# VARIABILI ALEATORIE MULTIVARIATE
(vers. 1/11/2013)

Daniela De Canditiis
modulo di CdP di teoria dei segnali - Ingegneria dell' informazione - (Sapienza - Latina)

## VARIABILI ALEATORIE MULTIVARIATE

Molto spesso nelle applicazioni siamo interessati a studiare il comportamento di piú v.a. simultaneamente e dunque da qui la necessitá di definire v.a. multivariate. Una v.a. multivariata $X = (X_1, \cdots, X_n)$ é una funzione

$$X : \omega \in \Omega \to X(\omega) \in R^n$$

per cui, comunque si scelga un punto $(x_1, \cdots, x_n) \in R^n$ é possibile calcolare la probabilitá $P(X_1 \leq x_1, \cdots, X_n \leq x_n)$, cioé l'evento $\{X_1 \leq x_1, \cdots, X_n \leq x_n\} = \bigcap_{i=1}^n \{X_i \leq x_i\} \subseteq \Omega$.

L'introduzione delle v.a. multivariate serve per poter affrontare diversi problemi reali, in particolare, quelli legati all'inferenza statistica. Dal punto di vista concettuale le v.a. multivariate non sono diverse dalle v.a. monovariate, ma la trattazione matematica rimane piú complicata per l'ovvio passaggio da $R$ a $R^n$. vediamo subito qualche esempio di v.a. multivariata

a) se l'esperimento casuale consiste nel lancio di due dadi, noi potremmo essere interessati alla v.a. $(X_1, X_2)$ dove la prima registra il risultato del lancio del primo dado e la seconda il risultato del lancio del secondo dado. E' chiaro che in questo caso le due v.a. che compongono il vettore 2-dimensionale sono indipendenti per definizione dell'esperimento, ma ció non preclude la possibilitá di poter parlare della variabile 2-variata (semmai, la cosa semplificherá la trattazione). Un'altra v.a. che si potrebbe definire é la v.a. $(X_1, X_2, X_3)$ in cui la prima v.a. registra la somma dei due risultati, la seconda registra la differenza tra il primo e il secondo e la terza registra la differenza tra il secondo ed il primo. In questo secondo esempio le v.a. che compongono la v.a. 3-variata non sono indipendenti e i possibili valori in $R^3$ assunti dalla v.a. sono punti dello spazio $R^3$ del tipo $\{(2, 0, 0), (3, 1, -1)...\}$

b) se l'esperimento casuale é individuare a caso un punto $(x, y)$ all'interno del quandrato $[0, 1] \times [0, 1]$, viene quasi naturale definire la v.a. 2-variata $(X, Y)$ in cui la prima v.a. rappresenta l'ascissa del punto scelto e la seconda v.a. rappresenta l'ordinata del punto scelto in un sistema di assi cartesiani.

Di seguito vediamo come si estendono i concetti base delle v.a. monovariate al caso delle v.a. multivariate.

La **funzione di ripartizione** di una v.a. $X$ n-variata é definita come

$$F_X(x) = F_{(X_1,...,X_n)}(x_1,...,x_n) = P(X_1 \leq x_1,...,X_n \leq x_n), \quad \text{per ogni } x \in R^n.$$

La f.r. consente il calcolo della probabilitá di un qualunque evento $B \subset R^n$ ma il suo utilizzo non é cosi immediato come nel caso $n = 1$. Infatti in $R^n$ gli intervalli (gli aperti della classe di Borel) sono piú complicati da gestire per esempio:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{(X,Y)}(a_2,b_2) - F_{(X,Y)}(a_1,b_2) - F_{(X,Y)}(a_2,b_1) + F_{(X,Y)}(a_1,b_1)$$

Per le multivariate allora, piuttosto che con la f.r., si preferisce lavorare direttamente con la distribuzione di probabilitá che permette di ricavare la prob di un qualunque evento in maniera piú diretta. Distinguiamo i due casi discreto e continuo:

- **v.a. discrete**

  La distribuzione di probabilitá di una v.a. discreta n-variata é la funzione

  $$p_{(X_1,...,X_n)}(x_1,...,x_n) = P(X_1 = x_1,...,X_n = x_n),$$

  definita per tutti i valori di $R^n$ che la v.a. $(X_1,...,X_n)$ puó assumere. Se il numero dei valori assumibili é finito parleremo di una n-variata discreta finita, se il numero di valori assumibili é numerabile parleremo di una n-variata discreta numerabile. Quindi, come nel caso monovariato, per definire una v.a. discreta abbiamo bisogno di conoscere i valori da essa assunti e le probabilitá con cui li assume. La funzione di ripartizione avrá la seguente forma

  $$F_{(X_1,...,X_n)}(x_1^*,...,x_n^*) = \sum_{\{(x_1,...,x_n):x_i \leq x_i^*\}} p_{(X_1,...,X_n)}(x_1,...,x_n), \tag{1}$$

  inoltre, la probabilitá di un qualunque evento $B \subset R^n$ si potrá calcolare sommando le probabilitá dei punti di $R^n$ assumibili dalla v.a. e contenuti in $B$, cioé

  $$P(B) = \sum_{\{(x_1,...,x_n) \in B\}} p_{(X_1,...,X_n)}(x_1,...,x_n), \tag{2}$$

