PROBABILITY 2017
HANDOUT 2: GAUSSIAN DISTRIBUTION

GIOVANNI PISTONE

Contents

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 $J\phi^{-1} = \left(J\phi \circ \phi^{-1}\right)^{-1}$. For each non-negative $f \colon \mathcal{B} \to \mathbb{R}^n$,

$$\int_{\mathcal{B}} f(\boldsymbol{y})\ d\boldsymbol{y} = \int_{\mathcal{A}} f \circ \phi(\boldsymbol{x})\ \ |\det\left(J\phi(\boldsymbol{x})\right)|\ \ d\boldsymbol{x} \tag{1}$$

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

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

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

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

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

Date: Moncalieri, March 2017.

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

$$\phi_\#\ell = \left|\det\left(J\phi^{-1}\right)\right|\cdot\ell = \left|\det\left(J\phi\circ\phi^{-1}\right)\right|^{-1}\cdot\ell \tag{2}$$

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

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

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

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. 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. All moments $\mu(n) = \int z^n\gamma(z)\ dz$ exists. If $f\colon \mathbb{R}\to\mathbb{R}$ absolutely continuous with $\int|f'(z)|\ \gamma(z)dz < +\infty$ then $\int|zf(z)|\ \gamma(z)dz < +\infty$ and $\int zf(z)\ \gamma(z)dz = \int f'(z)\ \gamma(z)dz$. The *Stein operator* $\delta f(z) = zf(z) - f'(z)$. We have

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

It follows $(1+n)\mu(n) = \mu(n+2)$. We define the *Hermite polynomials* to be $H_n(z) = \delta^n 1$; they are orthogonal with respect to $\gamma\cdot\ell$.

4. Let $Z \sim \mathrm{N}_1(0,1)$, $Y = b + aZ$, $a,b\in\mathbb{R}$. Then $\mathrm{E}(X) = b$, $\mathrm{E}(X^2) = a^2+b^2$, $\mathrm{Var}(X) = a^2$. 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 $Z = \sigma^{-1}(X-\mu) \sim \mathrm{N}_1(0,1)$.

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

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

For the standard Gaussian measure we have

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

If two probability measure have the same characteristic function, then they are equal. See i.e., [2, Ch. 13]. The characteristic function is non-negative definite. The characteristic function of $X \sim \mathrm{N}_1(\mu, \sigma^2)$ is

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

6. The product of absolutely continuous probability measures is

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

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

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

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

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

M_Z is everywhere strictly convex and analytic. The *characteristic function* $\zeta \mapsto \widehat{\gamma}_n(\zeta) = \mathrm{E}\left(\exp\left(\sqrt{-1}\zeta \cdot Z\right)\right) \in \mathbb{C}$ 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 and analytic.

2. Positive Definite Matrices

We collect here useful properties of matrices. The algebra of matrices used in Gaussian statistical models is discussed in the monograph [1, Appendix A]. Calculus on the space of matrices is treated in [3]. Find below a check-list of relevant facts.

(1) Denote by $\mathrm{Mat}(m \times n)$ the *vector* space of $m \times n$ real matrices. We have $M_{n,1} \leftrightarrow \mathbb{R}^n$. Let $\mathrm{Mat}(n \times n)$ be the vector space of $n \times n$ real matrices, $\mathrm{GL}(n)$ the group of invertible matrices, $\mathrm{Sym}(n)$ the vector space of real symmetric matrices.
(2) By identifying each matrix $A \in \mathrm{Mat}(m \times n)$ with its vectorized form $\mathrm{vec}(A) \in \mathbb{R}^{mn}$, the vector space $\mathrm{Mat}(m \times n)$ is an Hilbert space for the scalar product $\langle A, B \rangle = \mathrm{vec}(A)^* \mathrm{vec}(B) = \mathrm{Tr}(AB^*)$. The general linear group $\mathrm{GL}(n)$ is an open subset of $\mathrm{Mat}(n \times n)$.
(3) The mapping $f\colon \mathrm{Mat}(n \times n) \to \mathbb{R}$, $f(A) = \det(A)$, has derivative at A in the direction H (that is derivative at zero of $t \mapsto \det(A + tH) \in \mathbb{R}$), equal to $\mathrm{Tr}(\mathrm{adj}(A)H)$.
(4) The mapping $f\colon \mathrm{GL}(n) \to \mathrm{GL}(n)$, $f(A) = A^{-1}$, has derivative at A in the direction H, that is the derivative at zero of $t \mapsto (A+tH) \in \mathrm{GL}_n$, equal to $-A^{-1}HA^{-1}$.
(5) 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 $S^* = S^{-1}$.
(6) Each symmetric matrix $A \in \mathbb{S}_n$ has n real eigen-values λ_i, $i = 1, \dots, n$ and correspondingly an orthonormal basis of eigen-vectors $\boldsymbol{u}_i$, $i = 1, \dots, n$.

