



COMPLEMENTI DI STATISTICA

Parte I

Fulvio Ricci

Dipartimento di Fisica, Università di Roma *La Sapienza*



INDICE

La funzione generatrice dei momenti.

Esempio di una funzione generatrice: il caso della distribuzione di Gauss

La funzioni caratteristica

Le distribuzioni di probabilità bivariate.

Linee di regressione

Linee di regressione della media

Un caso di grande importanza: la distribuzione normale bivariata.

LA FUNZIONE GENERATRICE DEI MOMENTI

Le caratteristiche di fluttuazione di una variabile aleatoria \mathbf{X} possono essere dedotte a partire da una particolare funzione, la funzione generatrice dei momenti. Questa funzione dipende dalla distribuzione di probabilità secondo cui fluttua la variabile aleatoria considerata ed è definita nel caso di distribuzioni discrete come

$$g(t) = E[e^{Xt}] = \sum_k p_k \exp(x_k t)$$

e nel caso di distribuzioni continue

$$g(t) = E[e^{Xt}] = \int_{-\infty}^{+\infty} \exp(xt) f(x) dx$$

\mathbf{X} è la variabile aleatoria considerata, t è un' arbitraria variabile reale che non ha alcun particolare significato statistico.

Per $t = 0$ $g(0) = 1$ quindi la funzione g converge sempre a $t = 0$, mentre per $t \neq 0$ non è sempre verificata la convergenza della $g(t)$.

La conoscenza della funzione generatrice è sufficiente a determinare in modo univoco la corrispondente distribuzione di probabilità (**Teorema di unicità**). Infatti, se ci riferiamo per semplicità al caso continuo, sviluppiamo in serie di Mac Laurin dell'esponenziale presente nell'integrale che definisce la $g(t)$. Otterremo allora

$$g(t) = E[X^0] + tE[X] + \dots + \frac{t^n}{n!} E[X^n] + \dots = \sum_{i=0}^{\infty} \frac{\alpha_i}{i!} t^i$$

I coefficienti di questo sviluppo in serie di potenze della $g(t)$

$$\alpha_i = E[X^i]$$

sono i così detti *momenti* della distribuzione della variabile aleatoria \mathbf{X} . Quindi il *momento* di ordine n coincide con la derivata n -esima della funzione generatrice calcolata nel punto $t = 0$.

$$\alpha_i = \left[\frac{d^i}{d^t} g(t) \right]_{t=0}$$

Se consideriamo una funzione di una variabile aleatoria

$$Y = \varphi(X)$$

avremo che la sua funzione generatrice sarà definita da

$$G(t) = E[e^{t\varphi(x)}] = \int_{-\infty}^{+\infty} e^{t\varphi(x)} f(x) dx$$

Quanto scritto sopra è una semplice conseguenza del fatto che la probabilità di avere per Y una realizzazione tale da cadere nell'intervallo $(y, y + dy)$ è pari alla probabilità che x sia compreso nel corrispondente intervallo $(x, x + dx)$, e pertanto sia $f(x)dx$.

In particolare se la funzione $\varphi(x)$ è lineare $\varphi(x) = ax + b$ o più semplicemente $\varphi(x) = x - c$ avremo (nel caso delle variabili aleatorie continue)

$$G(t) = e^{ct} \int_{-\infty}^{+\infty} e^{xt} f(x) dx \quad \longrightarrow \quad G(t) = e^{-ct} g(t)$$

Sia ora $c = E[X] = \mu$.

In tal caso la derivata di ordine n della $G(t)$ calcolata in $t = 0$, è pari al **momento centrale** della distribuzione di probabilità

$$\alpha_n = E[(X - \mu)^n] = \left[\frac{d^n G(t)}{dt^n} \right]_{t=0}$$

Riportiamo qui di seguito senza dimostrarle alcune proprietà della funzione generatrice dei momenti.

a) - Sia $Y=aX+b$, allora e' facile verificare che

$$G(t) = E[e^{tY}] = E[e^{t(aX+b)}] = e^{bt}g(at)$$

b) - Siano S , X e Y variabili aleatorie legate tra loro dalla relazione $S = X + Y$ con X e Y tra loro indipendenti. Ne segue che

$$E[e^{St}] = E[e^{t(X+Y)}] = E[e^{tX}]E[e^{tY}]$$

cioe' la funzione generatrice $g(t)$ di S è ottenibile dal prodotto delle funzioni generatrici $a(t)$ e $b(t)$ di X e Y .

$$g(t) = a(t)b(t)$$

Piú in generale se S è definita come una somma di n variabili aleatorie indipendenti X_i aventi funzione generatrice $a_i(t)$, allora la funzione generatrice $g(t)$ di S è

$$S = \sum_i X_i \qquad g(t) = \prod_i a_i(t)$$

c) - $g(t)$ determina i momenti della distribuzione di probabilità e questi, a loro volta, determinano in modo

univoco la funzione di densità di probabilità $f(x)$. Infatti supponiamo per assurdo che $f_1(x)$ ed $f_2(x)$ sia diverse ma abbiano $\alpha_i^{(1)} = \alpha_i^{(2)}$. Lo sviluppo della funzione differenza $f_1(x) - f_2(x)$ è dato da

$$f_1(x) - f_2(x) = c_0 + c_1x + \dots + c_nx^n + \dots$$

In particolare, consideriamo la quantità

$$\Gamma = \int_{-\infty}^{+\infty} [f_1(x) - f_2(x)]^2 dx$$

Sviluppando la funzione differenza, avremo

$$\Gamma = c_0 \int_{-\infty}^{+\infty} [f_1(x) - f_2(x)] dx + \dots + c_n \int_{-\infty}^{+\infty} x^n [f_1(x) - f_2(x)] dx + \dots$$

che è equivalente a scrivere

$$\int_{-\infty}^{+\infty} [f_1(x) - f_2(x)]^2 dx = c_0(\alpha_1^{(1)} - \alpha_1^{(2)}) + \dots + c_i(\alpha_i^{(1)} - \alpha_i^{(2)})^n$$

Ma, essendo $\alpha_i^{(1)} = \alpha_i^{(2)}$, questo sviluppo risulta identicamente nullo e quindi deve essere $f_1(x) = f_2(x)$.

