PROBABILITY 2018
HANDOUT 2: GAUSSIAN DISTRIBUTION

GIOVANNI PISTONE

Contents

1. Standard Gaussian Distribution 1
2. Positive Definite Matrices 5
3. General Gaussian Distribution 6
4. Independence of Jointly Gaussian Random Variables 7

References 8

This handout covers multivariate Gaussian distributions and the relevant matrix theory. Two classical references are [3] and [1] (many reprints available). A modern advanced reference for positive definite matrices is [2].

1. Standard Gaussian Distribution

1 (Change of variable formula in $\mathbb{R}^d$)**.** Let $\mathcal{A}, \mathcal{B} \subset \mathbb{R}^d$ be open and ϕ be a diffeomerphism from $\mathcal{A}$ onto $\mathcal{B}$. Let $J\phi \colon \mathcal{A} \to \operatorname{Mat}(d \times d)$ be the Jacobian mapping of ϕ and $J\phi^{-1} \colon \mathcal{B} \to \operatorname{Mat}(d \times d)$ the Jacobian mapping of ϕ^{-1}, so that $J\phi^{-1} = (J\phi \circ \phi^{-1})^{-1}$. For each non-negative $f \colon \mathcal{B} \to \mathbb{R}^n$,

$$\int_{\mathcal{B}} f(\boldsymbol{y})\ d\boldsymbol{y} = \int_{\mathcal{A}} f \circ \phi(\boldsymbol{x})\ \left|\det\left(J\phi(\boldsymbol{x})\right)\right|\ d\boldsymbol{x}$$

Exercise 1. $\mathcal{A} =]0, 2\pi[\times]0, +\infty[$, $\mathcal{B} = \mathbb{R}^2_* = \mathbb{R}^2 \setminus \{(x,y) \in \mathbb{R}^2 | x \geqslant 0, y = 0\}$, $\phi(\theta, \rho) = (\rho\cos\theta, \rho\sin\theta)$.

$$J\phi(\theta,\rho) = \begin{bmatrix} -\rho\sin\theta & \cos\theta \\ \rho\cos\theta & \sin\theta \end{bmatrix}, \quad \det\left(J\phi(\theta,\rho)\right) = -\rho$$

$$\iint\limits_{\mathbb{R}^2_*} \mathrm{e}^{-(x^2+y^2)/2}\ dxdy = \iint\limits_{]0,2\pi[\times]0,+\infty[} \mathrm{e}^{-(\rho^2\cos^2\theta + \rho^2\sin^2\theta)/2}\ \rho\ d\theta d\rho =$$

$$\iint\limits_{]0,2\pi[\times]0,+\infty[} \mathrm{e}^{-\rho^2/2}\ \rho\ d\theta d\rho = 2\pi$$

2. (Image of an absolutely continous measure) Let $(S, \mathcal{F}, \mu)$ be measure space, $p \colon S \to \mathbb{R}_{>0}$ a probability density, $(\mathbb{X}, \mathcal{G})$ a measurable space, $\phi \colon S \to \mathbb{X}$ a measurable function. If ϕ has a measurable inverse, then the image measure is characterised by

$$\int f\ d\phi_\#(p \cdot \mu) = \int (f \circ \phi) p\ d\mu = \int (f \circ \phi)(p \circ \phi^{-1} \circ \phi)\ d\mu = \int f p \circ \phi^{-1}\ d\phi_\# \mu$$

Date: Version March 22, 2018.

hence $\phi_\#(p \cdot \mu) = (p \circ \phi^{-1}) \cdot \mu$. Eq. (1) applied to $f \circ \phi$ and the diffeomorphism ϕ^{-1} gives

$$\int_{\mathcal{B}} f \; d(\phi_\# \ell) = \int_{\mathcal{A}} f \circ \phi(\boldsymbol{x}) \; d\boldsymbol{x} = \int_{\mathcal{B}} f \circ \phi \circ \phi^{-1}(\boldsymbol{y}) \; \left|\det\left(J\phi^{-1}(\boldsymbol{y})\right)\right| \; d\boldsymbol{y} =$$

$$\int_{\mathcal{B}} f(\boldsymbol{y}) \; \left|\det\left(J\phi^{-1}(\boldsymbol{y})\right)\right| \; d\boldsymbol{y} = \int_{\mathcal{B}} f(\boldsymbol{y}) \; \left|\det\left(J\phi \circ \phi^{-1}(\boldsymbol{y})\right)\right|^{-1} \; d\boldsymbol{y}$$

