Contents

List of Figures

Chapter 1

Avvertenza

1.1 Altro materiale (sul web)

Per alcuni materiali didattici relativi a questo argomento dovreste trovare sul web (lì dove avete trovato questo scritto...)

- Un mio software didattico MLANP
- Notebook didattici con diversi esempi
- Su Microsoft Stream sono presenti le lezioni dei miei corsi dell'a.a. 2019-2020 (quello del covid19, e se state leggendo vuol dire che le cose in qualche modo sono andate...). Credo siano accessibili solo per gli utenti `unipa`.

1.2 Nota esclusiva per gli studenti dei miei corsi

Per gli studenti del terzo corso di Statistica del corso di laurea in Statistica per l'Analisi dei Dati

Gli appunti di questa sezione relativi alla normale bivariata e, in particolare, quelli sulla normale multivariata, sono molto abbondanti rispetto a quanto effettivamente riportato sulla scheda trasparenza e a quanto fatto in aula (*e a quanto sarà utile ai fini dell'esame*).

Tuttavia, almeno ad oggi , ho preferito lasciare tutto il materiale, anche quello non finalizzato al nostro corso. Chiariremo in aula, e in una successiva redazione di questi appunti, le parti essenziali.

Aggiungo con sincerità che ogni anno mi chiedo se è il caso di eliminare del tutto questa parte sulla normale bivariata e multivariata da questo corso e ogni anno, fino ad ora, mi sono sempre risposto di no!

Chapter 2

La distribuzione normale bivariata e multivariata.

href in ingresso con normale:biv

La distribuzione normale multivariata svolge un ruolo centrale in statistica ed in particolare nell'analisi dei **modelli lineari** :

Normale bivariata e multivariata nei modelli lineari

La conoscenza di tale modello e delle sue proprietà è indispensabile per potere studiare l'inferenza esatta nei modelli lineari classici, che sono il tema centrale delle mie note (ovunque le stiate leggendo in questo momento: sul web, su carta, etc.).
Inoltre, ma è un'opinione strettamente personale, è una bella teoria, completa, con delle parti formalmente molto interessanti.

I modelli lineari (regressione lineare semplice, multipla, analisi della varianza, etc...) si possono anche studiare senza assumere la normalità della componente accidentale, ma verrebbero a cadere alcune proprietà delle distribuzioni campionarie, e cadrebbero anche delle caratteristiche di ottimalità.

Il capitolo inizia con la presentazione della normale bivariata, come estensione della normale bivariata a componenti indipendenti.

Dopo passo all'estensione multivariata, che ricavo come distribuzione di combinazioni lineari di variabili normali indipendenti, usando noti teoremi di calcolo delle probabilità per vettori aleatori e con una notazione vettoriale compatta [1] link da creare con matrici2020

Personalmente ritengo che si possa passare dallo studio della normale univariata direttamente alla distribuzione normale multivariata, con l'opportuno bagaglio di conoscenza di algebra matriciale, tuttavia riporto comunque inizialmente la distribuzione bivariata, che risulta immediatamente accessibile senza utilizzare una

[1] gli strumenti tecnici eventualmente necessari per affrontare questa parte sono richiamati nell'appendice sul calcolo matriciale e sui vettori aleatori

particolare strumentazione di calcolo vettoriale e matriciale, e permette oltretutto di visualizzare i risultati graficamente, cosa che agevola la comprensione di molti concetti in modo intuitivo. Si apprezzeranno così i risultati sulla dipendenza e l'interdipendenza a due variabili, sulle funzioni di regressione, prima di passare alle strutture a tre e più variabili.

Le proprietà sulle forme quadratiche in variabili normali rendono poi molto scorrevole lo studio delle distribuzioni campionarie esatte degli stimatori del modello lineare almeno nei casi standard.

2.1 Dalla distribuzione normale univariata alla distribuzione normale bivariata

La distribuzione normale univariata svolge un ruolo centrale in calcolo delle probabilità ed in statistica: è la distribuzione limite di alcuni schemi teorici (per esempio quello bernoulliano), è la distribuzione limite della media standardizzata di una successione di variabili aleatorie con media uguale e varianza finita; la distribuzione normale è anche una distribuzione impiegata per la descrizione di distribuzioni osservate di variabili statistiche quantitative.

Adesso supponiamo di avere osservato due fenomeni quantitativi sulle stesse unità, e che su ciascuna delle distribuzioni marginali sia ragionevole adattare una distribuzione normale; ci chiediamo quale modello sia possibile adattare alla distribuzione congiunta dei due caratteri quantitativi. Ad esempio possiamo avere osservato che la distribuzione normale si adatta discretamente a descrivere le altezze di un gruppo di soggetti come pure a descrivere la distribuzione dei pesi degli stessi soggetti: esiste un modello che permette di ottenere un adattamento soddisfacente alla distribuzione congiunta osservata delle due variabili?

O esiste un modello che descrive le altezze dei padri e le altezze dei figli, come si chiese Galton?

Nei complementi di questo capitolo farò un cenno alla distribuzione normale bivariata come limite della distribuzione trinomiale, e poi alla normale multivariata come limite della distribuzione multinomiale (non essenziali per lo sviluppo del discorso generale).

Possiamo chiederci in che modo si può estendere la *distribuzione normale univariata* ad una variabile doppia: iniziamo con una generalizzazione semplice e 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, σ_2^2. Per le densità delle due distribuzioni si ha dunque:

distribuzione normale univariata

$$f_{X_1}(x_1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}},$$

$$f_{X_2}(x_2) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \tag{2.1}$$

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

e quindi se X_1 e X_2 sono indipendenti, la distribuzione congiunta delle due variabili normali ha densità data da: href in ingresso con label presente

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

Normale bivariata a componenti indipendenti

$$= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}. \tag{2.2}$$

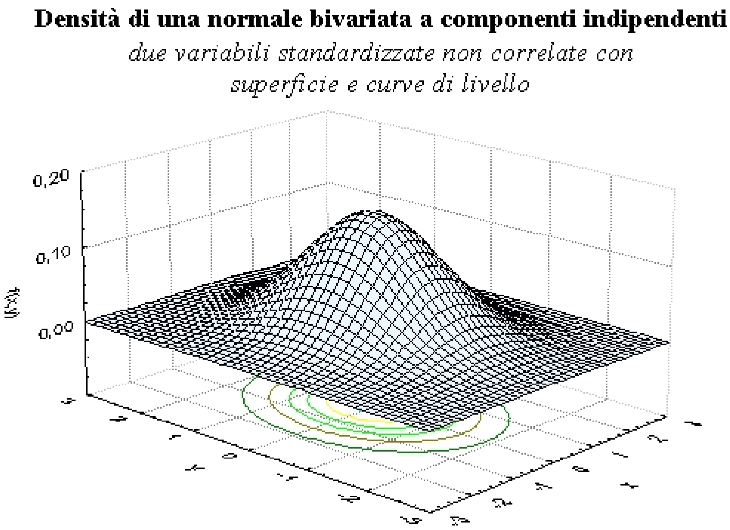

Figure 2.1: Variabili normali standardizzate indipendenti

chiaramente questo modello a componenti indipendenti non si presta alla descrizione del comportamento simultaneo di variabili in qualche modo correlate; negli esempi riportati prima, è evidente che le altezze e i pesi sono correlati; anche le due componenti di una variabile aleatoria trinomiale sono correlate.

Occorrerà quindi generalizzare in qualche modo il modello precedente introducendo almeno un parametro che tenga conto della correlazione fra le due variabili.