In conclusione le due funzioni $f_i(x)$ coincidono se hanno lo stesso insieme dei momenti. Quindi la funzione generatrice dei momenti è *sufficiente* a determinare in modo unico la corrispondente distribuzione di probabilità.

ESEMPIO DI UNA FUNZIONE GENERATRICE: IL CASO DELLA DISTRIBUZIONE DI GAUSS

Applichiamo la definizione della funzione generatrice dei momenti al caso di una variabile aleatoria che segua la distribuzione di Gauss.

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left(tx - \frac{(x - \mu)^2}{2\sigma^2}\right) dx$$

Consideriamo l'argomento dell'esponenziale; ad esso sommiamo e sottraiamo la quantità $\mu t + (1/2)\sigma^2 t^2$. Ne segue che avremo

$$\begin{aligned} & \mu t + (1/2)\sigma^2 t^2 - (1/2)\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 2t(x - \mu) + \sigma^2 t^2\right] = \\ & = \mu t + (1/2)\sigma^2 t^2 - (1/2)\left[\frac{x-\mu-\sigma^2 t}{\sigma}\right]^2 \end{aligned}$$

Ponendo $z = \left[\frac{x-\mu-\sigma^2 t}{\sigma}\right]$, otteniamo

$$g(t) = (1/\sqrt{2\pi}) \exp(\mu t + (1/2)\sigma^2 t^2) \int_{-\infty}^{+\infty} \exp\left(\frac{z}{2}\right) dz$$

ed infine

$$g(t) = \exp(\mu t + (1/2)\sigma^2 t^2)$$

Poichè la funzione generatrice dei momenti centrali è

$$G(t) = e^{-\mu t} g(t)$$

avremo che

$$G(t) = e^{\frac{\sigma^2 t^2}{2}}$$

Lasciamo verificare al lettore che

$$\left[\frac{d^2 G}{dt^2}\right]_{t=0} = \sigma^2$$

LA FUNZIONE CARATTERISTICA.

La funzione generatrice dei momenti di una distribuzione di probabilità non esiste necessariamente in quanto non è sempre possibile determinare un intervallo di variazione del parametro t , $-\delta \leq t \leq \delta$, centrato quindi attorno all'origine, nel quale la $g(t)$ sia sempre comunque definita.

Per ovviare a questo problema si introduce la funzione caratteristica. A questo scopo associamo alla variabile aleatoria X una variabile ausiliaria definita nel campo complesso

$$e^{itX} = \cos(tX) + i\sin(tX)$$

dove t è un'arbitraria variabile reale e $i^2 = -1$

Definiamo allora la Funzione caratteristica $\psi(X)$

$$\psi(X) = E[e^{itX}]$$

$E[e^{itx}]$ è il valore medio di una funzione complessa il cui significato è qui esplicitato:

$$W(X) = u(X) + iv(X)$$

$$E[w] = E[u] + iE[v]$$

Quindi la funzione caratteristica è il valore medio della variabile aleatoria ausiliaria e^{itX} che nel caso di variabili discrete è calcolabile come

$$\psi(t) = E[e^{itX}] = \sum_{\nu} p_{\nu} e^{itX_{\nu}}$$

mentre nel caso di variabile continua avremo

$$\psi(t) = E[e^{itX}] = \int_{-\infty}^{+\infty} e^{itX} f(X) dX$$

La sua esistenza è ora assicurata dal fatto che per qualsunque valore di t e^{itX} è limitata

$$|e^{itX}| = 1 \quad \forall t$$

La $\psi(t)$ è nota come la trasformata di Fourier della funzione densità di probabilità $f(X)$ detta anche *Trasformata di Fourier- Stieltjes*.

Dalla teoria delle trasformata di Fourier ne consegue che nota la $\psi(t)$ è possibile ricavare la $f(X)$ in modo univoco (teorema d'unicità) mediante l'antitrasformata di Fourier

$$f(X) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itX} \psi(t) dt$$

Inoltre se la funzione generatrice dei momenti $g(t)$ esiste, allora

$$g(t) = \psi(-it)$$

Proprietà della funzione caratteristica $\psi(t)$

1) $\psi(t)$ è una funzione continua

Infatti: scelto a caso un intervallo h qualsivoglia nel dominio di definizione della variabile aleatoria X

$$|e^{i[t(X+h)]} - e^{itX}| \leq |e^{ith} - 1|$$

Poichè si ha che $|e^{itX}| \leq 1$ e per h che tende a zero il membro di destra della disuguaglianza tende a zero cioè

$$\lim_{h \rightarrow 0} |e^{ith} - 1| = 0$$

2) $\psi(0) = 1$ (ovvia) e $|\psi(t)| \leq 1$ (infatti è il valore aspettato di una quantità compresa tra -1 e 1).

3) Sia $Y = aX + b$, allora la funzione caratteristica di Y è

$$\Psi(t) = E[e^{i(aX+b)t}] = e^{ibt}\psi(at)$$

Tale proprietà è analoga a quella della funzione generatrice.

4) Siano X ed Y due variabili aleatorie complesse indipendenti con funzioni caratteristiche $\psi_x(t)$ $\psi_y(t)$, e sia S la variabile somma $S = X + Y$ Avremo allora

$$\psi_s(t) = \psi_x(t)\psi_y(t) = E[e^{itS}] = E[e^{it(X+Y)t}]$$

$$\psi_s(t) = E[e^{itx}] = E[e^{ity}]$$

5) Due distribuzioni di probabilità sono distinte solo se hanno funzioni caratteristiche distinte. Ciò è una diretta conseguenza della cosiddetta relazione di *Parseval* che ora ricaveremo.

Sia $F(X)$ la funzione di distribuzione integrale di probabilità della variabile X (detta anche funzione di

ripartizione, che ammette $\psi(t)$ come una funzione caratteristica

$$\psi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx = \int_{-\infty}^{+\infty} e^{-itx} dF(X)$$

e sia $G(t)$ una seconda funzione di distribuzione che ammette una funzione caratteristica $\omega(x)$

$$\omega(x) = \int_{-\infty}^{+\infty} e^{itx} dG(t)$$

Integriamo la $\psi(t)$ rispetto alla funzione $G(t)$,

$$\int_{-\infty}^{+\infty} \psi(t) dG(t)$$

Sostituendo $\psi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$ avremo

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{itx} dG(t) dF(x)$$

Ne segue la relazione di Parseval

$$\int_{-\infty}^{+\infty} \psi(t) dG(t) = \int_{-\infty}^{+\infty} \omega(x) dF(x)$$

Concludiamo che le distribuzioni F e G sono distinte solo se la ψ e la ω sono distinte.