- **v.a. continue**

  La distribuzione di probabilitá di una v.a. continua é la sua densitá di probabilitá (di solito indicata con la lettera minuscola $f_{(X_1,...,X_n)}(x_1,...,x_n)$), definita come:

  $$F_{(X_1,...,X_n)}(x_1,...,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{(X_1,...,X_n)}(x_1,...,x_n) dx_1 \cdots dx_n, \tag{3}$$

  Attraverso la f. densitá, piuttosto che attraverso la f.r., si puó allora calcolare la probabilitá di un qualunque evento $B \subset R^n$ (sottinsieme di $R^n$) con la seguente formula:

  $$P(B) = \int \cdots \int_{\{(x_1,...,x_n) \in B\}} f_{(X_1,...,X_n)}(x_1,...,x_n) dx_1 \cdots dx_n, \tag{4}$$

osservazione: limitiamoci al caso $n = 2$ e pensiamo ad una v.a. 2-variata definita su un certo dominio $D \subset R^2$; $D$ é l'insieme dei valori assunti dalla v.a. bivariata. Se $D$ é finito oppure numerabile siamo nel caso discreto, invece se $D$ ha la cardinalitá del continuo siamo nel caso continuo. In entrambe le situazioni la distribuzione di probabilitá descrive come la massa totale $P(\Omega) = 1$ viene distribuita sul dominio $D$: per cui, se $D$ é finito o numerabile, la massa viene spacchettata in pezzettini puntuali di massa non nulla posizionati nei punti $(x, y) \in D$ assumibili (con prob appunto non nulla) dalla variabile bidimensionale, se invece $D$ é continuo allora la massa di probabilitá viene *spalmata* sul domionio $D$ in maniera che in nessun punto si concentri. In entrambe i casi peró la prob di un evento $B \subset D$ viene calcolata misurando la massa di probabilitá contenuta in $B$; in altri termini, aggregando i contributi (in probabilitá) dei punti di $D$ che appartengono a $B$ e ció equivale a sommare nel primo caso (vedi formula (2)) ed ad integrare nel secondo caso (vedi formula (4)); comunque, seguendo l' osservazione 2 fatta nel foglio esercitazioni3, le due formule (2) e (4) si possono uniformare utilizzando la definizione di integrale di Stieltjes, per cui in entrambe i casi $P(B) = \int \cdots \int_B dF_{X_1,\cdots,X_n}(x_1, \cdots, x_n)$ .

Le distribuzioni multivariate godono di diverse proprietá analoghe a quelle delle v.a. monovariate con adeguato aggiornamento delle notazioni, rimandiamo al libro di testo per una loro trattazione. Qui invece riportiamo una proprietá specifica per le multivariate: la possibilitá di ottenere la distribuzione marginale di una delle variabili aleatorie componenti il vettore a partire da quella congiunta del vettore aleatorio. Il viceversa, cioé la costruzione della congiunta a partire dalle marginali, non é invece possibile a meno che le v.a. componenti non siano indipendenti. Questa osservazione ci dice quanto conoscere la distribuzione congiunta di $(X, Y)$ sia molto di piú che conoscere la distribuzione di $X$ e di $Y$, a meno che appunto le due non siano indipendenti. Di seguito le formule da usare per ricavare le densitá marginali:

$$P(X_i = x^*) = \sum_{\{(x_1,\dots,x_n):x_i=x^*\}} p_{(X_1,\dots,X_n)}(x_1,...,x^*,...,x_n)$$

$$f_{(X_i)}(x^*) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1,\dots,X_n)}(x_1,...,x^*,...,x_n)dx_1 \cdots dx_{i-1}dx_{i+1} \cdots dx_n,$$

Di seguito alcune definizoni fondamentali per lo studio e l'analisi di una v.a.

- **VALOR MEDIO**

  Il valore medio di una v.a. n-variata é per definizione il vettore numerico $E(X) = (E(X_1), \cdots, E(X_n))$ in cui la componente $i$-esima é il valor medio della v.a. componente $i$-esima.

  per il valor medio valgono tutte le proprietá viste per il valor medio di v.a. monovariate con l'accortezza di estenderle a tutte le componenti del vettore esempio:

  -) Il valor medio di una v.a. n-variata costante (cioe' di una n-pla di numeri) é la n-pla stessa

  -) Se $a$ e $b$ sono due vettori costanti allora, $E(aX + b) = aE(X) + b$ da interdersi che ogni componente del vettore $E(aX + b)$ si ottiene moltplicando elemento per elemento $aE(X)$ ed aggiungendo il vettore b

  -) Il valor medio di una funzione $g(X) : X \in R^n \longrightarrow R$ di v.a. n-variata é il numero (perché $g(X)$ é univariata, cioé a valori in $R$)

  $$E(g(X)) = \sum_{\{(x_1,\dots,x_n)\}} g(x_1,...,x_n)p_{(X_1,\dots,X_n)}(x_1,...,x_n) \tag{5}$$
  $$E(g(X)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1,...,x_n)f_{(X_1,\dots,X_n)}(x_1,...,x_n)dx_1 \cdots\cdots dx_n \tag{6}$$

  per v.a. discreta e continua rispettivamente.