2.2 La densità della distribuzione normale bivariata

La densità di una variabile aleatoria $\mathbf{X} = (X_1, X_2)$ con distribuzione normale bivariata è data da: href in ingresso con label presente
href in ingresso con label presente

Normale bivariata a componenti correlate

$$f_{X_1,X_2}(x_1,x_2) = \\ = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]} \tag{2.3}$$

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

Ho utilizzato la coppia di variabili (X_1, X_2) , invece di (X, Y) , perchè questo renderà più semplice poi il passaggio alla normale multivariata, in cui useremo $\{X_1, X_2, \ldots, X_i, \ldots, X_p\}$. Tuttavia nella normale bivariata ho mantenuto la parametrizzazione classica con la correlazione ρ piuttosto che con la covarianza σ_{12} , che userò invece nella normale multivariata.

Distribuzioni marginali della normale bivariata

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

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

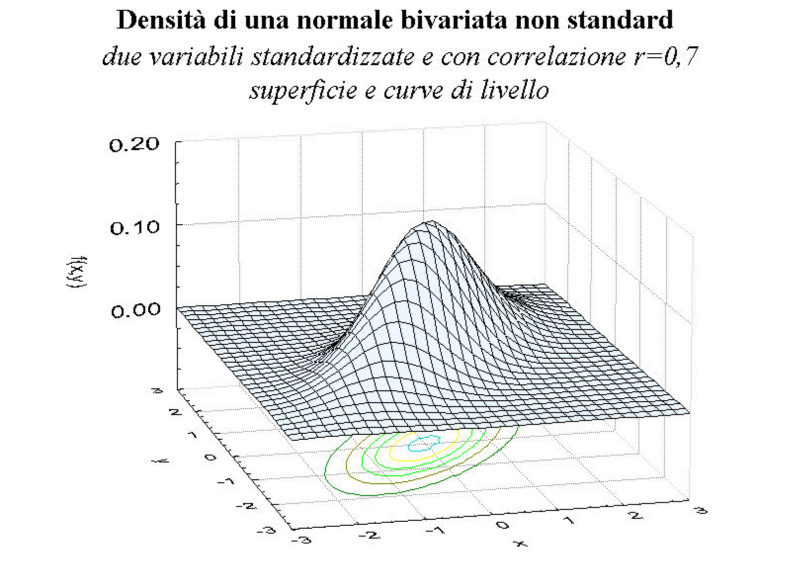

Figure 2.2: Densità di una normale bivariata

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

Momenti della normale bivariata

$$\begin{array}{lll} \text{Speranze matematiche:} & \mathrm{E}\left[X_1\right] = \mu_1, & \mathrm{E}\left[X_2\right] = \mu_2 \\ \text{Varianze:} & \mathrm{V}\left[X_1\right] = \sigma_1^2, & \mathrm{V}\left[X_2\right] = \sigma_2^2 \\ \text{Covarianza:} & \mathrm{Cov}\left[X_1, X_2\right] = & \rho\sigma_1\sigma_2. \end{array}$$

In termini matriciali:

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

href in ingresso con correlazione

Correlazione nella normale bivariata

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

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

2.2.1 Normale bivariata a componenti indipendenti

Si ha l'importantissima proprietà:

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** .

Dalla densità (2.3) si vede immediatamente che la condizione $\rho = 0$ è necessaria e sufficiente per la fattorizzazione della densità nelle due densità marginale, in modo da avere l'indipendenza [2]:

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

2.2.2 Ellissi di concentrazione nella normale bivariata

Si guardi la figura 2.3: un'intersezione della densità bivariata con un piano orizzontale di equazione $z = k$ è un contorno ellissoidale (in uno spazio di riferimento (x_1, x_2, z)) come si vede facilmente risolvendo l'equazione $f(x_1, x_2) = k$ (passando ai logaritmi):

$$
\begin{aligned}
& f(x_1,x_2) = k \qquad \Rightarrow \\
& -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] = \\
& = \log\left\{k\, 2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right\}; \qquad (2.4)
\end{aligned}
$$

che porta, in pochissimi elementari passaggi, alla relazione:

$$
\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 = \qquad (2.5)
$$

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

E' facile vedere che il membro sinistro della relazione (2.4) è una forma quadratica che può essere espressa in una forma più compatta:

$$
\begin{pmatrix} \dfrac{x_1-\mu_1}{\sigma_1} \\ \dfrac{x_2-\mu_2}{\sigma_2} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \dfrac{x_1-\mu_1}{\sigma_1} \\ \dfrac{x_2-\mu_2}{\sigma_2} \end{pmatrix} \qquad (2.6)
$$

da cui è facile vedere che la 2.4 è l'equazione di un ellisse, di centro in (μ_1, μ_2)

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

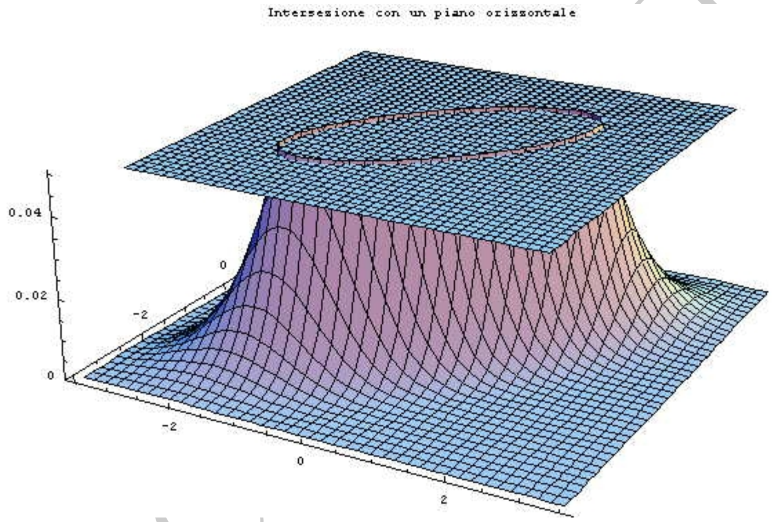

Figure 2.3: intersezioni con la normale bivariata

2.3 Distribuzioni condizionate nella normale bivariata

Le distribuzioni condizionate si ricavano facilmente, impiegando la relazione che permette di ricavare, per variabili aleatorie $\{X_1, X_2\}$ con distribuzione congiunta dotata di densità, la densità di X_1, condizionatamente ai valori x_2 assunti da X_2:

$$f_{X_1}(x_1|X_2 = x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}; \tag{2.7}$$

nel nostro caso è più comodo passare ai logaritmi, e porre:

$$z_1 = \frac{x_1 - \mu_1}{\sigma_1}; \qquad z_2 = \frac{x_2 - \mu_2}{\sigma_2}$$

e la (2.7) diventa:

$$\begin{aligned}
&\log(f_{X_1}(x_1|X_2 = x_2)) = \log(f_{X_1,X_2}(x_1, x_2)) - \log(f_{X_2}(x_2)) = \\
&\quad = \log\left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right) - \frac{(z_1^2 - 2\rho z_1 z_2 + z_2^2)}{2(1-\rho^2)} - \log\left(\frac{1}{\sqrt{2\pi}\sigma_2}\right) + \frac{1}{2}z_2^2 = \\
&\quad = \log\left(\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\right) - \frac{(z_1^2 - 2\rho z_1 z_2 + z_2^2 - (1-\rho^2)z_2^2)}{2(1-\rho^2)} = \\
&\quad = \log\left(\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\right) - \frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + \rho^2 z_2^2\right) = \\
&\quad = \log\left(\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\right) - \frac{1}{2}\frac{(z_1 - \rho z_2)^2}{1-\rho^2}
\end{aligned}$$

Adesso per ottenere il risultato finale, passiamo dal logaritmo $\log(f_{X_1}(x_1|X_2 = x_2))$ alla densità $f_{X_1}(x_1|X_2 = x_2)$ ed inoltre ritorniamo dalle variabili z_1, z_2 alle variabili originali x_1, x_2:

$$\begin{aligned}
f_{X_1}(x_1|X_2 = x_2) &= \exp\left(\log\left(f_{X_1}(x_1|X_2 = x_2)\right)\right) = \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\frac{\left(\frac{x_1-\mu_1}{\sigma_1} - \rho\frac{x_2-\mu_2}{\sigma_2}\right)^2}{1-\rho^2}\right\} = \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\frac{\left[x_1 - \mu_1 - \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right]^2}{\sigma_1^2(1-\rho^2)}\right\}.
\end{aligned} \tag{2.8}$$

(dopo qualche passaggio entro l'esponente)

Dall'espressione finale della 2.8 risulta evidente che la densità della distribuzione di X_1 condizionatamente ad un particolare valore x_2 di X_2 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, sinteticamente:

Distribuzioni condizionate nella normale bivariata

`href in ingresso  con distribuzioni condizionate`

$$X_1|\,\{X_2 = x_2\} \sim \mathcal{N}(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),\, \sigma_1^2(1 - \rho^2)) \tag{2.9}$$

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

Ovviamente si ha anche:

$$X_2|\,\{X_1 = x_1\} \sim \mathcal{N}(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1),\, \sigma_2^2(1 - \rho^2)) \tag{2.10}$$

Si guardi anche la figura 2.4, in cui si considera l'intersezione di una densità normale bivariata con un piano verticale di equazione $X_2 = x_2$.

