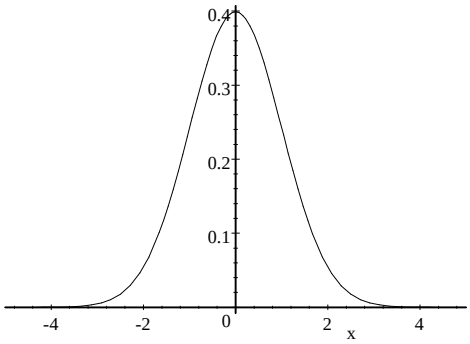
1 Variabili gaussiane

1.1 Gaussiane unidimensionali canoniche

Ricordiamo alcuni fatti elementari sulle gaussiane unidimensionali.

Si chiama gaussiana (normale) canonica $N(0,1)$ una v.a. con densità

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

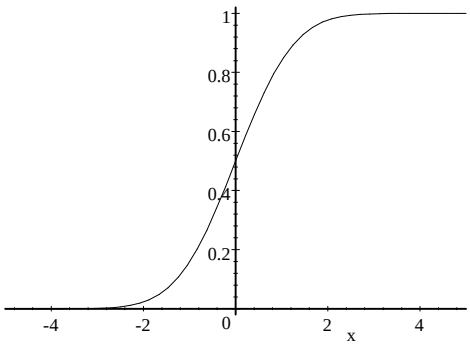


Gaussiana canonica

La sua funzione di distribuzione $F(x) = P(X \leq x)$, che indicheremo anche con $\Phi(x)$,

$$\Phi(x) = \int_{-\infty}^{x} f(t)dt$$

è tabulata su molti testi ed è calcolata da tutti i software.



Funzione di distribuzione gaussiana canonica

Data un'area (o probabilità) $\alpha \in (0,1)$, il numero q_α tale che $\Phi(q_\alpha) = \alpha$ si dice *quantile* (gaussiano canonico) di ordine α. Anch'esso è calcolato da tutti i software. Spesso al suo posto si usa il numero z_α tale che

$$\int_{z_\alpha}^{\infty} f(t)dt = \alpha.$$

1.2 Gaussiane generiche

Si chiama gaussiana (normale) $N(\mu, \sigma^2)$ una v.a. con densità

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}.$$

La sua forma è simile a quella della gaussiana canonica, ma è centrata nel punto μ ed ha un'ampiezza proporzionale a σ. La sua funzione di distribuzione si può calcolare a partire da quella canonica:

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} dt = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Si dimostra che

$$E[X] = \mu, \quad Var[X] = \sigma^2.$$

Se X è una gaussiana $N(\mu, \sigma^2)$, allora $aX+b$ è una gaussiana $N(a\mu + b, \sigma^2 a^2)$. Dando per buona la gaussianità di $aX+b$, si verifichi (con le usuali proprietà del valor medio) che i parametri sono quelli affermati.

Se X è una gaussiana $N(\mu, \sigma^2)$, la si può rappresentare nella forma $\sigma W + \mu$ rispetto ad una opportuna gaussiana canonica W:

$$X = \sigma W + \mu$$

(basta porre $W = \frac{Y-\mu}{\sigma}$ e verificare che W è $N(0,1)$, usando l'affermazione precedente). Geometricamente l'operazione $\sigma W + \mu$ si interpreta come una deformazione di ampiezza σ del grafico della densità di W seguita da una traslazione di ampiezza μ.

1.3 Il teorema limite centrale

Data una successione di v.a. $X_1, ..., X_n, ...$ indipendenti ed identicamente distribuite (i.i.d.), di media μ e varianza σ^2 finita, la v.a.

$$\frac{X_1 + ... + X_n - n\mu}{\sqrt{n}\sigma}$$

ha media 0 e varianza 1 (si verifica con le solite proprieta del valor medio). Il teorema limite centrale dice che essa è approssimativamente gaussiana $N(0,1)$, per n grande. Il significato rigoroso di questa frase è che per $n \to \infty$

$$P\left(a < \frac{X_1 + ... + X_n - n\mu}{\sqrt{n}\sigma} < b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

per ogni $a < b$.