- **VARIANZA / COVARIANZA**

  Al posto della varianza per v.a. n-variate si parla di matrice di varianza/covarianza. Se $X = (X_1, ..., X_n)$ é una v.a. n-variata allora la sua matrice di covarianza sará una matrice in cui sulla diagonale ci saranno le varianze delle v.a. componenti del vettore ed i cui valori extra diagonali saranno dati dalle covarianze. Piú specificamente, la martrice di varianza/covarianza indicata con $\Sigma$ é una matrice quadrata $n \times n$ tale che $\Sigma_{ii} = var(X_i)$ e $\Sigma_{ij} = cov(X_i, X_j)$, per $i \neq j$.

  Di seguito il concetto di covarianza

  La covarianza tra due v.a. $X$ e $Y$ congiuntamente distribuite é per definizione il numero (perché é il valor medio di una funzione $(X,Y) \rightarrow R$)

definito come $cov(X,Y) = E((X - E(X))(Y - E(Y)))$ che per i due casi discreto e continuo si calcola con le seguenti formule:

$$
\begin{aligned}
cov(X,Y) = & \quad \sum_{i,j}(x_i - E(X))(y_j - E(Y))p_{X,Y}(x_i, y_j) \\
cov(X,Y) = & \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E(X))(y - E(Y))f_{X,Y}(x,y)dxdy
\end{aligned}
$$

in pratica la covarianza é il valor medio dei prodotti degli scarti delle due variabili aleatorie dalle loro rispettive medie. Se gli scarti sono *in media* dello stesso segno allora si dice che le due variabili sono correlate positivamente (cioé quando aumenta l'una aumenta anche l'altra), viceversa se *in media* il segno degli scarti di X e degli scarti di Y sono opposti (cioé quando X si allontana dalla sua media per valori in eccesso allora Y si allontana dalla sua media per valori in difetto) allora si dice che le variabili sono correlate negativamente. La correlazione tra due v.a. é peró il numero adimenisonale

$$\rho(X,Y) = cov(X,Y)/\sigma_X\sigma_Y \quad \in [-1,1].$$

Una correlazione in modulo vicina ad 1 suggerisce un legame lineare tra le due v.a. invece una correlazione vicina a zero ci dice che le due v.a. non sono legate linearmente e dunque potrebbero essere legate da relazioni piú complicate di quella lineare. Una correlazione esattamente uguale a zero si ha quando le due v.a. sono indipendenti. Ma attenzione non vale il viceversa, cioé se $\rho(X,Y) = 0$ non é detto che le due v.a. siano indipendenti, a meno che esse non siano normali. Riassumiamo alcune proprietá:

-) L'unitá di misura della covarianza é il prodotto delle unitá di misura delle due v.a., perció si introduce la correlazione che invece é adimensionale.

-) vale la seguente uguaglianza $cov(X,Y) = E(XY) - E(X)E(Y)$. Di solito questo é il modo in cui la si calcola.

-) se $X$ e $Y$ sono indipendenti $\Rightarrow cov(X,Y) = 0$

-) se $X$ e $Y$ sono normali allora, $cov(X,Y) = 0 \Leftrightarrow$ X,Y indipendenti

-) $cov(X,X) = var(X)$

-) $var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} var(X_i) + \sum_{i \neq j} cov(X_i, X_i)$

- **MODA**

Esattamente come per una v.a. univariata, per una v.a. multivariata la moda é definita come il punto di massimo della sua funzione massa di probabilitá (nel caso discreto) oppure della sua densitá di probabilitá (nel caso continuo). Dunque nel primo caso, la moda é il valore piú probabile , mentre nel secondo caso la moda é il punto attorno cui é piú probabile che la v.a. continua assuma valori. In entrambe i casi la moda non é detto che coincida con il valor medio. Il valore della moda puó non essere unico perché ci possono essere piú punti in cui la funzione massa di prob. oppure densitá di prob. assume il suo valore massimo ed in questo caso si parla di distribuzioni multimodali.

- **distribuzioni condizionate**

  Date due variabili aleatorie congiuntamente distribuite, si puó parlare di distribuzione condizionata dell'una rispetto all'altra. Distinguiamo i due casi discreto e continuo.

  Nel caso discreto la distribuzione di $X$, condizionata dall'evento $\{Y = y\}$ é per definizione la funzione massa di probabilitá $p_{X|Y=y}(x) = \frac{p_{X,Y}(x,y)}{p(Y=y)}$, che associa ad ogni valore $x$ assumibile da $X$ la probabilitá con cui la variabile $X$ assume quel valore nell'universo delle coppie $(x, y)$, con $y$ fissato. Piú precisamente si puó ritrovare nella definizione appena data la definizione di probabilitá condizionata tra i due eventi $\{X = x\}$ e $\{Y = y\}$: $P(\{X = x\}|\{Y = y\}) = \frac{P(\{X=x\} \bigcap \{Y=y\})}{P(\{Y=y\})}$. Dunque la distribuzione di probabilitá di $X$ condizionata dall'evento $\{Y = y\}$, rimane proporzionale alla distribuzione congiunta $p_{(X,Y)}(x, y)$ con $y$ fissato, rinormalizzata con una costante cioé con la marginale di $Y$ valutata in $y$, quindi le probabilitá condizionate di $X$ conservano le proporzioni tra loro, poiché conservano la forma della congiunta ristretta alla retta $Y = y$.