LE DISTRIBUZIONI DI PROBABILITÀ BIVARIATE

Nello stesso campo di probabilità siano definite due variabili aleatorie distinte X ed Y che possono assumere insieme i valori

$$x_1, x_2, x_j \qquad y_1, y_2, y_k$$

Sia allora $p(x_j, y_k)$ la probabilità congiunta che X ed Y assumano i valori x_j ed y_k .

ESEMPIO TIPICO: X e Y sono i possibili risultati del lancio congiunto di 2 dadi. Il lettore noti bene che in questo caso specifico X e Y sono **indipendenti**. La proprietà fondamentale della probabilità congiunta è che

$$\sum_{jk} p(x_j y_k) = 1$$

dove la sommatoria è estesa a tutto il campo di probabilità. Inoltre dobbiamo avere che per qualunque i e k

$$p(x_j y_k) \geq 0$$

Nel caso delle distribuzioni bivariate si definiscono anche le probabilità marginali $f(x_j)$, $g(y_k)$

$$P[X = x_j] = \sum_k p(x_j y_k) = f(x_j)$$

$$P[Y = y_k] = \sum_j p(x_j y_k) = g(y_k)$$

La $f(x_j)$ (o la $g(y_k)$) rappresenta la probabilità che la variabile aleatoria X (o Y) assuma i valori x_j (o y_k) per tutti i possibili valori assunti dalla Y (o X).

Se, come nel caso del lancio di due dadi, gli eventi sono indipendenti, la probabilità congiunta di avere come risultato x_j ed y_k è pari al prodotto delle probabilità marginali, $p(x_j y_k) = f(x_j)g(y_k)$, altrimenti in generale avremo

$$p(x_j y_k) \neq f(x_j)g(y_k)$$

PA titolo d'esempio si consideri il caso di tre palline di diverso colore collocate a caso in tre diverse urne. Le variabili aleatorie oggetto di studio sono:

$X =$ numero totale di palline nella prima urna

$Y =$ numero totale di palline nella seconda urna

se il lettore enumera tutti i casi possibili, può facilmente rendersi conto che X dipende da Y e viceversa. Lasciamo al lettore l'onere di calcolarsi le probabilità marginale e dimostrandolo che $p(x_j y_k) \neq f(x_j)g(y_k)$.

Per distribuzioni continue si definirà la funzione densità di probabilità congiunta come la probabilità che X e Y assumano un valore compreso tra x , $x+dx$ e y , $y+dy$

$$P[x < X < x + dx, y < Y < y + dy] = p(x, y)dx dy$$

Ovviamente

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = 1$$

Le funzioni marginali di ripartizione e di probabilità sono così definite:

$$F(x) = \int_{-\infty}^x \int_{-\infty}^{+\infty} p(t, y) dt dy ; G(y) = \int_{-\infty}^y \int_{-\infty}^{+\infty} p(x, t) dx dt$$

$$f(x) = F'(x) = \int_{-\infty}^{+\infty} p(x, y) dy ; g(y) = G'(y) = \int_{-\infty}^{+\infty} p(x, y) dx$$

Se le variabili sono indipendenti, allora si ha

$$p(x, y) = f(x)g(y)$$

Vediamo ora di descrivere quali siano le quantità caratteristiche delle distribuzioni bivariate. Definiamo i valori medi delle singole variabili aleatorie. Nel caso di variabili continue, il valore aspettato della X è

$$\mu_x = E[X] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xp(x, y) dx dy = \int_{-\infty}^{+\infty} xf(x) dx$$

Nel caso di variabile discreta si ha che

$$\mu_x = E[X] = \sum_j \sum_k x_j p(x_j, y_k) = \sum_j x_j \sum_k p(x_j, y_k)$$

Le definizioni per $E[y] = \mu_y$ sono analoghe alle precedenti: è sufficiente scambiare x con y .

Consideriamo ora una nuova variabile aleatoria Z , funzione sia di X che Y : $Z = \psi(X, Y) = Z$. Possiamo dedurre il valore aspettato di Z applicando la definizione:

$$E[Z] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \psi(x, y) dx dy$$

ESEMPIO IMPORTANTE: $Z = X + Y$

$$\begin{aligned} E[X + Y] &= \int_{-\infty}^{+\infty} (x + y)p(x, y)dxdy = \\ &= \int_{-\infty}^{+\infty} xf(x)dx + \int_{-\infty}^{+\infty} yg(y)dy = E[X] + E[Y] \end{aligned}$$

abbiamo così dedotto l'importante *proprietà di additività dei valori medi*

$$E \left[\sum_i X_i \right] = \sum_i E[X_i]$$

Supponiamo che X ed Y siano due variabili indipendenti e consideriamo la variabile prodotto $Z = X \cdot Y$. Poichè $p(x, y) = f(x)g(y)$, avremo:

$$E[Z] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyp(x, y)dxdy = E[X] \cdot E[Y]$$

In conclusione, se

$$Z = \Pi_i X_i$$

con le variabili X_i tutte tra loro indipendenti, allora si ha

$$E[\Pi_i X_i] = \Pi_i E[X_i]$$

Vediamo ora cosa accade per la varianza. Nel caso di distribuzioni bivariate, definiamo la varianza di X nel caso di variabili discrete

$$Var[X] = \sum_j \sum_k (x_j - \mu_x)^2 p(x_j, y_k) = \sum_j (x_j - \mu_x)^2 f(x_j)$$

e nel caso di variabili continue

$$Var[X] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_x)^2 p(x, y)dxdy$$

ovvero

$$Var[X] = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx$$

Le definizioni relative alla $Var[Y]$ sono analoghe.

