STOCHASTIC PROCESSES 2015

2. MULTIVARIATE GAUSSIAN DISTRIBUTION

GIOVANNI PISTONE

Contents

Books

DD: Textbook by D. Dacunha-Castelle and M. Duflo [2]
JP: Textbook by J. Jacod and Protter [3]
RB: Monograph by R. Bhatia on *Positive Definite Matrices* [1]
GM: Monograph by S. Lauritzen on *Graphical Models* [4]

1. Recap

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

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

with respect to the Lebesgue measure.

2. *The* $\mathbb{R}^n$*-valued random variable* $Z = (Z_1, \dots, Z_n)$ *is* standard Gaussian, $Z \sim \mathrm{N}_n(0_n, I_n)$ *if its components are IID* $\mathrm{N}_1(0,1)$.

3.

(1) *The distribution* $\gamma_n = \gamma^{\otimes n}$ *of* $Z \sim \mathrm{N}_n(0, I)$ *has the product density*

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

(2) *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.*

Date: Moncalieri, May 25, 2015.

(3) *The* characteristic function $\zeta \mapsto \check{\gamma}_n(\zeta) = \mathrm{E}\left(\exp\left(\sqrt{-1}\zeta \cdot Z\right)\right) \in \mathbb{C}$ *is*

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

$\check{\gamma}_n$ *is nonnegative definite and analytic.*

4.

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

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

2. Conditioning

Proposition 1. *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* $\Sigma_{ii} = U_i\Lambda_iU_i^T$ *with* $U_i \in \mathbb{R}^{n_1\times r_i}$, $\Lambda_i \in \mathbb{R}^{r_i\times r_i}$ *positive diagonal,* $i = 1, 2$.

(1) *The blocks* Y_1, Y_2 *are independent, say* $Y_1 \perp\!\!\!\perp Y_2$, *if, and only if,* $\Sigma_{12} = 0$ *and* $\Sigma_{21} = 0$. *More precisely, if, and only if, there exist two independent standard Gaussian* $Z_i \sim \mathrm{N}_{r_1}(0, I)$ *and* $Z_2 \sim \mathrm{N}_{r_2}(0, I)$ *and matrices* A_1, A_2 *such that*

$$\begin{cases} Y_1 = b_1 + A_1Z_1 , \\ Y_2 = b_2 + A_2Z_2 . \end{cases}$$

(2) *Define* $\Sigma_{22}^+ = U_2\Lambda_2^{-1}U_2^T$. *The Gaussian random vector with components*

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

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

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

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

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

(4) *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*

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

then $K_{11} = \Sigma_{1|2}^{-1}$ *and* $K_{11}^{-1}K_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$, *so that 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 of (1). If the blocks are independent, they are uncorrelated. Viceversa, assume $\Sigma_{12} = 0$ and $\Sigma_{21} = \Sigma_{12}^T = 0$. As Σ_{ii} are nonnegative definite, $i = 1, 2$, there are spectral decompositions $\Sigma_{11} = U_1 \Lambda_1 U_1^T$, $\Sigma_{22} = U_2 \Lambda_2 U_2^T$, with $U_i \in \mathbb{R}^{n_1 \times r_1}$, $U_i U_i^T = I_{r_i}$ and Λ_i positive diagonal, $i = 1, 2$. We define

$$A_i = U_i \Lambda_i^{1/2}, \quad A_i^+ = \Lambda_i^{-1/2} U_i^T \ ,$$

so that $A_i A_i^T = \Sigma_{ii}$ and $A_i^+ \Sigma_{ii} A_i^{+T} = I_{r_i}$, $i = 1, 2$. The Gaussian random vector

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} A_1^+ & 0 \\ 0 & A_2^+ \end{bmatrix} \begin{bmatrix} Y_1 - b_1 \\ Y_2 - b_2 \end{bmatrix}$$

is $\mathrm{N}_{r_1+r_2}(0, I_{r_1+r_2})$, in particular $Z_1 \perp\!\!\!\perp Z_2$. We have

$$A_i Z_i = A_i A_i^+ (Y_i - b_i) = U_i U_i^T (Y_i - b_i) = Y_i - b_i$$

because $U_i U_i^T$ is the orthogonal projection of $\mathbb{R}^{n_i}$ on the subspace of the values of the random vector Y_i, $i = 1, 2$. In conclusion, for $i = 1, 2$ there exist independent white noise presentations. $\square$

Proof of (2). We start with an algebraic property sometimes called *Schur complement lemma.* Write $\Sigma_{22}^+ = U_2 \Lambda_2^{-1} U_2^T$ and compute

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

where we have used the equalities $\Sigma_{22}\Sigma_{22}^+\Sigma_{22}^+ = \Sigma_{22}^+$, $\Sigma_{22}^+\Sigma_{22}\Sigma_{22} = \Sigma_{22}$, $(I - \Sigma_{22}\Sigma_{22}^+)\Sigma_{21} = 0$. In particular, the last one depends on $U_2 U_2^T$ being the orthogonal projection on the support of Y_2.

The matrix $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^+\Sigma_{21}$ is sometimes called the Schur complement of the partitioned matrix. From the computation above we see that the Schur complement is nonnegative definite and that

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

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

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

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

$\square$

Proof of (3). 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 $f \colon \mathbb{R}^{n_1}$

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

We have

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

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 $x \mapsto x + (b_1 + L_{12}(y_2 - b_2))$. $\square$

Proof of (4). 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

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

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

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

□

3. 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:

$$\begin{aligned} \mathrm{P}(A \cap C|B) &= \mathrm{P}(A|B)\,\mathrm{P}(C|B) \\ \mathrm{P}(A|B \cap C) &= \mathrm{P}(A|B) \\ \mathrm{P}(A \cap B \cap C)\,\mathrm{P}(B) &= \mathrm{P}(A \cap B)\,\mathrm{P}(B \cap C) \end{aligned}$$

(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,

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

Errata: the equality

$$\mathrm{E}(f_1(Y_1)f_2(Y_2)f_3(Y_3))\,\mathrm{E}(f_2(Y_2)) = \mathrm{E}(f_1(Y_1)f_2(Y_2))\,\mathrm{E}(f_2(Y_2)f_3(Y_3))$$

is not implied by conditional independence for all f_2, as it can be seen by taking $f_2 = 1$. It is true for some $f_2(Y_2)$.

Proposition 2. *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. Rajendra Bhatia, *Positive definite matrices*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2007. MR 2284176 (2007k:15005)
2. Didier Dacunha-Castelle and Marie Duflo, *Probabilités et statistiques. tome 1: Problèmes à temps fixe*, Collection Mathématiques Appliqées pour la Matrise, Masson, Paris, 1982.
3. Jean Jacod and Philip Protter, *Probability essentials*, second ed., Universitext, Springer-Verlag, Berlin, 2003. MR MR1956867 (2003m:60002)
4. Steffen L. Lauritzen, *Graphical models*, The Clarendon Press Oxford University Press, New York, 1996, Oxford Science Publications. MR 98g:62001

Home work

Read all the texts below, then chose and solve 2 exercises. The paper is due Fri May 22.

Exercise 1. Derive in detail the conditional density of Proposition 1(4) for the bi- and tri-variate Gaussian distribution.

Exercise 2. Prove the equivalences in Definition 1.

Exercise 3. Derive in detail the two forms of the conditional independence in Proposition 2 for the tri-variate Gaussian distribution.

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