Capitolo 1
Richiami sulle variabili aleatorie

per gli studenti del corso di
Stima e identificazione

Luigi Chisci, 28 Febbraio 2019

PDF e CDF

Una variabile aleatoria vettoriale $X = [X_1, X_2, \ldots, X_n]^T$, definita su un opportuno *spazio di probabilità*, è completamente caratterizzata dalla sua *densità di probabilità* (PDF = *Probability Density Function*) $f_X : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ tale che, per ogni sottoinsieme $\mathbb{S}$ di $\mathbb{R}^n$, la probabilità che X assuma valori in $\mathbb{S}$ è data da

$$Prob\,(X \in \mathbb{S}) \;=\; \int_{\mathbb{S}} f_X(x)\,dx\,. \tag{1.1}$$

Da (1.1), posto $\mathbb{S} = \mathbb{R}^n$, è immediato constatare che la PDF $f_X(\cdot)$ deve soddisfare la condizione di normalizzazione

$$Prob\,(X \in \mathbb{R}^n) \;=\; \int_{\mathbb{R}^n} f_X(x)\,dx \;\stackrel{\triangle}{=}\; \int f_X(x)\,dx \;=\; 1\,. \tag{1.2}$$

Nella precedente formula, come nelle successive, l'assenza dell'insieme di integrazione viene implicitamente interpretata come integrazione su tutto $\mathbb{R}^n$. Nel seguito si userà la notazione $X \sim f_X(\cdot)$ a indicare che X è una variabile aleatoria con PDF $f_X(\cdot)$.

In alternativa alla PDF, la variabile aleatoria X può essere equivalentemente caratterizzata dalla sua *distribuzione cumulativa* (CDF = *Cumulative Distribution Function*) $F_X : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ definita nel seguente modo

$$F_X(x) = Prob\,(X \leq x) \stackrel{\triangle}{=} Prob\,(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(x)\,dx_1 \cdots dx_n \tag{1.3}$$

Si noti che la CDF $F_X(x)$, in corrispondenza del valore $x = [x_1, \ldots, x_n]^T$ non è altro che l'integrale della PDF $f_X(\cdot)$ sull'insieme $\mathbb{S} = (-\infty, x_1] \times \cdots \times (-\infty, x_n] \subset \mathbb{R}^n$. Viceversa, data la CDF $F_X(\cdot)$, è possibile ricavare da essa la PDF $f_X(\cdot)$ mediante derivazione, ovvero

$$f_X(x) \;=\; \frac{\partial^n F_X(x)}{\partial x_1 \cdots \partial x_n}\,. \tag{1.4}$$

Operatore di attesa

L'*operatore di attesa* (in inglese: *expectation*) $E[\cdot]$, applicabile ad una generica funzione $g(X)$ della variabile aleatoria X, è definito come segue:

$$E[g(X)] \triangleq \int g(x)\, f_X(x)\, dx\,. \tag{1.5}$$

Si noti che il valore atteso della funzione $g(X)$ è ottenuto integrando, su tutto $\mathbb{R}^n$, il prodotto della funzione $g(\cdot)$ per la PDF $f_X(\cdot)$ della variabile aleatoria X. È immediato constatare che l'operatore $E[\cdot]$ risulta lineare data la linearità dell'integrale.

Momenti di una variabile aleatoria

L'operatore di attesa definito in (1.5) consente di definire i *momenti* della variabile aleatoria X. In particolare, si definisce il *momento del primo ordine* o *media* $m_X \in \mathbb{R}^n$ della variabile aleatoria X come

$$m_X \triangleq E[X] = \int x\, f_X(x)\, dx \tag{1.6}$$

ed il *momento del secondo ordine* o *matrice di covarianza* $\Sigma_X \in \mathbb{R}^{n\times n}$ di X mediante

$$\Sigma_X = E\left[(X - m_X)(X - m_X)^T\right] = \int (x - m_X)(x - m_X)^T\, f_X(x)\, dx. \tag{1.7}$$