This shows that the image of the Lebesgue measure ℓ under a diffeomorphism is

$$\phi_\# \ell = \left|\det\left(J\phi^{-1}\right)\right| \cdot \ell = \left|\det\left(J\phi \circ \phi^{-1}\right)\right|^{-1} \cdot \ell$$

Exercise 2. $\mathcal{A} =]0,1[\times]0,1[$, $\mathcal{B} = \mathbb{R}^2_*$, $\phi(u,v) = (\sqrt{-2\log u}\cos(2\pi v), \sqrt{-2\log u}\sin(2\pi v))$,

$$J\phi(u,v) = \begin{bmatrix} -\frac{1}{2}(-2\log u)^{-1/2}\frac{2}{u}\cos(2\pi v) & -2\pi\sqrt{-2\log u}\sin(2\pi v) \\ -\frac{1}{2}(-2\log u)^{-1/2}\frac{2}{u}\sin(2\pi v) & 2\pi\sqrt{-2\log u}\cos(2\pi v) \end{bmatrix},$$

$$\det\left(J\phi(u,v)\right) = -\frac{2\pi}{u}, \quad \det\left(J\phi \circ \phi^{-1}(x,y)\right) = \frac{2\pi}{\mathrm{e}^{(x^2+y^2)/2}} \; .$$

The image of the uniform probability measure on $]0,1[^2$ under ϕ is $(2\pi)^{-1}\mathrm{e}^{-(x^2+y^2)/2} \; dxdy$.

3 (Marginalization)**.** The previous argument does not apply when Φ is not 1-to-1. We will show in the chapter on conditioning that in such a case

$$\Phi_\#(p \cdot \mu) = \hat{p} \cdot \Phi_\#(\mu)$$

where $\hat{p}$ is the conditional expectation of p with respect to Φ.

However, there are two common and simple cases namely, the finite state space case and the marginalisation. Assume $\mu = \mu_1 \otimes \mu_2$ on $S = S_1 \times S_2$ and consider the marginal projection $\Phi \colon (x_1, x_2) \mapsto x_1$. Then $\Phi^{-1}(A_1) = A_1 \times S_2$ and $\mu(\Phi^{-1}(A_1)) = \mu(A_1 \times S_2) = \mu_1(A_1)$ hence, $\Phi_\#(\mu) = \mu_1$. Let p be a density on S with respect to μ. For each positive $f \colon S_1$ we have

$$\int f \; d\Phi_\#(p \cdot \mu) = \int f \circ \Phi \; d(p \cdot \mu) = \iint f(x_1)p(x_1,x_2) \; \mu(dx_1, dx_2) =$$

$$\int f(x_1)\left(\int p(x_1,x_2) \; \mu_2(dx_2)\right) \; \mu_1(dx_1)$$

so that

$$\Phi_\#(p \cdot \mu) = p_1(x_1) \cdot \mu_1, \quad p_1(x_1) = \int p(x_1,x_2) \; \mu_2(dx_2)$$

For example, if $p(x_1,x_2) = (2\pi)^{-1}\mathrm{e}^{-(x_1^2+x_2^2)/2}$, then

$$\int p(x_1,x_2) \; dx_2 = (2\pi)^{-1/2}\mathrm{e}^{-x_1^2/2}\int (2\pi)^{-1/2}\mathrm{e}^{-x_2^2/2} \; dx_2 = c(2\pi)^{-1/2}\mathrm{e}^{-x_1^2/2}$$

with $c = \int (2\pi)^{-1/2}\mathrm{e}^{-x_2^2/2} \; dx_2 = 1$ as the further integration with respect to dx_1 shows. Notice that the argument applies to all $p(x_1,x_2) = cf(x_1)f(x_2)$.

4. The real random variable Z is *standard Gaussian*, $Z \sim \mathrm{N}_1(0,1)$, if its distribution ν has density

$$\mathbb{R} \ni z \mapsto \gamma(z) = (2\pi)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}z^2\right)$$

with respect to the Lebesgue measure. It is in fact a density, see above the computation of its two-fold product.