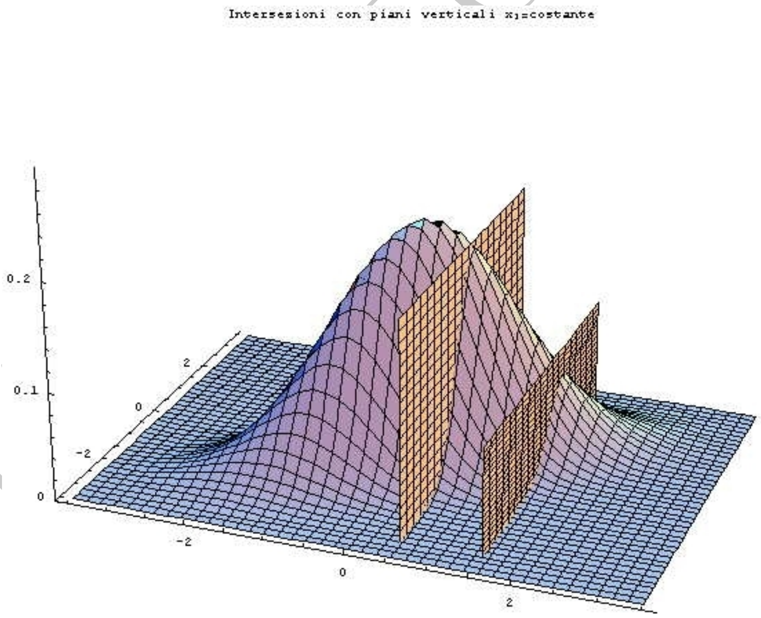

Figure 2.4: Distribuzioni Condizionate: intersezioni di un piano verticale perpendicolare alla variabile *condizionante* con la normale bivariata

Nei casi estremi si ha:

- 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) [3]; 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:
$$x_1 = \mu_1 + \text{sign}\,(\rho)\,\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

A questo punto siamo in grado di dare un preciso significato al quadrato dell'indice di correlazione lineare semplice (ρ)

Il coefficiente di determinazione

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.

2.3.1 Funzioni di regressione

In una qualsiasi distribuzione bivariata, la speranza matematica della distribuzione di una variabile, diciamo X_1, condizionata ai valori dell'altra variabile, $X_2 = x_2$, vista come funzione di x_2, ossia $\mathrm{E}\,[X_1|X_2 = x_2]$, va sotto il nome di *funzione di regressione*. (per qualsiasi variabile doppia: il concetto si può estendere alle distribuzioni di variabili multiple, ma non è questo il momento di farlo)

[3] 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

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:

Varianza della distribuzione condizionata

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

Evidentemente quanto visto in questa sezione sulle distribuzioni condizionate si estende immediatamente alla distribuzione di X_2 condizionatamente a $X_1 = x_1$; è sufficente scambiare gli indici "1" e "2" nelle formule viste finora.

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:

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

Momenti di X2 condizionatamente a X1

$$\mathrm{E}\left[X_2|X_1 = x_1\right] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1).$$
$$\mathrm{V}\left[X_2|X_1 = x_1\right] = \sigma_2^2(1 - \rho^2) \qquad \forall\, x_1$$

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) $\rho = 0$.

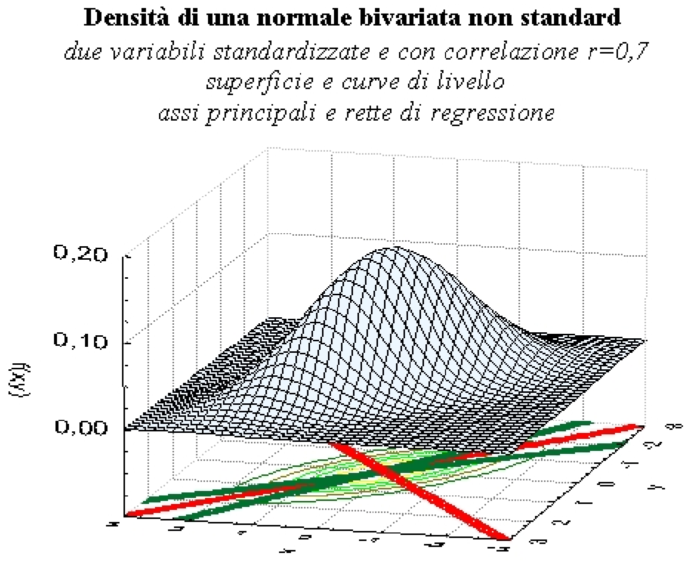

Figure 2.5: Variabili normali standardizzate con correlazione 0.7

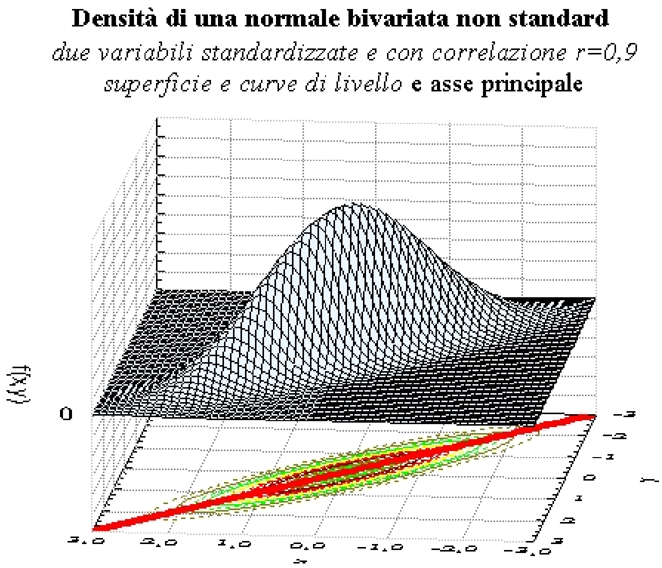

Figure 2.6: Variabili normali standardizzate con correlazione 0.9

2.4 Probabilità relativa al primo quadrante in funzione di ρ

Un'altra interpretazione di ρ si può avere riflettendo sulla probabilità integrale relativa al primo quadrante.

Se abbiamo una variabile doppia con distriubuzione normale bivariata, ci possiamo chiedere:

qual è la probabilità di osservare un'unità con valori superiori alla media per entrambe le componenti X_1 e X_2?

In termini geometrici questo corrisponde a chiederci qual è il volume sotteso dalla densità nel primo quadrante:

$$\text{Prob}\{X_1 \geq \mu_1 \cap X_2 \geq \mu_2\} = \int_{\mu_1}^{+\infty} \int_{\mu_2}^{+\infty} f(x_1, x_2) dx_1\, dx_2$$

Ovviamente questa probabilità è $\frac{1}{4}$ se $\rho = 0$; è maggiore o minore di $\frac{1}{4}$ secondo il segno di ρ ed anzi tale probabilità è funzione monotona di ρ, come si vede dalla figura 2.7

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.