Si noti che la media e la covarianza di X non sono altro che i valori attesi delle funzioni $g(X) = X$ e, rispettivamente, $g(X) = (X - m_X)(X - m_X)^T$. In modo analogo, si potrebbero definire momenti di ordine superiore al secondo della variabile aleatoria X che, tuttavia, non saranno mai considerati in questo corso. È immediato verificare che le componenti del vettore media $m_X = [m_1, \ldots, m_n]^T$ non sono altro che le medie delle rispettive componenti di X, ovvero

$$m_i = E[X_i] \qquad i = 1, \ldots, n\,. \tag{1.8}$$

Viceversa, le componenti σ_{ij} $(1 \le i, j \le n)$ della matrice di covarianza $\Sigma_X \in \mathbb{R}^{n\times n}$ hanno il seguente significato:

- $\sigma_{ii} \ge 0$ è la varianza della componente X_i del vettore X, i.e.

$$\sigma_{ii} = var(X_i) \triangleq E\left[(X_i - m_i)^2\right] \qquad i = 1, \ldots, n;$$

- σ_{ij}, $i \ne j$, è la covarianza delle componenti X_i e X_j del vettore X, i.e.

$$\sigma_{ij} = cov(X_i, X_j) \triangleq E[(X_i - m_i)(X_j - m_j)] \qquad i = 1, \ldots, n \quad j \ne i\,.$$

Pertanto, per la commutatività della moltiplicazione di scalari, risulta banalmente che $\sigma_{ij} = \sigma_{ji}$, ovvero che la matrice di covarianza $\Sigma_X = \Sigma_X^T$ è *simmetrica*. Si può inoltre verificare che la matrice di covarianza Σ_X è *semi-definita positiva* o *non-negativa definita* nel senso che $v^T \Sigma_X v \geq 0$ per ogni vettore $v \in \mathbb{R}^n$. A tale proposito si consideri, per un generico vettore $v \in \mathbb{R}^n$, la variabile aleatoria scalare $Y \triangleq v^T X$ con media $m_Y = v^T m_X$ e varianza

$$var(Y) = E\left[(Y - m_Y)^2\right] = v^T E\left[(X - m_X)(X - m_X)^T\right] v = v^T \Sigma_X v.$$

Poiché tale varianza deve certamente risultare non negativa a prescindere da v, ne consegue che la forma quadratica $v^T \Sigma_X v$ deve risultare non negativa per ogni v e che, di conseguenza Σ_X è semi-definita positiva, i.e. $\Sigma_X \geq 0$.

Riassumendo, la matrice di covarianza deve risultare simmetrica semi-definita positiva ($\Sigma_X = \Sigma_X^T \geq 0$) pertanto, per le proprietà di tali matrici, i suoi autovalori devono tutti essere reali non-negativi, i.e.

$$sp\,(\Sigma_X) = \{\lambda_1, \ldots, \lambda_n\} \implies \lambda_i \in \mathbb{R}_{\geq 0}, \quad i = 1, \ldots, n\,.$$

Inoltre, le componenti della matrice Σ_X devono soddisfare il seguente *criterio di Sylvester*

$$det \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1i} \\ \vdots & & \vdots \\ \sigma_{1i} & \cdots & \sigma_{ii} \end{bmatrix} \geq 0 \qquad\qquad i = 1, \ldots, n\,.$$

Con $\sigma_i = \sqrt{\sigma_{ii}}$ (radice positiva di $\sigma_{ii} \geq 0$) si denota la *deviazione standard* di X_i. Si noti che, in virtù della non-negatività di Σ_X, non solo le componenti diagonali σ_{ii} devono risultare non-negative ma, per ogni coppia di componenti distinte X_i e X_j del vettore X, deve valere la seguente relazione

$$\varrho_{ij} \triangleq \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \frac{cov(X_i, X_j)}{\sqrt{var(X_i)\, var(X_j)}} \in [-1, 1]\,. \tag{1.9}$$