Exercise 3. All moments $\mu(n) = \int z^n \gamma(z)\ dz$ exists. As $z\gamma(z) = -\gamma'(z)$, integration by parts produces a recurrent relation for the moments. [Hint: Write $\int z^n\gamma(z)\ dz = \int z^{n-1} z\gamma(z)\ dz = \int z^{n-1}(-\gamma'(z))\ dz$ and perform an integration by parts]

Exercise 4. If $f\colon \mathbb{R} \to \mathbb{R}$ absolutely continuous i.e., $f(z) = f(0) + \int_0^z f'(u)\ du$, with $\int |f'(u)|\,\gamma(u)\ du < +\infty$ then $\int |zf(z)|\,\gamma(z)\ dz < +\infty$. In fact,

$$\int |zf(z)|\,\gamma(z)\ dz = \int \left| z \left(f(0 + \int_0^z f'(u)\ du \right) \right| \gamma(z)\ dz \leqslant$$
$$|f(0)| \int |z|\,\gamma(z)\ dz + \int \left| z \int_0^z f'(u)\ du \right| \gamma(z)\ dz\ .$$

The first term in the RHS equals $\sqrt{2/\pi}\,|f(0)|$, while in the second term we have for $z \geqslant 0$,

$$\left| \int_0^z f'(u)\ du \right| \leqslant \int (0 \leqslant u \leqslant z)\,|f'(u)|\ du\ .$$

We have

$$\int \left| z \int_0^z f'(u)\ du \right| \gamma(z)\ dz \leqslant \int |z| \left(\int (0 \leqslant u \leqslant z)\,|f'(u)|\ du \right) \gamma(z)\ dz =$$
$$\int |f'(u)| \int_u^\infty z\gamma(z)\ dz\ du = \int |f'(u)| \int_u^\infty (-\gamma'(z))\ dz\ du =$$
$$\int |f'(u)|\,\gamma(u)\ du < \infty\ .$$

A similar argument applies to the case $z \leqslant 0$. This implies

$$\int zf(z)\gamma(z)\ dz = \int f(z)(-\gamma'(z))\ dz = \int f'(z)\ \gamma(z) dz\ .$$

Exercise 5. The *Stein operator* is $\delta f(z) = zf(z) - f'(z)$. We have

$$\int f(z)g'(z)\gamma(z)\ dz = \int \delta f(z)g(z)\gamma(z)dz$$

We define the *Hermite polynomials* to be $H_n(z) = \delta^n 1$. For example, $H_1(z) = z$, $H_2(z) = z^2 - 1$, $H_3(z) = z^3 - 3z$. Hermite polynomials are orthogonal with respect to γ,

$$\int H_n(z)H_m(z)\gamma(z)\ dz = 0 \quad \text{if } n > m\ .$$

5. Let $Z \sim \mathrm{N}_1(0,1)$, $Y = b + aZ$, $a, b \in \mathbb{R}$. Then $\mathbb{E}(X) = b$, $\mathbb{E}(X^2) = a^2 + b^2$, $\mathrm{Var}(X) = a^2$. If $a \neq 0$, then $\phi(z) = b + az$ is a diffeomorphism with inverse $\phi^{-1}(x) = a^{-1}(x - b)$, hence the density of X is

$$\gamma(a^{-1}(x-b))\,|a|^{-1} = (2\pi a^2)^{-1/2} \exp\left(\frac{1}{2a^2}(x-b)^2 \right)$$

If $a = 0$ then the distribution of $X = b$ is the Dirac measure at b. We say that X is Gaussian with mean b and variance a^2, $X \sim \mathrm{N}_1(b, a^2)$. Viceversa, if $X \sim \mathrm{N}_1(\mu, \sigma^2)$ and $\sigma^2 \neq 1$, then $Z = \sigma^{-1}(X - \mu) \sim \mathrm{N}_1(0,1)$.

6. The *characteristic function* of a probability measure μ is

$$\widehat{\mu}(t) = \int \mathrm{e}^{itx}\ \mu(dx) = \int \cos(tx)\ \mu(dx) + i \int \sin(tx)\ \mu(dx), \quad i = \sqrt{-1}$$

If two probability measure have the same characteristic function, then they are equal.

