la fonction *svd* (Singular Value Decomposition) qui, utilisée comme suit : $[A, \Delta, B] = \text{svd}(\Sigma)$, rend la matrice diagonale Δ dont les éléments sont rangés par ordre décroissant, et A, B sont des matrices unitaires telles que $\Sigma = A\Delta B^\top$, c'est-à-dire que dans notre cas, $A = B = P$.
2. Générer une réalisation z de Z en simulant des gaussiennes $\mathcal{N}_1(0, \lambda_i)$ pour les indices i tels que $\lambda_i > 0$, et compléter les autres composantes de z par des 0.
3. Calculer la réalisation correspondante $x = m + Pz$ de $X = m + PZ$.

Le conditionnement dans un vecteur gaussien possède des propriétés très particulières, comme en témoigne le résultat ci-dessous :

Théorème 1.3 Soit $(Y, X_1, \dots, X_n)^\top$ un vecteur gaussien de \mathbb{R}^{n+1} tel que $X = (X_1, \dots, X_n)^\top$ possède une matrice de variance inversible. Si $(a_1, \dots, a_n)^\top = \mathbb{V}(X)^{-1}(\text{cov}(Y, X_1), \dots, \text{cov}(Y, X_n))^\top$, on a :

$$\mathbb{E}(Y|X_1, \dots, X_n) = \sum_{i=1}^n a_i(X_i - \mathbb{E}X_i) + \mathbb{E}Y.$$

Preuve On suppose que (Y, X_1, \dots, X_n) est centré, et on note $\hat{Y} = \sum_{i=1}^n a_i X_i$. On vérifie facilement que pour tout $i = 1, \dots, n$:

$$\text{cov}(Y - \hat{Y}, X_i) = \mathbb{E}(Y - \hat{Y})X_i = 0.$$

Puisque le vecteur $(X_1, \dots, X_n, Y - \hat{Y})$ est gaussien, $Y - \hat{Y} \perp (X_1, \dots, X_n)$. En conséquence,

$$\mathbb{E}(Y|X_1, \dots, X_n) = \mathbb{E}(Y - \hat{Y}|X_1, \dots, X_n) + \hat{Y} = \mathbb{E}(Y - \hat{Y}) + \hat{Y} = \hat{Y}. \bullet$$

2. PROJECTION DE VECTEURS GAUSSIENS

[RÉF. : DACUNHA-CASTELLE ET DUFLO, OUVRARD, TOULOUSE]

Le théorème ci-dessous est essentiel dans toute la théorie des modèles gaussiens. Il intervient dans la plupart des problèmes d'estimation de paramètres issus d'une loi gaussienne : estimation et tests pour un échantillon gaussien, modèles linéaires gaussiens, filtrage linéaire gaussien (en particulier le filtre de Kalman-Bucy),...

Théorème 2.1 [COCHRAN] Soit $X \sim \mathcal{N}_n(0, \sigma^2 \text{Id})$ avec $\sigma > 0$ et $L_1 \oplus \dots \oplus L_p$ une décomposition de \mathbb{R}^n en sous-espaces orthogonaux de dimensions r_1, \dots, r_p . Les projections orthogonales π_1, \dots, π_p de X sur L_1, \dots, L_p sont des vecteurs gaussiens indépendants, et pour chaque $i = 1, \dots, p$:

$$\frac{1}{\sigma^2} \|\pi_i\|^2 \sim \chi_{r_i}^2,$$

avec $\|\cdot\|$ la norme euclidienne.

Preuve Soit $(e_j^i)_{i,j}$ une base orthonormée de \mathbb{R}^n telle que pour chaque $i = 1, \dots, p$, $(e_j^i)_{j=1, \dots, r_i}$ est une base orthonormée de L_i . Pour chaque $i = 1, \dots, p$: $\pi_i = \sum_{j=1}^{r_i} \langle X, e_j^i \rangle e_j^i$. Les vecteurs $(e_j^i)_{i,j}$ étant orthogonaux, on trouve pour tout $i \neq k$: $\text{cov}(\pi_i, \pi_k) = 0$. Comme $(\pi_1, \dots, \pi_p)^\top$ est un vecteur gaussien (car toute combinaison linéaire des v.a.r. $(\langle X, e_j^i \rangle)_{i,j}$ est gaussienne), π_1, \dots, π_p