(7) Let $A \in \mathrm{Mat}\,(m \times n)$ 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 $S \in \mathrm{Mat}\,(m \times r)$, $T \in \mathrm{Mat}\,(n \times r)$, 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 Image A. In fact Image $SS^* =$ Image A, $SS^*A = A$, and SS^* is a projection. Similarly, TT^* is the ortogonal projection unto Image A^*.

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

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

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

(11) A symmetric matrix B is positive definite, if, and only if, there exist an upper triangular matrix T such that $A = T'T$. T can be chosen to have nonnegative diagonal entries and it is unique if A is invertible.

(12) A symmetric matrix is positive definite, respectively strictly positive definite, if and only if all leading principal minors are nonnegative.

(13) 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.

(14) A symmetric matrix A is positive definite, respectively strictly positive definite, if there exist an Hilbert space $\mathbb{H}$ and vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, respectively linear independent vectors, with $a_{ij} = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$.

3. General Gaussian Distribution

Proposition 1. (1) *Let $Z \sim \mathrm{N}_n\,(0, I)$, $A \in \mathrm{Mat}\,(m \times n)$, $b \in \mathbb{R}^m$, $\Sigma = AA^T$. 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) *Given any non-negative definite Σ, there exists matrices A such that $\Sigma = AA^T$.*

(3) *If $\det(\Sigma) \neq 0$, then the distribution of $Y = b + AZ \sim \mathrm{N}_m\,(b, \Sigma)$, $A \in \mathbb{R}^{m \times m}$, $AA^T = \Sigma$, has a density given by*

$$\mathbb{R}^m \ni y \mapsto p_Y(y) = \left|\det\left(A^{-1}\right)\right| p_Z(A^{-1}(y-b)) = (2\pi)^{-\frac{m}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-b)^T\Sigma^{-1}(y-b)\right)$$

(4) *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 $\mathbb{R}^m$.*

(5) *$Y \sim \mathrm{N}_m\,(b, \Sigma)$ 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 $b_1, b_2 \in \mathbb{R}^m$, $A_i \in \mathrm{Mat}\,(m \times n_i)$, $Y_i = b_i + A_iZ_i$, $Z_i \sim \mathrm{N}_{n_1}\,(0, I)$, $i = 1, 2$. 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 $Y_i - b = A_iZ_i$, $i = 1, 2$. The singular value decomposition $A_i = S_i\Lambda_i^{1/2}T_i^*$ implies $\Sigma = S_i\Lambda S_i^*$, hence $S_1 = S_2 = S$ and $\Lambda_1 = \Lambda_2 = \Lambda$ (a part the order), and we are reduced to the case $Y_i - b = S\Lambda T_i^* Z_i$, $T_i \in \mathrm{Mat}\,(n_i \times r)$ and orthogonal, $i = 1, 2$. The conclusion follows from $T_1^* Z_1 \sim T_2^* Z_2$.

(2) Take $A = \Sigma^{1/2}$.

(3) Use the change of variable formula.

(4) From the singular value decomposition.

(5) The "if" part is a computation, the "only if" part requires the injection property of characteristic function, see for example [2, Ch. 13].

$\square$

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

(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*

$$\begin{cases} Y_1 = b_1 + A_1 Z_1 \ , \\ Y_2 = b_2 + A_2 Z_2 \ . \end{cases}$$

(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* $K_{11}^{-1} K_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$

Proof. (1) If the blocks are independent, they are uncorrelated. Viceversa,

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

(2) Computations.

