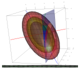
Modelli Lineari
(ed altro)

Laurea in Statistica per l'Analisi dei dati

Marcello Chiodi
marcello.chiodi@unipa.it

Dipartimento di Scienze Economiche, Aziendali e Statistiche (DSEAS)
Università di Palermo

Palermo, 2021



Indice

La distribuzione normale bivariata e multivariata

- La distribuzione normale multivariata svolge un ruolo centrale in statistica ed in particolare nell'analisi dei **modelli lineari** :
- E' indispensabile per potere studiare l'inferenza esatta nei modelli lineari classici
- è una *bella* teoria (opinione personale)
- Si inizierà con l'introduzione della distribuzione normale bivariata e poi con quella multivariata

- Lo studio della distribuzione normale bivariata ovviamente implica la conoscenza della distribuzione normale univariata e le nozioni di indipendenza di variabili aleatorie e di funzione di regressione
- Lo studio della normale multivariata necessita di una conoscenza solida di algebra vettoriale e matriciale
- La distribuzione normale multivariata rappresenta un modello di base per lo studio di relazioni fra variabili con funzioni di regressione lineari
- Le proprietà sulle forme quadratiche in variabili normali hanno un ruolo fondamentale nello sviluppo della teoria dell'inferenza esatta nei modelli lineari

La distribuzione normale bivariata

- Come estendere la **distribuzione normale univariata** ad una variabile doppia?
- Intanto iniziamo con una generalizzazione semplice a componenti indipendenti.
- Supponiamo di avere due variabili aleatorie X_1 e X_2 indipendenti e distribuite normalmente con speranze matematiche μ_1, μ_2 e varianze σ_1^2 e σ_2^2:

$$f_{X_1}(x_1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}, \quad f_{X_2}(x_2) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \tag{1}$$

$$\text{con:} \quad \{x_1, x_2\} \in \Re^2, \quad \{\mu_1, \mu_2\} \in \Re^2, \quad \sigma_1 > 0, \quad \sigma_2 > 0$$

Normale bivariata a componenti indipendenti

- e quindi se X_1 e X_2 sono indipendenti, la distribuzione congiunta delle due variabili normali ha densità data da:

$$f_{X_1,X_2}(x_1, x_2) \;=\; f_{X_1}(x_1) f_{X_2}(x_2) =$$

$$\frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{1}{2} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}. \tag{2}$$

- Ovviamente questo modello a componenti indipendenti non può descrivere il comportamento simultaneo di variabili correlate;
- Occorre quindi estendere il modello precedente introducendo la possibilità di correlazione fra le due variabili.

Normale bivariata a componenti correlate

- Densità di una variabile aleatoria $X = (X_1, X_2)$ con distribuzione normale bivariata:

Densità della Normale bivariata a componenti correlate

$$f_{X_1,X_2}(x_1, x_2) = \\ = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \; e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]} \tag{3}$$

con: $\{x_1, x_2\} \in \Re^2,\ \{\mu_1, \mu_2\} \in \Re^2,\ \sigma_1 > 0,\ \sigma_2 > 0,\ |\rho| < 1$

- Ho utilizzato (X_1, X_2) , invece di (X, Y) .
- Uso la parametrizzazione classica con la correlazione ρ (e non con la covarianza σ_{12} , usata invece nella normale multivariata).

Distribuzioni marginali della normale bivariata

- Le distribuzioni marginali sono palesemente ancora normali con densità date da:

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\} \qquad i=1,2$$

- Il risultato di ottiene integrando rispetto a ciascuna delle variabili (operazione non velocissima).
- Meglio ottenere il risultato dalla funzione caratteristica (il caso generale verrà ripreso per la normale multivariata).
- Per ora può bastare prenderlo come risultato dimostrato.