L'uso pratico del teorema spesso viene effettuato direttamente sulla grandezza $X_1 + ... + X_n$. Essa si ottiene da $\frac{X_1+...+X_n-n\mu}{\sqrt{n}\sigma}$ moltiplicando per $\sqrt{n}\sigma$ e sommando poi $n\mu$. Quindi approssimativamente $X_1 + ... + X_n$ è come una gaussiana $N(0,1)$ moltiplicata per $\sqrt{n}\sigma$, più $n\mu$, ovvero una $N(n\mu, n\sigma^2)$. Si usa quindi enunciare il teorema anche nella seguente forma, a volte più vicina all'uso pratico:

data una successione di v.a. $X_1, ..., X_n, ...$ i.i.d., di media μ e varianza σ^2 finita, la v.a. $X_1 + ... + X_n$ è approssimativamente gaussiana $N(n\mu, n\sigma^2)$.

1.4 Gaussiane multivariate canoniche

Ricordiamo che una v.a. Y, gaussiana $N(\mu, \sigma^2)$, si può rappresentare nella forma $\sigma X + \mu$ rispetto ad una opportuna gaussiana canonica X. Seguiremo

questa strada per introdurre le gaussiane multivariate (o multidimensionali). Cominciamo quindi con le gaussiane canoniche multivariate.

Un vettore aleatorio $X = (X_1, ..., X_n)$ si dice *gaussiano canonico* se $X_1, ..., X_n$ sono v.a. gaussiane canoniche indipendenti. Usando il fatto che la densità congiunta di un vettore aleatorio avente componenti indipendenti è uguale al prodotto delle densità delle componenti (le marginali), la densità di un vettore gaussiano canonico è

$$\begin{aligned} f_X(x) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) &= \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \\ &= (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\langle x,x\rangle}, \qquad x = (x_1, ..., x_n). \end{aligned}$$

Qui e nel seguito indichiamo con $\langle x, y\rangle$ il prodotto scalare tra i vettori x ed y: $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$.

Il grafico, in dimensione $n = 2$, di questa funzione è una superficie a campana, invariante per rotazioni attorno all'asse verticale coordinato; ogni sezione di questa superficie passante per tale asse fornisce una curva che è il grafico a campana di una gaussiana canonica.

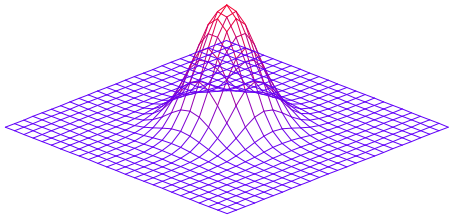


Fig. 3. Grafico di
$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)$.

1.5 Gaussiane multivariate generiche

Dato un vettore gaussiano canonico X a n dimensioni, una matrice A, $n \times k$ (quindi non necessariamente quadrata), ed un vettore m a k dimensioni,

consideriamo il vettore aleatorio $Y = AX + m$, che ha k dimensioni. Ogni vettore così ottenuto sarà detto *vettore gaussiano*. Intuitivamente abbiamo prima operato su X una trasformazione lineare, descritta dalla matrice A, trasformazione che può essere una rotazione unita ad una deformazione, poi abbiamo traslato di m il risultato. Ecco il grafico dopo una deformazione differenziata a seconda degli assi (i valori lungo un asse sono stati moltiplicati per un numero minore di uno, per cui si sono concentrati, mentre lungo l'altro asse sono stati moltiplicati per un numero maggiore di uno, aumentandone la dispersione):

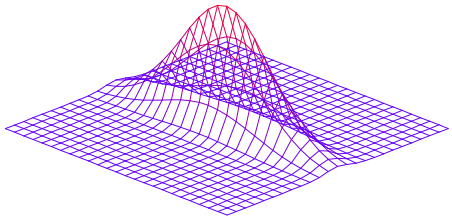


Fig. 4. Grafico di
$f(x,y) = \frac{1}{8\pi} \exp\left(-\frac{4x^2+\frac{1}{4}y^2}{2}\right)$.