Una proprietà importante della varianza (di facile dimostrazione) per variabili indipendenti è la seguente

$$Var[X \pm Y] = E[(X \pm Y) - (\mu_x \pm \mu_y)]^2 = Var[X] + Var[Y]$$

Quindi se abbiamo n variabili X_i **indipendenti**, allora

$$Var[\sum_i X_i] = \sum_i Var[X_i]$$

Vediamo quindi di generalizzare la definizione dei momenti di una distribuzione al caso della distribuzione bivariata. I momenti di ordine $n = i + k$ della variabile aleatoria bivariata X, Y sono definiti

$$\alpha_{ik} = E[X^i Y^k]$$

per tutti i valori possibili di i ed k tali che si ha $i + k = n$. Quindi nel caso di variabili discrete avremo

$$\alpha_{ik} = \sum_{l,m} x_l^i y_m^k p(x_l, y_m)$$

e in quello di variabili continue

$$\alpha_{ik} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^i y^k p(x, y) dx dy$$

Specifichiamo il caso dei momenti di ordine 1.

$$\alpha_{10} = E[X] = \mu_x \qquad \alpha_{01} = E[Y] = \mu_y$$

che localizzano il baricentro della distribuzione.

Definiamo ora i momenti centrali della distribuzione di ordine $i + k$

$$\alpha_{ik} = \mu_{ik} = E[(X - \mu_x)^i (Y - \mu_y)^k]$$

Riconosciamo subito che due dei momenti centrali di ordine 2 sono le varianze della distribuzione

$$\mu_{20} = Var[X] \quad \mu_{02} = Var[Y]$$

Il terzo momento centrale di ordine 2

$$\mu_{11} = E[(X - \mu_x)(Y - \mu_y)] = Cov[X, Y]$$

è detto **covarianza**. Essa stabilisce il grado di dipendenza delle due variabili aleatorie X e Y . Infatti se X ed Y sono indipendenti

$$E[XY] = E[X]E[Y]$$

e

$$Cov[XY] = 0$$

La covarianza può essere espressa anche nella forma

$$\mu_{11} = Cov[XY] = E[XY] - E[X]E[Y]$$

Essa dà altresì l'informazione del grado di dispersione delle due variabili aleatorie: ad esempio se una di esse fluttua poco attorno al valore medio allora la covarianza è bassa anche se le variabili sono dipendenti. Così come alla varianza viene associata la deviazione standard, accanto alla covarianza viene introdotto il coefficiente di correlazione

$$\rho[X, Y] = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}} = \frac{Cov[X, Y]}{\sigma_x \cdot \sigma_y} = E[X^*, Y^*]Cov[X^*, Y^*]$$

Si noti che il coefficiente di correlazione è indipendente dalla scelta delle unità di misura e dall'origine delle variabili X e Y , essendo

$$X^* = \frac{X - \mu_x}{\sigma_x} \quad Y^* = \frac{Y - \mu_y}{\sigma_y}$$

Attenzione: se le variabili X e Y sono indipendenti, allora $\mu_{11} = 0$, ma **non è vero il viceversa**. In altre parole aver provato che $Cov[XY] = 0$, non ci assicura l'indipendenza delle variabili X e Y .

ESEMPIO: date due variabili aleatorie indipendenti U e V che seguono la stessa distribuzione di probabilità; costruiamo le variabili aleatorie X, Y tali che

$$X = U + V \quad e \quad Y = U - V$$

Riferiamoci al caso discreto del lancio dei dadi.

U → lancio di un dado

V → lancio di un altro dado

X → somma dei risultati del lancio del dado U e del dado V

Y → differenza dei risultati del lancio del dado U e del dado V

Calcoliamo la covarianza di X e Y ed otteniamo

$$E[XY] = E[U^2] - E[V^2]$$

Poichè U e V seguono la stessa distribuzione allora $E[U^2] = E[V^2]$, quindi

$$Cov[XY] = E[XY] = 0$$

Ma, per come le abbiamo definite X e Y assumeranno sempre simultaneamente valori o entrambi pari o entrambi dispari per ogni lancio, quindi sono variabili dipendenti.

Dimostriamo ora che $|\rho[X, Y]| \leq 1$

Allo scopo consideriamo le variabili aleatorie normalizzate

$$X^* = \frac{X - \mu_x}{\sigma_x} \quad Y^* = \frac{Y - \mu_y}{\sigma_y}$$

e calcoliamo la varianza della variabile $X^* \pm Y^*$

$$Var[X^* \pm Y^*] = E \left[\left(\frac{X - \mu_x}{\sigma_x} \pm \frac{Y - \mu_y}{\sigma_y} \right)^2 \right]$$

Sviluppando l'algebra, potremo riscrivere la relazione precedente nel modo seguente

$$E[X^{*2}] + E[Y^{*2}] \pm 2 E[X^*Y^*] = 1 + 1 \pm 2 Cov[X^*Y^*]$$

Quindi si ha che

$$Var[X^* \pm Y^*] = 2(1 \pm Cov[X^*Y^*])$$

Osserviamo che

$$Cov[X^* \pm Y^*] = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} = \rho[X, Y]$$

In conclusione, poichè

$$Var[X^* \pm Y^*] = 2(1 \pm \rho[X, Y])$$

e poichè la varianza è comunque ≥ 0

$$-1 \leq \rho(x, y) \leq 1 \quad \text{ovvero} \quad |\rho(x, y)| \leq 1$$

Osserviamo che quando non c'è dispersione attorno al valore medio della variabile aleatoria, la varianza è nulla, cioè la variabile non fluttua e coincide con il valore atteso. Esprimiamo tutto questo in termini analitici:

