STOCHASTIC PROCESSES AND CALCULUS 2016

2. MULTIVARIATE GAUSSIAN DISTRIBUTION

GIOVANNI PISTONE

Contents

References

- The importance of the Gaussian distribution depend largely on the Central Limit Theorem, see [2, Part C]
- A classic on Multivariate Statistics is T. W. Anderson's monograph [1].

1. Standard Gaussian Distribution

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 $Z = (Z_1, \ldots, Z_n)$ is *standard Gaussian*, $Z \sim \mathrm{N}_n(0_n, I_n)$ if its components are IID $\mathrm{N}_1(0,1)$. We write $\gamma^{\otimes n}$ to denote the n-fold product measure.

3. For each multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geqslant}$ define Hermite polynomial $H_\alpha(x_1, \ldots, x_n) = H_{\alpha_1}(x_1) \cdots H_{\alpha_n}(x_n)$. The sequence $\frac{1}{\alpha!}H_\alpha$ is orthonormal and total in $L^1(\gamma^{\otimes n})$.

4.

(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 2016 v02.

(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.

2. Positive Definite Matrices

We collect here useful properties of matrices. See [1, Appendix A].

(1) Denote by $\mathbb{M}_{m\times n}$ the *vector* space of $m \times n$ real matrices. We have $M_{n,1} \leftrightarrow \mathbb{R}^n$. Let $\mathbb{M}_n$ be the vector space of $n \times n$ real matrices and by GL_n the group of invertible matrices. We denote by $\mathbb{S}_n$ its sub-vector space of real symmetric matrices.

(2) The vector space $\mathbb{M}_{m,n}$ is an Hilbert space for the scalar product $\langle A, B\rangle = \mathrm{Tr}\,(B'A)$. The general linear group GL_n is an open subset of $\mathbb{M}_{m.n}$.

(3) The mapping $f\colon \mathbb{M}_n$, $f(A) = \det(A)$ 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 $f\colon \mathrm{GL}_n$, $f(A) = A^{-1}$ 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) 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 $\boldsymbol{u}_i$, $i = 1, \dots, n$.

(6) A matrix $A \in \mathbb{M}_n$ is positive definite, respectively strictly positive definite, if $\boldsymbol{x} \in \mathbb{R}^n \neq 0$ implies $\boldsymbol{x}'A\boldsymbol{x} \geqslant 0$, respectively > 0. We denote by $\mathbb{S}_n^+$ the closed cone of $\mathbb{S}_n$ of positive definite matrices. A positive definite matrix is strictly positive definite if it is invertible. The set of strictly positive symmetric matrices is the interior of the cone $\mathcal{S}_n$.

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

(8) 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, $B \in \mathrm{GL}_n$.

(9) 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.

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

(11) 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.

(12) 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* $Z \sim \mathrm{N}_n(0, I)$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\Sigma = AA^T$. *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*

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

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

(4) *If* $\det(\Sigma) = 0$ *the distribution of* $N(b, \Sigma)$ *has no density w.r.t. the Lebesgue measure on* $\mathbb{R}^n$.

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* $\Sigma_{ii} = U_i \Lambda_i U_i^T$ *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*

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

(2) *Define* $\Sigma_{22}^{+} = U_2 \Lambda_2^{-1} U_2^T$. *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*

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

(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*

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

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

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

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

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

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

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

□

Proof of (3). The conditional distribution of Y_1 given Y_2 is a transition probability $\mu_{Y_1|Y_2} : \mathcal{B}(\mathbb{R}^{n_1}) \times \mathbb{R}^{n_2}$ such that for all bounded $f : \mathbb{R}^{n_1}$

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

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

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

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

□

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:

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

(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,

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

(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 3. *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. David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.

Home work

Read all the texts below, then pick up and solve 2 exercises. The paper is due Mon May 16.

Exercise 1. Derive in detail the conditional density of Proposition 2(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 3 for the tri-variate Gaussian distribution.

Exercise 4. Compute the joint density of $(Y_1, Y_2, Y_3) \sim \mathrm{N}(b, \Sigma)$, $\det(\Sigma) \neq 0$, if the process is Markov.

Exercise 5. Compute the joint density of $(Y_1, Y_2, Y_3) \sim \mathrm{N}(b, \Sigma)$, $\det(\Sigma) \neq 0$, if the process is a martingale, i.e. $\mathrm{E}(Y_2|Y_1) = Y_1$ and $\mathrm{E}(Y_3|Y_1, Y_2) = Y_2$.

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