

# Capitolo 2

## Variabili aleatorie Gaussiane

per gli studenti del corso di  
**Stima e identificazione**

Luigi Chisci, 28 Febbraio 2019

Si richiamano brevemente definizioni e proprietà riguardanti variabili aleatorie Gaussiane, dette anche normali.

### Definizione

Facendo riferimento alla variabile aleatoria vettoriale  $X = [X_1, \dots, X_n]^T$  definita su  $\mathbb{R}^n$ , si dice che tale variabile è Gaussiana (normale) di media  $m_X \in \mathbb{R}^n$  e covarianza  $\Sigma_X \in \mathbb{R}^{n \times n}$ ,  $\Sigma_X = \Sigma_X^T > 0$ , se la sua densità di probabilità (PDF) è

$$f_X(x) = \mathcal{N}(x; m_X, \Sigma_X) \triangleq \frac{1}{\sqrt{\det(2\pi\Sigma_X)}} \exp\left[-\frac{1}{2}(x - m_X)^T \Sigma_X^{-1}(x - m_X)\right] \quad (1.1)$$

Si userà la notazione  $X = \mathcal{N}(m_X, \Sigma_X)$  a indicare che  $X$  è una variabile aleatoria Gaussiana (normale) di media  $m_X$  e covarianza  $\Sigma_X$ , ovvero  $X \sim \mathcal{N}(\cdot; m_X, \Sigma_X)$  a indicare che  $X$  ha la PDF definita in (1.1). Indicata con  $m_i$  la  $i$ -esima componente di  $m_X$  e con  $\sigma_{ij}$  la componente  $(i, j)$  della matrice  $\Sigma_X$ , risulta che

$$m_i = E[X_i], \quad \sigma_{ii} = \sigma_i^2 = \text{var}(X_i) \triangleq E[\tilde{X}_i^2], \quad \sigma_{ij} = \text{cov}(X_i, X_j) \triangleq E[\tilde{X}_i \tilde{X}_j]$$

con  $\tilde{X} \triangleq X - m_X = [\tilde{X}_1, \dots, \tilde{X}_n]^T$ . Poiché la matrice di covarianza deve risultare, per definizione, simmetrica e non-negativa definita, le sue componenti devono soddisfare le seguenti condizioni:

$$\sigma_{ii} = \sigma_i^2 \geq 0, \forall i \quad \rho_{ij} \triangleq \frac{\sigma_{ij}}{\sigma_i \sigma_j} \in [-1, 1], \forall i \neq j.$$

Nel caso, detto *non degenera*, in cui  $\Sigma_X > 0$  si ha, in particolare, che  $\sigma_{ii} > 0$  e  $\rho_{ij} \in (-1, 1)$ .

## Decomposizione spettrale della matrice di covarianza

Come si vedrà in seguito, risulta spesso conveniente adottare una opportuna decomposizione, nota come *decomposizione spettrale* o anche *autovalori-autovettori*, della matrice di covarianza  $\Sigma_X = \Sigma_X^T \geq 0$ . A tale proposito si premette che una generica matrice simmetrica gode delle seguenti proprietà.

**Fatto 1** - Una matrice simmetrica  $\Sigma_X = \Sigma_X^T \in \mathbb{R}^{n \times n}$

- ha tutti gli autovalori reali, cioè  $sp(\Sigma_X) = \{\lambda_1, \dots, \lambda_n\} \subset \mathbb{R}$ ;
- ha  $n$  autovettori  $\{v_1, \dots, v_n\}$  linearmente indipendenti (quindi è diagonalizzabile) e mutuamente ortogonali, i.e.  $v_i^T v_j = 0$  per ogni  $i \neq j$ .

Se in particolare  $\Sigma_X \geq 0$  ( $\Sigma_X > 0$ ), i predetti autovalori sono tutti non negativi, i.e.  $\lambda_i \geq 0$  (positivi, i.e.  $\lambda_i > 0$ ).

Alla luce di questo risultato di algebra lineare, si definiscono la matrice  $V = [v_1, \dots, v_n]$  degli autovettori e la matrice diagonale  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$  degli autovalori. Senza perdita di generalità, si possono scegliere gli autovettori  $v_i$  di norma unitaria (autovettori) cioè tali che  $v_i^T v_j = \delta_{ij}$ , dove  $\delta_{ij}$  denota la delta di Kronecker uguale ad 1 quando  $i = j$  ed a 0 altrimenti. In virtù della indipendenza lineare degli autovettori, la matrice  $V$  è invertibile cioè esiste  $V^{-1}$  e per la ortonormalità degli autovettori tale matrice soddisfa  $VV^T = V^T V = I$  (cioè  $V$  è ortogonale), da cui risulta  $V^{-1} = V^T$ . Dalla definizione di autovalori ed autovettori, si ha:

$$\Sigma_X V = V \Lambda \implies \Sigma_X = V \Lambda V^{-1}$$

da cui si ottiene la *decomposizione spettrale* della matrice di covarianza  $\Sigma_X$

$$\Sigma_X = V \Lambda V^T = \sum_{i=1}^n \lambda_i v_i v_i^T \quad (1.2)$$

e (se  $\det \Sigma_X \neq 0$ ) per la sua inversa, detta *matrice di informazione*,

$$\Sigma_X^{-1} = V \Lambda^{-1} V^T = \sum_{i=1}^n \lambda_i^{-1} v_i v_i^T. \quad (1.3)$$