Exercise 6. For the standard Gaussian probability measure we have

$$\widehat{\gamma}(t) = \int \cos(tz)\ \gamma(z)dz = \mathrm{e}^{-\frac{t^2}{2}}\ .$$

In fact, by derivation under the integral

$$\frac{d}{dt}\widehat{\gamma}(t) = -\int z\sin(tz)\ \gamma(z)dz = \int \sin(tz)\gamma'(z)\ dz = -t\gamma(t)$$

and $\widehat{\gamma}(0) = 1$. The characteristic function of $X \sim \mathrm{N}_1\left(\mu, \sigma^2\right)$ is

$$\mathbb{E}\left(\mathrm{e}^{itX}\right) = \mathbb{E}\left(\mathrm{e}^{it(\mu+\sigma Z)}\right) = \mathrm{e}^{it\mu}\,\mathbb{E}\left(\mathrm{e}^{i(\sigma^t)Z}\right) = \mathrm{e}^{-t\mu+\frac{1}{2}\sigma^2t^2}$$

Exercise 7. The characteristic function $\widehat{\mu}$ of the probability measure μ on $\mathbb{R}$ is *non-negative definite*. Take $t_1, \dots, t_n$ in $\mathbb{R}$ with $n = 1, 2, \dots$. The matrix

$$T = \left[\widehat{\mu}(t_i - t_j)\right]_{i,j=1}^n = \left[\int \mathrm{e}^{i(t_i-t_j)x}\ \mu(dx)\right]_{i,j=1}^n$$

is *Hermitian*, that is the transposed matrix is equal to the conjugate matrix equivalently, T is equal to its adjoint T^*. An Hermitian matrix T is non-negative definite if for all complex vector $\boldsymbol{\zeta} \in \mathbb{C}^n$ it holds $\boldsymbol{\zeta}^* T \boldsymbol{\zeta} \geqslant 0$. In our case

$$\boldsymbol{\zeta}^*\left[\int \mathrm{e}^{i(t_i-t_j)x}\ \mu(dx)\right]\boldsymbol{\zeta} = \sum_{i.j=1}^n \int \overline{\zeta_i}\zeta_j \mathrm{e}^{i(t_i-t_j)x}\ \mu(dx) =$$

$$\sum_{i.j=1}^n \int \overline{\zeta_i}\mathrm{e}^{it_ix}\overline{\overline{\zeta_j}\mathrm{e}^{it_jx}}\ \mu(dx) = \int \left\|\sum_{i=1}^n \overline{\zeta_i}\mathrm{e}^{it_ix}\right\|^2\ \mu(dx) \geqslant 0\ .$$

Exercise 8. let $X \sim \mathrm{N}_1\left(b, \sigma^2\right)$ and $f\colon \mathbb{R} \to \mathbb{R}$ continuous and bounded. Show that $\lim_{\sigma\to 0}\mathbb{E}\left(f(X)\right) = f(b)$.

Exercise 9. Let X be a real random variable with density p with respect to the Lebesgue measure, and let $Z \sim \mathrm{N}_1\left(0, 1\right)$. Assume X and Z are independent i.e., the joint random variable (X, Z) has density $p \otimes \gamma$ with respect to the Lebesgue measure of ^{2}. Compute the density of $X + Z$. [Hint: make a change of variable $(x, z) \mapsto (x + z, z)$ then marginalize.]

7. The product of absolutely continuous probability measures is

$$(p_1 \cdot \mu_1) \otimes (p_2 \cdot \mu_2) = (p_1 \otimes p_2) \cdot \mu_1 \otimes \mu_2$$

The $\mathbb{R}^d$-valued random variable $Z = (Z_1, \dots, Z_d)$ is multivariate *standard Gaussian*, $Z \sim \mathrm{N}_n\left(0_d, I_d\right)$ if its components are IID $\mathrm{N}_1\left(0, 1\right)$. We write $\nu_d = \nu^{\otimes d}$ to denote the d-fold product measure. The distribution $\nu_d = \gamma^{\otimes d}$ of $Z \sim \mathrm{N}_n\left(0, I\right)$ has the product density

$$\mathbb{R}^n \ni \boldsymbol{z} \mapsto \gamma(\boldsymbol{z}) = \prod_{j=1}^n \phi(z_j) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\left\|\boldsymbol{z}\right\|^2\right)$$

Exercise 10. The *moment generating function* $t \mapsto \mathbb{E}\left(\exp\left(t \cdot Z\right)\right) \in \mathbb{R}_>$ is

$$\mathbb{R}^n \ni t \mapsto M_Z(t) = \prod_{j=1}^n \exp\left(\frac{1}{2}t_i^2\right) = \exp\left(\frac{1}{2}\left\|t\right\|^2\right)$$

M_Z is everywhere strictly convex and analytic.

Exercise 11. The *characteristic function* $\zeta \mapsto \widehat{\gamma}_n(\zeta) = \mathbb{E}\left(\exp\left(\sqrt{-1}\zeta \cdot Z\right)\right)$ is

$$\mathbb{R}^n \ni \zeta \mapsto \widehat{\gamma}_n(\zeta) = \prod_{j=1}^{2} \exp\left(-\frac{1}{2}\zeta_i^2\right) = \exp\left(-\frac{1}{2}\|\zeta\|^2\right)$$

$\widehat{\gamma}_n$ is non-negative definite.