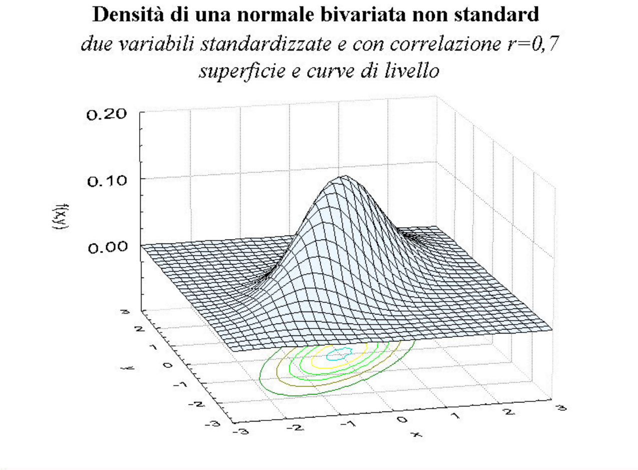


Figure: Densità di una normale bivariata

Momenti della normale bivariata

I primi due momenti identificano completamente la distribuzione, in quanto si ha:

Momenti della normale bivariata

Speranze matematiche:	$\mathrm{E}[X_1] = \mu_1,$	$\mathrm{E}[X_2] = \mu_2$
Varianze:	$\mathrm{V}[X_1] = \sigma_1^2,$	$\mathrm{V}[X_2] = \sigma_2^2$
Covarianza	$\mathrm{Cov}[X_1, X_2] =$	$\rho\sigma_1\sigma_2.$

In termini matriciali:

$$\mathrm{E}[\mathrm{X}] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \mathrm{V}[\mathrm{X}] = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Correlazione nella normale bivariata

Correlazione nella normale bivariata

Per cui la correlazione lineare fra le due componenti X_1 e X_2 è data da ρ; infatti:

$$\mathrm{Corr}\left[X_1, X_2\right] = \frac{\mathrm{Cov}\left[X_1, X_2\right]}{\sqrt{\mathrm{V}\left[X_1\right]\mathrm{V}\left[X_2\right]}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho.$$

Non correlazione $\Longleftrightarrow$ indipendenza nella normale bivariata

Si ha l'importantissima proprietà:

Non correlazione $\Longleftrightarrow$ indipendenza nella normale bivariata

In una normale bivariata:

$$X_1 \perp X_2 \Longleftrightarrow \rho = 0$$

ossia **l'assenza di correlazione lineare implica l'indipendenza, per due variabili con distribuzione congiunta normale bivariata** .

Cosa implica la correlazione nulla nella normale bivariata?

Dalla (3) si vede che la condizione $\rho = 0$ è necessaria e sufficiente per la fattorizzazione della densità nelle due densità marginali, in modo da avere l'indipendenza [1]:

$$
\begin{aligned}
\rho = 0 \qquad & \Rightarrow \\
f_{X_1,X_2}(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \times \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} = \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{2\sigma_1}\right)^2\right\} \times \\
& \quad \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{2\sigma_2}\right)^2\right\} = \\
&= f_{X_1}(x_1) \times f_{X_2}(x_2)
\end{aligned}
$$

[1] Superfluo ricordare che invece l'indipendenza implica sempre l'assenza di correlazione lineare, per qualsiasi variabile aleatoria doppia

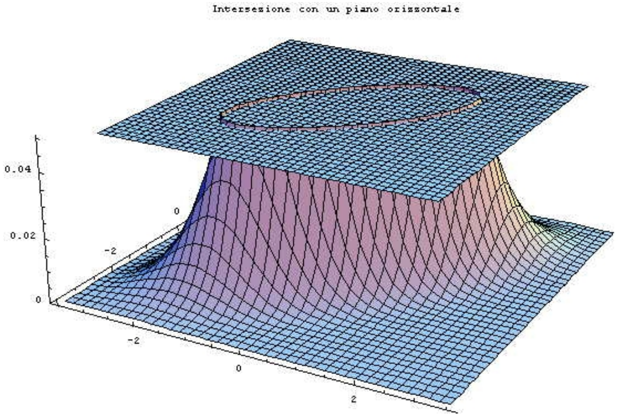


Figure: intersezioni con la normale bivariata

Ellissi di equiprobabilità

Un'intersezione (fig. 2) della densità bivariata con un piano orizzontale di equazione $z = k$ ($k \geq 0$) è un contorno ellissoidale. Risolvendo l'equazione $f(x_1, x_2) = k$ (e passando ai logaritmi):