La quantità scalare ϱ_{ij}, rapporto fra la covarianza di X_i e X_j ed il prodotto delle rispettive deviazioni standard, prende il nome di *coefficiente di correlazione* di X_i e X_j e deve assumere valori nell'intervallo $[-1, 1]$. Per verificare (1.9) è sufficiente considerare il vettore bi-dimensionale $Y = [X_i, X_j]^T$ e verificare che la sua matrice di covarianza $\Sigma_Y = \Sigma_Y^T \geq 0$ è data da

$$\Sigma_Y = \begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{bmatrix}$$

da cui, applicando il criterio di Sylvester, si deduce che $\sigma_{ii} \geq 0$ e

$$\sigma_{ii}\sigma_{jj} - \sigma_{ij}^2 \geq 0 \implies \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}} \leq 1 \implies -1 \leq \frac{\sigma_{ij}}{\sigma_i \sigma_j} \triangleq \varrho_{ij} \leq 1\,.$$

Si può dunque affermare che, date due variabili aleatorie scalari X e Y, vale quanto segue

$$-1 \leq \varrho_{XY} \stackrel{\triangle}{=} \frac{cov(X,Y)}{\sqrt{var(X)\, var(Y)}} \leq 1\,. \tag{1.10}$$

Momenti incrociati e variabili aleatorie incorrelate

Date due variabili aleatorie (vettoriali) X e Y, si definisce la loro *matrice di cross-covarianza* Σ_{XY} come segue:

$$\Sigma_{XY} \stackrel{\triangle}{=} E\left[(X - m_X)(Y - m_Y)^T\right]\,. \tag{1.11}$$

Si noti che

$$\Sigma_{XY} = \Sigma_{YX}^T\,.$$

Le variabili aleatorie X e Y si dicono *incorrelate* se $\Sigma_{XY} = 0$.

PDF congiunte, marginali e condizionate

Date due variabili aleatorie X, definita su $\mathbb{R}^n$, e Y, definita su $\mathbb{R}^p$, si definisce la loro PDF congiunta $f_{X,Y}(\cdot,\cdot)$ come la PDF della variabile aleatoria $Z = [X^T, Y^T]^T$, definita su $\mathbb{R}^{n+p}$, tale che per ogni sottoinsieme $\mathbb{S} \subseteq \mathbb{R}^{n+p}$ risulta che

$$Prob\,(Z \in \mathbb{S}) \;=\; \int_{\mathbb{S}} f_{X,Y}(x,y)\,dx\,dy\,.$$

Dalla PDF congiunta $f_{X,Y}(\cdot,\cdot)$ si può ricavare la PDF marginale $f_X(\cdot)$ ragionando nel seguente modo. Si fissi un sottoinsieme $\mathbb{S}_X \subseteq \mathbb{R}^n$; allora deve risultare

$$\begin{aligned}
Prob\,(X \in \mathbb{S}_X) &= \int_{\mathbb{S}_X} f_X(x)\,dx &&= \\
Prob\,(X \in \mathbb{S}_X, Y \in \mathbb{R}^p) &= \int_{\mathbb{S}_X}\int_{\mathbb{R}^p} f_{X,Y}(x,y)\,dx\,dy &&= \\
&= \int_{\mathbb{S}_X}\left[\int_{\mathbb{R}^p} f_{X,Y}(x,y)\,dy\right]\,dx
\end{aligned} \tag{1.12}$$

da cui, per confronto, si ha

$$f_X(x) \;=\; \int f_{X,Y}(x,y)\,dy, \tag{1.13}$$

ovvero la PDF marginale di X è ottenuta dalla PDF congiunta di X e Y mediante integrazione rispetto all'altra variabile Y. Ragionando nello stesso modo (scambiando X con Y) si ha:

$$f_Y(y) \;=\; \int f_{X,Y}(x,y)\,dx. \tag{1.14}$$

Si vuole adesso definire la distribuzione di probabilità condizionata di X dato Y come la distribuzione di probabilità della variabile aleatoria X sapendo che la variabile aleatoria Y assume un certo valore y. Per caratterizzare tale distribuzione, si adotta la PDF condizionata $f_{X|Y}(x|y)$ che ovviamente deve soddisfare la condizione di normalizzazione

$$\int f_{X|Y}(x|y)dx \;=\; 1 \tag{1.15}$$

e deve ragionevolmente essere proporzionale alla PDF congiunta di X e Y secondo la relazione

$$f_{X|Y}(x|y) \;=\; c(y)\,f_{X,Y}(x,y). \tag{1.16}$$

Sostituendo (1.16) in (1.15) ed utilizzando (1.13), si deduce immediatamente che $c(y) = 1/f_Y(y)$ da cui l'espressione della PDF condizionata

$$f_{X|Y}(x|y) \;=\; \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{1.17}$$