**Fatto 2** -  $\Sigma_X \in \mathbb{R}^{n \times n}$  è simmetrica se e solo se esistono una matrice ortogonale  $V \in \mathbb{R}^{n \times n}$  ed una matrice diagonale  $\Lambda \in \mathbb{R}^{n \times n}$  per cui vale la decomposizione spettrale (1.2). Le colonne  $v_i$  di  $V$  non sono altro che gli autovettori di norma unitaria di  $\Sigma_X$  mentre le corrispondenti componenti diagonali  $\lambda_i$  di  $\Lambda$  sono i rispettivi autovalori di  $\Sigma_X$ .

## Variabile Gaussiana standard

Come prima applicazione della decomposizione spettrale, si desidera determinare da  $X = \mathcal{N}(m_X, \Sigma_X)$ , mediante trasformazione lineare, una variabile aleatoria  $Z = AX + b$  tale che  $Z$  abbia media nulla, i.e.  $m_Z = 0$ , e covarianza unitaria, i.e.  $\Sigma_Z = I$ . In seguito, si vedrà come una trasformazione lineare non alteri la Gaussianità per cui  $Z = \mathcal{N}(0, I)$  così determinata è una variabile aleatoria Gaussiana standard (normalizzata), alla quale risulta utile ricondursi in quanto ad essa si riferiscono tutte le tabelle relative alla distribuzione Gaussiana. In primo luogo, si osserva che imponendo  $E[Z] = E[AX + b] = Am_X + b = 0$  si ricava  $b = -Am_X$  da cui  $Z = A(X - m_X) = A\tilde{X}$ . Successivamente, per imporre la matrice di covarianza unitaria, si ha

$$\Sigma_Z \triangleq E[ZZ^T] = E[A\tilde{X}\tilde{X}^T A^T] = AE[\tilde{X}\tilde{X}^T]A^T = A\Sigma_X A^T = I$$

da cui, sfruttando la decomposizione spettrale  $\Sigma_X = V\Lambda V^T = \Gamma\Gamma^T$  con  $\Gamma \triangleq V\Lambda^{1/2}$  e  $\Lambda^{1/2} \triangleq \text{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\}$ , si ottiene

$$I = A\Gamma\Gamma^T A^T \implies A\Gamma = I \implies A = \Gamma^{-1} = (V\Lambda^{1/2})^{-1} = \Lambda^{-1/2}V^T.$$

**Fatto 3** - Se  $X = \mathcal{N}(m_X, \Sigma_X)$ , allora  $Z = \Lambda^{-1/2}V^T(X - m_X)$  è una variabile aleatoria Gaussiana standard, cioè  $Z = \mathcal{N}(0, I)$ .

## Regioni di confidenza

Data la variabile aleatoria  $X \sim f_X(\cdot)$  definita su  $\mathbb{R}^n$  ed il livello di confidenza  $\alpha \in (0, 1)$ , si definisce *regione di confidenza di  $X$  al  $100\alpha\%$*  come un sottoinsieme  $\mathcal{R}_\alpha$  di  $\mathbb{R}^n$  di volume minimo tale che la probabilità di appartenenza di  $X$  ad  $\mathcal{R}_\alpha$  sia proprio  $\alpha$ . Matematicamente,  $\mathcal{R}_\alpha$  deve soddisfare le seguenti due condizioni:

1.  $Prob(X \in \mathcal{R}_\alpha) = \int_{\mathcal{R}_\alpha} f_X(x) dx = \alpha;$

2. Per ogni  $\mathcal{S} \subset \mathbb{R}^n$  tale che  $Prob(X \in \mathcal{S}) = \int_{\mathcal{S}} f_X(x) dx = \alpha$  deve risultare che  $vol(\mathcal{S}) \triangleq \int_{\mathcal{S}} dx \geq vol(\mathcal{R}_\alpha) \triangleq \int_{\mathcal{R}_\alpha} dx.$