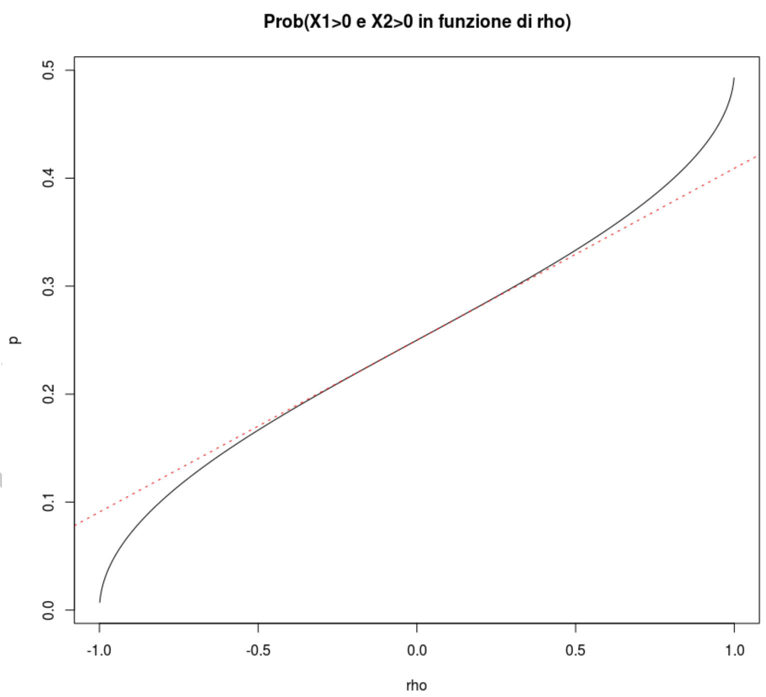

Figure 2.7: Probabilità integrale relativa al primo quadrante in funzione di ρ

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.

2.5 Trasformazioni lineari e assi principali nella normale bivariata

Se vogliamo ottenere la densità di una nuova variabile aleatoria doppia $\mathbf{Z} = (Z_1, Z_2)^{\mathsf{T}}$, ottenuta con una trasformata lineare della variabile aleatoria doppia $\mathbf{X} = (X_1, X_2)^{\mathsf{T}}$, distribuita secondo una generica normale bivariata di parametri $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$:

$$\mathbf{Z} = \mathbf{B} \begin{pmatrix} \frac{X_1-\mu_1}{\sigma_1} \\ \frac{X_2-\mu_2}{\sigma_2} \end{pmatrix} \qquad \text{con} \qquad \mathbf{B} = \frac{1}{\sqrt{2}} \begin{pmatrix} +\frac{1}{\sqrt{1+\rho}} & +\frac{1}{\sqrt{1+\rho}} \\ -\frac{1}{\sqrt{1-\rho}} & +\frac{1}{\sqrt{1-\rho}} \end{pmatrix} \tag{2.11}$$

occorre:

1. sostituire nella densità 2.3 (x_1, x_2) con (z_1, z_2), tenendo conto che le variabili compaiono solo nella forma quadratica a esponente;
2. moltiplicare la densità per lo Jacobiano della trasformazione da $\mathbf{Z}$ a $\mathbf{X}$, $\left(\frac{\partial mX}{\partial mZ}\right)$ oppure dividere per lo Jacobiano della trasformazione da $\mathbf{X}$ a $\mathbf{Z}$.

E' evidente che la variabile aleatoria doppia

$$\begin{pmatrix} \frac{X_1-\mu_1}{\sigma_1} \\ \frac{X_2-\mu_2}{\sigma_2} \end{pmatrix}$$

ha matrice di varianza e covarianza:

$$\mathrm{V}\begin{bmatrix} \frac{X_1-\mu_1}{\sigma_1} \\ \frac{X_2-\mu_2}{\sigma_2} \end{bmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

(infatti si tratta palesemente di due variabili standardizzate con correlazione ρ)

Omettendo per brevità i passaggi, è facile vedere che si ottiene:

$$\mathbf{Z} \sim \mathcal{N}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \tag{2.12}$$

Il risultato sulla matrice di varianze e covarianze di $\mathbf{Z}$ si ottiene facilmente dalla relazione valida per trasformate lineari di v.a.:

$$\mathrm{V}[\mathbf{AX}] = \mathbf{A}\mathrm{V}[\mathbf{X}]\mathbf{A}^{\mathsf{T}}$$

e nel nostro caso si ottiene:

$$\frac{1}{\sqrt{2}}\begin{pmatrix} +\frac{1}{\sqrt{1+\rho}} & +\frac{1}{\sqrt{1+\rho}} \\ -\frac{1}{\sqrt{1-\rho}} & +\frac{1}{\sqrt{1-\rho}} \end{pmatrix}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} +\frac{1}{\sqrt{1+\rho}} & -\frac{1}{\sqrt{1-\rho}} \\ +\frac{1}{\sqrt{1+\rho}} & +\frac{1}{\sqrt{1-\rho}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Questa trasformazione dunque trasforma una normale bivariata qualsiasi in una normale bivariata a componenti indipendenti e standardizzate.

La trasformazione $\mathbf{B}$ è una trasformazione *ortogonalizzante* perchè rende le variabili ortogonali ma ovviamente non è unica. Ad esempio un altra matrice che rende le variabili X indipendenti con la trasformazione definita nella 2.11, è la matrice $\mathbf{B}_1$:

$$\mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} \tag{2.13}$$

2.5.1 Altre trasformazioni

In effetti nelle trasformazioni (2.11) e (2.13) si può trasformare direttamente lo scarto originario, anzichè lo scarto standardizzato, e inglobare la standardizzazione, ossia la divisione per σ_1 e σ_2 *direttamente nella matrice di trasformazione* $\mathbf{B}$. Ad esempio usando:

$$\mathbf{B}^* = \frac{1}{\sqrt{2}}\begin{pmatrix} +\frac{1}{\sigma_1\sqrt{1+\rho}} & +\frac{1}{\sigma_2\sqrt{1+\rho}} \\ -\frac{1}{\sigma_1\sqrt{1-\rho}} & +\frac{1}{\sigma_2\sqrt{1-\rho}} \end{pmatrix} \quad \text{e quindi} \quad \mathbf{Z} = \mathbf{B}^*\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix} \tag{2.14}$$

(la prima colonna di $\mathbf{B}$ è stata divisa per σ_1 e la seconda per σ_2) oppure si può agire sulle colonne della matrice $\mathbf{B}_1$ con:

$$\mathbf{B}_1^* = \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}} & \frac{1}{\sigma_2\sqrt{1-\rho^2}} \end{pmatrix} \quad \text{e quindi} \quad \mathbf{Z} = \mathbf{B}_1^*\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix} \tag{2.15}$$

Supponiamo di avere una variabile doppia (X_1, X_2) con distribuzione normale bivariata di parametri:

$$\mu_1 = 3, \quad \mu_2 = 5; \quad \sigma_1 = 2, \quad \sigma_2 = 5; \quad \rho = 0,6$$

Una trasformazione che ortonormalizza $\mathbf{X}$, ossia che la rende a componenti *indipendenti* e *standardizzate*, è applicando la 2.15:

$$\begin{aligned} \mathbf{Z} &= \mathbf{B}_1^*\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{0,6}{2\sqrt{1-(0,6)^2}} & \frac{1}{5\sqrt{1-(0,6)^2}} \end{pmatrix}\begin{pmatrix} X_1-3 \\ X_2-5 \end{pmatrix} = \\ &= \begin{pmatrix} 0,5\times(X_1-3) \\ 0,375\times(X_1-3)+0,25\times(X_2-5) \end{pmatrix}. \end{aligned}$$