In modo analogo, scambiando X con Y, si ha

$$f_{Y|X}(y|x) \;=\; \frac{f_{X,Y}(x,y)}{f_X(x)}. \tag{1.18}$$

Confrontando (1.17) e (1.18), si ottiene la relazione

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

da cui si deduce immediatamente la formula di Bayes

$$f_{X|Y}(x|y) \;=\; \frac{f_{Y|X}(y|x)\,f_X(x)}{f_Y(y)}. \tag{1.19}$$

Utilizzando (1.14) e (1.18), (1.19) può essere equivalentemente riscritta come

$$f_{X|Y}(x|y) \;=\; \frac{f_{Y|X}(y|x)\,f_X(x)}{\displaystyle\int f_{Y|X}(y|x)\,f_X(x)\,dx}. \tag{1.20}$$

Si noti come la divisione in (1.19) oppure in (1.20) non è altro che una normalizzazione necessaria a soddisfare (1.15). Esaminando (1.20) si può comprendere appieno ed intuitivamente il significato di condizionamento della distribuzione di probabilità di X all'evento $Y = y$. A tale proposito, conviene osservare che le varie PDF che compaiono in (1.20) hanno il seguente significato:

- $f_{X|Y}(x|y)$ è la PDF condizionata (*a posteriori*, i.e. dopo il condizionamento);
- $f_X(x)$ è la PDF non condizionata (*a priori*, i.e. prima del condizionamento);
- $f_{Y|X}(y|x)$, detta *verosimiglianza*, è una funzione che, fissato y, per un certo valore x di X quantifica quanto questo valore risulta *verosimile* con l'evento $Y = y$.

Pertanto, (1.20) sancisce che la PDF a posteriori è ottenuta moltiplicando quella a priori per la verosimiglianza e normalizzando il risultato di tale prodotto per una opportuna costante in modo da avere integrale unitario (rispetto a x). Tenuto conto che la normalizzazione non ha effetto, ovviamente, sulla forma della PDF, questa operazione non fa altro che aumentare la densità di probabilità in quei valori x che risultano più verosimili con y e, viceversa, diminuirla in quelli che sono meno verosimili con y. Si anticipa che uno dei metodi adottati per affrontare il problema della stima di una variabile X sulla base dell'osservazione di un'altra variabile Y (il cosiddetto metodo Bayesiano che verrà esplorato in seguito) è il condizionamento di X dato Y assumendo che X e Y siano variabili aleatorie congiuntamente distribuite.

Variabili aleatorie indipendenti

Due variabili aleatorie X e Y si dicono indipendenti se la loro PDF congiunta soddisfa la seguente relazione

$$f_{X,Y}(x, y) \;=\; f_X(x)\, f_Y(y) \tag{1.21}$$

ovvero è fattorizzabile come prodotto di due funzioni $f_X(\cdot)$ e $f_Y(\cdot)$ nelle sole indeterminate x e, rispettivamente, y. Si constata immediatamente come in virtù di (1.13), (1.14) e (1.21), tali funzioni $f_X(\cdot)$ e $f_Y(\cdot)$ non sono altro che le PDF marginali di X e, rispettivamente, Y. L'indipendenza di variabili aleatorie, come lo stesso nome suggerisce, comporta che il condizionamento di una all'altra risulta ininfluente. Infatti - utilizzando (1.17), (1.18) e (1.21) - è immediato verificare che l'indipendenza di X e Y implica

$$f_{X|Y}(x|y) = f_X(x), \qquad f_{Y|X}(y|x) = f_Y(y)\,. \tag{1.22}$$

È altresì semplice dimostrare quanto segue

$$X \text{ e } Y \text{ indipendenti} \iff F_{X,Y}(x, y) = F_X(x)F_Y(y) \tag{1.23}$$

$$X \text{ e } Y \text{ indipendenti} \implies E\left[g(X)\, h(Y)\right] = E\left[g(X)\right]\, E\left[h(Y)\right] \tag{1.24}$$