$$f(x_1, x_2) = k \qquad \Rightarrow$$

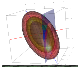
$$-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] =$$

$$= \log\left\{k\, 2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right\}; \tag{4}$$

che porta, in pochissimi elementari passaggi, alla relazione:

$$\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 = \tag{5}$$

$$= \quad -2(1-\rho^2)\log\left\{k\, 2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right\}$$



Fome quadratiche ed ellisi di equiprobabilità

Il membro sinistro della relazione (5) è una forma quadratica che può essere espressa in una forma più compatta:

$$\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2=$$

$$=\begin{pmatrix}\dfrac{x_1-\mu_1}{\sigma_1}\\ \dfrac{x_2-\mu_2}{\sigma_2}\end{pmatrix}^{\mathsf{T}}\begin{pmatrix}1 & -\rho\\ -\rho & 1\end{pmatrix}\begin{pmatrix}\dfrac{x_1-\mu_1}{\sigma_1}\\ \dfrac{x_2-\mu_2}{\sigma_2}\end{pmatrix} \tag{6}$$

quindi la (5) è l'equazione di un ellisse di centro in (μ_1, μ_2)

Distribuzioni condizionate nella normale bivariata

La densità di X_1, condizionatamente ai valori x_2 assunti da X_2, si ha da:

$$f_{X_1}(x_1|X_2 = x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}; \tag{7}$$

Passando ai logaritmi: $z_1 = \frac{x_1 - \mu_1}{\sigma_1}$; $\quad z_2 = \frac{x_2 - \mu_2}{\sigma_2}$ la (7) diventa:

$$\log(f_{X_1}(x_1|X_2 = x_2)) = \log(f_{X_1,X_2}(x_1, x_2)) - \log(f_{X_2}(x_2)) =$$

$$= \log\left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right) - \frac{\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)}{2(1-\rho^2)} - \log\left(\frac{1}{\sqrt{2\pi}\sigma_2}\right) + \frac{1}{2}z_2^2 =$$

$$= \log\left(\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\right) - \frac{\left(z_1^2 - 2\rho z_1 z_2 + z_2^2 - (1-\rho^2)z_2^2\right)}{2(1-\rho^2)} =$$

$$= \log\left(\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\right) - \frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + \rho^2 z_2^2\right) =$$

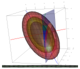
$$= \log\left(\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\right) - \frac{1}{2}\frac{(z_1 - \rho z_2)^2}{1-\rho^2}$$

Distribuioni condizionate - cont.

Passiamo dal logaritmo $\log(f_{X_1}(x_1|X_2 = x_2))$ alla densità $f_{X_1}(x_1|X_2 = x_2)$, ritorniamo dalle variabili z_1, z_2 alle variabili originali x_1, x_2:

$$
\begin{aligned}
f_{X_1}(x_1|X_2 = x_2) &= \exp\left(\log\left(f_{X_1}(x_1|X_2 = x_2)\right)\right) = \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2} \frac{\left(\frac{x_1-\mu_1}{\sigma_1} - \rho\frac{x_2-\mu_2}{\sigma_2}\right)^2}{1-\rho^2} \right\} = \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2} \frac{\left[x_1 - \mu_1 - \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right]^2}{\sigma_1^2(1-\rho^2)} \right\}. \qquad (8)
\end{aligned}
$$

(dopo qualche passaggio entro l'esponente). Che densità è?
Quella di una normale con valore atteso $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$ e varianza $\sigma_1^2(1-\rho^2)$, ossia:



Distribuioni condizionate - cont.