2. Positive Definite Matrices

8. We collect here a few useful properties of matrices. * denotes transposition.

(1) Denote by $\operatorname{Mat}(m \times n)$ the *vector* space of $m \times n$ real matrices. We have $\operatorname{Mat}(m \times 1) \leftrightarrow \mathbb{R}^m$. Let $\operatorname{Mat}(n \times n)$ be the vector space of $n \times n$ real matrices, $\operatorname{GL}(n)$ the group of invertible matrices, $\operatorname{Sym}(n)$ the vector space of real symmetric matrices.

(2) Given $A \in \operatorname{Mat}(n \times n)$, a real eigen-value of A is a real number λ such that $A - \lambda I$ is singular i.e., $\det(A - \lambda I) = 0$. If λ is an eigen-value of A, $\boldsymbol{u}$ an eigen-vector of A associated to λ if $A\boldsymbol{u} = \lambda\boldsymbol{u}$.

(3) By identifying each matrix $A \in \operatorname{Mat}(m \times n)$ with its vectorized form $\operatorname{vec}(A) \in \mathbb{R}^{mn}$, the vector space $\operatorname{Mat}(m \times n)$ is an Euclidean space for the scalar product $\langle A, B\rangle = \operatorname{vec}(A)^* \operatorname{vec}(B) = \operatorname{Tr}(AB^*)$. The general linear group $\operatorname{GL}(n)$ is an open subset of $\operatorname{Mat}(n \times n)$.

(4) A square matrix whose columns form an orthonormal system, $S = [\boldsymbol{s}_1 \cdots \boldsymbol{s}_n]$, $\boldsymbol{s}_i^* \boldsymbol{s}_j = (i = j)$, has determinant ± 1. The property is characterised by $S^* = S^{-1}$. The set of such matrices is the orthogonal group $\operatorname{O}(n)$.

(5) *Each symmetric matrix $A \in \operatorname{Sym}(n)$ has n real eigen-values λ_i, $i = 1, \dots, n$ and correspondingly an orthonormal basis of eigen-vectors $\boldsymbol{u}_i$, $i = 1, \dots, n$.*

(6) Let $A \in \operatorname{Mat}(m \times n)$ and let $r > 0$ be its rank i.e., the dimension of the space generated by its columns, equivalently by its rows. There exist matrices $S \in \operatorname{Mat}(m \times r)$, $T \in \operatorname{Mat}(n \times r)$, and a positive diagonal $r \times r$ matrix Λ, such that $S^*S = T^*T = I_r$, and $A = S\Lambda^{1/2}T^*$. The matrix SS^* is the orthogonal projection onto $\operatorname{image} A$. In fact $\operatorname{image} SS^* = \operatorname{image} A$, $SS^*A = A$, and SS^* is a projection. Similarly, TT^* is the ortogonal projection unto $\operatorname{image} A^*$.

(7) A symmetric matrix $A \in \operatorname{Sym}(n)$ is positive definite, $A \in \operatorname{Sym}_+(n)$, respectively strictly positive definite, $A \in \operatorname{Sym}_{++}(n)$, if $\boldsymbol{x} \in \mathbb{R}^n \neq 0$ implies $\boldsymbol{x}' A \boldsymbol{x} \geqslant 0$, respectively > 0. $\operatorname{Sym}_+(n)$ is a closed pointed cone of $\operatorname{Sym}(n)$, whose interior is $\operatorname{Sym}_{++}(n)$. A positive definite matrix is strictly positive definite if it is invertible.

(8) A symmetric matrix A is positive definite, respectively strictly positive definite, if, and only if, all eigen-values are non-negative, respectively positive.

(9) A symmetric matrix B is positive definite if, and only if, $A = B'B$ for some $B \in \mathbb{M}_n$. Moreover, $A \in \operatorname{GL}_n$ if, and only if, $B \in \operatorname{GL}_n$.

(10) A symmetric matrix A is positive definite if, and only if $A = B^2$ and B is positive definite. We write $B = A^{\frac{1}{2}}$ and call B the *positive square root* of A.