sont des vecteurs gaussiens indépendants. Enfin, pour tout $i = 1, \dots, r_i$, les v.a.r. $\langle X, e_1^i \rangle, \dots, \langle X, e_{r_i}^i \rangle$ sont indépendantes et de même loi $\mathcal{N}_1(0, \sigma^2)$. Par suite,

$$\frac{1}{\sigma^2} \|\pi_i\|^2 = \sum_{j=1}^{r_i} \left(\frac{\langle X, e_j^i \rangle}{\sigma} \right)^2 \sim \chi_{r_i}^2. \bullet$$

Le théorème de Cochran permet d'obtenir facilement des informations sur les estimateurs de la moyenne ou de la variance dans un échantillon gaussien. Pour un échantillon X_1, \dots, X_n de v.a.r., on note \bar{X}_n et S_n^2 sa moyenne et variance empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

L'estimateur naturel $(1/n) \sum_{i=1}^n (X_i - \bar{X}_n)^2$ de $\text{var}(X_1)$ est *biaisé*. En revanche, S_n^2 à le mérite d'estimer *sans biais* la variance $\text{var}(X_1)$.

On rappelle que la *loi de Student* à n degrés de liberté, notée T_n , est la loi du quotient $\sqrt{n}X/\sqrt{Y}$, où $X \perp Y$, $X \sim \mathcal{N}_1(0, 1)$ et $Y \sim \chi_n^2$.

Théorème 2.2 [FISHER] Soit $X = (X_1, \dots, X_n)^\top \sim \mathcal{N}_n(me, \sigma^2 \text{Id})$, avec $\sigma > 0$ et $e = (1, \dots, 1)^\top$. Alors, $\bar{X}_n \perp S_n$. De plus, $\sqrt{n}(\bar{X}_n - m)/\sigma \sim \mathcal{N}_1(0, 1)$, $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$ et $\sqrt{n}(\bar{X}_n - m)/S_n \sim T_{n-1}$.

D'après la loi forte des grands nombres, $S_n \rightarrow \sigma$ p.s. La dernière assertion du théorème de Fisher, le théorème de la limite centrale unidimensionnel et le lemme de Slutsky montrent que pour les grandes valeurs de n , T_n est proche de la loi $\mathcal{N}_1(0, 1)$.

A titre de complément, on peut montrer (cf [OUVRARD]) que $\bar{X}_n \perp S_n \Leftrightarrow X$ est gaussien. La preuve, qui utilise la fonction caractéristique, peut judicieusement être insérée dans une leçon d'analyse et probabilités.

Preuve Cas $m = 0$ et $\sigma^2 = 1$. Soit L le s.e.v. de \mathbb{R}^n engendré par e . Le projecteur orthogonal P sur L est la matrice $n \times n$ ne contenant que des $1/n$. On a alors $PX = \bar{X}_n e$ et $(\text{Id} - P)X = (X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^\top$. Comme $(\text{Id} - P)X$ est la projection orthogonale de X sur l'orthogonal de L , on déduit du théorème de Cochran que $PX \perp (\text{Id} - P)X$, et en particulier que $\bar{X}_n \perp S_n^2$. Enfin, $(n-1)S_n^2 = \|(\text{Id} - P)X\|^2 \sim \chi_{n-1}^2$ d'après le théorème de Cochran. \bullet

3. STATISTIQUE DES ECHANTILLONS GAUSSIENS

[RÉF. : DACUNHA-CASTELLE ET DUFLO, OUVRARD]

3.1 LE MODÈLE

On dispose d'*observations* réelles x_1, \dots, x_n . En terme de modélisation, la première étape consiste à considérer que ces réels sont des *réalisations* de n v.a.r.i.i.d. notées X_1, \dots, X_n , c'est-à-dire que pour une certaine éventualité ω , $x_i = X_i(\omega)$. Ces observations sont supposées être issues d'une loi gaussienne. On peut donc considérer dans la suite que $X_1 \sim \mathcal{N}_1(m, \sigma^2)$, avec m et $\sigma > 0$ inconnus. L'enjeu est maintenant de donner des valeurs approchées pour les paramètres m et σ^2 (cadre paramétrique). On note \bar{x}_n et s_n^2 la moyenne et la variance des observations x_1, \dots, x_n :