Consideriamo adesso un'altra trasformata lineare:

$$\mathbf{W} = \mathbf{C}\begin{pmatrix} \frac{X_1-\mu_1}{\sigma_1} \\ \frac{X_2-\mu_2}{\sigma_2} \end{pmatrix} \quad \text{con} \quad \mathbf{C} = \frac{1}{\sqrt{2}}\begin{pmatrix} +1 & +1 \\ -1 & +1 \end{pmatrix} \tag{2.16}$$

E' facile vedere intanto che questa è una rotazione ortogonale degli assi. Infatti si vede facilmente che:

$$\mathbf{C}\mathbf{C}^{\mathsf{T}} = \frac{1}{\sqrt{2}}\begin{pmatrix} +1 & +1 \\ -1 & +1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} +1 & -1 \\ +1 & +1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ancora si vede:

$$\mathbf{W} \sim \mathcal{N}_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}\right]$$

Infatti:

$$\begin{aligned} \mathrm{V}[\mathbf{W}] &= \mathbf{C}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\mathbf{C}^{\mathsf{T}} = \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix} +1 & +1 \\ -1 & +1 \end{pmatrix}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} +1 & +1 \\ -1 & +1 \end{pmatrix}^{\mathsf{T}} = \\ &= \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix} \end{aligned}$$

Questa rotazione porta gli assi coordinati lungo gli *assi principali* degli ellissi di concentrazione. W_1 è la combinazione lineare delle variabili standardizzate, di massima varianza fra tutte le trasformazioni normalizzate. W_2 invece è di minima varianza.

Dimostrazione Supponiamo per semplicità di avere una variabile normale bivariata $\mathbf{Z} = \{Z_1, Z_2\}$ a componenti standardizzate , ossia con valori attesi nulli e varianze unitarie delle due componenti, e quindi con matrice di varianze e covarianze:

$$\mathrm{V}[\mathbf{Z}] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Cerchiamo una combinazione lineare W delle componenti di $\mathbf{Z}$, a coefficienti normalizzati, che abbia la massima varianza.

$$W = c_1 Z_1 + c_2 Z_2; \qquad \mathrm{V}[W] = c_1^2 + 2\rho c_1 c_2 + c_2^2; \quad \text{col vincolo:} \quad c_1^2 + c_2^2 = 1$$

Perchè il vincolo sia rispettato basta porre:

$$c_2 = \sqrt{1-c_1^2} \qquad \text{oppure} \qquad c_2 = -\sqrt{1-c_1^2} \tag{2.17}$$

Sviluppiamo i passaggi per ora solo per il caso di c_2 positivo e quindi per la varianza di W abbiamo, sostituendo a c_2 il valore $\sqrt{1-c_1^2}$:

$$\mathrm{V}[W] = c_1^2 + 2\rho c_1\sqrt{1-c_1^2} + (1-c_1^2) = 1 + 2\rho c_1\sqrt{1-c_1^2}$$

Questa è una funzione da massimizzare rispetto a c_1, per cui calcoliamone la derivata ed eguagliamola a zero:

$$\frac{\partial \mathrm{V}[W]}{\partial c_1} = 2\rho\left\{\sqrt{1-c_1^2} - \frac{c_1^2}{\sqrt{1-c_1^2}}\right\}$$

$$\frac{\partial \mathrm{V}[W]}{\partial c_1} = 0 \qquad \Rightarrow \qquad \rho \left\{ \sqrt{1-c_1^2} - \frac{c_1^2}{\sqrt{1-c_1^2}} \right\} = 0$$

Si vede subito che se $\rho = 0$ qualsiasi valore di $c1$ annulla la derivata prima: è ovvio perchè in questo caso avremmo due variabili standardizzate e ortogonali e qualsiasi combinazione lineare a coefficienti normalizzati conduce a variabili con la stessa varianza:

$$\rho = 0 \qquad \Rightarrow \qquad \mathrm{V}[W] = 1$$

Se $\rho \neq 0$ si annulla la derivata con le soluzioni:

$$c_1 = \mp \frac{1}{\sqrt{2}}$$

Controlliamo il segno della derivata seconda in questi due punti:

$$\left. \frac{\partial^2 \mathrm{V}[W]}{\partial {c_1}^2} \right|_{c_1 = \frac{1}{\sqrt{2}}} = -8\rho \quad \text{e} \quad \left. \frac{\partial^2 \mathrm{V}[W]}{\partial {c_1}^2} \right|_{c_1 = -\frac{1}{\sqrt{2}}} = 8\rho$$

Quindi se ρ è positivo il punto di massimo si ha in corrispondenza di $c_1 = \frac{1}{\sqrt{2}}$, mentre in corrispondenza di $c_1 = -\frac{1}{\sqrt{2}}$ si ha il minimo, ossia la combinazione lineare a coefficienti normalizzati di minima varianza; ovviamente vanno invertiti i segni dei valori di c_1 se $\rho < 0$.

In definitiva, ricordando che $c_2 = \sqrt{1-c_1^2} = \sqrt{1-\frac{1}{\sqrt{2}}^2} = \frac{1}{\sqrt{2}}$, si ottiene la combinazione lineare di massima varianza:

$$\text{se} \quad \rho > 0 \quad \Rightarrow \quad W = \frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2$$

$$\text{se} \quad \rho < 0 \quad \Rightarrow \quad W = -\frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2$$

Se per soddisfare la condizione di normalizzazione dei coefficienti, nella (2.17) prendiamo il radicale col segno negativo ($c_2 = -\sqrt{1-c_1^2}$), si ottengono le stesse soluzioni $\{c_1, c_2\}$, ma cambiate di segno [4]; per brevità ometto la dimostrazione, semplice ma noios.

Se invece prendiamo, per annullare la derivata prima, la soluzione che porta ad un valore positivo della derivata seconda, e quindi ad un minimo di $\mathrm{V}[W]$, si ha la combinazione di minima varianza, fra tutte quelle a coefficienti normalizzati:

$$\text{se} \quad \rho > 0 \quad \Rightarrow \quad W = \frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2$$

$$\text{se} \quad \rho < 0 \quad \Rightarrow \quad W = -\frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2$$

Si noti che le varianze di $\mathbf{W}$ sono gli autovalori della matrice di correlazione $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

[4] E' ovvio dalla definizione di W che $\mathrm{V}[W]$ rimane inalterata cambiando segno sia a c_1 che a c_2.

Supponiamo adesso di avere una variabile normale bivariata generica, per comodità con valori attesi nulli delle due componenti, e con matrice di varianze e covarianze:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Per risolvere lo stesso problema occorre procedere in modo simile, anche se stavolta il trattare con variabili con varianze non unitarie non consente tutte le semplificazioni del caso precedente

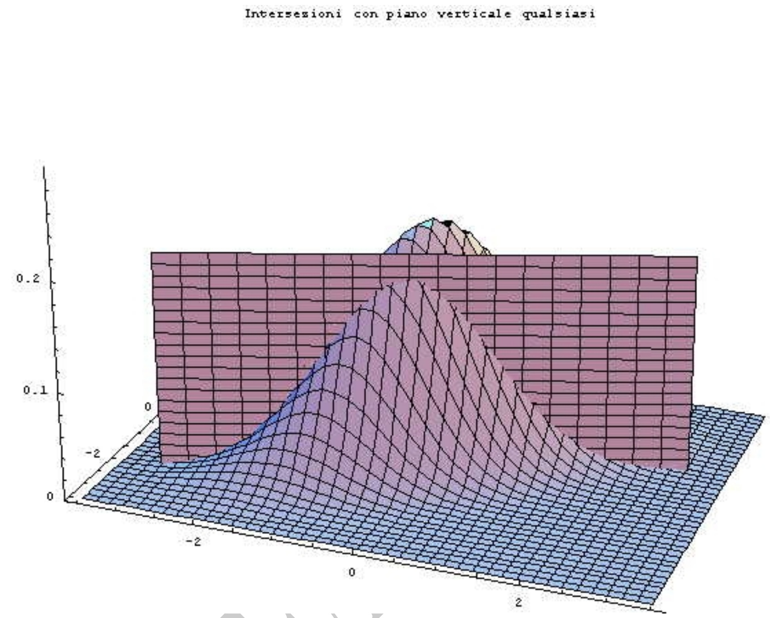

Figure 2.8: Combinazioni lineari qualsiasi di variabili normali: intersezioni di un piano verticale con la normale bivariata (non ortogonale rispetto alle due variabili)

livello avanzato

2.6 Altre proprietà della normale bivariata

href in ingresso con complementi

2.6.1 L'esponente della distribuzione normale bivariata

In questa sezione mostro un risultato relativo all'*esponente della normale bivariata* , che risulta distribuito proporzionalmente ad una χ^2; più avanti vedremo il risultato analogo per distribuzioni normali multivariate, secondo un'impostazione esclusivamente matriciale. *esponente della normale bivariata*