$$Var[X^* \pm Y^*] = 0 \quad \Rightarrow \quad X^* \pm Y^* = \text{costante}$$

ovvero

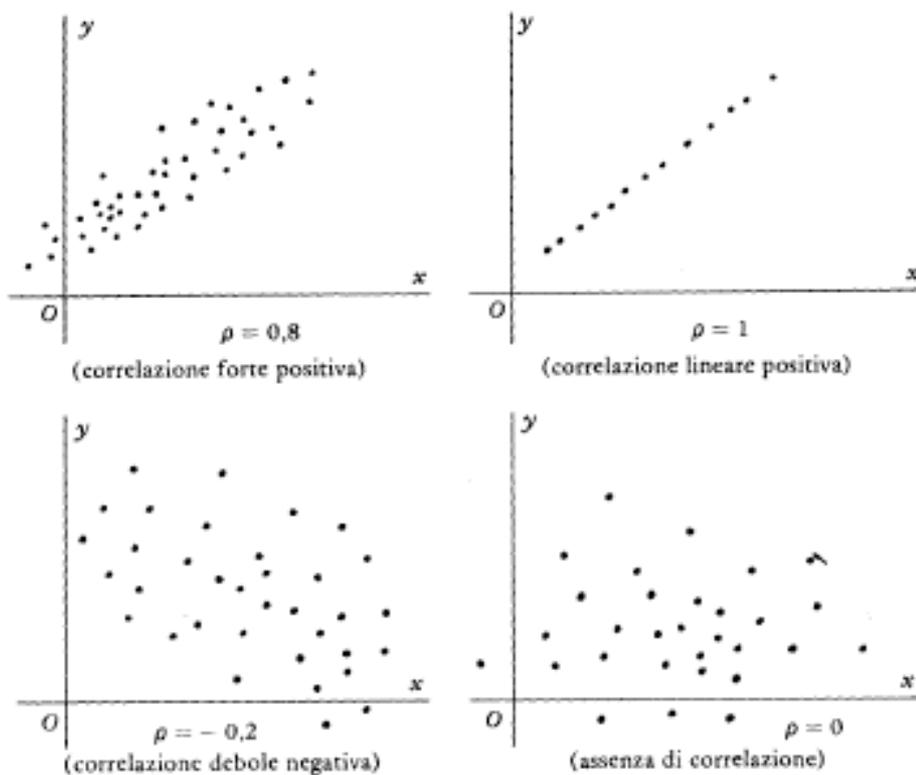
$$\frac{X - \mu_x}{\sigma_x} \pm \frac{Y - \mu_y}{\sigma_y} = \text{costante}$$

$$\rho[X, Y] = \pm 1 \quad \Leftrightarrow \quad Y = \pm \frac{\sigma_y}{\sigma_x} X + \text{costante}$$

Quindi nel caso di dipendenza lineare delle due variabili si ha $\rho[X, Y] = \pm 1$

$\rho = +1$ correlazione positiva

$\rho = -1$ correlazione negativa



I concetti fin qui introdotti per una popolazione bivariata possono essere estesi al caso di più di due variabili aleatorie: parleremo allora di *popolazione multivariata*. Sia

$$S_n = \sum_{i=1}^n X_i$$

dove

$$E[S_n] = \sum_{i=1}^n \mu_i$$

con $\mu_i = E[X_i]$ e

$$Var[S_n] = \sum_{i=1}^n Var[X_i] + 2 \sum_{j=1}^n \sum_{i=1}^n Cov[X_j X_k]$$

con $j \neq k$. Tale insieme di di variabili aleatorie è descritto dalle seguenti grandezze

- n valori medi μ_i che localizzano le singole variabili
- n deviazioni standard σ_i che caratterizzano la dispersione delle singole variabili
- $n(n - 1)$ coefficienti di correlazioni ρ_i che caratterizzano la correlazione a due a due delle variabili

Interpretando la varianza come la correlazione di una variabile con se stessa, possiamo introdurre la **matrice di correlazione** del sistema di n variabili aleatorie.

$$Cov[X_j, X_k]$$

Notiamo che tale matrice è simmetrica

$$Cov[X_j, X_k] = Cov[X_k, X_j]$$

e quindi, per definirla univocamente, basta individuare soltanto metà dei suoi elementi.

LINEE DI REGRESSIONE

È interessante studiare se esiste una qualche relazione tra due variabili aleatorie definite in un campo di probabilità, esplicitabile tramite espressione matematica.

Siano allora X e Y le due variabili aleatorie. Per fissare le idee sia X la variabile casuale indipendente. Cerchiamo allora di esprimere Y , variabile aleatoria dipendente, sulla base della conoscenza del campo di probabilità in cui varia la X e sulla base dell'equazione di una curva del tipo $y = \psi(x)$ che definisce il legame matematico tra le due variabili.

$$y = \psi(x) \quad \rightarrow \quad \text{curva di regressione}$$

Supponiamo per semplicità che la $y = \psi(x)$ rappresenti la retta

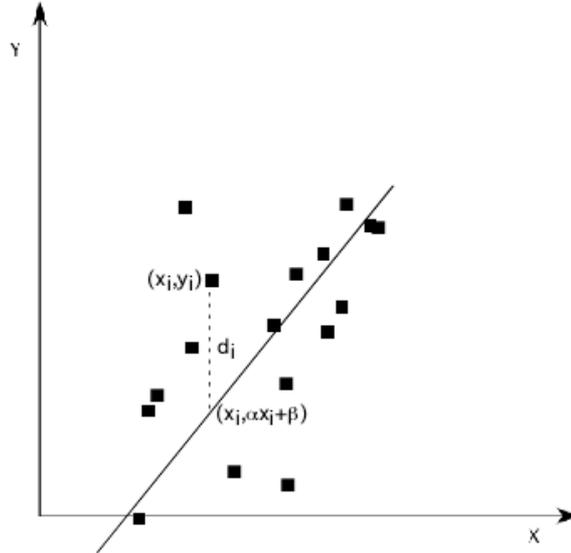
$$Y = \alpha X + \beta$$

Parleremo allora di **retta di regressione**.

X ed Y sono tali per cui ad ogni valore di una di esse corrispondono diversi valori dell'altra variabile. Tali valori saranno dispersi attorno a questa retta di regressione.

Le rette di regressione sono definite sulla base dei valori assunti da α e da β ; esisteranno quindi infinite rette di regressione. Definiremo come migliore retta di regressione quella che meglio si accorda ai punti (x_i, y_i) della distribuzione di probabilità, e stabiliamo come criterio di scelta della retta quello di rendere minima *la somma dei quadrati delle proiezioni sull'asse Y delle distanze dei singoli punti del campo di probabilità dalla migliore*

retta di regressione:



$$\left(\sum_i |(y_i - \alpha x_i - \beta)| \right)_{\text{minimo}}$$

o in modo del tutto analogo,

$$\left[\sum_i (y_i - \alpha x_i - \beta)^2 \right]_{\text{minimo}}$$

La sommatoria è estesa a tutto il campo di probabilità, quindi cercare questo minimo è equivalente ad individuare per quali valori di α e β è minima la quantità

$$G(\alpha, \beta) = \{E[(Y - \alpha X - \beta)^2]\}$$

Il metodo seguito per determinare le rette di regressione è detto dei **Minimi Quadrati**.

