# Probablità, Statistica e Processi Stocastici

Franco Flandoli, Università di Pisa

Corso per la Scuola di Dottorato in Ingegneria

# Programma del corso

- vettori gaussiani, PCA ed fPCA
- altri metodi di analisi e previsione di serie storiche
- catene di Markov e simulazioni MCMC
- equazioni differenziali stocastiche

# PCA ed fPCA

- PCA = principal component analysis
- fPCA = functional principal component analysis

Sono due tecniche statistiche, basate su interessanti teorie matematiche, la prima adatta all'esplorazione di tabelle, la seconda all'esplorazione di serie storiche, immagini ed altro
N. ore: circa 6

# Altri metodi di analisi e previsione di serie storiche

- Metodi regressivi (simili ai modelli ARIMAX)
- Holt Winters e suoi precursori (smorzamento esponenziale, versione con trend)
- corredati di altri elementi utili, come acf, algoritmi di decomposizione.

N. ore: circa 4

# Catene di Markov e simulazioni MCMC

- un po' di teoria, teorema ergodico e di convergenza all'equilibrio
- algoritmo di Metropolis, misure di Gibbs
- esempio: il modello di Ising

N. ore: circa 5

# Equazioni differenziali stocastiche

- elementi di teoria dei processi stocastici
- moto browniano
- integrali stocastici e calcolo stocastico
- equazioni differenziali stocastiche
- simulazioni

N. ore: circa 5-9

# Software

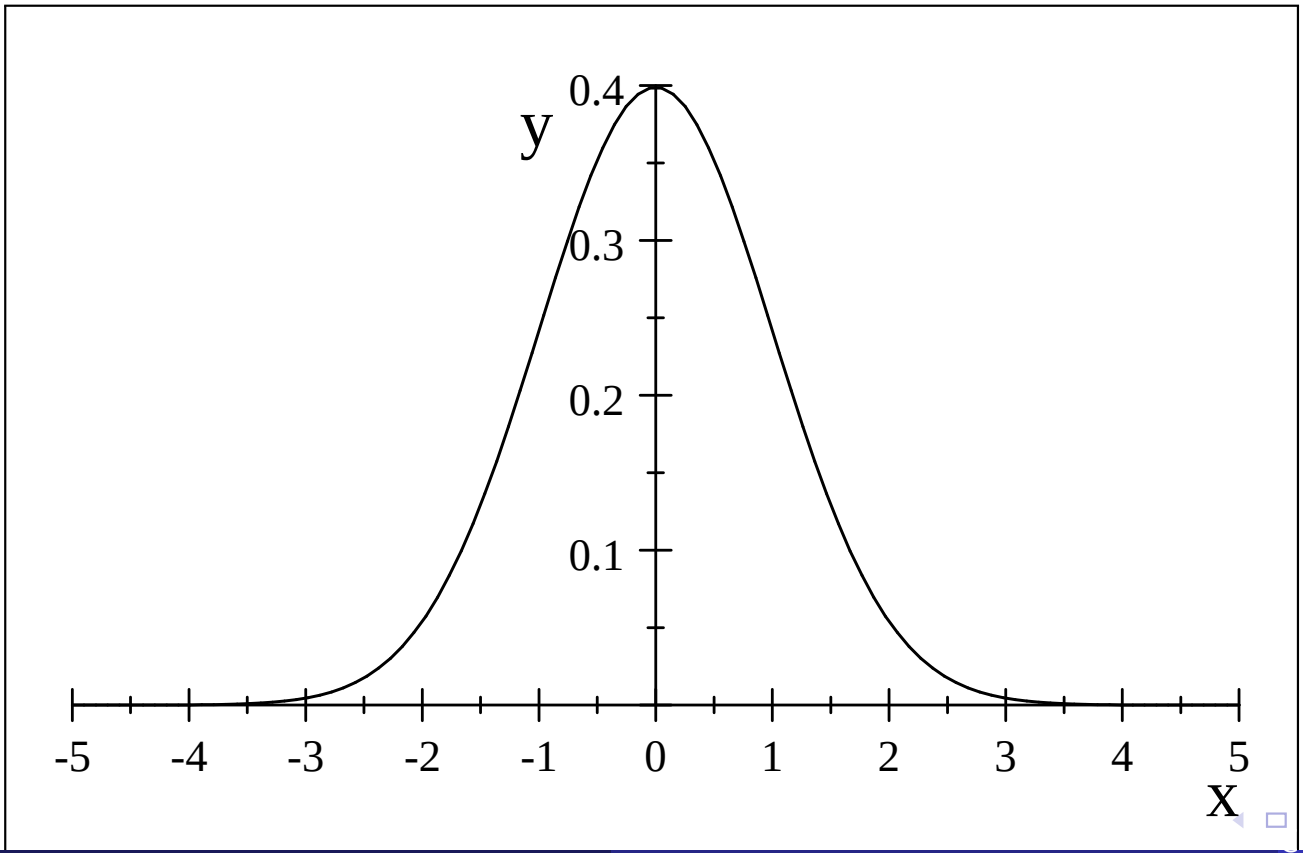
Tutto il corso sarà basato sull'uso del software R.