  esempio: Un'urna contiene 5 palline bianche e 8 rosse. Estraiamo 2 palline senza reinserimento. Sia $X_i$ la v.a. che assume valore 1 se la $i$-esima pallina estratta é bianca e 0 altrimenti. Si determini la funzione massa di probabilitá congiunta della variabile $(X_1, X_2)$ e poi si calcoli la prob condizionata di $X_1$ dato che $X_2 = 1$.

  soluzione: i possibili valori assunti dalla v.a. 2-variata $(X_1, X_2)$ sono le coppie $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, determiniamo le probabilitá con cui li assume. Poiché le estrazioni avvengono senza reinserimento $p_{(X_1,X_2)}(0, 0) = \frac{8}{13}\frac{7}{12}$; $p_{(X_1,X_2)}(0, 1) = \frac{8}{13}\frac{5}{12}$; $p_{(X_1,X_2)}(1, 0) = \frac{5}{13}\frac{8}{12}$; $p_{(X_1,X_2)}(1, 1) = \frac{5}{13}\frac{4}{12}$. Prima di valutare la distribuzione condizionata di $X_1$ dall'evento $X_2 = 1$, valutiamo la marginale di $X_2$ nel punto 1 e cioé $p_{X_2}(1) = P(\{X_2 = 1\}) = P((0, 1)) + P((1, 1) = \frac{8}{13}\frac{5}{12} + \frac{5}{13}\frac{4}{12}$. Adesso possiamo valutare la distribuzione condizionata di $X_1$ dall'evento $X_2 = 1$, otteniamo:

  $$p_{X_1|X_2=1}(0) = \frac{p_{(X_1,X_2)}(0, 1)}{p_{X_2}(1)} = \frac{8 \cdot 5}{8 \cdot 5 + 5 \cdot 4}$$
  $$p_{X_1|X_2=1}(1) = \frac{p_{(X_1,X_2)}(1, 1)}{p_{X_2}(1)} = \frac{5 \cdot 4}{8 \cdot 5 + 5 \cdot 4}$$

  come si vede dai numeri, la distribuzione condizionata di $X_1$ sui suoi due possibili valori 0 e 1 conserva le proporzioni della congiunta ristretta alla retta $X_2 = 1$, cioé vale l'uguaglianza $\frac{p_{X_1|X_2=1}(0)}{p_{X_1|X_2=1}(1)} = \frac{p_{X_1,X_2}(0,1)}{p_{X_1,X_2}(1,1)}$.

Nel caso continuo, se $X$ e $Y$ sono congiuntamnete distribuite, la densitá condizionata di $X$ dato $Y = y$ é la funzione densitá di probabilitá $f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, definita per tutti gli $y$ tali che $f_Y(y) > 0$. Poiché la probabilitá $P(Y = y) = 0$, in questo caso per ritrovare la definizione di probabilitá condizionata di due eventi dobbiamo moltiplicare entrambe i membri dell'eguaglianza per $dx$ ottenendo

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

che per $dx$ e $dy$ piccolini si puó tradurre nel seguente rapporto in cui si riconosce chiaramente la definizione di probabilitá condizionata:

$$P(\{x < X \le x+dx\})|\{y < Y \le y+dy\}) = \frac{P(\{x < X \le x + dx\}) \cap \{y < Y \le y + dy\})}{P(\{y < Y \le y + dy\})}.$$

Come per le v.a. discrete anche per le v.a. continue la distribuzione di probabilitá di $X$ condizionata dall'evento $\{Y = y\}$, rimane proporzionale alla distribuzione congiunta $f_{(X,Y)}(x, y)$ con $y$ fissato, rinormalizzata con una costante cioé con la marginale di $Y$ valutata in $y$, quindi le probabilitá condizionate di $X$ conservano le proporzioni tra loro, poiché conservano la forma della congiunta ristretta alla retta $Y = y$.

Osservazione: poiché le distribuzioni condizionate sono distribuzioni (solo definite su un universo ristretto), per loro valgono tutte le definizioni e le proprietá che valgono per le distribuzioni non condizionate. Esempio

$$E(X|Y = y) = \int x f_{X|Y=y}(x)dx$$

$$var(X|Y = y) = \int (x - E(X|Y = y))^2 f_{X|Y=y}(x)dx$$

esempio: (Ross, cap 6. n. 43) La densitá congiunta di $X$ e $Y$ é data da $f_{(X,Y)}(x, y) = ce^{-x}(x^2 - y^2)$, per $0 < x < \infty$, $-x < y < x$. Determinare la densitá condizionata di $Y$ dato $X = x$.

soluzione: per calcolare la densitá condizionata dobbiamo prima valutare la marginale di $X$ nel generico $x \in (0, +\infty)$. Applicando la formula $f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)dy$ avremo

$$f_X(x) = ce^{-x} \int_{-x}^{x} (x^2 - y^2)dy = ce^{-x} \left[x^2y - y^3/3\right]_{-x}^{x} = \frac{4}{3}ce^{-x}x^3$$

da cui applicando la formula $f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$ ricaviamo:

$$f_{Y|X=x}(y) = \frac{3(x^2 - y^2)}{4x^3}, \quad \text{per} - x < y < x$$

da notare che $f_{Y|X=x}(y)$ é un polinomio di secondo grado in $y$ e dunque avrá la forma di una parabola, i coefficienti del polinomio dipendono da $x$, inoltre anche il suppporto di $f_{Y|X=x}(y)$ cambia dato $X = x$, diventando sempre piú grande al crescere di x.