Cerchiamo il minimo di questa funzione. A questo scopo, aggiungiamo e togliamo $\mu_y + \alpha\mu_x$ nell'espressione in parentesi

$$G(\alpha, \beta) = E \left[\{(Y - \mu_y) - \alpha(X - \mu_x) + (\mu_y - \alpha\mu_x - \beta)\}^2 \right]$$

Riscriviamo tale relazione in modo più compatto:

$$G(\alpha, \beta) = \mu_{02} - 2\alpha\mu_{11} + \alpha^2\mu_{20} + (\mu_y - \alpha\mu_x - \beta)^2$$

Per calcolare il minimo di una funzione di due variabili procederemo nel modo seguente:

$$\frac{dG(\alpha, \beta)}{d\alpha} = 0 = -2\mu_{11} + 2\alpha\mu_{20} - 2\mu_x(\mu_y - \alpha\mu_x - \beta)$$

$$\frac{dG(\alpha, \beta)}{d\beta} = 0 = -2(\mu_y - \alpha\mu_x - \beta)$$

$$\begin{cases} (\mu_y - \alpha\mu_x - \beta) = 0 \\ -\mu_{11} + \alpha\mu_{20} - \mu_x(\mu_y - \alpha\mu_x - \beta) = 0 \end{cases}$$

$$\begin{cases} \alpha = \frac{\mu_{11}}{\mu_{20}} \\ \beta = \mu_y - \frac{\mu_{11}}{\mu_{20}}\mu_x \end{cases}$$

Poichè si ha che

$$\mu_{11} = \rho\sigma_x\sigma_y \quad e \quad \mu_{20} = \sigma_x^2$$

allora possiamo scrivere in modo compatto le espressioni per α e β

$$\alpha = \rho \frac{\sigma_y}{\sigma_x} \quad \beta = \mu_y - \frac{\sigma_y}{\sigma_x} \rho \mu_x$$

ed infine l'equazione della migliore retta di regressione

$$y = \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) + \mu_y$$

o in modo equivalente

$$\frac{y - \mu_y}{\sigma_y} = \rho \frac{x - \mu_x}{\sigma_x}$$

Quindi la retta passa per il baricentro della distribuzione (μ_x, μ_y) ed il suo coefficiente angolare ha lo stesso segno del coefficiente di correlazione ρ

Vediamo ora di calcolare quanto vale, al minimo, *la somma dei quadrati delle proiezioni sull'asse Y delle distanze dei punti dalla migliore retta di regressione*. Sostituiamo ad α e β le formule riportate sopra ed otteniamo

$$E[(Y - \alpha_{min}X - \beta_{min})^2] = \frac{\mu_{02}\mu_{20} - \mu_{11}^2}{\mu_{20}} = \sigma_y^2(1 - \rho^2)$$

Un criterio alternativo per definire la migliore retta di regressione è quello di minimizzare *la somma dei quadrati delle proiezioni sull'asse X delle distanze dei punti dalla migliore retta di regressione*, partendo cioè dalla quantità

$$[G(\gamma, \delta)]_{min} = \{E[(X - \gamma Y - \delta)^2]\}_{min}$$

Seguendo un analogo sviluppo algebrico troveremo la retta definita dall'equazione

$$\frac{y - \mu_y}{\sigma_y} = \frac{1}{\rho} \frac{x - \mu_x}{\sigma_x}$$

che passa ancora per il baricentro (μ_x, μ_y) e per cui si ha

$$E \left[\left(X - \frac{1}{\alpha_{min}} Y + \frac{\beta}{\alpha_{min}} \right)^2 \right] = \sigma_x^2 (1 - \rho^2)$$

Per $\rho = \pm 1$ le due rette di regressione coincidono e la distribuzione dei punti del campo di probabilità giace esattamente sulla retta d'equazione

$$Y^* = \pm X^*$$

Per $\rho = 0$ le rette sono parallele agli assi e sono equazioni

$$y = \mu_x \quad , \quad x = \mu_y$$

LINEE DI REGRESSIONE DELLA MEDIA.

Estendiamo ora le considerazioni precedenti al caso di distribuzioni *subordinate* di probabilità tra due variabili aleatorie: in questo caso troveremo le così dette **Linee** di **Regressione** della **Media**

Supponiamo infatti che una delle due variabili assuma un ben preciso valore $X = x$. Dovremo allora considerare il valore medio subordinato di Y , subordinato cioè al fatto che la variabile aleatoria X assuma lo specifico valore x . Per definizione di valore aspettato subordinato avremo

$$\mu(Y|x) = E[Y|x] = \int_{-\infty}^{+\infty} yg(y|x)dy$$

dove

$$g(y|x) \equiv \frac{p(x, y)}{f(x)}$$

è la funzione densità di probabilità subordinata della variabile aleatoria Y che utilizzeremo come funzione peso per calcolare i vari momenti subordinati della distribuzione. Osserviamo che questi risultano tutti essere funzioni di x .

Se calcoliamo il valore aspettato medio del $\mu(Y|x)$ su tutti i possibili valori della x troveremo allora

$$E[\mu(Y|x)] = \int_{-\infty}^{+\infty} f(x) \int_{-\infty}^{+\infty} y \frac{p(x, y)}{f(x)} dx dy = \mu_y = E[Y]$$

$\mu(Y|x)$ ha il significato di parametro di localizzazione (baricentro della distribuzione) relativo a quel particolare valore x assunto dalla variabile aleatoria X .

Analogamente la varianza subordinata della Y è definita come

$$\sigma^2(Y|x) = Var[Y|x] = \int_{-\infty}^{+\infty} [y - \mu(Y|x)]^2 g(y|x) dy$$

ESEMPIO: consideriamo una distribuzione avente probabilità congiunta $p(x, y) = 8xy$ nell'intervallo $0 \leq x \leq 1$; $0 \leq y \leq x$. Applicando la definizione possiamo trovare l'espressione analitica delle funzioni marginali di probabilità per x e y . Otterremo

$$f(x) = 4x^3 \qquad g(y) = 4y$$

$$\mu(Y|x) = \int_{-0}^{+x} y \frac{2y}{x^2} dy = \int_{-0}^{+x} 2 \frac{y^2}{x^2} dx = \frac{2}{3}x$$

$$Var[Y|x] = \int_{-0}^{+x} \left(y - \frac{2}{3}x\right)^2 \frac{2y}{x^2} dx = \frac{x^2}{18}$$

Vediamo ora se $E[Y|x]$ è effettivamente una linea di regressione nel senso stabilito dal metodo dei minimi quadrati. A questo scopo consideriamo un'altra curva $I = h(x)$ tale che minimizzi la quantità:

$$E[\{Y - h(x)\}^2] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y - h(x))^2 p(x, y) dx dy$$

$$E[\{Y - h(x)\}^2|x] = \int_{-\infty}^{+\infty} f(x) \int_{-\infty}^{+\infty} [y - h(x)]^2 g(y|x) dy$$

Questa quantità rappresenta il momento del secondo ordine della distribuzione subordinata di Y rispetto al punto $h(x)$.