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \text{ et } s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Le modèle statistique est $\{P_{m,\sigma}; m \in \mathbb{R}, \sigma > 0\}$, où $P_{m,\sigma} = \mathcal{N}_1(m, \sigma^2)^{\otimes n}$, et (X_1, \dots, X_n) désigne un échantillon de la loi $P_{m,\sigma}$. Les intervalles de confiance et les tests sont construits à l'aide du théorème de Fisher, qui a l'avantage de fournir des lois à n fixé (tests et intervalles de confiance non asymptotiques).

On note $\alpha \in]0, 1[$ le *niveau* du test, c'est-à-dire le maximum de l'*erreur de 1ère espèce* lorsque le paramètre (m ou σ) parcourt l'ensemble défini par l'hypothèse nulle H_0 (rappelons que l'*erreur de 1ère espèce* est la probabilité de rejeter H_0 à tort). On choisit souvent $\alpha = 1\%, 5\%$ ou 10% , de manière à considérer en priorité des tests de niveau faible (c'est le *principe de Neyman*). On ne considère que les cas les plus courants en pratique, i.e. σ (resp. m) est inconnu lorsque l'on veut tester m (resp. σ). Sinon, il suffit d'utiliser l'égalité $\sqrt{n}(\bar{X}_n - m)/\sigma \sim \mathcal{N}_1(0, 1)$ sous $P_{m,\sigma}$.

3.2 LE TEST DE STUDENT (OU t-TEST)

On fixe m_0 , et on veut tester par exemple $H_0 : m \leq m_0$ contre $H_1 : m > m_0$ au niveau α . La statistique de test $\sqrt{n}(\bar{X}_n - m)/S_n \sim T_{n-1}$ sous $P_{m,\sigma}$. La région de rejet est du type $\{(z_1, \dots, z_n) \in \mathbb{R}^n : \bar{z}_n > a\}$ car H_0 est rejetée à tort dès que \bar{X}_n prend des valeurs anormalement grandes. Notons $t_{n-1,\alpha}$ le quantile d'ordre $1 - \alpha$ de la loi T_{n-1} . Alors, sous H_0 :

$$P_{m,\sigma} \left(\bar{X}_n \in \left[m_0 + t_{n-1,\alpha} \frac{S_n}{\sqrt{n}}, \infty \right] \right) \leq P_{m,\sigma} \left(\bar{X}_n \in \left[m + t_{n-1,\alpha} \frac{S_n}{\sqrt{n}}, \infty \right] \right) = \alpha.$$

Pour ce test, une *région de rejet* au niveau α est donc $R_\alpha := \{(z_1, \dots, z_n) \in \mathbb{R}^n : \bar{z}_n > m_0 + t_{n-1,\alpha} s_n / \sqrt{n}\}$. Autrement dit, la procédure de décision est définie ainsi : on rejette H_0 au niveau α si $\bar{x}_n \in R_\alpha$.

3.3 LE TEST DE FISHER

On fixe σ_0 , et on veut tester par exemple $H_0 : \sigma \leq \sigma_0$ contre $H_1 : \sigma > \sigma_0$ au niveau α . La statistique de test $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$ sous $P_{m,\sigma}$. Soit $\chi_{n-1,\alpha}^2$ le quantile d'ordre $1 - \alpha$ de la loi χ_{n-1}^2 . On montre comme dans le cas précédent que pour ce test, la région de rejet au niveau α est $R_\alpha := \{(z_1, \dots, z_n) \in \mathbb{R}^n : \bar{z}_n > \chi_{n-1,\alpha}^2 \sigma_0^2 / (n-1)\}$. Autrement dit, on rejette H_0 au niveau α si $s_n^2 \in R_\alpha$.