Dato un vettore  $c \in \mathbb{R}^n$ , una matrice  $\Sigma \in \mathbb{R}^{n \times n}$  simmetrica e definita positiva, ed uno scalare  $\gamma > 0$ , si definisce l'iper-ellissoide, in  $\mathbb{R}^n$ ,  $\mathcal{E}_\gamma(c, \Sigma)$  nel seguente modo:

$$\mathcal{E}_\gamma(c, \Sigma) \triangleq \left\{ x \in \mathbb{R}^n : (x - c)^T \Sigma^{-1} (x - c) \leq \gamma \right\}. \quad (1.4)$$

Si noti che il vettore  $c$  è il centro di simmetria dell'iper-ellissoide, la matrice  $\Sigma$  caratterizza la forma dell'iper-ellissoide (direzioni degli assi e proporzioni fra le loro lunghezze) mentre  $\gamma$  è semplicemente un fattore di scala che espande/contracte isotropicamente l'iper-ellissoide senza alterarne la forma. È immediato constatare, sfruttando la decomposizione spettrale  $\Sigma = \sum_{i=1}^n \lambda_i v_i v_i^T$ , che gli assi principali dell'iper-ellissoide sono allineati con gli autovettori  $v_i$  della matrice  $\Sigma$  e che la lunghezza del semi-asse principale  $i$ -esimo (allineato con  $v_i$ ) coincide con  $\sqrt{\gamma \lambda_i}$ , dove  $\lambda_i$  è l'autovalore di  $\Sigma_X$  associato a  $v_i$ .

Per una variabile aleatoria Gaussiana, le regioni di confidenza sono iper-ellipsoidi univocamente determinati, come mostrato nel teorema che segue.

**Teorema 1** - Se  $X = \mathcal{N}(m_X, \Sigma_X)$ , allora per ogni  $\alpha \in (0, 1)$  la corrispondente regione di confidenza al  $100\alpha\%$  è data da

$$\mathcal{R}_\alpha = \mathcal{E}_{\gamma(\alpha)}(m_X, \Sigma_X) = \{x : (x - m_X)^T \Sigma_X^{-1} (x - m_X) \leq \gamma(\alpha)\} \quad (1.5)$$

con  $\gamma(\alpha) = \chi_{n,\alpha}^2$  dove il valore  $\chi_{n,\alpha}^2$  è ricavato dalla tabella dei valori percentili per la distribuzione chi-quadrato (con  $n = \dim X$  gradi di libertà e probabilità  $\alpha$ ).

*Dimostrazione* - Si consideri la famiglia di iper-ellipsoidi  $\mathcal{E}_\gamma \triangleq \mathcal{E}_\gamma(m_X, \Sigma_X)$  al variare del parametro di scala  $\gamma \geq 0$ . È evidente che, indicata con

$$\alpha(\gamma) = \text{Prob}(X \in \mathcal{E}_\gamma) = \int_{\mathcal{E}_\gamma} \mathcal{N}(x; m_X, \Sigma_X) dx = \text{Prob}\left((x - m_X)^T \Sigma_X^{-1} (x - m_X) \leq \gamma\right)$$

la probabilità che  $X$  appartenga a  $\mathcal{E}_\gamma$ ,  $\alpha(\gamma)$  è una funzione monotona crescente di  $\gamma$  con  $\alpha(0) = 0$  e  $\lim_{\gamma \rightarrow \infty} \alpha(\gamma) = 1$ . Pertanto, fissato  $\alpha \in (0, 1)$ , esisterà un unico valore  $\gamma(\alpha)$  per cui

$$\text{Prob}(X \in \mathcal{E}_{\gamma(\alpha)}) = \text{Prob}\left(\tilde{X}^T \Sigma_X^{-1} \tilde{X} \leq \gamma(\alpha)\right) = \alpha.$$

Introdotta la decomposizione spettrale  $\Sigma_X = \Gamma \Gamma^T = V \Lambda V^T$  e la v.a. Gaussiana standard (normalizzata)  $Z \triangleq \Gamma^{-1} \tilde{X} = \mathcal{N}(0, I)$  tale che  $\tilde{X}^T \Sigma_X^{-1} \tilde{X} = Z^T Z$ , si ha

$$\begin{aligned} \text{Prob}(X \in \mathcal{E}_{\gamma(\alpha)}) &= \text{Prob}\left(\tilde{X}^T \Sigma_X^{-1} \tilde{X} \leq \gamma(\alpha), X = \mathcal{N}(m_X, \Sigma_X)\right) \\ &= \text{Prob}\left(Z^T Z = \|Z\|^2 \leq \gamma(\alpha), Z = \mathcal{N}(0, I)\right) = \alpha. \end{aligned} \quad (1.6)$$

È ben noto che la variabile aleatoria  $\|Z\|^2 = \sum_{i=1}^n Z_i^2$ , dove  $Z = \mathcal{N}(0, I)$ , ha una distribuzione chi-quadrato a  $n$  gradi di libertà. Pertanto, il valore  $\gamma(\alpha)$  in (1.6) coincide con il valore  $\chi_{n,\alpha}^2$  riportato nella tabella dei valori percentili di tale distribuzione in corrispondenza della probabilità  $\alpha$ .