Scaricarlo da rete (gratuito).

# PCA (principal component analysis)

L'idea geometrica parte dalle gaussiane multidimensionali.
Iniziamo con un breve richiamo sulle gaussiane unidimensionali.
Densità gaussiana standard

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



# Gaussiane unidimensionali

La gaussiana standard ha media 0 e deviazione standard 1.
Se $Z$ è una variabile aleatoria gaussiana standard, la nuova variabile aleatoria

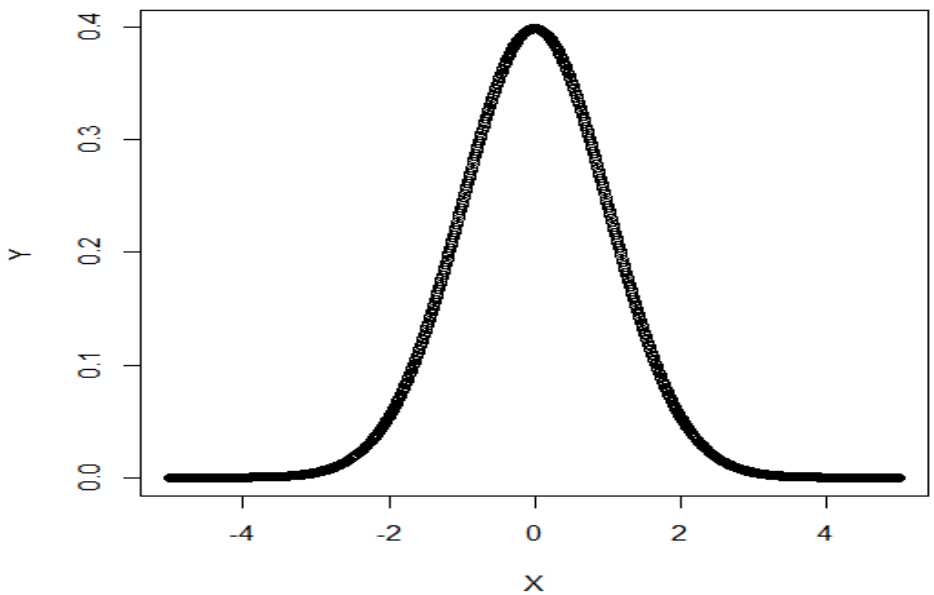
$$X = \sigma Z + \mu$$

ha media $\mu$ e deviazione standard $\sigma$ (la sua densità si scrive facilmente, vedi Wiki). Verrà abbreviata $N\left(\mu, \sigma^2\right)$.
Col software R possiamo generare campioni $N\left(\mu, \sigma^2\right)$, col comando:
`rnorm(n,mu,sig)`

```
> rnorm(10,7,1)
[1] 6.349231 7.608871 7.474523 7.664420 5.027992 6.748387 6.341020
6.948569
[9] 7.359332 7.368808
```