Exercise 12. If you are not familiar with the previous items, try the following exercise.
Consider the matrices

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \theta \in \mathbb{R} .$$

Check that $R(\theta)^* R(\theta) = I$, $\det R(\theta) = 1$, and $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$. Compute the matrix

$$\Sigma(\theta) = R(\theta) \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R(\theta)^* , \quad \lambda_1, \lambda_2 \geqslant 0 .$$

Chech that $\det \Sigma(\theta) = \lambda_1\lambda_2$, $\Sigma(\theta)^* = \Sigma(\theta)$, the eigenvalues of $\Sigma(\theta)$ are λ_1, λ_2, and $\Sigma(\theta)R(\theta) = R(\theta)\,\mathrm{diag}\,(\lambda_1, \lambda_2)$. Compute

$$A(\theta) = R(\theta) \begin{bmatrix} \lambda_1^{1/2} & 0 \\ 0 & \lambda_2^{1/2} \end{bmatrix} R(\theta)^* \ , \quad \lambda_1, \lambda_2 \geqslant 0 \ .$$

Check that $A(\theta)A(\theta)^* = A(\theta)A(\theta) = \Sigma(\theta)$.

Exercise 13. Let $A \in \mathrm{O}(n)$ and $Z \sim \mathrm{N}_n(0, I)$. Check that $AZ \sim \mathrm{N}_n(0, I)$. let $B \in \mathrm{Mat}\,(n \times r)$, $r < n$, and assume that the columns are orthonormal. Check that $BZ\,\mathrm{N}_r(0, I)$. [Hint: complete B to an orthogonal matrix by adding columns, $[B|C] \in \mathrm{O}(n)$ and use the marginalization.]

Exercise 14. Let $Z \sim \mathrm{N}_1(0, 1)$, $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathrm{Mat}\,(2 \times 1)$. Check that AZ has no density with respect to the Lebesgue measure.

Exercise 15. Let $Z \sim \mathrm{N}_2(0, I)$, $A = \begin{bmatrix} 1 & 1 \end{bmatrix} \in \mathrm{Mat}\,(1 \times 2)$. Compute the density od AZ.

3. General Gaussian Distribution

Proposition 1.

(1) ***Definition*** *Let* $Z \sim \mathrm{N}_n(0, I)$, $A \in \mathrm{Mat}\,(m \times n)$, $b \in \mathbb{R}^m$, $\Sigma = AA^*$. *Then* $Y = b + AZ$ *has a distribution that depends on* Σ *and* b *only. The distribution of* Y *is called* Gaussian with mean b and variance Σ, $\mathrm{N}_m(b, \Sigma)$.

(2) ***Statility*** *If* $Y \sim \mathrm{N}_m(b, \Sigma)$, $B \in \mathrm{Mat}\,(r \times m)$, $c \in \mathbb{R}^r$, *then* $c + BY \sim \mathrm{N}_r(c + Bb, B\Sigma B^*)$.

(3) ***Existence*** *Given any non-negative definite* $\Sigma \in \mathrm{Sym}_+(n)$ *and any vector* $b \in \mathbb{R}^n$, *the Gaussian distribution* $\mathrm{N}_n(b, \Sigma)$ *exists.*

(4) ***Density*** *If* $\Sigma \in \mathrm{Sym}_{++}(n)$ *e.g.,* $\Sigma \in \mathrm{Sym}_+(n)$ *and moreover* $\det(\Sigma) \neq 0$, *then the Gaussian distribution* $\mathrm{N}_m(b, \Sigma)$, *has a density with respect to the Lebesgue measure on* $\mathbb{R}^n$ *given by given by*

$$p_Y(y) = (2\pi)^{-\frac{m}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-b)^T\Sigma^{-1}(y-b)\right) \ .$$

(5) ***No density*** *If the rank of* Σ *is* $r < m$, *then the distribution of* $\mathrm{N}_m(b, \Sigma)$ *is supported by the image of* Σ. *In particular it has no density w.r.t. the Lebesgue measure on* $\mathbb{R}^n$.

(6) ***Characteristic function*** $Y \sim \mathrm{N}_m(b, \Sigma)$ *if, and only if, the characteristic function is*

$$\mathbb{R}^m \ni t \mapsto \exp\left(-\frac{1}{2}t^*\Sigma t + ib^*t\right)$$

Proof.