- **FUNZIONE DI V.A.**

Nel caso multivariato una funzione della v.a. $(X_1, ..., X_n)$ puó essere una v.a. univariata se $g : R^n \rightarrow R$ oppure puó essere ancora multivariata se $g : R^n \rightarrow R^n$. Il primo caso é piú semplice: se si é interessati solo al valor medio si applicano direttamente le formule (5) e (6), invece se si é interessati alla distribuzione allora si applica il procedimento seguente:

*Nel caso di v.a. continua, per ogni $z \in R$ si valuta $F_{g(X_1,...,X_n)}(z) = P(g(X_1, ..., X_n) \leq z)$ mediante un integrale n-dimensionale e poi dalla f.r. si ricava la densitá per derivazione; nel caso discreto si valuta direttamente la funzione massa di probabilitá calcolando $P(g(X_1, ..., X_n) = z)$ come somma delle probabilitá della v.a. multivariata $(X_1, .., X_n)$ sui valori che soddisfano il vincolo $g(X_1, ..., X_n) = z$.*

Sicuramente tra le funzioni univariate di una v.a. multivariata la somma riveste un ruolo centrale nelle applicazioni. In particolare si puó dimostrare che la distribuzione della v.a. somma di $n$ variabili aleatorie indipendenti e congiuntamente distribuite é la convoluzione delle distribuzioni.

Un altro esempio molto utile di funzione univariata di una v.a. multivariata é la funzione max oppure minimo. In particolare si puó dimostrare che la distribuzione delle v.a. massimo di $n$ variabili aleatorie indipendenti ed identicamente distribuite é $nF_X(x)^{n-1}f_X(x)$, mentre quella del minimo é $n(1 - F_X(x))^{n-1}f_X(x)$, dove $F_X(x)$ e $f_X(x)$ sono la f.r. e la densitá della generica v.a..

esempio: Se $X$ é una variabile uniforme in $(0, 1)$ e $Y$ una variabile esponenziale di parametro $\lambda = 1$, indipendente da $X$, si determini la funzione densitá di probabilitá di $Z = X + Y$.

soluzione: poiché le variabili sono indipendenti possiamo applicare la formula di convoluzione per cui

$$f_{(X+Y)(z)} = \int_B f_X(z - \tau)f_Y(\tau)d\tau$$

dove l'intervallo $B$ deve essere tale da rendere non nulle le due funzioni $f_X$ e $f_Y$. Piché $f_X$ é non nulla in $(0, 1)$ dove vale 1 e $f_Y$ é non nulla in $(0, +\infty)$ dove vale $e^{-y}$, allora gli argomenti delle due funzioni integrande dovranno soddisfare questi vincoli: $0 < z - \tau < 1$ e $\tau > 0$. Da questi vincoli si ricava l'inrtervallo di integrazione di $\tau$ che risulta essere $z - 1 < \tau < z$, per cui se $z - 1$ é giá maggiore di zero risulta automaticamente verificato anche $\tau > 0$, se invece $z - 1$ é minore di zero allora il vincolo $\tau > 0$ deve essere ulteriormente aggiunto; per cui si ha

$$f_{(X+Y)(z)} = \begin{cases} 0 & z \leq 0 \\ \int_0^z e^{-\tau} d\tau = 1 - e^{-z} & 0 < z < 1 \\ \int_{z-1}^z e^{-\tau} d\tau = e^{-z}(e-1) & z \geq 1 \end{cases}$$

Il caso in cui $g : R^n \to R^n$ é piú complicato perché bisogna ricavare una distribuzione multivariata da una multivariata, peró c' é un caso in cui la densitá di probabilitá multivariata si puó ottenere direttamente dalla densitá multivariata della v.a. originale utilizzando un risultato analogo a quello del Teorema presentato nel foglio esercitazioni 3 e che enunciamo di seguito:

Teorema

Sia $X_1, ..., X_n$ una v.a. n-variata continua con densitá $f_{X_1,...,X_n}(x_1, ..., x_n)$ e sia $g$ una mappa $g : R^n \to (y_1, ..., y_n) = (g_1(x_1, ..., x_n), ..., g_n(x_1, ..., x_n)) \in R^n$ continua e con Jacobiano $J(x_1, ..., x_n) \neq 0$ in ogni punto. Allora la mappa $g$ é invertibile e indicando con

$$h : (y_1, ..., y_n) \in R^n \to (x_1, ..., x_n) = (h_1(y_1, ..., y_n), ..., h_n(y_1, ..., y_n)) \in R^n$$

la sua mappa inversa, la densitá di probabilitá della v.a. $(Y_1, ..., Y_n) = g(X_1, ..., X_n)$ é data dalla seguente formula

$$f_{(Y_1,..,Y_n)}(y_1, .., y_n) = f_{(X_1,..,X_n)}(h_1(y_1, .., y_n), .., h_n(y_1, ..., y_n))|J(h_1(y_1, .., y_n), .., h_n(y_1, .., y_n))|^{-1}.$$

esempio: Siano $X_1$ e $X_2$ v.a. indipendenti e uniformemente distribuite su (0,1), si determini la densitá congiunta della v.a. $(Y_1, Y_2) = (X_1 + X_2, X_1/X_2)$.