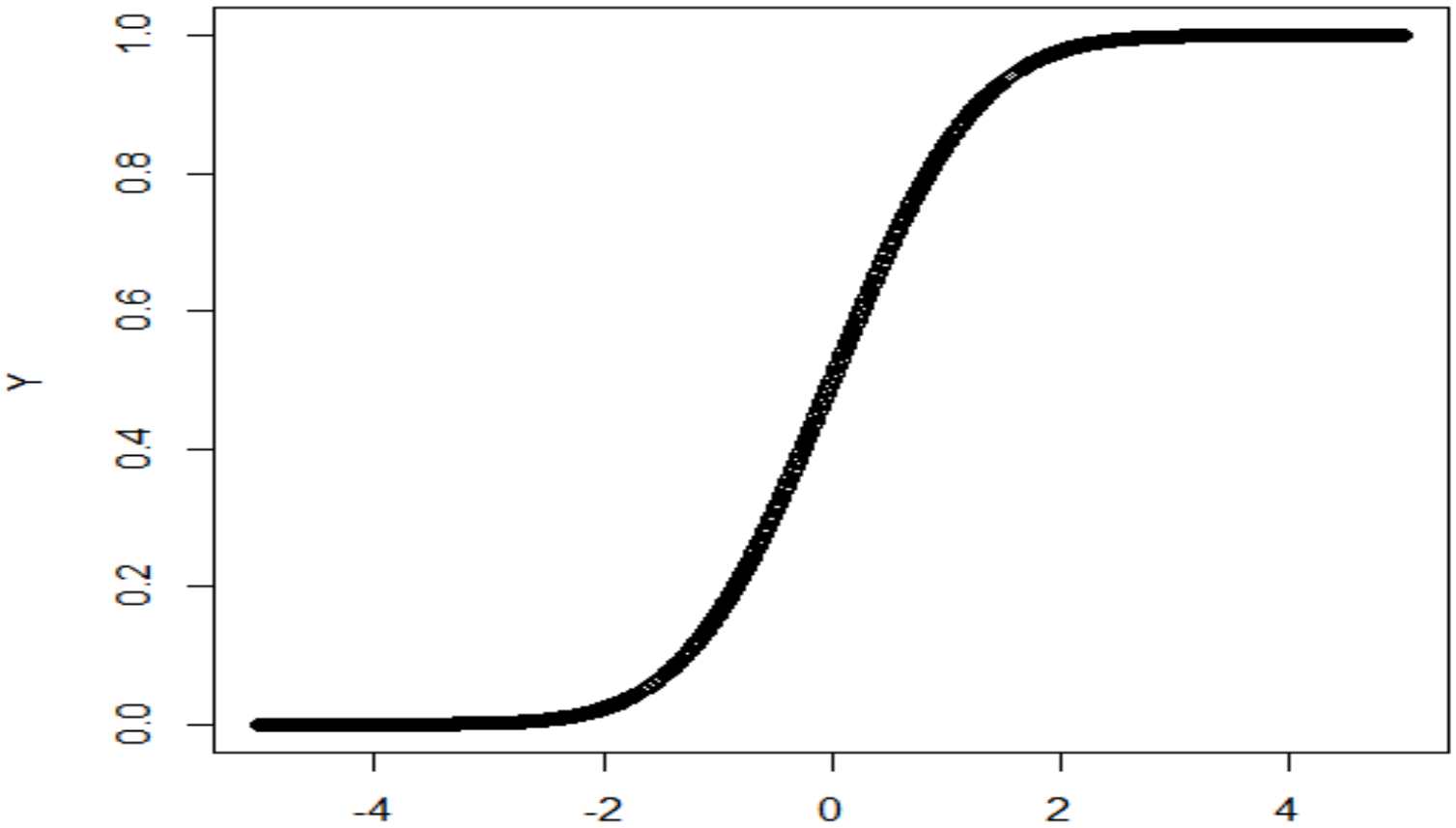
# Gaussiane unidimensionali

Coi comandi dnorm e pnorm si ottengono densità e cumulativa:
X =seq(-5,5,0.01)
Y= dnorm(X)
plot(X,Y,asp=8)