In particolare, da (1.24) si ha

$$\Sigma_{XY} \triangleq E\left[(X - m_X)(Y - m_Y)^T\right] = E\left[X - m_X\right]\, E\left[Y - m_Y\right]^T \;=\; 0 \tag{1.25}$$

da cui si deduce che l'indipendenza implica l'incorrelazione mentre l'implicazione inversa, in generale, non vale (due variabili aleatorie incorrelate non sono, in generale, indipendenti). Il lettore può facilmente constatare, dalla definizione, che le variabili aleatorie X, definita su $\mathbb{R}^n$, e Y, definita su $\mathbb{R}^p$, sono indipendenti se e solo se, qualunque siano i sottoinsiemi $\mathbb{S}_X \subseteq \mathbb{R}^n$ e $\mathbb{S}_Y \subseteq \mathbb{R}^p$, gli eventi $X \in \mathbb{S}_X$ e $Y \in \mathbb{S}_Y$ risultano indipendenti, nel senso che

$$Prob\,(X \in \mathbb{S}_X, Y \in \mathbb{S}_Y) \;=\; Prob\,(X \in \mathbb{S}_X)\;\; Prob\,(Y \in \mathbb{S}_Y)\;. \tag{1.26}$$

Da quest'ultima interpretazione del concetto di indipendenza, risulta che l'indipendenza delle variabili aleatorie X e Y implica l'indipendenza di $W = g(X)$ e $Z = h(Y)$ ottenute da quest'ultime mediante arbitrarie funzioni $g(\cdot)$ e $h(\cdot)$. Infatti,

$$\begin{aligned} F_{W,Z}(w,z) &\stackrel{\triangle}{=} Prob\,(W \le w, Z \le z) = Prob\,(g(X) \le w, h(Y) \le z) \\ &= Prob\,(X \in \mathbb{S}_X(w), Y \in \mathbb{S}_Y(z)) \end{aligned}$$

dove $\mathbb{S}_X(w) \stackrel{\triangle}{=} \{x \in \mathbb{R}^n : g(x) \le w\}$ e $\mathbb{S}_Y(z) \stackrel{\triangle}{=} \{y \in \mathbb{R}^p : h(y) \le z\}$. Quindi, sfruttando l'indipendenza di X e Y

$$\begin{aligned} F_{W,Z}(w,z) &= Prob\,(X \in \mathbb{S}_X(w))\;\; Prob\,(Y \in \mathbb{S}_Y(z)) \\ &= Prob\,(g(X) \le w)\;\; Prob\,(h(Y) \le z) \\ &= Prob\,(W \le w)\;\; Prob\,(Z \le z) \\ &\stackrel{\triangle}{=} F_W(w)\, F_Z(z) \end{aligned}$$

da cui, in virtù di (1.23), $W = g(X)$ e $Z = h(Y)$ risultano indipendenti.

PDF di funzioni di variabili aleatorie

Data la distribuzione di probabilità di una variabile aleatoria X, si è spesso interessati a determinare la distribuzione di un'altra variabile aleatoria $Y = g(X)$ da essa dipendente tramite una funzione nota $g(\cdot)$. Il seguente teorema consente di determinare la PDF di Y dalla PDF di X a condizione che la funzione $g(\cdot)$ sia invertibile.

Teorema - Data la v.a. $X \sim f_X(\cdot)$, sia $Y = g(X)$ con $g(\cdot)$ invertibile di inversa $h(\cdot) = g^{-1}(\cdot)$. Allora la v.a. Y ha PDF

$$f_Y(y) = f_X(h(y))\;\left| det\,\frac{\partial h}{\partial y}\,(y)\right| = \frac{f_X(h(y))}{\left| det\,\dfrac{\partial g}{\partial x}\,(h(y))\right|}\;. \tag{1.27}$$