Soluzione: poiché le v.a. sono indipendenti la densitá congiunta di $(X_1, X_2)$ la possiamo esprime come il prodotto delle marginali. Dunque $f_{(X_1,X_2)}(x_1, x_2) = 1$, per $(x_1, x_2) \in (0, 1) \times (0, 1)$. In questo caso abbiamo $g_1(x_1, x_2) = x_1 + x_2$ e $g_2(x_1, x_2) = x_1/x_2$. Lo Jacobiano dela trasfomazione vale $J = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_1} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} = -(x_1 + x_2)/x_2^2$. Calcoliamo, inoltre, la mappa inversa della funzione data, cioé risolviamo il sistema

$$\begin{cases} y_1 = & x_1 + x_2 \\ y_2 = & x_1/x_2 \end{cases} \Rightarrow \begin{cases} x_1 = & y_1 y_2/(1 + y_2) \\ x_2 = & y_1/(1 + y_2) \end{cases} . \tag{7}$$

Valutiamo lo Jacobiano in $(y_1, y_2)$ ed otteniamo $J(y_1, y_2) = -(1+y_2)^2/y_1$ Notiamo inoltre che se $x_1$ e $x_2$ sono in (0,1) allora $0 < y_1y_2/(1 + y_2) < 1$ e $0 < y_1/(1 + y_2) < 1$ utilizziamo adesso il teorema appena enunciato e scriviamo

$$f_{(Y_1,Y_2)}(y_1, y_2) = y_1/(1+y_2)^2 \quad \text{per } 0 < y_1y_2/(1+y_2) < 1 \text{ e } 0 < y_1/(1+y_2) < 1$$

Da notare che avremmo potuto ricavare $J^{-1}(y_1, y_2) = -y_1/(1 + y_2)^2$ direttamente calcolando lo Jacobiano della trasformazione inversa (7).

**ESERCIZI**

---

1) (Ross, cap.6 n. 21) Sia $f(x,y) = 24xy$ per $0 \le x \le 1$, $0 \le y \le 1$, $0 \le x+y \le 1$ e zero altrove.

a) si mostri che $f(x,y)$ é la densitá congiunta di un vettore aleatorio.

b) si determini $E(X)$

c) si determini $E(Y)$

soluzione 1) per rispondere alla domanda a) verifichiamo che $f(x,y) \ge 0$ nel suo dominio di definizione, rimane da verificare che integra ad uno. Verifichiamolo

$$\int_0^1 \int_0^{1-x} 24xydydx = \int_0^1 12xy^2]_{y=0}^{y=1-x} dx = 12\left[\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4}\right]_0^1 = 1$$

valutiamo adesso $E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)dydx$ nel nostro caso si ha:

$$E(X) = \int_0^1 x \int_0^{1-x} 24xydydx = \int_0^1 12x^2(1-x)^2 dx = 2/5$$

analogamente

$$E(y) = \int_0^1 y \int_0^{1-y} 24xydxdy = \int_0^1 12y^2(1-y)^2 dy = 2/5$$

---

2)(dall'aglio cap III, n. 13) Data la v.a. $(X,Y)$ con f. di densitá $f(x,y) = x+y$, per $0 < x < 1$, $0 < y < 1$, trovare la distribuzione della v.a. $Z = Y/X$

soluzione 2) utilizziamo il procedimento descritto nel ripasso. Scriviamo $F_Z(z) = P(Y/X \le z) = \int\int_B (x+y)dxdy$, dove $B = \{(x,y) \in (0,1)\times(0,1) : y \le zx$. Intanto notiamo che se $z < 0$ il semipiano $y \le zx$ non interseca il quadrato $(0,1)\times(0,1)$ pertanto $B = \emptyset$; se invece $z \ge 0$, allora il semipiano interseca il quadrato in meno della sua metá se $0 < z < 1$, in piú della sua metá se $1 \le z$, e dunque

$$F_Z(z) = \begin{cases} \int_0^1 \int_{y=0}^{y=xz} (x+y)dydx = & (z + z^2/2)/3 & 0 < z < 1 \\ \int_0^{1/z} \int_{y=0}^{y=xz} (x+y)dydx + \int_{1/z}^1 \int_{y=0}^{y=1} (x+y)dydx = & 1 - 1/3z - 1/6z^2 & z \ge 1 \end{cases}$$

da cui derivando si ottine

$$f_Z(z) = \begin{cases} 1/3 + z/3 = & 0 < z < 1 \\ -1/3 + 1/3z^3 & z \ge 1 \end{cases}$$

---

3) (Ross cap 6 n.47) Se 3 camion vanno in panne in punti aleatori di una strada di lunghezza L, si trovi la probabilitá che non ci siano 2 camion che distino meno di d tra loro, quando $d \leq L$.

soluzione 3): sia $X_i$ la v.a. distribuita uniformente su $(0, L)$, che rappresenta il punto in cui si ferma il camion $i = 1, 2, 3$. Suopponiamo che il primo camion si fermi prima del secondo che si ferma prima del terzo, allora la probabilitá richiesta dal testo é la seguente $P(X_3 > X_2 + d, X_2 > X_1 + d)$. Poiché le 3 variabili peró sono interscambiabili, la probabilitá dovrá essere moltiplicata per 3! che é il numero diverso di disposizioni dei 3 camion lungo la strada.