(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 $\det(\Sigma_{22}) \neq 0$.

□

Proposition 3. (1) *Define The Gaussian random vector with components*

$$\begin{aligned}\widetilde{Y}_1 &= Y_1 - (b_1 + L_{12}(Y_2 - b_2)), \quad L_{12} = \Sigma_{12}\Sigma_{22}^{+}\\ \widetilde{Y}_2 &= Y_2 - b_2\end{aligned}$$

is such that $\mathrm{E}\left(\widetilde{Y}_1\right) = 0$, $\mathrm{Var}\left(\widetilde{Y}_1\right) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21}$, *and* $\widetilde{Y}_1 \perp\!\!\!\perp \widetilde{Y}_2$. *It follows*

$$\mathrm{E}\left(Y_1|Y_2\right) = b_1 + L_{12}(Y_2 - b_2)$$

(2) *The conditional distribution of* Y_1 *given* $Y_2 = y_2$ *is Gaussian with*

$$Y_1|(Y_2 = y_2) \sim \mathrm{N}_{n_1}\left(b_1 + L_{12}(y_2 - b_2), \Sigma_{11} - L_{12}\Sigma_{21}\right)$$

(3) *The conditional density of* Y_1 *given* $Y_2 = y_2$ *in terms of the partitioned concentration is*

$$p_{Y_1|Y_2}(y_1|y_2) = (2\pi)^{-\frac{n_1}{2}}\det\left(K_{1|2}\right)^{\frac{1}{2}} \times \\ \exp\left(-\frac{1}{2}(y_1 - b_1 - K_{11}^{-1}K_{12}(y_2 - b_2))^T K_{11}(y_1 - b_1 - K_{11}^{-1}K_{12}(y_2 - b_2))\right)$$

Proof. (1) We have

$$\begin{bmatrix}\widetilde{Y}_1\\ \widetilde{Y}_2\end{bmatrix} = \begin{bmatrix}I & -\Sigma_{12}\Sigma_{22}^{+}\\ 0 & I\end{bmatrix}\begin{bmatrix}Y_1 - b_1\\ Y_2 - b_2\end{bmatrix} \sim \mathrm{N}_{n_1+n_2}\left(0, \begin{bmatrix}\Sigma_{1|2} & 0\\ 0 & \Sigma_{22}\end{bmatrix}\right)$$

It follows

$$\mathrm{E}\left(Y_1|Y_2\right) = \mathrm{E}\left(\widetilde{Y}_1 + b_1 + L_{12}(Y_2 - b_2)\middle|Y_2\right) = \mathrm{E}\left(\widetilde{Y}_1\right) + b_1 + L_{12}(Y_2 - b_2)$$

(2) The conditional distribution of Y_1 given Y_2 is a transition probability $\mu_{Y_1|Y_2}\colon \mathcal{B}(\mathbb{R}^{n_1}) \times \mathbb{R}^{n_2}$ such that for all bounded $f\colon \mathbb{R}^{n_1}$

$$\mathrm{E}\left(f(Y_1)|Y_2\right) = \int f(y_1)\ \mu_{Y_1|Y_2}(dy_1|Y_2).$$

We have

$$\mathrm{E}\left(f(Y_1)|Y_2\right) = \mathrm{E}\left(f(\widetilde{Y}_1 + \mathrm{E}\left(Y_1|Y_2\right))\middle|Y_2\right) = \int f(x + \mathrm{E}\left(Y_1|Y_2\right))\ \gamma(dx; 0, \Sigma_{1|2})$$

where $\gamma(dx; 0, \Sigma_{1|2})$ is the measure of $\mathrm{N}_{n_1}\left(0, \Sigma_{1|2}\right)$. We obtain the statement by considering the effect on the distribution $\mathrm{N}_{n_1}\left(0, \Sigma_{1|2}\right)$ of the translation $x \mapsto x + (b_1 + L_{12}(y_2 - b_2))$.