Resta da dimostrare che ogni altra regione  $S \subset \mathbb{R}^n$  per cui  $\text{Prob}(X \in S) = \alpha$  è tale che il suo volume non è inferiore a quello di  $\mathcal{E}_{\gamma(\alpha)}$ , cioè  $\text{vol}(S) \geq \text{vol}(\mathcal{E}_{\gamma(\alpha)})$ . A tale proposito,

si definiscono gli insiemi disgiunti

$$A \triangleq S \cap \mathcal{E}_{\gamma(\alpha)}, \quad B \triangleq \mathcal{E}_{\gamma(\alpha)} \setminus A, \quad C \triangleq S \setminus A$$

tali che  $\mathcal{E}_{\gamma(\alpha)} = A \cup B$  e  $S = A \cup C$ . Poiché

$$f_X(x) = \mathcal{N}(x; m_X, \Sigma_X) = c \exp \left[ -\frac{1}{2} (x - m_X)^T \Sigma_X^{-1} (x - m_X) \right]$$

si ha:

$$\begin{cases} f_X(x) \geq c \exp(-\gamma/2), & \forall x \in \mathcal{E}_{\gamma(\alpha)} \\ f_X(x) \leq c \exp(-\gamma/2), & \forall x \notin \mathcal{E}_{\gamma(\alpha)}. \end{cases}$$

Quindi,

$$\begin{cases} f_X(x) \geq c \exp(-\gamma/2), & \forall x \in B \\ f_X(x) \leq c \exp(-\gamma/2), & \forall x \in C. \end{cases} \quad (1.7)$$

Per ipotesi:

$$\begin{aligned} \alpha &= \int_{\mathcal{E}_{\gamma(\alpha)}} f_X(x) dx = \int_A f_X(x) dx + \int_B f_X(x) dx \\ &= \int_S f_X(x) dx = \int_A f_X(x) dx + \int_C f_X(x) dx \end{aligned}$$

da cui

$$\int_B f_X(x) dx = \int_C f_X(x) dx. \quad (1.8)$$

Da (1.7) si deduce che

$$\int_B f_X(x) dx \geq \int_B c e^{-\gamma/2} dx = c e^{-\gamma/2} \text{vol}(B) \quad (1.9)$$

$$\int_C f_X(x) dx < \int_C c e^{-\gamma/2} dx = c e^{-\gamma/2} \text{vol}(C). \quad (1.10)$$

Combinando (1.9)-(1.10), si verifica immediatamente che  $\text{vol}(B) < \text{vol}(C)$ , da cui

$$\text{vol}(\mathcal{E}_{\gamma(\alpha)}) = \text{vol}(A) + \text{vol}(B) < \text{vol}(A) + \text{vol}(C) = \text{vol}(S)$$

come volevasi dimostrare. Quindi  $\mathcal{R}_\alpha = \mathcal{E}_{\gamma(\alpha)}$ .  $\square$

Dal precedente teorema si evince come le regioni di confidenza della distribuzione Gaussiana siano iper-ellissoidi centrati nel valor medio della Gaussiana con matrice di forma proporzionale alla covarianza della Gaussiana e fattore di scala dipendente dal livello di confidenza stabilito.

## Variabile aleatoria Gaussiana degenera

Si assuma la matrice di covarianza  $\Sigma_X$  singolare, per la precisione  $\text{rank} \Sigma_X = r < n$ . Allora  $\Sigma_X$  ha  $n - r$  autovalori nulli e ammette la decomposizione spettrale

$$\Sigma_X = \underbrace{[V_1 \ V_2]}_V \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \quad (1.11)$$

con  $\Lambda_1 \in \mathbb{R}^{r \times r}$  diagonale e definita positiva. Da (1.11) si deducono le relazioni

$$V_1^T \Sigma_X V_1 = \Lambda_1, \quad V_2^T \Sigma_X = 0.$$

Si consideri la variabile aleatoria trasformata

$$Z \triangleq V^T (X - m_X) = V^T \tilde{X} = \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tilde{X} = \begin{bmatrix} V_1^T \tilde{X} \\ V_2^T \tilde{X} \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}.$$