# Gaussiane unidimensionali

X =seq(-5,5,0.01)
Y= pnorm(X)
plot(X,Y)



# Gaussiana multidimensionale standard

Densità gaussiana (o normale) standard in dimensione $n$:

$$\begin{aligned} f(x_1, ..., x_n) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}} \\ &= (2\pi)^{-n/2} \exp\left( -\frac{1}{2} \left( x_1^2 + ... + x_n^2 \right) \right) \end{aligned}$$

o in notazione vettoriale

$$f(\mathbf{x}) = (2\pi)^{-n/2} \exp\left( -\frac{1}{2} \|\mathbf{x}\|^2 \right), \qquad \mathbf{x} = (x_1, ..., x_n).$$

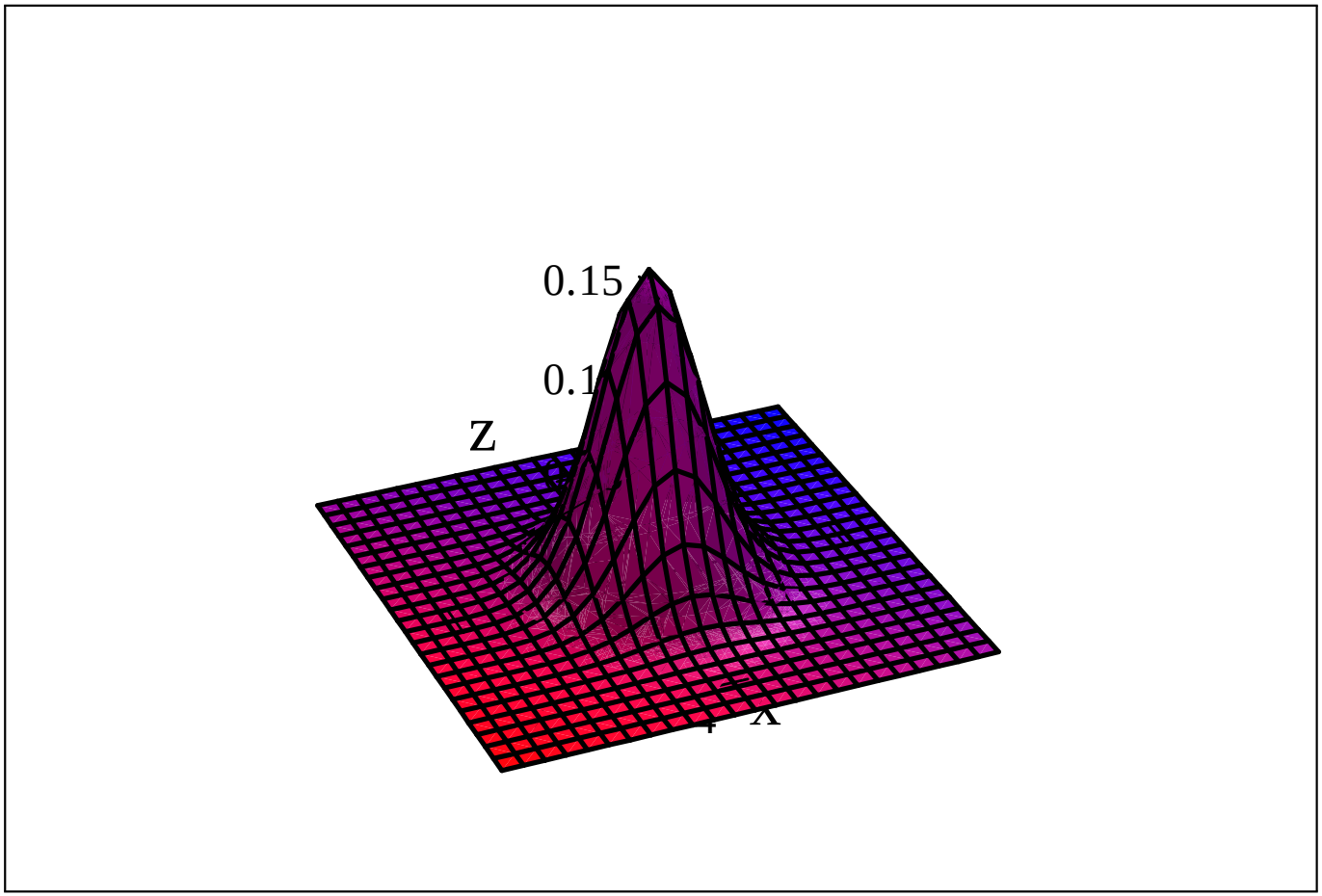
Questa è la densità di un *vettore aleatorio*

