Appunti delle lezioni di Identificazione dei Sistemi Incerti (R. Costanzi)

Filippo F.

Capitolo 1: RICHIAMI SULLE VARIABILI ALEATORIE

Chiamiamo $X = [X_1, X_2, \dots, X_n]^T$ una variabile aleatoria con *funzione di densità di probabilità (PDF)* $X \sim f_X : \mathbb{R}^n \to \mathbb{R}_+$ tale che per ogni sottoinsieme $S \in \mathbb{R}^n$ la probabilità è:

$$\mathbb{P}(X \in S) = \int_S f_X(x)dx$$

Deve essere valida la <u>*condizione di normalizzazione*</u>: $\mathbb{P}(X \in \mathbb{R}^n) = \int f_X(x)dx = 1$
<u>N.B.</u> Quando si integra in tutto $\mathbb{R}^n$ nell'integrale non sono presenti gli estremi di integrazione.

In alternativa alla funzione di densità di probabilità si può utilizzare la *funzione di densità cumulativa (CDF)* $F_X : \mathbb{R}^n \to \mathbb{R}_+$.

$$F_X(x) = \mathbb{P}(X < x) \triangleq \mathbb{P}(X_1 < x_1, \dots, X_n < x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(x)dx_1 \dots dx_n$$

La relazione tra la PDF e la CDF è $F_X(x) = \int f_X(x)dx$ e $\frac{dF_X(x)}{dx} = f_X(x)$.

- ***Operatore di attesa*** $E[g(X)]$ per $g(X)$ funzione di X v.a. con $X \sim f_X(x)$

$$E[g(X)] = \int g(X) f_X(x)dx$$

- ***Momenti di una variabile aleatoria*** $X \sim f_X(x)$:

1) Momento di ordine 1: ***media*** $m_X = E[X] = \int X f_X(x)dx$ con $m_X \in \mathbb{R}^n$
Gli elementi di m_X sono le medie di tutte le componenti di $X \to m_i = E[X_i]$.
2) Momento di ordine 2: ***matrice di covarianza*** Σ_X

$$\Sigma_X = E[(X - m_X)(X - m_X)^T], \Sigma_X \in \mathbb{R}^{nxn}$$

Gli elementi di Σ_X sono:
- σ_{ii} varianza di $X_i \to \sigma_{ii} = E[(X_i - m_i)^2] \; \forall \, i = 1, \dots, n$
- σ_{ij} covarianze di X_i e $X_j \to \sigma_{ij} = E[(X_i - m_i)(X_j - m_j)]$

<u>N.B.</u> σ_{ij} è simmetrica $\to \sigma_{ij} = \sigma_{ji}$ e di conseguenza anche la matrice di covarianza è simmetrica

$$\Sigma_X = {\Sigma_X}^T$$

Inoltre, la matrice di covarianza è *simmetrica e semi-definita positiva*:

$$v^T \Sigma_X v \geq 0 \; \forall v \in \mathbb{R}^n$$

ossia ha tutti autovalori $\lambda_i \in \mathbb{R}_+ \; \forall \, i = 1, \dots, n$.

- ***Coefficiente di correlazione:***

$$\rho_{ij} = \frac{cov(X_i, X_j)}{\sqrt{var(X_i)var(X_j)}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad con \; \rho_{ij} \in [-1,1]$$

Con $\sigma_i = \sqrt{\sigma_{ii}}$ deviazione standard.

- ***Momenti incrociati***: Considero una coppia (X, Y) di v.a.

Definisco la <u>*matrice di cross-covarianza*</u> come: $\Sigma_{X,Y} = E[(X - m_X)(Y - m_Y)^T]$
$\Sigma_{X,Y}$ è *simmetrica* $\to \Sigma_{X,Y} = {\Sigma_{X,Y}}^T$ e se (X, Y) sono "*non correlate*" allora $\Sigma_{X,Y} = 0_{nxn}$.

- ***PDF congiunte (Probability Density Function):***

Sia X una variabile aleatoria di n dimensioni e Y un'altra di p dimensioni, la PDF congiunta è $f_{X,Y}(x, y)$. Definiamo $Z = [X^T, Y^T]^T$ e il sottoinsieme $S \in \mathbb{R}^{n+p}$, allora la probabilità che $\mathbb{P}(Z \in S) = \int_S f_{X,Y}(x, y)\, dxdy$.