Valutiamo $P(X_3 > X_2 + d, X_2 > X_1 + d) = \int\int\int_B \frac{1}{L^3} dx_3 dx_2 dx_3$, con $B = \{(x_1, x_2, x_3) \in (0, L)^3 : x_3 > x_2 + d, x_2 > x_1 + d\}$, otteniamo

$$3!P(X_3 > X_2 + d, X_2 > X_1 + d) = 6 \int_{2d}^{L} \int_{d}^{x_3-d} \int_{0}^{x_2-d} \frac{1}{L^3} dx_3 dx_2 dx_1 =$$

$$6 \int_{2d}^{L} \int_{d}^{x_3-d} \frac{x_2 - d}{L^3} dx_3 dx_2 = \frac{6}{L^3} \int_{2d}^{L} \frac{x_3^2}{2} - 2x_3 d + 2d^2 dx_3 = \frac{1}{L^3}(L^3 - 6dL^2 + 12d^2 L - 8d^3) =$$

$$\frac{(L - 2d)^3}{L^3}$$

---

4) (Ross cap. 6, n.22) la densitá congiunta di $X$ e $Y$ é data da

$$f_{(X,Y)}(x, y) == \begin{cases} x + y & 0 < x < 1, 0 < y < 1 \\ 0 & \text{altrove} \end{cases}$$

dire se $X$ e $Y$ sono indipendenti. Calcolare $f_X(x)$, e determinare $P(X + Y < 1)$.

soluzione 4) le due variabili non sono indipendenti perché la distribuzione congiunta non puó fattorizzarsi nel prodotto $f_X(x)f_Y(y)$. Determiniamo la $f_X(x) = \int_0^1 f_{(X,Y)}(x, y)dy = \int_0^1 (x+y)dy = x+1/2$, per $0 < x < 1$. Rispondiamo adesso alla seconda domanda, definiamo $B = \{(x, y) \in (0, 1)^2 : x + y < 1\}$

$$P(X+Y < 1) = \int\int_B (x+y)dxdy = \int_0^1 \int_0^{1-x} (x+y)dydx = \int_0^1 (\frac{1}{2} - \frac{x^2}{2})dx = 1/3$$

---

5) (Ross cap.6, n.27) Se $X$ é una variabile unifrome in $(0, 1)$ e $Y$una variabile esponenziale di parametro $\lambda = 1$, indipendente da $X$, si determini la funzione di distribuzione di $Z = X + Y$ e di $Z = X/Y$.

soluzione 5) I dati del problema dicono che $f_{(X,Y)}(x, y) = e^{-\lambda}$, nella striscia di semipiano $0 < x < 1, 0 < y$. Valutiamo la funzione di distribuzione di $Z = X + Y$:

$$F_{(X+Y)}(z) = P(X+Y \le z) = \begin{cases} 0 & z \le 0 \\ \int_0^z \int_0^{z-x} e^{-y} dy dx & 0 < z < 1 \\ \int_0^1 \int_0^{z-x} e^{-y} dy dx & z \ge 1 \end{cases}$$

$$F_{(X+Y)}(z) = P(X+Y \le z) = \begin{cases} 0 & z \le 0 \\ z - 1 + e^{-z} & 0 < z < 1 \\ 1 + e^{-z}(1-e) & z \ge 1 \end{cases}$$

da notare che derivando questa funzione si riottiene la funzione densitá calcolata nel testo come esempio di applicazione della convoluzione tra due densitá.

Calcoliamo adesso la funzione di distribuzione di $Z = X/Y$, si ha per definizione di funzione di distribuzione:

$$F_{(X/Y)}(z) = P(X/Y \le z) = P(Y \ge X/z) = \int_0^1 \int_{x/z}^{+\infty} e^{-y} dy dx = z(1 - e^{-1/z})$$

---

6) (Ross cap.6 n.39) Si scelga un numero a caso dall'insieme $\{1, 2, 3, 4, 5\}$. Ora si scelga un numero a caso tra i numeri dello stesso insieme non maggiori di $x$, ovvero tra quelli $\{1, 2.., X\}$. Si chiami questo numero $Y$. Si determini la densitá discreta congiunta di $X$ e $Y$, la condizionata di $X$ dato che $Y = i$, per $i = 1, 2, 3, 4, 5$. Sono indipendenti $X$ e $Y$?

soluzione 6) la distribuzione discreta congiunta vale

$$P(X = j, Y = i) = \frac{1}{5} \cdot \frac{1}{j}, \quad \text{per} j = 1, ..., 5 \text{ e } i = 1, ..., j$$

per calcolare la probabilitá condizionata di $X$ dato $Y = i$, calcoliamo prima la marginale di $Y$ e cioé valutiamo $P(Y = i) = P(\bigcup_{l=i}^5 \{X = l, Y = i\}) = \sum_{l=i}^5 \frac{1}{5} \frac{1}{l}$, e poi con la formula delle probabilitá condizionate valutiamo

$$P(X = j | Y = i) = \frac{P(X = j, Y = i)}{p(Y = i)} = \frac{1}{j5} \left( \sum_{l=i}^{5} \frac{1}{5l} \right)^{-1}, \text{ per } j \ge i$$

Le due variabili non sono indipendenti basta anche solo osservare che $P(X = j) = \frac{1}{5} \neq P(X = j | Y = i)$.