Si rifletta sul significato geometrico dell'operazione AX (sempre nel caso bi-dimensionale, in cui la visualizzazione è semplice). Ogni punto x del piano viene mandato nel punto Ax. Di conseguenza ogni possibile realizzazione x del vettore aleatorio X subisce questa trasformazione. La densità di probabilità, che descrive la probabilità attribuita alle diverse realizzazioni, si modifica di conseguenza, concentrandosi o dilatandosi a seconda della trasformazione subita dai punti x del piano.

Molto utile è visualizzare la trasformazione subita dalle *curve di livello* della densità. Prendiamo una curva di livello della densità canonica $f_X(x)$, ovvero una curva del piano descritta dall'equazione $f_X(x) = C$, con C costante positiva. Per la forma della densità canonica, questa curva è anche descritta dall'equazione $\sum_{i=1}^{2} x_i^2 = C'$ per una opportuna costante positiva C', quindi è una circonferenza centrata nell'origine, fatto comunque ovvio dalla visualizzazione grafica di $f_X(x)$. Quindi le curve di livello della den-

sità gaussiana canonica sono circonferenze concentriche. Quando operiamo la trasformazione lineare descritta dalla matrice A, per la quale i punti x si trasformano nei punti Ax, le circonferenze centrate nell'origine si trasformano in *ellissi* centrate nell'origine. Queste sono le curve di livello della densità del vettore aleatorio AX.

Queste affermazioni sono vere se A è non degenere (invertibile), altrimenti il piano viene addirittura trasformato in una retta dall'applicazione A, e l'interpretazione geometrica, pur possibile, è meno semplice da descriversi a parole. In conclusione, se A è non degenere, le curve di livello della densità del vettore aleatorio AX sono ellissi concentriche. Si riguardino i grafici precedenti sotto questa luce.

Chiariamo un punto per capire ancor meglio l'immagine grafica. Abbiamo detto che A trasporta circonferenze in ellissi. I punti x tali che $\sum_{i=1}^{2} x_i^2 = C'$ si trasformano nei punti $y = Ax$. Questi punti y hanno la proprietà che le controimmagini $x = A^{-1}y$ soddisfano l'equazione precendente, quindi

$$\sum_{i=1}^{2} \left(\left(A^{-1}\right)_{ij} y_j \right)^2 = C'.$$

Detta Λ la matrice $\left(A^{-1}\right)^T A^{-1}$, l'equazione precedente ha la forma

$$\langle \Lambda y, y \rangle = C'.$$

Intanto è chiaro che questa è l'equazione di un'ellisse, essendo Λ una matrice simmetrica definita positiva. I suoi assi sono dati dai due autovettori della matrice simmetrica Λ (o da due qualsiasi degli autovettori nel caso particolare in cui Λ fosse un multiplo dell'identità). Siccome gli autovettori di Λ coincidono con gli autovettori di Λ^{-1} (come fatto generale di algebra lineare), e siccome $\Lambda^{-1} = AA^T$, arriviamo all'importantissima conclusione: *gli assi delle ellissi* (curve di livello della densità di AX) *sono gli autovettori di* AA^T. Ricordiamo che ci stiamo riferendo al caso di A non degenere. Questi autovettori rappresentano una base privilegiata in cui leggere le informazioni contenute nel vettore gaussiano (su di essi si basa ad esempio il metodo delle *componenti principali*, di grande interesse applicativo). Torneremo tra breve su questo argomento, quando interpreteremo AA^T come covarianza di X.

Concludiamo questo paragrafo eneunciando un fatto molto importante, ovvia conseguenza della definizione da noi scelta: se Y è un vettore gaussiano k-dimensionale, B è una matrice $k \times h$ e b è un vettore h-dimensionale,

allora $Z = BY + b$ è un vettore gaussiano h-dimensionale. In altre parole, **trasformazioni lineari o affini di vettori gaussiani producono vettori gaussiani**. La dimostrazione è ovvia: Y si può scrivere, per definizione, nella forma $AX + m$, con X gaussiano canonico. Allora vale

$$Z = B(AX + m) + b = (BA)X + (Bm + b) = A'X + m'$$

con $A' = BA$ e $m' = Bm + b$. Quindi Z è gaussiano.