(3) A further application of the Schur complement gives

$$\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} = \begin{bmatrix}I & \Sigma_{12}\Sigma_{22}^{-1}\\ 0 & I\end{bmatrix}\begin{bmatrix}\Sigma_{1|2} & 0\\ 0 & \Sigma_{22}\end{bmatrix}\begin{bmatrix}I & 0\\ \Sigma_{22}^{-1}\Sigma_{21} & I\end{bmatrix}$$

whose inverse is

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix} \begin{bmatrix} \Sigma_{1|2}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{1|2}^{-1} & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{1|2}^{-1} & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{1|2}^{-1} & -\Sigma_{1|2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{1|2}^{-1} & \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{1|2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} + \Sigma_{22}^{-1} \end{bmatrix}$$

In particular, we have $K_{11} = \Sigma_{1|2}^{-1}$ and $K_{11}^{-1}K_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$, hence

$$Y_1|(Y_2 = y_2) \sim \mathrm{N}_{n_1}\left(b_1 - K^{-1}K_{12}(y_2 - b_2), K_{11}^{-1}\right)$$

so that the exponent of the Gaussian density has the factor

$$(y_1 - b_1 + K_{11}^{-1}K_{12}(y_2 - b_2))^T K_{11}(y_1 - b_1 + K_{11}^{-1}K_{12}(y_2 - b_2))$$

□

5. Conditional independence

Conditional independence is a key property in Statistics e.g. Graphical Models, in Stochastic Processes e.g., Markov processes, in Random Fields, in Machine Learning.

Definition 1.

(1) The nonzero events A, B, C are such that A and C are independent given B, $A \perp\!\!\!\perp C|\, B$, if each one of the following equivalent conditions are satisfied:

$$\mathrm{P}(A \cap C|B) = \mathrm{P}(A|B)\,\mathrm{P}(C|B)$$
$$\mathrm{P}(A|B \cap C) = \mathrm{P}(A|B)$$
$$\mathrm{P}(A \cap B \cap C)\,\mathrm{P}(B) = \mathrm{P}(A \cap B)\,\mathrm{P}(B \cap C)$$

(2) Random variables Y_1, Y_3 are conditionally independent given the random variable Y_2, $Y_1 \perp\!\!\!\perp Y_3|\, Y_2$ if each one of the following equivalent conditions are satisfied. If f_i, $i = 1, \dots, 3$, are bounded,

$$\mathrm{E}(f_1(Y_1)f_3(Y_3)|Y_2) = \mathrm{E}(f_1(Y_1)|Y_2)\,\mathrm{E}(f_3(Y_3)|Y_2)$$
$$\mathrm{E}(f_1(Y_1)|Y_2, Y_3) = \mathrm{E}(f_1(Y_1)|Y_2)$$

(3) A stochastic process $Y_1, \dots, Y_N$ is a *Markov Process* if $(Y_1, \dots, Y_k) \perp\!\!\!\perp Y_k, \dots, Y_N|\, Y_k$, $k = 1, 2, \dots, N$.

Proposition 4. *Let be given*

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim \mathrm{N}_{n_1+n_2+n_3}\left(\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}\right)$$

We have $Y_1 \perp\!\!\!\perp Y_3|\, Y_2$ *if, and only if,* $\Sigma_{13} = \Sigma_{12}\Sigma_{22}^{+}\Sigma_{23}$. *In such a case,*

(1)

$$\begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix}\Bigg|(Y_2 = y_2) \sim \mathrm{N}_{n_1+n_3}\left(\begin{bmatrix} b_1 \\ b_3 \end{bmatrix} + \begin{bmatrix} \Sigma_{12} \\ \Sigma_{32} \end{bmatrix}\Sigma_{22}^{+}(y_2 - b_2), \begin{bmatrix} \Sigma_{1|2} & 0 \\ 0 & \Sigma_{3|2} \end{bmatrix}\right)$$

(2)

$$Y_1|(Y_2 = y_2, Y_3 = y_3) = Y_1|(Y_2 = y_2) \sim \mathrm{N}_{n_1}\left(b_1 + \Sigma_{1,2}\Sigma_{22}^{+}(y_2 - b_2), \Sigma_{1|2}\right)$$

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] Jean Jacod and Philip Protter, *Probability essentials*, second ed., Universitext, Springer-Verlag, Berlin, 2003.

[3] Jan R. Magnus and Heinz Neudecker, *Matrix differential calculus with applications in statistics and econometrics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1999, Revised reprint of the 1988 original. MR 1698873

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