Distribuzioni condizionate nella normale bivariata

$$X_1 | \{X_2 = x_2\} \sim \mathcal{N}(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),\ \sigma_1^2(1 - \rho^2)) \quad (9)$$

1. La speranza matematica della distribuzione condizionata di X_1 varia **linearmente** in funzione di x_2 (con pendenza data da $\rho\frac{\sigma_1}{\sigma_2}$)
2. La varianza è costante (**omoschedasticità**) e la sua riduzione è proporzionale a ρ^2 (**coefficiente di determinazione**)
3. La distribuzione condizionata è **normale** ;

Si noti l'analogia con le assunzioni del modello di regressione lineare semplice (e poi del modello lineare generale)
Ovviamente si ha anche:

$$X_2 | \{X_1 = x_1\} \sim \mathcal{N}(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1),\ \sigma_2^2(1 - \rho^2)) \quad (10)$$

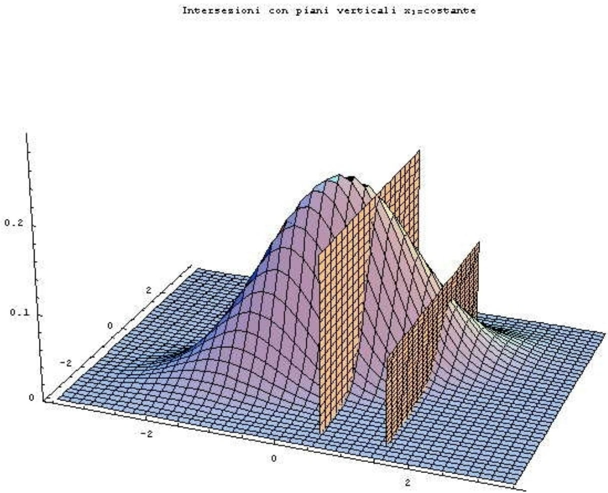


Figure: Distribuzioni Condizionate: intersezioni di piani verticali perpendicolare alla variabile *condizionante* $X_2 = x_2$ con la densità normale bivariata

Casi limite

- se $\rho = 0$: la conoscenza di X_2 è ininfluente sulla determinazione di X_1 (è il caso di indipendenza)
- se $\rho^2 = 1$: la conoscenza di X_2 determina senza incertezza X_1: è il caso degenere di due variabili linearmente dipendenti in modo esatto, che danno luogo ad una distribuzione degenere che concentra la massa su una retta (e non è dotata di densità congiunta) [2]; si vede facilmente perchè la distribuzione condizionata (di X_1 a X_2, ma si può fare il ragionamento anche sull'altra distribuzione condizionata, di X_2 a X_1) ha varianza nulla attorno ad una retta di equazione:

$$x_1 = \mu_1 + \operatorname{sign}(\rho) \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

[2] Ricordo che due variabili casuali X, Y possono benissimo non avere una distribuzione congiunta dotata di densità, anche se sono dotate di densità le rispettive distribuzioni marginali: si pensi a tutti i casi in cui $Y = f(X)$, con $f(\cdot)$ monotona

Il coefficiente di determinazione

Significato del quadrato dell'indice di correlazione lineare semplice (ρ)

Il coefficiente di determinazione ρ^2

Dall'ultima proprietà risulta chiara l'interpretazione del coefficiente di determinazione ρ^2: indica di *quanto la conoscenza di X_2 aumenta la nostra informazione su X_1* (perchè indica di quanto si riduce la varianza di X_1 in virtù del condizionamento ad un particolare valore $X_2 = x_2$).

Vale la pena di notare che è irrilevante a quale particolare valore di X_2 ci si condiziona: la riduzione di varianza è sempre proporzionale a ρ^2.
La stessa spiegazione si ottiene per X_2 in funzione di X_1: la riduzione della varianza di X_2 in virtù del condizionamento ad un particolare valore $X_1 = x_1$ è sempre proporzionale a ρ^2.