4. LE THEOREME DE LA LIMITE CENTRALE SUR \mathbb{R}^d

[RÉF. : TOUTES]

Soient X_1, X_2, \dots des vecteurs aléatoires sur \mathbb{R}^d , supposés indépendants et de même loi intégrable. La moyenne empirique des n premiers vecteurs aléatoires est $\bar{X}_n = (1/n) \sum_{k=1}^n X_k$. D'après la *loi forte des grands nombres*, $\bar{X}_n \rightarrow \mathbb{E}X_1$ p.s. lorsque $n \rightarrow \infty$. Le *théorème de la limite centrale* est un principe d'invariance qui précise la vitesse de convergence dans la loi forte des grands nombres.

Théorème 4.1 [TLC] Soient X_1, X_2, \dots des vecteurs aléatoires sur \mathbb{R}^d , supposés indépendants et de même loi de carré intégrable. Alors, lorsque $n \rightarrow \infty$, on a :

$$\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \mathbb{V}(X_1)).$$

La preuve est directe à partir de la version unidimensionnelle du TLC et l'astuce de Cramèr-Wold.

LE TLC ET LA THÉORIE DES PROBABILITÉS. La position "centrale" de ce théorème réside dans le fait suivant : dès lors que les vecteurs aléatoires X_1, \dots, X_n sont indépendants et de même loi de carré intégrable, la loi de \bar{X}_n est proche de $\mathcal{N}_d(\mathbb{E}X_1, \mathbb{V}(X_1)/n)$ lorsque n est grand (avec toutes les précautions d'usage !). En termes de modélisation, l'impact de ce résultat est considérable : il signifie que l'on peut raisonnablement considérer que la somme de petites perturbations indépendantes est la réalisation d'une loi qui est proche d'une loi gaussienne.

LA VITESSE DE CONVERGENCE DANS LE TLC (CAS $d = 1$). Lorsque l'on utilise le TLC, que ce soit pour calculer des intervalles de confiance asymptotiques, faire des tests asymptotiques, ou bien pour justifier le fait qu'une erreur puisse être raisonnablement considérée comme étant issue d'une loi normale, on est toujours amené à dire : pour les grandes valeurs de n , $\sqrt{n}(\bar{X}_n - m)$ suit à peu près une loi normale. Que signifie ici "pour les grandes valeurs de n " ? Autrement dit, comment peut-on contrôler l'erreur commise en remplaçant $\sqrt{n}(\bar{X}_n - m)$ par la loi normale correspondante ? Un élément de réponse est donné par l'*inégalité de Berry-Esséen* : sous réserve que X_1 admette un moment d'ordre 3, et sous les conditions du TLC :

$$\sup_{x \in \mathbb{R}^d} \left| \mathbb{P}(\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \leq x) - \mathbb{P}(G \leq x) \right| \leq Cn^{-1/2},$$

avec $G \sim \mathcal{N}_1(0, \mathbb{V}(X_1))$ et C une constante indépendante de n , dépendant des 3 premiers moments de X_1 . Pour les applications, il est important de disposer d'une valeur de C qui soit la plus petite possible, ce qui fut l'objet d'une longue quête...

REFERENCES

- Dacunha-Castelle D. et Duflo M. *Probabilités et statistiques, 1. Problèmes à temps mobile*. Masson, 1993.
- Foata D. et Fuchs A. *Calcul des probabilités, 2ème édition*. Dunod, 1998.
- Ouvrard J.-Y. *Probabilités 2 - Maîtrise Agrégation*. Cassini, 2000.
- Toulouse P.S. *Thèmes de probabilités et statistique*. Dunod, 1999.

APPENDICE : GENERALITES SUR LES MATRICES ALEATOIRES

[RÉF. : DACUNHA-CASTELLE ET DUFLO]

Une *matrice* (resp. *vecteur*) *aléatoire* est une matrice (resp. vecteur) dont les éléments sont des v.a.r. Elle est intégrable (resp. de carré intégrable) si ses composantes le sont, et son *espérance mathématique* est la matrice (resp. vecteur) des espérances des éléments qui la composent. Si A et B sont des matrices déterministes de tailles $n \times p$ et $k \times l$ et si X est une matrice aléatoire de taille $p \times k$, on a $\mathbb{E}(AX) = A\mathbb{E}(X)$, $\mathbb{E}(XB) = \mathbb{E}(X)B$ et $\mathbb{E}(X^\top) = \mathbb{E}(X)^\top$.