Quanto detto fino ad ora dovrebbe chiarire lo stretto legame che sussiste tra la teoria dei vettori gaussiani e l'algebra lineare. La teoria gaussiana è in un certo senso la parte "lineare" del calcolo delle probabilità.

1.6 Densità di probabilità di un vettore gaussiano

Poniamo $Q = AA^T$. Quando la matrice Q è invertibile ($\det Q \neq 0$) si può dimostrare che la densità di un vettore gaussiano $Y = AX + m$, X canonico, è data da

$$f_Y(x) = \frac{1}{\sqrt{(2\pi)^n \det Q}} \exp\left(-\frac{1}{2}\left\langle Q^{-1}(x-m), x-m\right\rangle\right) \qquad x \in \mathbb{R}^n. \tag{1}$$

In analogia col caso unidimensionale, diremo che il vettore aleatorio Y è gaussiano $N(m, Q)$.

La matrice Q è una matrice simmetrica, ed è definita positiva nelle ipotesi precedenti:

$$\langle Qx, x\rangle > 0 \qquad \forall x \neq 0, \qquad x \in \mathbb{R}^n.$$

Osservazione 1. Se $n = 1$, Q è un singolo numero reale positivo, quindi della forma

$$Q = \sigma^2,$$

ed analogamente m è un singolo numero reale. Vale $\det Q = \sigma^2$, per cui $f_X(x)$ assume la forma

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \qquad x \in \mathbb{R}.$$

Questa è l'usuale densità gaussiana $N(m, \sigma^2)$. Quindi l'espressione $f_X(x)$ definita dalla formula (1) è una generalizzazione del caso unidimensionale.

1.7 Media e covarianza di un vettore gaussiano

Proseguendo l'osservazione precedente, nel caso uni-dimensionale m e $Q = \sigma^2$ sono media e varianza di X. Nel caso multidimensionale vale un fatto simile (omettiamo la dimostrazione, ma si veda la prossima osservazione):

$$m_i = E[X_i], \qquad Q_{ij} = Cov(X_i, X_j), \tag{2}$$

ovvero m è il vettore dei valori medi mentre Q è la *matrice di covarianza*. Ricordiamo che

$$Cov(X_i, X_j) = E[(X_i - m_i)(X_j - m_j)],$$

per cui in particolare gli elementi della diagonale sono le varianze:

$$Q_{ii} = Var[X_i].$$

Osservazione 2. Supponiamo che $X_1, ..., X_n$ siano v.a. gaussiane indipendenti, X_i con legge $N(m_i, \sigma_i^2)$. Allora la densità congiunta del vettore $X = (X_1, ..., X_n)$ è data dal prodotto delle marginali, per cui

$$\begin{aligned} f_X(x_1, ..., x_n) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \frac{1}{\sqrt{(2\pi)^n \sigma_1^2 \cdots \sigma_n^2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{(x_i - m_i)^2}{\sigma_i^2}\right). \end{aligned}$$

Introduciamo il vettore $m = (m_1, ..., m_n)$ e la matrice

$$Q = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \sigma_n^2 \end{pmatrix}. \tag{3}$$

Vale $\det Q = \sigma_1^2 \cdots \sigma_n^2$ e $Q_{ij}^{-1} = \sigma_i^2 \delta_{ij}$, per cui $\sum_{i=1}^{n} \frac{(x_i - m_i)^2}{\sigma_i^2} = \langle Q^{-1}(x - m), x - m \rangle$. Pertanto $f_X(x)$ ha la forma (1). In questo caso particolare, è automaticamente dimostrata l'affermazione (2).

Osservazione 3. Il conto appena fatto si presta ad una visualizzazione geometrica notevole. Restringiamoci al caso $n = 2$, $m = 0$. La funzione $f_X(x) = f_{X_1}(x_1) f_{X_2}(x_2)$ ha un grafico "a campana", ottenuto "modulando"

la curva a campana $f_{X_1}(x_1)$ mediante la $f_{X_2}(x_2)$, o viceversa. Le curve di livello $f_X(x) = C^2$ sono ellissi rispetto agli assi coordinati, concentriche, coincidenti con le curve di livello della forma quadratica $\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}$.