(1) Assume $b_1, b_2 \in \mathbb{R}^m$, $A_i \in \mathrm{Mat}\,(m \times n_i)$, $Y_i = b_i + A_iZ_i$, $Z_i \sim \mathrm{N}_{n_i}(0, I)$, $i = 1, 2$. If $b_1 \neq b_2$ then the expected values of Y_1 and Y_2 are different, hence the distribution is different. Assume $b_1 = b_2 = b$, and consider the distribution of $Y_i - b = A_iZ_i$, $i = 1, 2$. We can write $A_i = S_i\Lambda_i^{1/2}T_i^*$, which in turn implies implies $\Sigma = S_i\Lambda S_i^*$, but $\Sigma = S\Lambda S^*$, hence $S_1 = S_2 = S$ and $\Lambda_1 = \Lambda_2 = \Lambda$ (a part the order). We are reduced to the case $Y_i - b = S\Lambda T_i^* Z_i$, $T_i \in \mathrm{Mat}\,(n_i \times r)$ with both with orthonormal columns. The conclusion follows from $T_1^*Z_1 \sim T_2^*Z_2$.

(2) $Y \sim \mathrm{N}_m(b, \Sigma)$ means $Y = b + AZ$ with $Z\,\mathrm{N}_n(0, I)$ and $AA^* = \Sigma$. It follows

$$c + BY = c + B(b + AZ) = (c + Bb) + (BA)Z \ ,$$

wth $(BA)(BA)^* = BAA^*B^* = B\Sigma B^*$.

(3) Take $Y = b + \Sigma^{1/2}Z$, $Z \sim \mathrm{N}_n(0, I)$.

(4) Use the change of variable formula in $Y = b + AZ$ with $A = \Sigma^{1/2}$ to get

$$p_Y(y) = \left|\det\left(A^{-1}\right)\right| p_Z(A^{-1}(y - b)) \ .$$

The express each term with Σ.

(5) use the decomposition $\Sigma = S\Lambda S^*$ and note that some elements on the diagonal of Λ are zero.

(6) The "if" part is a computation, the "only if" part requires the injection property of characteristic function.

□

Exercise 16 (Linear interpolation of the Brownian motion). Let Z_n, $n = 1, 2\dots$ be IID $\mathrm{N}_1(0, 1)$. Given $0 < \sigma \ll 1$, define recursively the times $t_0 = 0$ and $t_{n+1} = t_n + \sigma^2$. Let $T = \{t_n | n = 0, 1, \dots\}$. Define recursively $B(0) = 0$, $B(t_{n+1}) = B(t_n) + \sigma Z_n$. As $B(t_n) = \sum_{i=1}^n \sigma Z_i = \sigma \sum_{i=1}^n Z_i$, then $\mathrm{Var}(B(t_n)) = \sigma^2 \mathrm{Var}(\sum_{i=1}^n Z_i) = n\sigma^2 = t_n$. For each $t \in \mathbb{R}_{>0} \setminus T$, define $B(t)$ by linear interpolation i.e.,

$$B(t) = \frac{t_{n+1} - t}{t_{n+1} - t_n} B(t_n) + \frac{t - t_n}{t_{n+1} - t_n} B(t_{n+1}) \ , \quad t \in [t_n, t_{n+1}] \ .$$

Compute the variance of $B(t)$ and the density of $B(t)$.

4. Independence of Jointly Gaussian Random Variables

Proposition 2. *Consider a partitioned Gaussian vector*

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathrm{N}_{n_1+n_2}\left(\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) \ .$$

Let $r_i = \mathrm{Rank}(\Sigma_{ii})$, $\Sigma_{ii} = S_i \Lambda_i S_i^*$ *with* $S_i \in \mathrm{Mat}(n_i \times r_i)$, $S_i^* S = I_{r_i}$, *and* $\Lambda_i \in \mathrm{diag}_{++}(r_i)$, $i = 1, 2$.

(1) *The blocks* Y_1, Y_2 *are independent,* $Y_1 \perp\!\!\!\perp Y_2$*, if, and only if,* $\Sigma_{12} = 0$*, hence* $\Sigma_{21} = \Sigma_{12}^* = 0$*. More precisely, if, and only if, there exist two independent standard Gaussian* $Z_i \sim \mathrm{N}_{r_i}(0, I)$ *and matrices* $A_i \in \mathrm{Mat}(n_i \times r_i)$, $i = 1, 2$, *such that*

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \begin{bmatrix} A_1 & 0 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \ .$$