È immediato constatare che

$$Z_1 = V_1^T \tilde{X} = \mathcal{N}(0, \Lambda_1) \quad (1.12)$$

$$Z_2 = V_2^T \tilde{X} = \mathcal{N}(0, 0). \quad (1.13)$$

In particolare, da (1.13) risulta che il vettore aleatorio  $X$  soddisfa la seguente relazione lineare

$$V_2^T \tilde{X} = V_2^T (X - m_X) = 0 \quad (1.14)$$

che equivale a  $n - r$  vincoli di uguaglianza lineari deterministici sulle componenti di  $X$ . In termini geometrici, (1.14) sancisce che il vettore  $\tilde{X} = X - m_X$  è ortogonale agli autovettori relativi agli  $n - r$  autovalori nulli della matrice  $\Sigma_X$ , ovvero è ortogonale al sottospazio nullo della matrice di covarianza  $\Sigma_X$ . Viceversa,  $Z_1$  è distribuito normalmente con media nulla e matrice di covarianza diagonale  $\Lambda_1$ . Ricapitolando,  $X = \mathcal{N}(m_X, \Sigma_X)$ , con  $\Sigma_X$  singolare, viene detta variabile aleatoria Gaussiana degenera ed è normalmente distribuita nella varietà lineare  $\{X : V_2^T (X - m_X) = 0\}$  di  $\mathbb{R}^n$ .

*Esempio* - A titolo di esempio si consideri la variabile aleatoria Gaussiana degenera

$$X = \mathcal{N}\left(\begin{bmatrix} 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\right).$$

Per tale variabile si desidera valutare la regione di confidenza al 90% nonché la probabilità che assuma valori nel rettangolo  $\mathcal{R} = \{x = [x_1, x_2]^T : -1 \leq x_1 \leq 2, -2 \leq x_2 \leq 1\}$ . In

primo luogo si osserva che  $\text{rank } \Sigma_X = r = 1 < n = 2$  e che  $\Sigma_X$  ammette la decomposizione spettrale

$$\Sigma_X = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}.$$

Pertanto  $X = [X_1, X_2]^T$  è normalmente distribuita sulla retta, nel piano  $(x_1, x_2)$ , di equazione

$$v_2^T (x - m_X) = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} x_1 - 4 \\ x_2 - 2 \end{bmatrix} = \frac{1}{\sqrt{5}}(x_1 - 4) - \frac{2}{\sqrt{5}}(x_2 - 2) = 0 \implies x_1 = 2x_2.$$

La regione di confidenza al 90% coinciderà dunque con il segmento di tale retta compreso fra le rette di equazioni

$$v_1^T (x - m_X) = \pm \sqrt{\lambda_1 \gamma} \quad (1.15)$$

con  $\gamma \cong 2.71$  (corrispondente ad un grado di libertà e livello di confidenza del 90%). Pertanto, combinando (1.15) con l'equazione della retta (allineata con l'autovettore associato all'autovalore non nullo)  $x_1 = 2x_2$ , si ha

$$\begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2x_2 - 4 \\ x_2 - 2 \end{bmatrix} = \pm \sqrt{5\gamma} \implies x_2 = 2 \pm \sqrt{\gamma} = 2 \pm 1.645$$

da cui risulta che l'intervallo di confidenza al 90% coincide con il segmento della retta  $x_1 = 2x_2$  compreso fra le ordinate 0.355 e 3.645, o equivalentemente fra le ascisse 0.710 e 7.290, ovvero

$$x_1 = 2x_2 \in [0.710, 7.290], \quad x_2 \in [0.355, 3.645].$$

Per calcolare la probabilità che  $X$  assuma valori nel rettangolo  $\mathcal{R} = [-1, 2] \times [-2, 1]$ , si determinano le intersezioni di tale rettangolo con la retta  $x_1 = 2x_2$  dove risulta concentrata la distribuzione di probabilità di  $X$ . Con semplici calcoli, tali intersezioni risultano nei punti  $x = [-1, -1/2]^T$  e  $x = [2, 1]^T$ . Definendo la variabile aleatoria Gaussiana standard

$$Z = \frac{v_1^T \tilde{X}}{\sqrt{\lambda_1}} = \begin{bmatrix} \frac{2}{5} & \frac{1}{5} \end{bmatrix} \begin{bmatrix} x_1 - 4 \\ x_2 - 2 \end{bmatrix} = \mathcal{N}(0, 1)$$

si verifica che ai suddetti punti  $x = [-1, -1/2]^T$  e  $x = [2, 1]^T$  corrispondono i valori della Gaussiana normalizzata  $Z = -5/2$  e, rispettivamente,  $Z = -1$ . Pertanto,

$$\begin{aligned} \text{Prob}(X \in \mathcal{R}) &= \text{Prob}(-2.5 < Z < 0) - \text{Prob}(-1 < Z < 0) \\ &= \text{Prob}(0 < Z < 2.5) - \text{Prob}(0 < Z < 1) \\ &= 0.4938 - 0.3413 \\ &= 0.1525. \end{aligned}$$

## Proprietà di invarianza della distribuzione Gaussiana

In questo paragrafo si fa riferimento a variabili aleatorie Gaussiane non degeneri.