Indichiamo con $h(x_1, x_2)$ la forma quadratica che compare ad esponente della densità della normale bivariata, in modo che:

$$h(x_1, x_2): \qquad f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2}h(x_1, x_2)\right\}$$

Operiamo ora su $h(x_1, x_2)$:

$$\begin{aligned} h(x_1, x_2) &= \frac{1}{(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \left(-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] = \\ &= \frac{1}{(1-\rho^2)}\begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix} \end{aligned}$$

Adesso è facile vedere che i fattori di quest'ultima espressione dipendenti da ρ possono essere espressi in una forma ancora più compatta:

$$\frac{1}{(1-\rho^2)}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} = \mathbf{R}^{-1}$$

Inoltre è possibile scomporre $\mathbf{R}^{-1}$ nel prodotto di una matrice $\mathbf{B}$ per la sua trasposta $\mathbf{B}^{\mathsf{T}}$; infatti:

$$\mathbf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{1}{1+\rho} & -\frac{1}{1-\rho} \\ \frac{1}{1+\rho} & \frac{1}{1-\rho} \end{pmatrix} \qquad \text{e quindi}$$

$$\begin{aligned} \mathbf{B}\mathbf{B}^{\mathsf{T}} &= \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{1}{1+\rho} & -\frac{1}{1-\rho} \\ \frac{1}{1+\rho} & \frac{1}{1-\rho} \end{pmatrix} \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{1}{1+\rho} & +\frac{1}{1+\rho} \\ -\frac{1}{1-\rho} & \frac{1}{1-\rho} \end{pmatrix} = \\ &= \frac{1}{2}\begin{pmatrix} \frac{1}{1+\rho} + \frac{1}{1-\rho} & \frac{1}{1+\rho} - \frac{1}{1-\rho} \\ \frac{1}{1+\rho} - \frac{1}{1-\rho} & \frac{1}{1+\rho} + \frac{1}{1-\rho} \end{pmatrix} = \frac{1}{2\,(1-\rho^2)}\begin{pmatrix} 2 & -2\rho \\ -2\rho & 2 \end{pmatrix} = \mathbf{R}^{-1}. \end{aligned}$$

E' facile vedere che la matrice $\mathbf{B}$ è ottenuta dagli autovettori di $\mathbf{R}$ divisi per le radici quadrate degli autovalori link con autovettori Tornando ora alla 2.18, otteniamo:

$$\begin{aligned} h(x_1, x_2) &= \frac{1}{(1-\rho^2)}\begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix} = \\ &= \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix}^{\mathsf{T}} \mathbf{R}^{-1} \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix} = \end{aligned}$$

$$= \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix}^{\mathsf{T}} \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{1+\rho} & -\frac{1}{1-\rho} \\ \frac{1}{1+\rho} & \frac{1}{1-\rho} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{1+\rho} & +\frac{1}{1+\rho} \\ -\frac{1}{1-\rho} & \frac{1}{1-\rho} \end{pmatrix} \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix}$$
$$= \mathbf{z}^{\mathsf{T}}\mathbf{z}$$

avendo posto:

$$\mathbf{z} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{1+\rho} & +\frac{1}{1+\rho} \\ -\frac{1}{1-\rho} & \frac{1}{1-\rho} \end{pmatrix} \begin{pmatrix} \frac{x_1-\mu_1}{\sigma_1} \\ \frac{x_2-\mu_2}{\sigma_2} \end{pmatrix} \tag{2.18}$$

Si confronti adesso questa trasformazione con la (2.11): $\mathbf{Z}$ si distribuisce secondo una normale bivariata a componenti standardizzate e indipendenti, ossia $\mathbf{Z} \sim \mathcal{N}_2\left[\mathbf{0}_2, \mathbf{I}_2\right]$.

Questa relazione è molto utile per diversi motivi: sostanzialmente ci fa vedere che attraverso una trasformazione lineare (2.11)

L'esponente della normale bivariata

Il numeratore dell'esponente della densità di una normale bivariata si distribuisce come una variabile aleatoria χ^2 con due gradi di libertà.

$$\frac{1}{(1-\rho^2)}\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(X_1-\mu_1)(X_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2\right] \sim \chi^2_2 \tag{2.19}$$

distanze di mahalanobis trasformata inversa della F(Chi2)

completare

2.6.2 Probabilità integrali nella normale bivariata

livello avanzato

Il calcolo delle probabilità integrali su domini rettangolari $x_{1a} < X_1 < x_{1b} \cap x_{2a} < X_2 < x_{2b}$ non è banale, e non è riconducibile al calcolo di integrali della normale univariata, tranne che per il caso di indipendenza $\rho = 0$, per il quale si ha:

$$\begin{aligned} \rho = 0 \quad \Rightarrow \quad & \text{Prob}\left\{x_{1a} < X_1 < x_{1b} \cap x_{2a} < X_2 < x_{2b}\right\} = \\ = \quad & \text{Prob}\left\{x_{1a} < X_1 < x_{1b}\right\} \times \text{Prob}\left\{x_{2a} < X_2 < x_{2b}\right\} \end{aligned}$$

Nel caso generale, quando $\rho \neq 0$, Il calcolo non è effettuabile in modo semplice ed occorre ricorrere a software particolari.

Solo per fare un cenno, possiamo limitarci al caso di una normale bivariata a componenti standardizzate {} e con correlazione ρ, infatti per mezzo di trasformate lineari elementari separate di ciascuna componente ci si può sempre ricondurre ad un dominio di integrazione rettangolare

2.6.3 Generazione di numeri casuali da una Distribuzione normale: formula di Box-Muller

livello avanzato

completare

E' un esempio classico nell'ambito della generazione di numeri casuali, ed è la *formula di Box-Muller* per *generare coppie di numeri pseudo-casuali normali standardizzati indipendenti*.

formula di Box-Muller

Non è il procedimento più efficiente per generare numeri pseudo-casuali da una distribuzione normale bivariata, ma è una formula *classica* e ci serve anche per vedere un'interessante proprietà della normale bivariata.

La sfrutteremo anche per qualche esempio realizzato mediante distribuzioni simulate.

Piuttosto che esporre prima la formula, e poi dimostrarne la validità con procedimenti complessi, preferisco qui fornire una dimostrazione, che ricalca da vicino quella originaria di Box-Muller (1958), fondata su proprietà geometriche della distribuzione normale bivariata, che fa scaturire la formula in modo naturale.

Supponiamo di voler generare un punto casuale P, di coordinate X_P e Y_P, da una normale bivariata standardizzata a componenti indipendenti [5].

Si rappresenti P in un sistema di riferimento in coordinate polari ed in questo sistema siano ρ_P il modulo di P e θ_P l'anomalia di P (vedere figura).