- ***Ricavare le pdf marginali dalla pdf congiunta:***

$$f_X(x) = \int f_{X,Y}(x,y)dy \quad e \quad f_Y(y) = \int f_{X,Y}(x,y)dx$$

- ***Distribuzione di probabilità condizionata:***

La distribuzione di probabilità della v.a. X (PDF condizionata al valore $Y = y$) sapendo che la v.a. Y assume il valore y è $f_{X|Y}(x|y)$. Analogamente alla distribuzione di probabilità deve rispettare la condizione di normalizzazione $\int f_{X|Y}(x|y)dx = 1$.
$f_{X|Y}(x|y)$ è proporzionale alla PDF congiunta:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$
$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Confrontando le due espressioni si ottiene la relazione:

$$f_{Y|X}(y|x)f_X(x) = f_{Y,X}(y,x)f_Y(y)$$

da cui deriva la formula di Bayes.

<u>Formula di Bayes:</u>

$$\boxed{f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x)dx}}$$

$\int f_{Y|X}(y|x)f_X(x)dx$ termine di normalizzazione indipendente da x.

Possiamo pensare di definire le PDF come:

- $f_{X|Y}(x|y)$ PDF *condizionata* o "a posteriori" del condizionamento
- $f_X(x)$ PDF *non condizionata* o "a priori" del condizionamento

dove consideriamo la $Y = y$ come il risultato dell'esperimento.

- $f_{Y|X}(y|x)$ *verosimiglianza di Y dato X*

$$f_{X|Y}(x|y) = f_X(x)f_{Y|X}(y|x)$$

Fissato y, per un certo valore x di X quantifica quanto questo valore risulta verosimile con l'evento $Y = y$.

<u>N.B.</u>

X e Y sono indipendenti se $f_{X,Y}(x,y) = f_X(x)f_Y(y) \rightarrow f_{X|Y}(x|y) = f_X(x)$
L'indipendenza di X e Y implica che $\Sigma_{X,Y} = 0_{nxn}$ ma <u>non il contrario</u>.
Se X e Y sono v.a. indipendenti $\rightarrow g(X)$ e $h(Y)$ sono v.a. indipendenti.

- ***PDF di funzioni di variabili aleatorie:***

Sia $X \sim f_X(x)$ e $Y = g(X)$ con $g(\cdot)$ invertibile $h(\cdot) = g^{-1}(\cdot)$, allora la PDF di Y ha:

$$f_Y(y) = f_X(h(y))\left|det\left(\frac{\partial h(y)}{\partial y}\right)\right| = \frac{f_X(h(y))}{\left|det\left(\frac{\partial g(h(y))}{\partial x}\right)\right|}$$

Esempi:

- $Y = X_1 + X_2$ con $X_1 \sim f_1(\cdot)$ e $X_2 \sim f_2(\cdot)$ v.a. indipendenti

$$f_Y(y) = (f_1(\cdot) * f_2(\cdot))(y) = \int f_1(y-x)f_2(x)dx = \int f_1(x)f_2(y-x)dx$$

con $*$ prodotto di convoluzione.

- $Y = h(X) + V$ con $X \sim f_X(\cdot)$ e $V \sim f_V(\cdot)$ v.a. indipendenti

$f_{Y|X}(y|x) = f_V\big(y - h(x)\big)$; $f_{X,Y}(x,y) = f_V\big(y - h(x)\big)f_X(x)$; $f_Y(y) = \int f_V\big(y - h(x)\big)f_X(x)dx$

Dalla formula di Bayes otteniamo:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_V(y - h(x))f_X(x)}{\int f_V\big(y - h(x)\big)f_X(x)dx}$$

master COPY

master COPY

Capitolo 2: VARIABILI ALEATORIE GAUSSIANE

Sia $X = [X_1, X_2, \dots, X_n]^T$ una variabile aleatoria definita su $\mathbb{R}^n$ "*gaussiana*" di media m_X e matrice di covarianza $\Sigma_X \in \mathbb{R}^{nxn}$. Le v.a. gaussiane hanno come PDF:

$$f_X(x) = \mathcal{N}(x, m_X, \Sigma_X) = \frac{1}{\sqrt{det(2\pi\Sigma_X)}} e^{-\frac{1}{2}(x-m_X)^T {\Sigma_X}^{-1}(x-m_X)}$$

Nella successiva trattazione useremo la notazione $X \sim \mathcal{N}(m_X, \Sigma_X)$ per indicare una variabile aleatoria vettoriale che ha vettore media m_X e matrice di covarianza Σ_X.