### Marginalizzazione

Il seguente teorema afferma che due variabili aleatorie congiuntamente Gaussiane sono anche marginalmente Gaussiane, ovvero che la Gaussianità viene preservata dall'operazione di *marginalizzazione*.

**Teorema 2** (Invarianza della distribuzione Gaussiana rispetto alla marginalizzazione)

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} \right) \implies \begin{cases} X = \mathcal{N}(m_X, \Sigma_X) \\ Y = \mathcal{N}(m_Y, \Sigma_Y) . \end{cases}$$

*Dimostrazione* - Posto

$$\begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} = \begin{bmatrix} \Omega_X & \Omega_{XY} \\ \Omega_{YX} & \Omega_Y \end{bmatrix}^{-1} = \begin{bmatrix} (\Omega_X - \Omega_{XY}\Omega_Y^{-1}\Omega_{YX})^{-1} & \times \\ \times & (\Omega_Y - \Omega_{YX}\Omega_X^{-1}\Omega_{XY})^{-1} \end{bmatrix}, \quad (1.16)$$

dove i blocchi  $\times$  sono volutamente non specificati, si ha

$$\begin{aligned} f_{X,Y}(x, y) &\propto \exp \left\{ -\frac{1}{2} \left[ (x - m_X)^T \Omega_X (x - m_X) + (y - m_Y)^T \Omega_Y (y - m_Y) \right. \right. \\ &\quad \left. \left. + 2(x - m_X)^T \Omega_{XY} (y - m_Y) \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[ y - m_Y + \Omega_Y^{-1} \Omega_{YX} (x - m_X) \right]^T \Omega_Y \left[ y - m_Y + \Omega_Y^{-1} \Omega_{YX} (x - m_X) \right] \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (x - m_X)^T (\Omega_X - \Omega_{XY} \Omega_Y^{-1} \Omega_{YX}) (x - m_X) \right\} . \end{aligned}$$

Quindi, grazie a (1.16),

$$f_X(x) = \int f_{X,Y}(x, y) dy \propto \exp \left[ -\frac{1}{2} (x - m_X)^T \Sigma_X^{-1} (x - m_X) \right].$$

Poiché  $\int f_X(x) dx = 1$ , deve risultare che

$$f_X(x) = \mathcal{N}(x; m_X, \Sigma_X) \implies X = \mathcal{N}(m_X, \Sigma_X) .$$

In modo analogo, scambiando  $X$  e  $Y$  nei precedenti sviluppi, si procede per dimostrare che  $Y = \mathcal{N}(m_Y, \Sigma_Y)$ .  $\square$

## Indipendenza e incorrelazione

I seguenti due risultati evidenziano alcune relazioni esistenti fra Gaussianità (marginale e congiunta), indipendenza ed incorrelazione.

**Teorema 3** - Si considerino due v.a. Gaussiane ed indipendenti

$$X = \mathcal{N}(m_X, \Sigma_X) \quad \text{e} \quad Y = \mathcal{N}(m_Y, \Sigma_Y)$$

allora  $X$  e  $Y$  sono congiuntamente Gaussiane, ovvero

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix} \right).$$

*Dimostrazione* - In virtù della ipotizzata indipendenza di  $X$  e  $Y$  risulta che  $f_{X,Y}(x, y) = f_X(x)f_Y(y)$  da cui, per la ipotizzata Gaussianità marginale di  $X$  e  $Y$ , si ha

$$\begin{aligned} f_{X,Y}(x, y) &\propto \exp \left[ -\frac{1}{2}(x - m_X)^T \Sigma_X^{-1} (x - m_X) \right] \exp \left[ -\frac{1}{2}(y - m_Y)^T \Sigma_Y^{-1} (y - m_Y) \right] \\ &= \exp \left\{ -\frac{1}{2} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^T \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix}^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} \right\}. \end{aligned}$$

Quindi, poiché  $\int \int f_{X,Y}(x, y) dx dy = 1$ ,

$$f_{X,Y}(x, y) = \mathcal{N} \left( \begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix} \right)$$

come volevasi dimostrare. □

**Teorema 4** - Due v.a. congiuntamente Gaussiane ed incorrelate sono indipendenti.

*Dimostrazione* - Se le variabili aleatorie  $X$  e  $Y$  sono congiuntamente Gaussiane ed incorrelate, si ha

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix} \right)$$

da cui

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{\sqrt{\det 2\pi \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix}}} \exp \left\{ -\frac{1}{2} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^T \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix}^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} \right\} \\ &= \frac{1}{\sqrt{\det 2\pi \Sigma_X}} \exp \left[ -(x - m_X)^T \Sigma_X^{-1} (x - m_X) \right] \cdot \\ &\quad \cdot \frac{1}{\sqrt{\det 2\pi \Sigma_Y}} \exp \left[ -(y - m_Y)^T \Sigma_Y^{-1} (y - m_Y) \right] \\ &= f_X(x) f_Y(y) \end{aligned}$$

come volevasi dimostrare. □

## Condizionamento

Il seguente risultato afferma che due variabili aleatorie congiuntamente Gaussiane sono anche condizionatamente Gaussiane l'una all'altra, ovvero che la Gaussianità viene preservata dall'operazione di condizionamento.