Dimostrazione - Si consideri la CDF della v.a. $Y = [Y_1, Y_2, \ldots, Y_n]'$:

$$\begin{aligned} F_Y(y) &\triangleq Prob\,(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) \\ &= Prob\,(g_1(X) \le y_1, g_2(X) \le y_2, \ldots, g_n(X) \le y_n) \\ &= Prob\left(X \in h(S), S \triangleq \{Y : Y_i \le y_i, i = 1, 2, \ldots, n\}\right) \\ &= \int_{h(S)} f_X(x)\; dx_1\; dx_2\; \cdots\; dx_n \qquad (1.28) \\ &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f_X(h(y)) \left| det\, \frac{\partial h}{\partial y}(y) \right| \; dy_1\; dy_2\; \cdots\; dy_n \\ &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f_Y(y)\; dy_1\; dy_2\; \cdots\; dy_n \end{aligned}$$

da cui, come volevasi dimostrare,

$$f_Y(y) \;=\; f_X(h(y)) \left| det\, \frac{\partial h}{\partial y}(y) \right| .$$

Si noti che in (1.28) si è fatto uso della formula di cambiamento delle variabili in un integrale multiplo

$$dx_1\; dx_1\; \cdots\; dx_n \;=\; \left| det\, \frac{\partial h}{\partial y}(y) \right| \; dy_1\; dy_2\; \cdots\; dy_n\,.$$

La seconda espressione in (1.29) deriva dal teorema della funzione inversa e dal fatto che det $M^{-1} = (\det\; M)^{-1}$. □

Nel caso scalare, il precedente risultato si riduce a quanto segue.

Corollario 1 - Nel caso di variabile aleatoria $X \sim f_X(\cdot)$ scalare e di funzione $g(\cdot)$ scalare e invertibile, con inversa $h(\cdot) = g^{-1}(\cdot)$, si ha

$$f_Y(y) \;=\; f_X\,(h(y)) \left| \frac{dh}{dy}(y) \right| \;=\; \frac{f_X\,(h(y))}{\left| \dfrac{dg}{dx}\,(h(y)) \right|}\,. \qquad (1.29)$$

Dimostrazione - Per dimostrare (1.29) è sufficiente applicare (1.27) al caso scalare e ricordare la proprietà di derivazione della funzione inversa

$$\left[g^{-1}\right]'(y) \;=\; \frac{1}{g'\left(g^{-1}(y)\right)}$$

dove f' indica la derivazione della funzione f rispetto al suo argomento. $\square$

Un'altra interessante applicazione del Teorema riguarda il calcolo della PDF della somma di variabili aleatorie indipendenti.

Corollario 2 - Sia $Y = X_1 + X_2$, con $X_1 \sim f_1(\cdot)$ e $X_2 \sim f_2(\cdot)$ variabili aleatorie indipendenti. Allora

$$f_Y(y) \;=\; (f_1 * f_2)(y) \triangleq \int f_1(y-x) f_2(x) dx = \int f_1(x) f_2(y-x) dx \,. \tag{1.30}$$

Dimostrazione - Si consideri la variabile aleatoria vettoriale

$$Z \;=\; \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} Y \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 + X_2 \\ X_2 \end{bmatrix} = \underbrace{\begin{bmatrix} I & I \\ 0 & I \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}}_{X} = TX = g(X)\,.$$

Si noti che in questo caso $g(X) = TX$ risulta banalmente invertibile con

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = g^{-1}(X) = T^{-1}X = \begin{bmatrix} Z_1 - Z_2 \\ Z_2 \end{bmatrix}.$$

Inoltre $\partial g / \partial x = T$ con $det\ T = 1$, per cui si ha

$$f_Z(z) = f_X(z_1 - z_2, z_2) = f_1(z_1 - z_2) f_2(z_2)$$

dove si è sfruttata l'indipendenza di X_1 e X_2. Ponendo $z_1 = y, z_2 = x$ ed estraendo dalla PDF congiunta $f_{Z_1,Z_2}(\cdot,\cdot)$ la PDF marginale $f_{Z_1}(\cdot) = f_Y(\cdot)$ mediante integrazione, si ottiene

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

o, equivalentemente, sfruttando la proprietà commutativa dell'integrale di convoluzione

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

come volevasi dimostrare. $\square$

I successivi due corollari mostrano applicazioni interessanti del precedente teorema a problemi di stima in cui si desidera stimare una variabile aleatoria X dall'osservazione di un'altra variabile aleatoria Y che, in generale, dipende sia da X che da un certo errore di osservazione V.