- ***Decomposizione spettrale della matrice di covarianza:***

$\Sigma_X = {\Sigma_X}^T \in \mathbb{R}^{nxn}$ simmetrica $\lambda_i \in \mathbb{R}$ con n autovalori linearmente indipendenti e mutualmente ortogonali $v_i^T v_j = 0 \; \forall \, i, j$.

Sia V matrice degli autovettori $V = [v_1, v_2 \dots, v_n]$ con v_i di norma unitaria.

$$v_i^T v_j = \delta_{ij} = \begin{cases} 1 \; con \; i = j \\ 0 \; con \; i \neq j \end{cases} \rightarrow V^{-1} = V^T \text{ con } \delta_{ij} \text{ delta di Kronecker}$$

Chiamiamo $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$ la matrice diagonale degli autovalori tale che $\Sigma_X V = V\Lambda$

$\rightarrow \Sigma_X = V\Lambda V^{-1}$ che grazie alla decomposizione spettrale della matrice di covarianza possiamo scrivere $\Sigma_X = V\Lambda V^{-1} = V\Lambda V^T = \sum_{i=1}^{n} \lambda_i v_i {v_i}^T$.

Definiamo la *matrice di informazione* ${\Sigma_X}^{-1}$ come ${\Sigma_X}^{-1} = V\Lambda^{-1}V^T = \sum_{i=1}^{n} {\lambda_i}^{-1} v_i {v_i}^T$.

- ***Variabile Gaussiana Standard***

Data $X = \mathcal{N}(m_X, \Sigma_X)$ trovare la v.a. $Z = AX + b$ tale che $Z \sim \mathcal{N}(0, I)$. Sfruttando la decomposizione spettrale si ricava A e b che soddisfa la richiesta:

$$b = -Am_X \text{ e } A = \Lambda^{-\frac{1}{2}} V^T$$

Imponendo $E[Z] = E[AX + b] = Am_X + b = 0 \rightarrow b = -Am_X$ da cui $Z = AX - Am_X = A(X - m_X) = A\tilde{X}$

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

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

$$I = A\Gamma\Gamma^T \mathrm{A}^T \rightarrow A\Gamma = \mathrm{I} \rightarrow \mathrm{A} = \Gamma^{-1} = \left(\mathrm{V}\Lambda^{1/2}\right)^{-1} = \Lambda^{-\frac{1}{2}} V^T$$

- ***Regioni di confidenza:***

Data la v.a. $X \sim f_X(\cdot)$ definita su $\mathbb{R}^n$ e il livello di confidenza $\alpha \in [0,1]$, si definisce la regione di confidenza di X al $100\alpha\%$ come un sottoinsieme $R_\alpha \subseteq \mathbb{R}^n$ di volume minimo tale che la probabilità di appartenenza di X a R_α sia proprio α.

$$\mathbb{P}(X \in R_\alpha) = \alpha \; \forall \, \alpha \in [0,1]$$

Per le "*PDF gaussiane*" la regione di confidenza R_α è univocamente determinata ed è rappresentata da un "*iper-ellissoide*" con assi allineati con gli autovettori e di lunghezza proporzionali ad $\alpha\sqrt{\lambda_i}$.

$X \sim \mathcal{N}(m_X, \Sigma_X) \; \forall \, \alpha \in [0,1]$ la regione di confidenza è rappresentata da questa relazione:

$$R_\alpha = \left\{x \in \mathbb{R}^n : \; (x - m_X)^T {\Sigma_X}^{-1}(x - m_X) \leq \gamma(\alpha)\right\} \text{ con } \gamma(\alpha) = {\chi^2}_{n,\alpha} \text{ e } n = dim(X).$$

$\gamma(\alpha)$ funzione scalare dipendente dalle dimensioni di X e dal valore di α.