Funzioni di regressione nella normale bivariata

La speranza matematica della distribuzione di X_1, condizionata ai valori di $X_2 = x_2$, ossia $\mathrm{E}\left[X_1 | X_2 = x_2\right]$, va sotto il nome di *funzione di regressione*.

Nella normale bivariata le funzioni di regressione sono lineari

Nel nostro caso *normale bivariato*:

$$\mathrm{E}\left[X_1 | X_2 = x_2\right] = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

e quindi è una funzione lineare di x_2.

Si ha anche l'importante risultato:

$$\mathrm{V}\left[X_1 | X_2 = x_2\right] = \sigma_1^2(1 - \rho^2) \qquad \forall\, x_2$$

Distribuzione di X_2 condizionatamente a $X_1 = x_1$

Per la distribuzione di X_2 condizionatamente a $X_1 = x_1$; è sufficente scambiare gli indici "1" e "2" nelle formule.
La funzione di regressione è allora:

$$\mathrm{E}\left[X_2 | X_1 = x_1\right] = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1).$$

La riduzione nella varianza di X_2 è sempre la stessa, ossia ρ^2:

$$\mathrm{V}\left[X_2 | X_1 = x_1\right] = \sigma_2^2(1 - \rho^2) \qquad \forall\, x_1$$

Entrambe le rette di regressione passano per il punto di coordinate $\{\mu_1, \mu_2\}$, e formano un angolo che si annulla quando $|\rho| = 1$.
Sono ortogonali e parallele agli assi coordinati se (e solo se) $\rho = 0$.