$$\mathbf{Z} = (Z_1, ..., Z_n)$$

fatto di componenti $Z_i$ gaussiane standard e indipendenti.

# Gaussiana bidimensionale standard

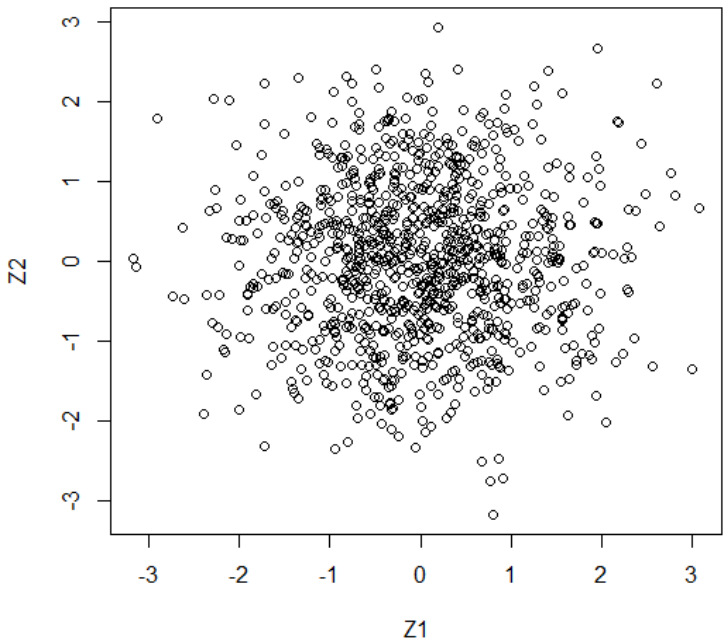
$f(x,y) = (2\pi)^{-1} \exp\left(-\frac{1}{2}\left(x^2+y^2\right)\right)$. Curve di livello: $x^2+y^2=R^2$ (circonferenze, centrate nell'origine)




(ottenuto con Maple)

# Simulazione di punti gaussiani standard nel piano

n=1000
Z1=rnorm(n)
Z2=rnorm(n)
plot(Z1,Z2)



# Gaussiana multidimensionale generica

Se $\mathbf{Z} = (Z_1, ..., Z_n)$ è un vettore gaussiano standard, $A : \mathbb{R}^n \to \mathbb{R}^k$ è una matrice, $\boldsymbol{\mu} \in \mathbb{R}^k$ è un vettore (deterministico), allora

$$\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$$

è un vettore gaussiano qualsiasi. I vettori gaussiani sono tutti e soli quelli ottenibili in questo modo.
Posto $Q = AA^T$, se $\det Q \neq 0$, allora il vettore $\mathbf{X}$ ha densità

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det Q}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T Q^{-1}(\mathbf{x}-\boldsymbol{\mu})\right).$$

Si verifica che $\boldsymbol{\mu}$ è il vettore delle medie (delle componenti di $\mathbf{X}$) e $Q$ è la *matrice di covarianza*.

# Visualizzazioni gaussiana bidimensionale generica

Il grafico della densità non è molto efficace. Proviamo a titolo di esempio.