Corollario 3 - Sia $Y = h(X,V)$, con $X \sim f_X(\cdot)$ e $V \sim f_V(\cdot)$ variabili aleatorie indipendenti. Assumendo che la funzione $h(x,v)$ sia invertibile rispetto a v, cioè dati x e y risulta univocamente determinato $v = h^{-1}(x,y)$ tale che $y = h(x,v)$, vale quanto segue:

$$f_{X,Y}(x,y) = \frac{f_V(h^{-1}(x,y))}{\left| det \dfrac{\partial h}{\partial v}(x, h^{-1}(x,y)) \right|} f_X(x) \tag{1.31}$$

$$f_{Y|X}(y|x) = \frac{f_V(h^{-1}(x,y))}{\left| det \dfrac{\partial h}{\partial v}(x, h^{-1}(x,y)) \right|} \tag{1.32}$$

$$f_Y(y) = \int \frac{f_V(h^{-1}(x,y))}{\left| \dfrac{det\, \partial h}{\partial v}(x, h^{-1}(x,y)) \right|} f_X(x)\, dx \tag{1.33}$$

$$f_{X|Y}(x|y) = \frac{f_V(h^{-1}(x,y))\, f_X(x)}{\int f_V(h^{-1}(x,y))\, f_X(x) dx}\,. \tag{1.34}$$

Dimostrazione - Si consideri la trasformazione invertibile

$$X' = \begin{bmatrix} X \\ V \end{bmatrix} \longrightarrow Y' = g(X') = \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X \\ h(X,V) \end{bmatrix}. \tag{1.35}$$

In virtù dell'ipotesi di invertibilità di $h(x,v)$ rispetto a v, risulta che $det\, \partial h / \partial v \neq 0$. Applicando (1.27) a $Y' = g(X')$ definita in (1.35) si ottiene quindi

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{f_{X,V}\,(x, h^{-1}(x,y))}{\left| det\, \frac{\partial h}{\partial v}(x, h^{-1}(x,y)) \right|} \\ &= \frac{f_V\,(h^{-1}(x,y))\;\; f_X(x)}{\left| det\, \frac{\partial h}{\partial v}(x, h^{-1}(x,y)) \right|} \end{aligned}$$

dove l'ultimo passaggio sfrutta l'indipendenza di X e V. Le successive formule (1.32), (1.33) e (1.34) sono ottenute ricordando che $f_{Y|X} = f_{X,Y}/f_Y$, $f_Y = \int f_{X,Y}\, dx$ e $f_{X|Y} = f_{Y|X} f_X / f_Y$. □

Corollario 4 - Sia $Y = h(X) + V$, con $X \sim f_X(\cdot)$ e $V \sim f_V(\cdot)$ variabili aleatorie indipendenti. Allora, vale quanto segue:

$$f_{X,Y}(x,y) = f_V\,(y - h(x))\;\; f_X(x) \tag{1.36}$$

$$f_{Y|X}(y|x) = f_V\,(y - h(x)) \tag{1.37}$$

$$f_Y(y) = \int f_V\,(y - h(x))\;\; f_X(x)\; dx \tag{1.38}$$

$$f_{X|Y}(x|y) = \frac{f_V(y - h(x) f_X(x)}{\int f_V(y - h(x)) f_X(x) dx}\,. \tag{1.39}$$

Dimostrazione - Questo risultato deriva direttamente dal precedente Corollario 4 relativamente al caso particolare $h(X,V) = h(X) + V$, per cui si ha

$$h^{-1}(x,y) = y - h(x), \ \frac{\partial h}{\partial v} = I \ \Rightarrow \ det \frac{\partial h}{\partial v} = 1 \,. \tag{1.40}$$

Infatti, le espressioni (1.36)-(1.39) sono immediatamente ottenute da (1.31)-(1.34) mediante le sostituzioni (1.40). $\square$