È semplice dimostrare che il minimo del momento del secondo ordine si ottiene solo se esso è calcolato rispetto al valore medio della distribuzione stessa (*proprietà fondamentale della varianza*). Infatti si ha

$$E[(X - c)^2] = E[(X - \mu + \mu - c)^2] = \mu_2 + (c - \mu)^2 \geq \mu^2$$

Dunque $E[\{y - h(x)\}^2]$ assume il valore minimo quando $h(x)$ tende al valore medio della distribuzione subordinata di Y . Analogamente alla linea di regressione definita da

$$E[Y|x] = \mu(y|x)$$

esisterà la linea di regressione della media

$$E[X|y] = \int_{-\infty}^{+\infty} x f(x|y) dx = \int_{-\infty}^{+\infty} x \frac{p(x, y)}{g(y)} dx = \mu(x|y)$$

ed in genere

$$\mu(Y|x) \neq \mu(X|y)$$

Se le linee di regressione sono **rette**, **queste coincidono con le rette ottenute con il metodo dei minimi quadrati**. In tal caso si dice che la distribuzione di probabilità bivariata ha una *regressione lineare*.

Nel caso in cui le curve di regressione non siano rette, si può stabilire un parametro che stabilisca il grado di

concentrazione della distribuzione di probabilità intorno ad esse. A questo scopo, poiché

$$Var[Y] = E[\{Y - \mu_y + \mu(Y|x) - \mu(Y|x)\}^2]$$

che sviluppato porta alla espressione

$$Var[Y] = Var[Y|x] + E[\{\mu(Y|x) - \mu_y\}^2]$$

da cui segue che

$$Var[Y] - Var[Y|x] = E[\{\mu(Y|x) - \mu_y\}^2]$$

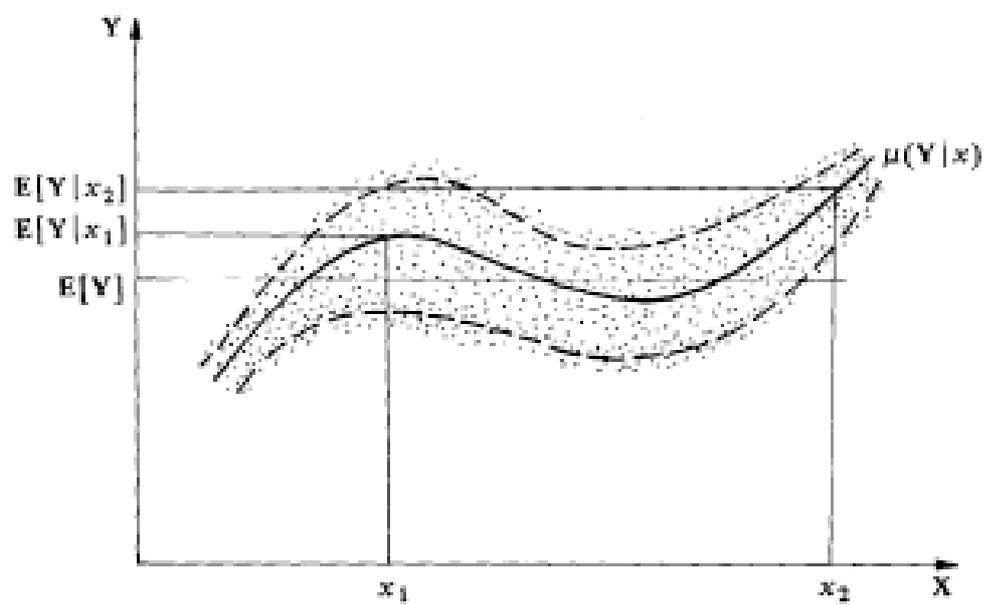
definiamo la quantità η **rapporto di correlazione** di Y su X

$$\eta^2 = 1 - \frac{Var[Y|x]}{Var[Y]} = \frac{E[\{\mu(Y|x) - \mu_y\}^2]}{Var[Y]}$$

che risulta essere $|\eta| \leq 1$

a) $\eta \pm 1$ quando $Var[Y|x] = 0$ (distribuzione tutta sulla linea di regressione della media)

b) $\eta = 0$ quando $E[\{\mu(Y|x) - \mu_y\}^2] = 0$ cioè $\mu(Y|x) = \mu_y$ quindi nel caso in cui le variabili X e Y sono indipendenti.



UN CASO DI GRANDE IMPORTANZA:
LA DISTRIBUZIONE NORMALE BIVARIATA.

Siano X ed Y due variabili aleatorie aventi distribuzioni marginali di probabilità **normali**.

Nel caso in cui X ed Y siano indipendenti, la probabilità congiunta è

$$p(x, y) = \frac{\exp \left[-\frac{1}{2} \left\{ \left(\frac{x - \mu_x}{\sigma_x} \right)^2 + \left(\frac{y - \mu_y}{\sigma_y} \right)^2 \right\} \right]}{2 \pi \sigma_x \sigma_y}$$

Nel caso in cui X ed Y sono **normali** ma dipendenti

$$p(x, y) = \frac{1}{d} e^{-(ax^2 + bxy + cy^2)}$$

dove le quantità a, b, c, d dobbiamo definirle

- sulla base dei parametri noti delle distribuzioni marginali delle variabili aleatorie X ed Y

$$\sigma_x^2 = \frac{1}{d} \int \int_{-\infty}^{+\infty} (x - \mu_x)^2 e^{-(ax^2 + bxy + cy^2)} dx dy$$

$$\sigma_y^2 = \frac{1}{d} \int \int_{-\infty}^{+\infty} (y - \mu_y)^2 e^{-(ax^2 + bxy + cy^2)} dx dy$$

$$\text{Cov}[X, Y] = E[(X - \mu_x)(Y - \mu_y)] = \sigma_x \sigma_y \rho$$

ovvero

$$\sigma_x \sigma_y \rho = \frac{1}{d} \int \int_{-\infty}^{+\infty} (x - \mu_x)(y - \mu_y) e^{-(ax^2 + bxy + cy^2)} dx dy$$