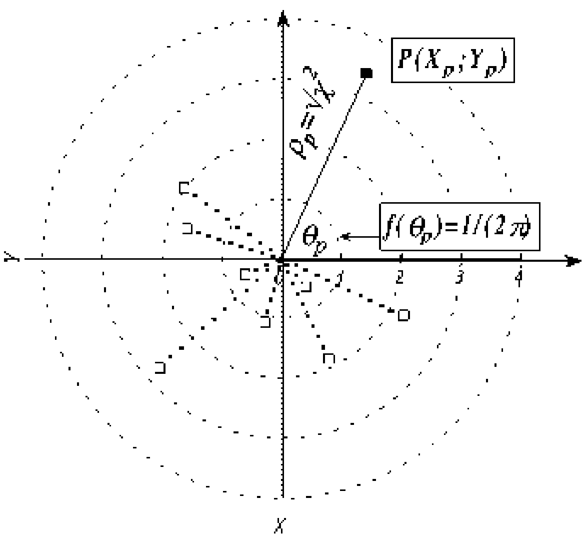

Figure 2.9: Rappresentazione in coordinate polare: generazione di numeri casuali da una normale bivariata

[5] Per comodità di notazione qui le due variabili sono X e Y

Invece di generare X_P e Y_P, generiamo ρ_P e θ_P: ρ_P e θ_P risultano ancora indipendenti per simmetria radiale della distribuzione normale bivariata a componenti indipendenti. Sappiamo che $\rho_P^2 = X_P^2 + Y_P^2$, per cui ρ_P è la radice quadrata di una v.a. χ^2 con due gradi di libertà, che possiamo facilmente generare, perchè è una v.a. esponenziale di parametro $\lambda = \frac{1}{2}$, per cui genereremo ρ_P mediante la relazione:

$$\rho_P = \sqrt{-2 \log U}$$

essendo U un numero casuale da una distribuzione uniforme standard.

Infatti, se Z è una variabile aleatoria con distribuzione esponenziale di parametro λ, la sua funzione di ripartizione è data da: $F(z) = 1 - e^{-\lambda z)}$; invertendo la relazione si ottiene: $Z = -\frac{1-log(F(.))}{\lambda}$. E' noto che $F()$, per qualsiasi variabile aleatoria, ha una distribuzione uniforme in (0-1), per cui anche $1 - F()$ ha distribuzione uniforme in (0-1) e quindi ponendo $\lambda = \frac{1}{2}$, per avere una determinazione pseudo aleatoria Z di una χ_2^2 occorre calcolare: $Z = -2 \log U$.

Per la generazione dell'anomalia θ_P, è sufficiente constatare che, per variabili normali standardizzate indipendenti, la densità dei punti lungo una circonferenza con centro nell'origine è costante, qualsiasi sia il raggio, per cui θ_P è distribuito uniformemente nell'intervallo $[0, 2\pi]$. Pertanto per generare θ_P si ricorre all'elementare trasformazione lineare:

$$\theta_P = 2\pi V$$

essendo V un numero casuale indipendente da U estratto da una uniforme standard.

Ottenute quindi due determinazioni casuali di ρ_P e θ_P, per ottenere i due numeri casuali X_P e Y_P, occorre ritrasformare in coordinate cartesiane:

$$X_P = \rho_P \cos \theta_P \qquad Y_P = \rho_P \sin \theta_P$$

Indicando ora semplicemente con X e Y le coordinate cartesiane, si ottiene infine la **formula di Box-Muller** **citazione**

Formula di Box-Muller

Se $\{U, V\}$ è una coppia di numeri pseudo-casuali generati da una distribuzione uniforme nell'intervallo (0-1), allora con la trasformazione:

$$\begin{aligned} X &= \sqrt{-2 \log U} \, \cos(2\pi V) \\ Y &= \sqrt{-2 \log U} \, \sin(2\pi V) \end{aligned} \qquad (2.20)$$

si ottiene una coppia di numeri pseudo-casuali $\{X, Y\}$ generati da una distribuzione normale bivariata standardizzata a componenti indipendenti.

La peculiarità di tale formula, che la rende a prima vista sorprendente, è che si ottengono *coppie* di numeri casuali normali indipendenti. Sebbene questo non sia il metodo più veloce per generare numeri casuali normali, è una formula molto compatta per la generazione di singoli numeri casuali normali.

Una dimostrazione della validità della formula di Box-Muller, più complessa e più lunga di quella fornita, si ottiene applicando i noti teoremi del calcolo delle probabilità per le trasformazioni di variabili aleatorie alla trasformazione dalle v.a. $\{U, V\}$ alle variabili $\{X, Y\}$. Si calcola prima lo Jacobiano di questa trasformazione; oppure, più comodamente, si calcola l'inverso dello Jacobiano della trasformazione diretta da $\{X, Y\}$ alle variabili $\{U, V\}$; risulta comunque comodo considerare $\frac{Y}{X} = \tan(2(\pi V)$ e $X^2+Y^2 = -2\log U$. Si noti come se si effettua il rapporto Y/X, si deve ottenere un numero casuale distribuito secondo una Cauchy, che è infatti ottenibile come rapporto fra due v.a. normali standard indipendenti.

In pratica $\{X, Y\}$ si può considerare come una determinazione di una v.a. con distribuzione $\mathcal{N}_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$. Pertanto per ottenere un campione *(simulato!)* di ampiezza n () $\{X_i, Y_i\}$, $(i = 1, 2, \ldots, n))$ da tale distribuzione , occorre applicare la (2.20) a n coppie indipendenti $\{U_i, V_i\}$ $(i = 1, 2, \ldots, n)$ di numeri pseudo-casuali generati da una distribuzione uniforme nell'intervallo (0-1). Ovviamente non sarà un *vero* campione, ma piuttosto un campione *simulato*, dal momento che la generica coppia $\{U_i, V_i\}$ $(i = 1, 2, \ldots, n)$ non è una coppia di determinazioni da una *vera distribuzione* uniforme, ma è generata attraverso opportuni algoritmi di generazione di sequenze di numeri *pseudo-casuali*.

Senza entrare qui in particolari sul concetto di sequenza di numeri *pseudo-aleatori*, basta segnalare che per algoritmo di generazione di numeri pseudo-casuali da una distribuzione uniforme, si intende un algoritmo che genera sequenze che *somigliano* ad analoghe sequenze estratte da una vera distribuzione uniforme. **citazione** link con pagine mi livello avanzato

2.6.4 La distribuzione normale bivariata come limite della trinomiale

La distribuzione normale bivariata può essere introdotta in molti modi, alcuni dei quali sono delle estensioni bivariate di casi univariati che conducono alla normalità.

Nell'ambito delle approssimazioni asintotiche elementari, è noto che per valori di p non vicini a zero o ad 1, la distribuzione normale univariata fornisce un'approssimazione asintotica delle probabilità binomiali per n grande.

Possiamo chiederci se un 'analoga approssimazione asintotica vale per la distribuzione trinomiale, e successivamente per la multinomiale.

In effetti si può dimostrare che la distribuzione normale bivariata scaturisce come limite di una distribuzione trinomiale, purché le probabilità non siano troppo vicine a zero o a uno. Si tratta di un'estensione del noto teorema limite di De Moivre-Laplace, che consente di approssimare le probabilità di una distribuzione binomiale mediante la densità di una normale.

Supponiamo di avere una distribuzione trinomiale (ossia una multinomiale con tre modalità A_1, A_2, A_3) con probabilità di verificarsi di ciascuna delle tre modalità: (p_1, p_2, p_3). Ovviamente si ha: $p_3 = 1 - p_1 - p_2$.