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\mu} = 0$$

ovvero

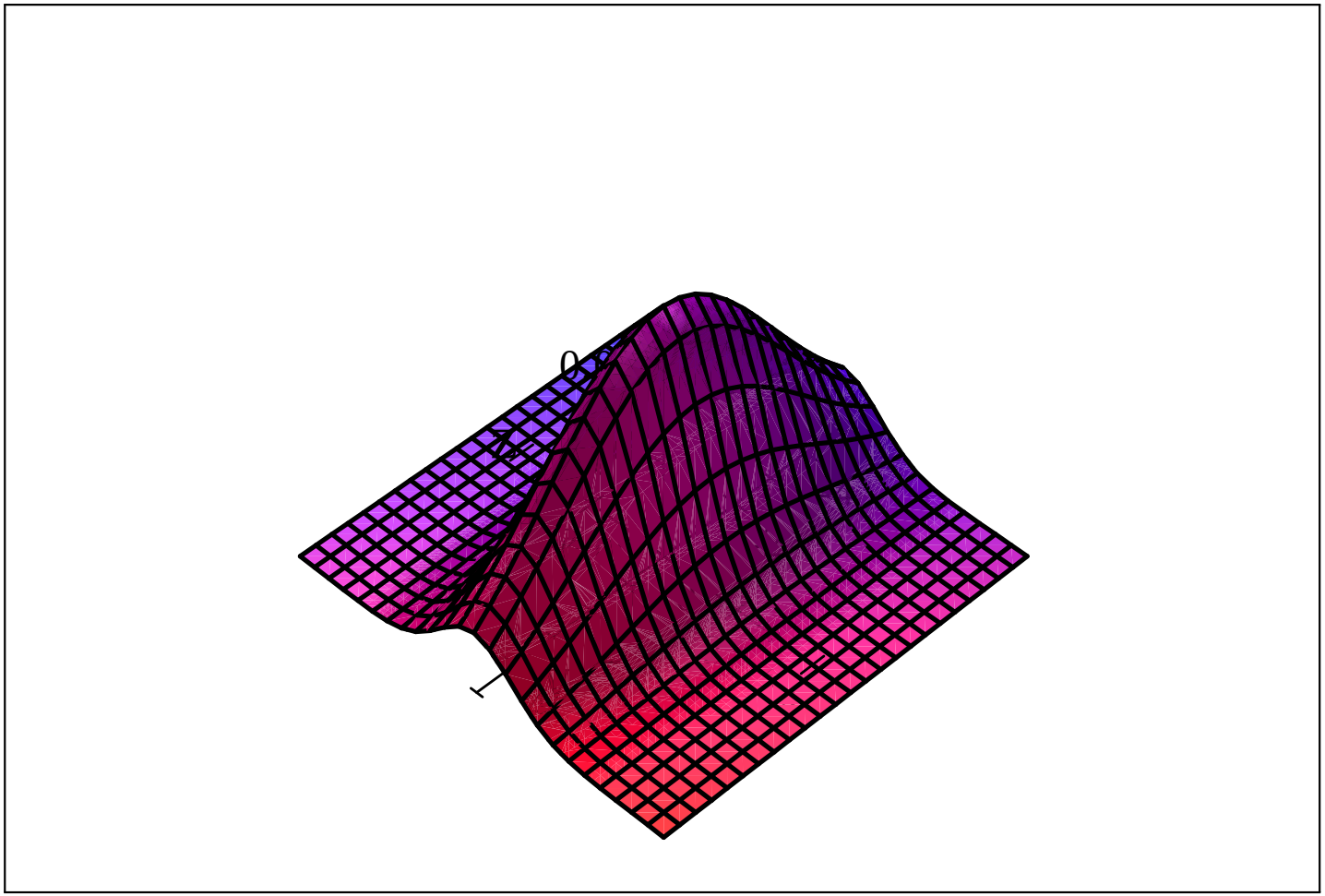
$$\begin{aligned} X_1 &= 3Z_1 \\ X_2 &= Z_2. \end{aligned}$$

$$Q = AA^T = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q^{-1} = \begin{pmatrix} 1/9 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \det Q = 9$$

$$f(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 \cdot 9}} \exp\left(-\frac{1}{2}\left(\frac{x_1^2}{9} + x_2^2\right)\right).$$