Pour les vecteurs aléatoires de carrés intégrables, la notion de variance se généralise ainsi : si X et Y sont des vecteurs aléatoires de même taille, on pose

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}[X - \mathbb{E}X)(X - \mathbb{E}X)^\top] \text{ la matrice de variance de } X; \\ \text{cov}(X, Y) &= \mathbb{E}[X - \mathbb{E}X)(Y - \mathbb{E}Y)^\top] \text{ la matrice de covariance de } X \text{ et } Y. \end{aligned}$$

Noter que $\mathbb{V}(X)$ est une matrice symétrique semi-définie positive.

Si X et Y sont des vecteurs aléatoires possédant les propriétés d'intégrabilité adéquates et A et B sont des matrices déterministes, de tailles permettant d'effectuer les sommes et les produits ci-dessous, on a les relations :

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}XX^\top - \mathbb{E}X\mathbb{E}X^\top & \mathbb{V}(X+Y) &= \mathbb{V}(X) + \mathbb{V}(Y) + \text{cov}(X, Y) + \text{cov}(Y, X) \\ \text{cov}(X, Y) &= \mathbb{E}XY^\top - \mathbb{E}X\mathbb{E}Y^\top & \text{cov}(AX, BY) &= A\text{cov}(X, Y)B^\top \\ \text{cov}(X, Y) &= \text{cov}(Y, X)^\top & \mathbb{V}(AX+B) &= A\mathbb{V}(X)A^\top \end{aligned}$$

Théorème Soit X un vecteur aléatoire de \mathbb{R}^n de carré intégrable. Alors, $\det \mathbb{V}(X) = 0$ si, et seulement si il existe une liaison affine p.s. entre les composantes de X . Dans ce cas, on a aussi $\mathbb{P}(X - \mathbb{E}X \in \text{Im}(\mathbb{V}(X))) = 1$.

Preuve Supposons que $\det \mathbb{V}(X) = 0$. Comme $\mathbb{V}(X)$ est une matrice symétrique semi-définie positive, il existe une matrice diagonale Δ et une matrice orthogonale Q telles que $Q\mathbb{V}(X)Q^\top = \Delta$. En posant $Y = QX$, on obtient un vecteur aléatoire de matrice de variance Δ . Soit v un vecteur propre correspondant à une valeur propre nulle de $\mathbb{V}(X)$. Comme $Y = QX$, on a pour une coordonnée Y_k de Y : $Y_k = v^\top X$. Alors, $\text{var}(Y_k) = v^\top \mathbb{V}(X)v = 0$, d'où p.s. $v^\top X = v^\top \mathbb{E}X$. On a donc p.s. $X - \mathbb{E}X \in \text{Im}(\mathbb{V}(X))$, ce qui entraîne l'existence d'une relation affine p.s. entre les composantes de X . La réciproque est immédiate : s'il existe une liaison affine p.s. entre les composantes de X , alors en considérant la variance dans cette relation, on établit l'existence d'une relation entre les colonnes de $\mathbb{V}(X)$. •

Si $\det \mathbb{V}(X) = 0$, on a $\mathbb{P}(X \in \mathbb{E}X + \text{Im}(\mathbb{V}(X))) = 1$ et le vecteur aléatoire X ne prend ses valeurs que dans un sous-espace affine de \mathbb{R}^n : on dit que la loi de X est *dégénérée*. Comme par ailleurs $\lambda(\mathbb{E}X + \text{Im}(\mathbb{V}(X))) = 0$, λ désignant la mesure de Lebesgue sur \mathbb{R}^n , le résultat suivant est immédiat.

Corollaire Soit X un vecteur aléatoire de \mathbb{R}^n dont les composantes sont de carrés intégrables. Si $\det \mathbb{V}(X) = 0$, la loi de X ne possède pas de densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n .