La distribuzione di probabilità, ossia la probabilità che in n esperimenti indipendenti ciascuna delle tre modalità si presenti n_1, n_2 ed n_3 volte, è data da:

$$\text{Prob}\left\{X_1 = n_1, X_2 = n_2, X_3 = n_3\right\} = \binom{n}{n_1\ n_2\ n_3} {p_1}^{n_1} {p_2}^{n_2} {p_3}^{n_3}$$

Per i primi due momenti della distribuzione si ha:

$$\begin{aligned} \mathrm{E}\left[X_i\right] &= np_i \quad i = 1,2,3 \\ \mathrm{V}\left[X_i\right] &= np_i(1-p_i) \quad i = 1,2,3 \\ \mathrm{Cov}\left[X_i, X_j\right] &= -np_ip_j \quad i \neq j \quad i = 1,2,3 \quad j = 1,2,3 \end{aligned} \tag{2.21}$$

(si veda eventualmente la sezione **??**, per le caratteristiche sommarie di una distribuzione multinomiale qualsiasi) `link con sec:multinomiale` .

Evidentemente $n_3 = n - n_1 - n_2$, e chiaramente la distribuzione, sebbene formalmente trivariata, si riconduce ad una distribuzione bivariata nella coppia $\{X_1, X_2\}$ dato il vincolo: $X_1 + X_2 + X_3 = n$.

Si può dimostrare che la ditribuzione di probabilità di $\{X_1, X_2\}$ è approssimabile, per valori di n grandi, dalla densità di probabilità normale bivariata che abbia gli stessi primi due momenti, purchè nè p_1 nè p_2 nè p_3 siano vicini a zero o ad uno.

Qui mi limito a dare un cenno alla dimostrazione solo per la probabilità massima che si raggiunge nel punto $n\,p_1$, $n\,p_2$.

Si parte dall'approssimazione di Stirling al fattoriale di un numero intero grande:

$$x! \approx x^x e^{-x} \sqrt{2\pi x} \tag{2.22}$$

Si tratta di un approssimazione asintotica nel senso che:

$$\lim_{x\to\infty} \frac{x!}{x^x e^{-x}\sqrt{2\pi x}} = 1$$

Calcoliamo adesso la probabilità massima, in corrispondenza dei valori attesi $n\,p_1$, $n\,p_2$ e $n\,p_3 = n - n\,p_1 - n\,p_2$

$$P_{\max} = \text{Prob}\,\{X_1 = n\,p_1, X_2 = n\,p_2, X_3 = n - n\,p_1 - n\,p_2\} =$$

$$= \frac{n!}{(n\,p_1)!(n\,p_2)!(n - np_1 - np_2)!} \times {p_1}^{n\,p_1}{p_2}^{n\,p_2} 1 - p_1 - {p_2}^{n - n\,p_1 - n\,p_2} \tag{2.23}$$

A questo punto si sostituiscono nella (2.23) i fattoriali con l'approssimazione asintotica (2.22), valida per valori grandi dell'argomento, e si ottiene il risultato [6]:

$$P_{\max} \approx \frac{1}{2\pi n \sqrt{p_1 p_2 (1 - p_1 - p_2)}} = \frac{1}{2\pi \sqrt{np_1}\sqrt{np_2}\sqrt{1 - p_1 - p_2}}. \tag{2.24}$$

Per ottenere il risultato finale, basta sfruttare alcune delle relazioni in (2.21) che esprimono i momenti della trinomiale in funzione dei parametri, e ricordando che sfruttiamo la relazione $p_3 = 1 - p_1 - p_2$, avendo parametrizzato rispetto alle componenti X_1, X_2. In particolare si ha per la correlazione fra le due componenti:

$$\begin{aligned}
\text{Corr}\,[X_1, X_2] &= \frac{\text{Cov}\,[X_1, X_2]}{\sqrt{\text{V}\,[X_1]\,\text{V}\,[X_2]}} = \frac{-np_1p_2}{\sqrt{np_1(1 - p_1)np_2(1 - p_2)}} = \\
&= -\sqrt{\frac{p_1\,p_2}{(1 - p_1)(1 - p_2)}}
\end{aligned}$$

Da questo si ricava che:

$$1 - (\text{Corr}\,[X_1, X_2])^2 = 1 - \frac{p_1\,p_2}{(1 - p_1)(1 - p_2)} = \frac{1 - p_1 - p_2}{(1 - p_1)(1 - p_2)}$$

e quindi:

$$1 - p_1 - p_2 = (1 - (\text{Corr}\,[X_1, X_2])^2)(1 - p_1)(1 - p_2).$$

Sostituendo quest'ultima relazione nella (2.24), e ricordando anche le espressioni per le varianze delle due componenti della binomiale X_1 e X_2, possiamo finalmente esprimere l'approssimazione asintotica di $P_{\max}$ solo in funzione dei momenti delle due componenti:

$$\begin{aligned}
P_{\max} &\approx \frac{1}{2\pi\sqrt{np_1}\sqrt{np_2}\sqrt{1 - p_1 - p_2}} = \\
&= \frac{1}{2\pi\sqrt{np_1}\sqrt{np_2}\sqrt{(1 - (\text{Corr}\,[X_1, X_2])^2)(1 - p_1)(1 - p_2)}} = \\
&= \frac{1}{2\pi\sqrt{np_1(1 - p_1)}\sqrt{np_2(1 - p_2)}\sqrt{1 - (\text{Corr}\,[X_1, X_2])^2}} = \\
&= \frac{1}{2\pi\sigma_{X_1}\sigma_{X_1}\sqrt{1 - \rho^2_{X_1,X_2}}}
\end{aligned}$$

[6] ometto i passaggi perchè soltanto noiosi e di scarsa utilità didattica, si tratta di semplificare tutti i termini con la funzione esponenziale e gli esponenti delle potenze con la stessa base a numeratore e denominatore

e finalmente quest'ultima espressione rende chiaro che la $P_{\max}$ è approssimata dalla probabilità massima di una distribuzione normale bivariata che ha gli stessi parametri della distribuzione trinomiale che si vuole approssimare, ossia una distribuzione normale bivariata di parametri:

$$\mu_1 = n\,p_1, \quad \mu_2 = n\,p_2, \quad \sigma_1^2 = n\,p_1(1-p_1),$$

$$\sigma_2^2 = n\,p_2(1-p_2), \quad \rho = -\sqrt{\frac{p_1\,p_2}{(1-p_1)(1-p_2)}}$$

2.6.5 Esempio numerico.

Supponiamo di avere una distribuzione trinomiale di parametri $n = 50$, $p_1 = 0,5$ e $p_2 = 0,3$. Si ha allora, applicando le relazioni (2.21):

$$\begin{aligned}
\mathrm{E}\,[X_1] &= 50 \times 0,5 = 25 \quad \mathrm{E}\,[X_2] = 50 \times 0,3 = 15 \\
\mathrm{V}\,[X_1] &= 50 \times 0,5 \times 0,5 = 12,5 \quad \mathrm{V}\,[X_2] = 50 \times 0,3 \times 0,7 = 10,5 \\
\mathrm{Corr}\,[X_i, X_j] &= -\sqrt{\frac{0,5 \times 0,3}{0,5 \times 0,7}} = -\sqrt{\frac{3}{7}} = 0,6547
\end{aligned}$$

La probabilità massima è dunque approssimabile dalla probabilità massima della distribuzione normale bivariata con gli stessi momenti primi e secondi, e quindi:

$$P_{\max} = \mathrm{Prob}\,\{X_1 = 25, X_2 = 15\} \approx \frac{1}{2\pi\sqrt{12,5}\sqrt{10,5}\sqrt{1-\frac{3}{7}}} = 0,01838$$

Tanto per farsi un'idea della qualità dell'apoprossimazione, il risultato esatto è:$P_{\max} = 0,01809$, con un errore relativo dell'$1,6\%$.

Per la dimostrazione completa, ossia delle approssimazioni a tutte le probabilità e non solo a quella massima, si può generalizzare per esempio quella di Feller per il caso univariato (teorema di DeMoivre Laplace), ma sarebbe troppo lunga e inutile per i fini di questo testo. **citazione**