---

7) Le variabili aleatorie discrete $X$ e $Y$ hanno la seguente distribuzione congiunta

$$P_{X,Y}(x, y) = \begin{cases} c(x^2 + y^2), & \text{se } x \in \{1, 2, 4\} \text{ e } y \in \{1, 3\} \\ 0, & \text{altrimenti} \end{cases}$$

(a) Quanto deve valere c?

(b) Calcola $P(Y < X)$
(c) Calcola $P(Y > X)$
(d) Calcola $P(Y = X)$
(e) Calcola $P(Y = 3)$
(f) Cerca le funzioni marginali $p_X(x)$ e $p_Y(y)$.
(g)Calcola $E[X]$, $E[Y]$ and $E[XY]$.
(h) perché $E[X]E[Y] \neq E[XY]$?

soluzione 7) Dalla distribuzione abbiamo che ci sono 6 coppie (x, y) che hanno probabilitá non nulla di occorrere: esse sono (1,1),(1,3),(2,1),(2,3),(4,1), e (4,3). La probabilitá di una coppia é proporzionale alla somma dei quadrati $x^2 + y^2$. poiché la probabilitá dell'intero spazio campione deve essere pari a 1 si ha:

$$(1 + 1)c + (1 + 9)c + (4 + 1)c + (4 + 9)c + (16 + 1)c + (16 + 9)c = 1 \Rightarrow c = 1/72$$

(b) Ci sono tre eventi elementari per cui $y < x$:

$$P(Y < X) = P((2, 1)) + P((4, 1)) + P((4, 3)) = (5 + 17 + 25)/72 = 47/72.$$

c) Ci sono due eventi elementari per cui $y > x$:

$$P(Y > X) = P((1, 3)) + P((2, 3)) = (10 + 13)/72 = 23/72.$$

d) C'e' solo un evento elementare per cui y = x: $P(Y = X) = P((1, 1)) = 2/72$. Notiamo che, usando la b) e la c) si ha $P(Y < X) + P(Y > X) + P(Y = X) = (47 + 23 + 2)/72 = 1$
e) Ci sono 3 esiti elementari per cui $y = 3$:

$$P(Y = 3) = P((1, 3)) + P((2, 3)) + P((4, 3)) = (10 + 13 + 25)/72 = 48/72.$$

f) prima bisogna valutare le distribuzioni marginali e poi applicare la formula del valor medio
$E(X) = 1\frac{12}{72} + 2\frac{12}{72} + 4\frac{12}{72} = 3;$
$E(Y) = E[Y] = 1\frac{24}{72} + 3\frac{48}{72} = 7/3$ .

$$E(XY) = \sum_{x,y} xyP_{(X,Y)}(x, y) = 1\cdot\frac{2}{72}+2\cdot\frac{5}{72}+4\cdot\frac{17}{72}+3\cdot\frac{10}{72}+6\cdot\frac{13}{72}+12\cdot\frac{25}{72} = 61/9$$

poiché le due variabili non sono indipendenti essendo $P_{(X,Y)}(x, y) \neq P_X(x)P_Y(y)$ il valor medio del prodotto non coincide con il prodotto dei valori medi.

---

8) Le variabili casuali $X$ e $Y$ sono distribuite congiuntamente secondo la seguente legge

$$f_{X,Y}(x, y) = \begin{cases} ax, & \text{se } 1 \leq x \leq 2 \text{ e } 0 \leq y \leq x \\ 0, & \text{altrimenti} \end{cases}$$

(a) Calcolare la costante $a$.
(b) Determinare la funzione di densitá marginale $f_Y(y)$
(c) Definita la v.a. $Z = Y - X$ determinare la sua funzione di densitá $f_Z(z)$.

soluzione 8)
a) Imponiamo la condizione di normalizzazione

$$\int_1^2 \int_0^x axdydx = \int_1^2 \left. axy\right]_{y=0}^{y=x} dx = \int_1^2 ax^2dx = a7/3 = 1 \quad \Rightarrow a = 3/7$$

b) la densitá marginale di Y é data dalla formula $f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)dx$ e dunque in questo caso si ha

$$f_Y(y) = \begin{cases} \int_1^2 \frac{3}{7}xdx & = 9/14 & \text{se} \quad 0 \le y \le 1 \\ \int_y^2 \frac{3}{7}xdx & = \frac{3}{14}(4 - y^2) & \text{se} \quad 1 < y \le 2 \\ 0 & & \text{altrimenti} \end{cases}$$

c) prima troviamo la funzione di ripartizione $F_Z(z) = P(Z \le z)$ e poi derivando rispetto a z troviamo la funzione di densitá di Z=Y-X.

$$F_Z(z) = \begin{cases} 0 & & \text{se} z < -2 \\ \int_{-z}^2 \int_0^{x+z} \frac{3}{7}xdydx & = \frac{8}{7} + \frac{6z}{7} - \frac{z^3}{14} & \text{se} - 2 \le z \le -1 \\ \int_1^2 \int_0^{x+z} \frac{3}{7}xdydx & = 1 + \frac{9z}{14} & \text{se} - 1 \le z \le 0 \\ 1 & & \text{se} z > 0 \end{cases}$$

da cui derivando

$$f_Z(z) = \begin{cases} 0 & & \text{se} z < -2 \\ \frac{6}{7} - \frac{3z^2}{14} & \text{se} - 2 \le z \le -1 & \\ \frac{9}{14} & \text{se} - 1 \le z \le 0 & \\ 0 & & \text{se} z > 0 \end{cases}$$

---