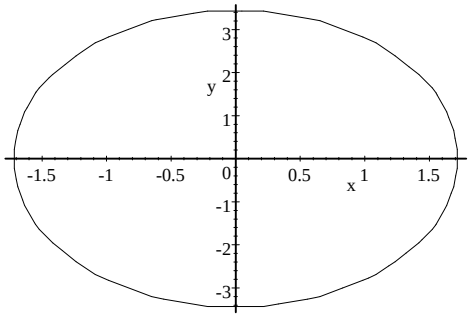


Fig. 5. Luogo dei punti $x^2 + \frac{1}{4}y^2 = 3$.

Nota: possiamo ora chiarire il seguente fatto fondamentale: *due v.a. X ed Y congiuntamente gaussiane* (ovvero tali che il vettore (X, Y) è gaussiano) *sono indipendenti se e asolo se sono non correlate*: $Cov(X, Y) = 0$. Un'implicazione è vera anche senza l'ipotesi di gaussianità. Per l'altra implicazione, sotto l'ipotesi di gaussianità, se vale $Cov(X, Y) = 0$, allora la matrice di covarianza Q del vettore gaussiano (X, Y) è una matrice diagonale, quindi vale

$$\begin{aligned} f(x, y) &= \frac{1}{\sqrt{(2\pi)^2 \sigma_X^2 \sigma_Y^2}} \exp\left(-\frac{(x - m_X)^2}{2\sigma_X^2} - \frac{(y - m_Y)^2}{2\sigma_Y^2} \right) \\ &= f_X(x) f_Y(y). \end{aligned}$$

Pertanto X ed Y sono indipendenti. Qui abbiamo usato il fatto generale che due v.a. X ed Y sono indipendenti *se e solo se* la loro densità congiunta $f(x, y)$ è uguale al prodotto delle due densità marginali $f_X(x) f_Y(y)$.

Osservazione 4 (interpretazione fondamentale degli autovettori ed autovalori di Q). Possiamo riformulare un fatto espresso in precedenza per la matrice AA^T in termini della covarianza Q (in quanto coincide con AA^T). La matrice simmetrica definita positiva Q possiede un sistema di autovettori $e_1, ..., e_n$ ortonormali che costituiscono una base di $\mathbb{R}^n$ (tale sistema non è sempre unico: si pensi al caso $Q = id$). *Questi autovettori sono gli assi*

degli ellissoidi che si ottengono come insiemi di livello della densità $f_X(x)$. La visualizzazione (anche solo mentale) di questi ellissoidi fa capire in quali direzioni la gaussiana X ha maggior dispersione. In molte applicazioni (che vedremo) le direzioni con maggior dispersione sono quelle con maggior energia o potenza media. Gli autovettori $e_1, ..., e_n$ giocano quindi un ruolo molto importante nello studio della distribuzione di energia o potenza tra le diverse componenti.

Inoltre, operando una trasformazione ortogonale di $\mathbb{R}^n$, si può diagonalizzare Q. In altre parole, se ci si pone nel sistema di coordinate ortonormali $e_1, ..., e_n$, la matrice Q assume la forma diagonale (3). Si vede allora che *gli autovalori* σ_i^2 *di* Q *misurano proprio la dispersione di* X *nelle direzioni* e_i.

Concludiamo enunciando un risultato fondamentale sulle trasformazioni lineari di vettori gaussiani. Abbiamo già visto che se X è un vettore gaussiano n-dimensionale $N(m_X, Q_X)$, B è una matrice $n \times k$ e b è un vettore k-dimensionale, allora il vettore aleatorio $Y = BX + b$ è gaussiano. Si dimostra che la sua media m_Y e matrice di covarianza Q_Y sono date da:

$$m_Y = Bm_X, \qquad Q_Y = BQ_XB^T.$$

In particolare si rifletta sul caso in cui B è la matrice che esprime il cambio di base che diagonalizza Q_X.