- sulla base della relazione di normalizzazione

$$\frac{1}{d} \int \int_{-\infty}^{+\infty} e^{-(ax^2+bxy+cy^2)} dx dy = 1$$

Risolvendo il sistema si ottiene

$$a = c = \frac{1}{2(1-\rho^2)} \quad b = -\frac{\rho}{(1-\rho^2)}$$

$$d = 2 \pi \sqrt{1-\rho^2} \sigma_x \sigma_y$$

Possiamo quindi esplicitare la densità di probabilità congiunta di una distribuzione bivariata gaussiana:

$$p(x, y) = \frac{1}{2 \pi \sigma_x \sigma_y \sqrt{1-\rho^2}} \cdot \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{(x-\mu_x)^2}{\sigma_x^2} - 2 \rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right\} \right]$$

Per verificare che le densità marginali sono normali, basta ad esempio sommare e sottrarre $\rho^2 x^2$ all'esponente e calcolare $\int_{-\infty}^{+\infty} p(x, y) dy = \frac{e^{-x^2/2}}{\sqrt{2} \pi}$.

Si vede anche che per $\rho = 0$ la funzione $p(x, y)$ si riduce al prodotto di due distribuzioni normali. Quindi *nel caso delle distribuzioni normali, la condizione $\rho = 0$ non solo è necessaria (condizione vera per qualunque distribuzione) ma è anche sufficiente* affinché le variabili aleatorie siano *indipendenti*.

Vediamo allora di calcolare anche le densità di probabilità marginali e poi di ottenere quelle condizionate. Consideriamo per semplicità il caso di variabili normali ridotte.

$$f(x) = \int_{-\infty}^{+\infty} p(x, y) dy = \frac{1}{2 \pi \sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2 \rho xy + y^2)}$$

Aggiungiamo e togliamo all'esponente un fattore pari a $\rho^2 x^2$

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{2 \pi \sqrt{1 - \rho^2}} \int_{-\infty}^{+\infty} e^{-\frac{(y - \rho x)^2}{2(1 - \rho^2)}} dy$$

Poniamo $Z = (y - \rho x) / \sqrt{1 - \rho^2}$ e di conseguenza $dy = \sqrt{1 - \rho^2} dz$

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{2 \pi} \int_{-\infty}^{+\infty} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Analogamente si ricava per $g(y)$. Questo dimostra che le distribuzioni di probabilità marginali sono normali anche per $\rho \neq 0$.

Abbiamo già ricordato che la funzione densità di probabilità condizionata in generale è definita

$$g(y|x) = \frac{p(x, y)}{f(x)}$$

Per variabili aleatorie non ridotte avremo che

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x - \mu_x)^2}{\sigma_x^2}}$$

per cui

$$g(y|x) = \frac{1}{2 \pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} e^{-\frac{1}{2} Q(x, y)} \cdot \left(\sqrt{2\pi} \sigma_x \cdot e^{\frac{1}{2} \frac{(x - \mu_x)^2}{\sigma_x^2}} \right)$$

dove $Q(x, y)$ è pari a

$$Q = \frac{1}{1 - \rho^2} \left[\frac{(x - \mu_x)^2}{\sigma_x^2} - 2 \rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right]$$

Riscriviamo allora la funzione $g(y|x)$ in modo più compatto

$$g(y|x) = \frac{1}{\sqrt{2\pi}(1-\rho^2)\sigma_y} e^{-\frac{1}{2} \frac{F(x,y)}{1-\rho^2}}$$

avendo introdotto l'argomento dell'esponenziale $F(x, y) / [2(1-\rho)]$

$$F(x, y) = \rho^2 \frac{(x - \mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x - \mu_x)}{\sigma_x} \frac{(y - \mu_y)}{\sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} =$$

o che riconosciamo essere semplicemente

$$F(x, y) = \left(\frac{(y - \mu_y)}{\sigma_y} - \rho \frac{(x - \mu_x)}{\sigma_x} \right)^2$$

Da questa relazione riconosciamo che la distribuzione di probabilità condizionata è ancora una gaussiana di valore atteso condizionato

$$E[Y|x] = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x)$$

e varianza

$$Var[Y|x] = \sigma_y^2 (1 - \rho^2)$$

Notiamo che $E[Y|x]$ è la funzione lineare che abbiamo ricavato nel precedente paragrafo. Concludiamo allora che la distribuzione normale bivariata è un caso in cui si ha regressione lineare con retta d'equazione

$$y - \mu_y = \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x)$$

Infine studiamo quale sia il luogo dei punti per cui si ha $p(x, y) = costante = Z_o$: tale luogo dei punti è una curva di equiprobabilità

Questo equivale ad imporre

$$Q(x, y) = -2 \ln \left[Z_o 2 \pi \sigma_x \sigma_y \sqrt{1 - \rho^2} \right] = \text{costante}$$

Poichè $Q(x, y) \geq 0$, allora abbiamo che

$$0 < \ln \left[Z_o 2 \pi \sigma_x \sigma_y \sqrt{1 - \rho^2} \right] \leq 1$$

Ne segue necessariamente che si deve avere

$$0 < Z_o \leq \frac{1}{2 \pi \sigma_x \sigma_y \sqrt{1 - \rho^2}}$$

Osserviamo che per $x = \mu_x$ e $y = \mu_y$, si ha la probabilità massima

$$Z_{max} = \frac{1}{2 \pi \sigma_x \sigma_y \sqrt{1 - \rho^2}}$$

Inoltre la curva $Z_o = p(x, y)$ è un'ellisse il cui centro cade in μ_x, μ_y e la cui eccentricità δ è

$$\delta = 2 \frac{|\rho|}{1 + |\rho|}$$

Concludiamo notando che

- per $\rho = 0$ ($\delta = 0$) l'ellisse è un cerchio e le variabili indipendenti

- per $\rho = \pm 1$ ($\delta = 0$) l'ellisse collassa in una linea retta: quindi la distribuzione collassa lungo la retta di regressione.