(2) *(The following property is sometimes called* Schur complement lemma.*) Write* $\Sigma_{22}^+ = S_2 \Lambda_2^{-1} S_2^*$*. Then,*

$$\begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^+ \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Sigma_{22}^+\Sigma_{21} & I \end{bmatrix} =$$

$$\begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^+\Sigma_{21} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Sigma_{22}^+\Sigma_{21} & I \end{bmatrix} =$$

$$\begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^+\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \ ,$$

hence the last matrix is non-negative definite. The Shur complement of the partitioned covariance matrix Σ *is*

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} \in \mathrm{Sym}_{+}(n_1) \ .$$

(3) *Assume* $\det(\Sigma) \neq 0$*. Then both* $\det\left(\Sigma_{1|2}\right) \neq 0$ *and* $\det(\Sigma)_{22} \neq 0$*. If we define the partitioned* concentration *to be*

$$K = \Sigma^{-1} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} ,$$

then $K_{11} = \Sigma_{1|2}^{-1}$ *and* $K_{11}^{-1}K_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$.

Exercise 17. Let $\Sigma \in \mathrm{Sym}_{+}(n)$ and let $r = \mathrm{Rank}(\Sigma)$. We know that $\Sigma = S\Lambda S^*$ with $S \in \mathrm{Mat}(n \times r)$, $S^*S = I_r$, $\lambda \in \mathrm{diag}_{++}(r)$. Let us define $\Sigma^{+} = S\Lambda^{-1}S^*$. Then it follows by simple computation that $\Sigma^{+}\Sigma = \Sigma\Sigma+ = SS^*$. Also, $\Sigma\Sigma^{+}\Sigma = \Sigma$ and $\Sigma^{+}\Sigma\Sigma^{+} = \Sigma^{+}$. If $Y \sim \mathrm{N}_n(0, \Sigma)$, then $Y = SS^*Y$. In fact, $Y - SS^*Y = (I - SS^*)Y$ is a Guassian random variable with variance $(I - SS^*)S\Lambda S^*(I - SS^*) = 0$ because $(I - SS^*)S = S - SS^*S = S - S = 0$.

Proof. (1) If the blocks are independent, they are uncorrelated. Conversely, if $\Sigma_{ii} = S_i\Lambda_i S_i^*$, $i = 1, 2$, define $A_i = S_i\Lambda_i^{1/2}$ to get

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}^* = \Sigma \ .$$

(2) Computations using Ex. 17.

(3) From the computation above we see that the Schur complement is positive definite and that

$$\det\left(\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) = \det\left(\Sigma_{1|2}\right)\det\left(\Sigma_{22}\right) \ .$$

It follows that $\det(\Sigma) \neq 0$ implies both $\det\left(\Sigma_{1|2}\right) \neq 0$ and $\det(\Sigma_{22}) \neq 0$. The condition

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

is equivalent to

$$\begin{aligned} I &= K_{11}\Sigma_{11} + K_{12}\Sigma_{21} \\ 0 &= K_{11}\Sigma_{12} + K_{12}\Sigma_{22} \\ &\vdots \end{aligned}$$

Right-multiply the second equation by Σ_{22}^{-1} and substitute in the first one, to get $K_{11}\Sigma_{1|2} = I$, hence $K_{11}^{-1} = \Sigma_{1|2}$. The other equality follows by left-multiplying the second equation by K_{11}^{-1}.

□

Exercise 18 (Whitening). Let $Y \sim \mathrm{N}_n(b, \Sigma)$. Assume Σ has rank r and decomposition $\Sigma = S\Lambda S^*$, $S^*S = I_r$, $\lambda \in \mathrm{diag}_{++}(r)$. Then $Z = \Lambda^{-1/2}S^*(Y - b)$ ia a white noise, $Z \sim \mathrm{N}_r(O, I)$. Moreover, $b + S\Lambda^{1/2}Z = Y$. In fact,

$$Y - (b + S\Lambda^{1/2}Z) = (Y - b) - S\Lambda^{1/2}\Lambda^{-1/2}S^*(Y - b) = (I - SS^*)(Y - b) = 0 \ .$$

References

[1] T. W. Anderson, *An introduction to multivariate statistical analysis*, third ed., Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003. MR 1990662

[2] Rajendra Bhatia, *Positive definite matrices*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2007. MR 2284176 (2007k:15005)

[3] Paul R. Halmos, *Finite-dimensional vector spaces*, The University Series in Undergraduate Mathematics, D. Van Nostrand Co., Inc., Princeton-Toronto-New York-London, 1958, 2nd ed. MR 0089819

Collegio Carlo Alberto
E-mail address: giovanni.pistone@carloalberto.org