**Teorema 5** (Invarianza della distribuzione Gaussianiana rispetto al condizionamento)

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} \right) \implies X|Y = \mathcal{N}(m_{X|Y}, \Sigma_{X|Y})$$

con

$$m_{X|Y} = m_X + \Sigma_{XY} \Sigma_Y^{-1} (Y - m_Y) \quad (1.17)$$

$$\Sigma_{X|Y} = \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX}. \quad (1.18)$$

*Dimostrazione* - Si noti che, usando la definizione (1.16),

$$\begin{aligned} f_{X,Y}(x|y) &\propto f_{X,Y}(x,y) \propto \exp \left\{ -\frac{1}{2} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix}^T \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [x^T \Omega_X x + 2x^T (\Omega_{XY}(y - m_Y) - \Omega_X m_X)] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [x - (m_X - \Omega_X^{-1} \Omega_{XY}(y - m_Y))]^T \Omega_X \right. \\ &\quad \left. [x - (m_X - \Omega_X^{-1} \Omega_{XY}(y - m_Y))] \right\} \end{aligned}$$

da cui

$$f_{X|Y}(x|y) = \mathcal{N}(x; m_X - \Omega_X^{-1} \Omega_{XY}(y - m_Y), \Omega_X^{-1} = \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX})$$

Poiché

$$\Omega_X \Sigma_{XY} + \Omega_{XY} \Sigma_Y = 0 \implies \Sigma_{XY} \Sigma_Y^{-1} + \Omega_X^{-1} \Omega_{XY} = 0,$$

si ha che  $-\Omega_X^{-1} \Omega_{XY} = \Sigma_{XY} \Sigma_Y^{-1}$ , da cui

$$f_{X|Y}(\cdot|y) = \mathcal{N}(\cdot; m_X + \Sigma_{XY} \Sigma_Y^{-1} (y - m_Y), \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX})$$

come volevasi dimostrare. □

## Trasformazione lineare

La distribuzione Gaussiana si preserva anche sotto trasformazioni lineari. Vale infatti il seguente risultato.

**Teorema 6** (Invarianza della distribuzione Gaussiana rispetto a trasformazioni lineari)  
 Si consideri una v.a. Gaussiana  $X = \mathcal{N}(m_X, \Sigma_X)$  e la trasformazione *lineare*  $Y = AX + b$ , con  $A \in \mathbb{R}^{p \times n}$  matrice a rango pieno cioè tale che  $\text{rank } A = \min(p, n)$ . Allora la v.a.  $Y$  risulta anch'essa Gaussiana, più precisamente  $Y = \mathcal{N}(m_Y, \Sigma_Y)$  con

$$m_Y = Am_X + b, \quad \Sigma_Y = A\Sigma_X A^T.$$

*Dimostrazione* - Si distinguono tre casi:

- (a)  $A$  è quadrata, i.e.  $p = n$ ;
- (b)  $A$  è rettangolare a rango pieno secondo le righe, i.e.  $p < n$ ;
- (c)  $A$  è rettangolare a rango pieno secondo le colonne, i.e.  $p > n$ .

Caso (a) - In questo caso,  $A$  è invertibile in virtù dell'ipotesi di rango pieno, per cui si può scrivere  $X = A^{-1}Y - b$  e, quindi,  $m_X = A^{-1}m_Y - b$ . Pertanto, usando il Teorema relativo alla PDF di funzione di variabile aleatoria, si ha

$$\begin{aligned} f_Y(y) &= f_X(A^{-1}y - b) |det A^{-1}| \\ &\propto \exp \left[ -\frac{1}{2} (A^{-1}y - b - m_X)^T \Sigma_X^{-1} (A^{-1}y - b - m_X) \right] \\ &= \exp \left\{ -\frac{1}{2} [A^{-1}(y - m_Y)]^T \Sigma_X^{-1} [A^{-1}(y - m_Y)] \right\} \\ &= \exp \left[ -\frac{1}{2} (y - m_Y)^T (A\Sigma_X A^T)^{-1} (y - m_Y) \right]. \end{aligned}$$

Quindi, poiché  $\int f_Y(y) dy = 1$ , risulta che

$$f_Y(y) = \mathcal{N}(y; m_Y = Am_X + b, \Sigma_Y = A\Sigma_X A^T) \implies Y = \mathcal{N}(m_Y, \Sigma_Y)$$

come volevasi dimostrare.