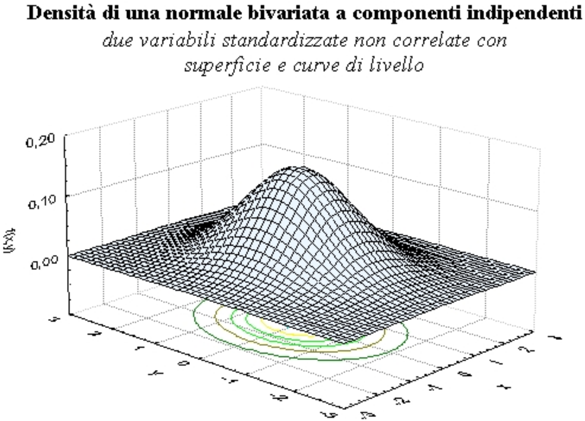


Figure: Variabili normali standardizzate indipendenti

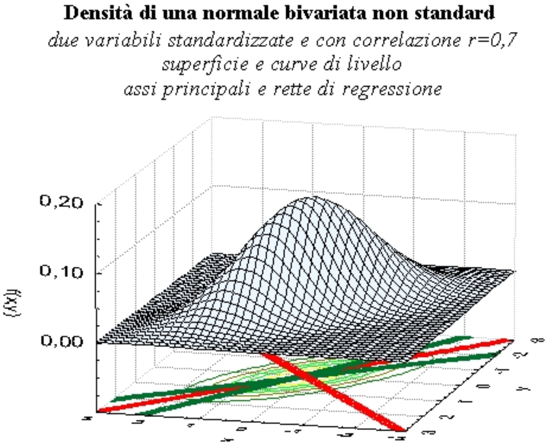


Figure: Variabili normali standardizzate con correlazione 0.7

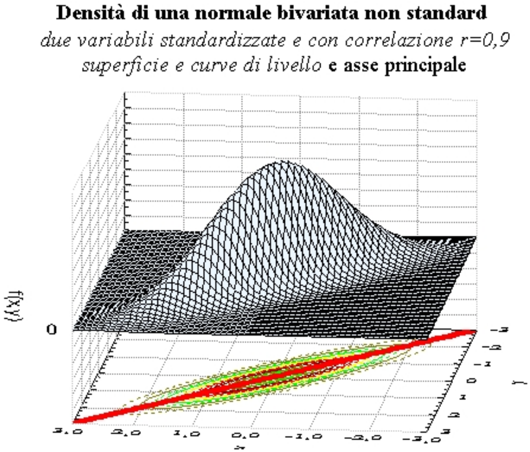


Figure: Variabili normali standardizzate con correlazione 0.9

Probabilità relativa al primo quadrante in funzione di ρ

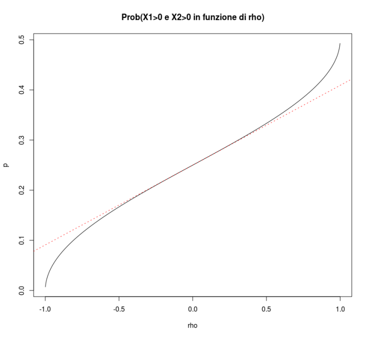
$\text{Prob}\{X_1 \geq \mu_1 \cap X_2 \geq \mu_2\}$

Qual è la probabilità di osservare un'unità con valori superiori alla media per entrambe le componenti X_1 e X_2? Ossia qual è il volume sotteso dalla densità nel primo quadrante?

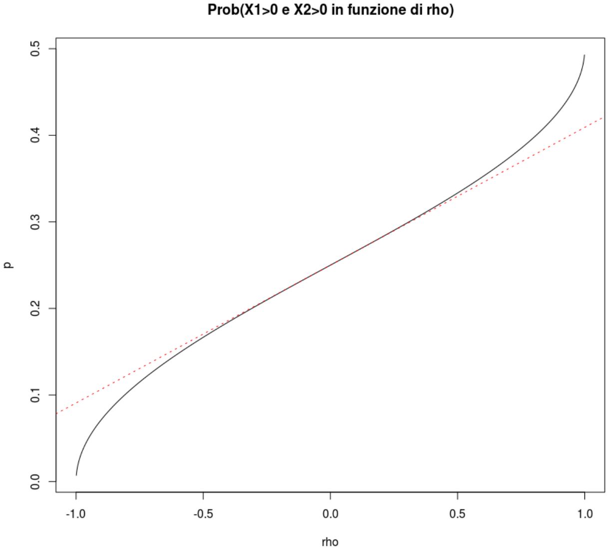
$$\text{Prob}\{X_1 \geq \mu_1 \cap X_2 \geq \mu_2\} =$$

$$\int_{\mu_1}^{+\infty} \int_{\mu_2}^{+\infty} f(x_1, x_2) dx_1 \; dx_2$$

Probabilità integrale relativa al primo quadrante in funzione di ρ



Probabilità relativa al primo quadrante in funzione di ρ



Questa probabilità è $\frac{1}{4}$ se $\rho = 0$; è maggiore o minore di $\frac{1}{4}$ secondo il segno di ρ ed è funzione monotona di ρ, come si vede dalla figura 7
Si vede dalla figura che per valori di $|\rho| < \frac{1}{2}$ l'andamento di tale funzione è pressocchè lineare: infatti la tangente alla curva nel punto $\rho = 0$ si sovrappone benissimo alla curva nel tratto centrale.
Dal punto di vista pratico si pensi ad un campione di n osservazioni estratto da una normale bivariata. Suddividendo le osservazioni in quattro classi secondo il fatto che ciascuna delle due variabili sia maggiore o minore della mediana (in pratica dicotomizzando le variabili in due categorie: alto/basso) otteniamo una classica tabella di associazione 2×2 :

	$X_2 < Me_2$	$X_2 \geq Me_2$	$tot.$
$X_1 < Me_1$	a	b	$\frac{n}{2}$
$X_1 \geq Me_1$	b	a	$\frac{n}{2}$
$tot.$	$\frac{n}{2}$	$\frac{n}{2}$	n

$\frac{a}{n}$ è una stima campionaria di tale probabilità.