Caso (b) - Poiché in questo caso  $\text{rank } A = p$ , si ha la seguente decomposizione a valori singolari

$$A = U [\Lambda_1 \ 0] V$$

con  $U \in \mathbb{R}^{p \times p}, V \in \mathbb{R}^{n \times n}$  ortogonali (e quindi invertibili) e  $\Lambda_1 \in \mathbb{R}^{p \times p}$  diagonale ed invertibile. Quindi

$$Y = AX + b = U [\Lambda_1 \ 0] V X + b = U [\Lambda_1 \ 0] Z + b$$

con

$$Z \triangleq VX = \mathcal{N}(Vm_X, V\Sigma_X V^T) \quad (1.19)$$

grazie al risultato precedentemente mostrato nel caso (a). Partizionando  $Z \in \mathbb{R}^n$  e  $V \in \mathbb{R}^{n \times n}$  nel seguente modo

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

con  $Z_1 \in \mathbb{R}^p$  e  $V_1 \in \mathbb{R}^{p \times n}$ , da (1.19) ed in virtù del Teorema 2, risulta quanto segue

$$Z_1 = V_1 X = \mathcal{N}(V_1 m_X, V_1 \Sigma_X V_1^T).$$

Ricordando che  $A_1 \triangleq U\Lambda_1$  è invertibile ed applicando nuovamente il risultato dimostrato nel caso (a), si ha che

$$Y = AX + b = U\Lambda_1 V_1 X + b = A_1 Z_1 + b = \mathcal{N}\left(\underbrace{U\Lambda_1 V_1 m_X}_A + b, \underbrace{U\Lambda_1 V_1 \Sigma_X (U\Lambda_1 V_1)^T}_{A^T}\right)$$

come volevasi dimostrare.

Caso (c) - Si noti che in questo caso  $Y$  è una variabile aleatoria degenera in quanto  $\Sigma_Y = A\Sigma_X A^T$  ha rango  $n = \dim X < p = \dim Y$ . La decomposizione a valori singolari di  $A$  assume, in questo caso, la forma

$$A = \underbrace{[U_1 \ U_2]}_U \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V$$

con  $\Lambda_1 \in \mathbb{R}^{n \times n}$  diagonale ed invertibile,  $U \in \mathbb{R}^{p \times p}$  e  $V \in \mathbb{R}^{n \times n}$  entrambe ortogonali (e quindi invertibili),  $U_1 \in \mathbb{R}^{p \times n}$ . Definito

$$Z \triangleq \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V X = \begin{bmatrix} \Lambda_1 V X \\ 0 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix},$$

risulta che  $Z_1 = \mathcal{N}(\Lambda_1 V m_X, \Lambda_1 V \Sigma_X V^T \Lambda_1)$  e  $Z_2 = \mathcal{N}(0, 0)$ . Inoltre, poiché  $Z_2$  è deterministicamente (con probabilità 1) nulla,  $Z_1$  e  $Z_2$  risultano indipendenti e quindi, grazie al Teorema 3, congiuntamente Gaussiane, ovvero

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V m_X, \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V \Sigma_X V^T \begin{bmatrix} \Lambda_1 & 0 \end{bmatrix}\right).$$

Notando che  $Y = UZ + b$ , con  $U$  invertibile, e ricorrendo al risultato mostrato nel caso (a), si ha

$$Y = \mathcal{N} \left( \underbrace{U \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V}_{A} m_X + b, \underbrace{U \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V}_{A} \Sigma_X \underbrace{\left( U \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} V \right)^T}_{A^T} \right) = \mathcal{N} (Am_X + b, A\Sigma_X A^T)$$

come volevasi dimostrare. □

## Teorema del limite centrale

La popolarità della distribuzione Gaussiana non è solamente dovuta alle sue proprietà di invarianza illustrate nel precedente paragrafo, ma è anche conseguenza del ben noto *teorema del limite centrale* di seguito enunciato.

**Teorema del limite centrale** - Si consideri una successione  $\{X_n\}$  di variabili aleatorie indipendenti e identicamente distribuite con  $E[X_n] = m$  e  $var X_n = \Sigma$ . Definita  $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ , si ha che

$$\text{per } n \rightarrow \infty : S_n \rightarrow \mathcal{N} \left( m, \frac{\Sigma}{n} \right)$$

In altri termini, una variabile aleatoria risultante dalla somma di variabili aleatorie indipendenti con la stessa distribuzione tende ad avere, quando il numero dei termini della somma cresce, una distribuzione Gaussiana. Grazie a questo risultato, si ricorre frequentemente alla distribuzione Gaussiana nelle applicazioni pratiche per modellare in modo semplice fenomeni complessi derivanti dalla sovrapposizione di un numero elevato di cause indipendenti.