# Esempio di gaussiana bidimensionale

$$\frac{1}{\sqrt{(2\pi)^2 \cdot 9}} \exp\left(-\frac{1}{2}\left(\frac{x_1^2}{9} + x_2^2\right)\right)$$



Si immagini un esempio analogo con rotazione:

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}.$$

# Curve di livello

I calcoli sulla densità in genere sono difficili. Più trasparente è visualizzare le curve di livello:

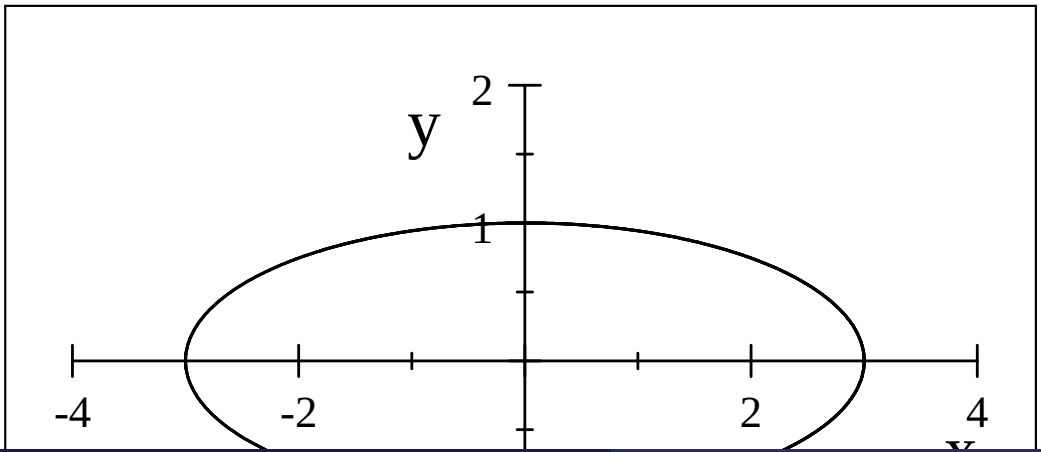
$$\left\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = a\right\}.$$

Si vede immediatamente che hanno la forma

$$\left\{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}-\boldsymbol{\mu})^T Q^{-1} (\mathbf{x}-\boldsymbol{\mu}) = r_a^2\right\}.$$

Questi sono ellissoidi. Ad esemio, per $A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{\mu} = 0$, sono le ellissi della forma

$$\left(\frac{x_1}{3}\right)^2 + x_2^2 = r_a^2.$$

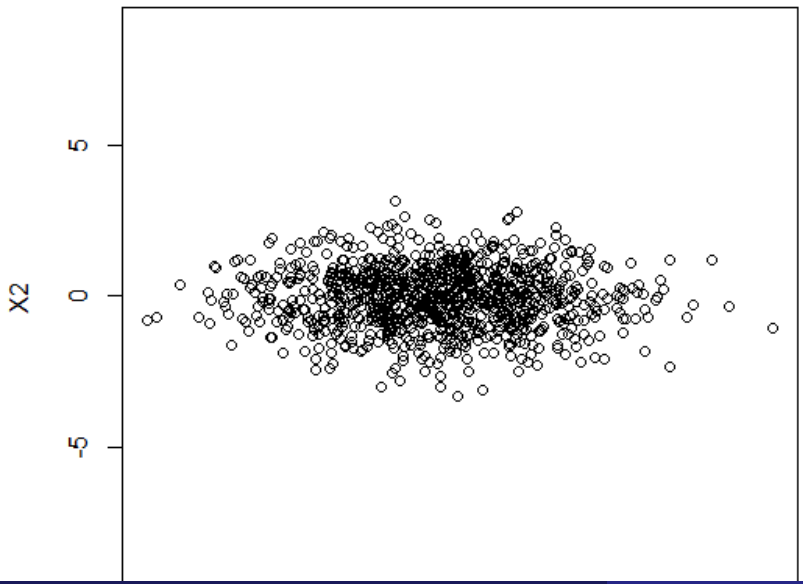


# Visualizzazione tramite punti aleatori

Ricordiamo che i comandi `n=1000; Z1=rnorm(n); Z2=rnorm(n); plot(Z1,Z2)` generavano punti gaussiani standard nel piano. Essendo $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$, per generare punti secondo la gaussiana $\mathbf{X}$, basta applicare la trasformazione affine ai punti standard generati sopra. Esempio:

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\mu} = 0, \qquad \begin{cases} X_1 = 3Z_1 \\ X_2 = Z_2 \end{cases}$$

X1=3*Z1; X2=Z2; plot(X1,X2,asp=1)



# Si parte da A o da Q?

- Nella nostra definizione, un vettore gaussiano è definito tramite una matrice $A$: $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$.
- Poi però la densità è definita tramite $Q$.
- Se si parte da $A$, si calcola $Q$ con la formula $Q = AA^T$.
- Se si parte da $Q$ e serve conoscere $A$, bisogna trovare una matrice $A$ tale che $AA^T = Q$.
- Negli esempi/applicazioni, di solito si parte da $Q$, è nota $Q$.

A questo scopo però è necessario capire meglio il concetto di *matrice di covarianza*.

# Matrice di covarianza

Dato un vettore aleatorio (anche non gaussiano) $\mathbf{X} = (X_1, ..., X_n)$, chiamiamo sua media il vettore

$$\boldsymbol{\mu} = (E[X_1], ..., E[X_n])$$

e matrice di covarianza la matrice $n \times n$ di componenti

$$Q_{ij} = Cov(X_i, X_j), \qquad i, j = 1, ..., n$$

Ricordiamo che

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ & \quad E[XY] - E[X]E[Y]. \end{aligned}$$

e che questa operazione è lineare in entrambi i sui argomenti.

**Theorem**

*La matrice $Q$ è simmetrica, definita non negativa.*

# Matrice di covarianza

- Media e matrice di covarianza riassumono solo alcune caratteristiche statistiche di un vettore aleatorio (come media e varianza di una variabile aleatoria).
- Però per i vettori gaussiani racchiudono tutte le informazioni.
- Gli elementi sulla diagonale di $Q$ sono le varianze delle componenti, $Q_{ii} = Cov(X_i, X_i) = Var[X_i]$ quindi catturano la dispersione delle singole componenti.
- Gli elementi fuori dalla diagonale $Q_{ij}$, per $i \neq j$, catturano in una certa misura il legame, la dipendenza reciproca, tra le variabili $X_i$ e $X_j$. (Modulato dalla dispersione.)
- Più precisamente, per esaminare il legame andrebbe usato il coefficiente di correlazione, che non dipende dalla scala

$$\rho(X_i, X_j) = \frac{Cov(X_i, X_i)}{\sqrt{Var[X_i]\, Var[X_j]}}.$$

# Dati sperimentali multidimensionali

Esamineremo nelle esercitazioni una tabella del tipo

| | SC | SA.SC | TD |
|---|---|---|---|
| Piem | 0.471 | -0.707 | -0.607 |
| Vaos | 0.348 | -0.642 | -0.813 |
| Lomb | 1.397 | -0.836 | -0.790 |
| TrAA | 0.435 | -1.269 | -0.966 |
| ... | ... | ... | ... |

dove SC sono le Spese Complessive per famiglia di quella regione in un certo anno, SA.SC le Spese per Alimenti rispetto al totale delle spese, TD il Tasso di Disoccupazione. Il vettore aleatorio è

$$\mathbf{X} = (X_1, X_2, X_3) = (SC, SA.SC, TD)\,.$$

# Fare un modello

Supponiamo di voler "conoscere la densità di probabilità" di $\mathbf{X}$.
Più realisticamente, supponiamo di voler trovare una densità di probabilità $f(x_1, x_2, x_3)$ che "descriva ragionevolmente" i dati.
Per semplicità, decidiamo di cercarla gaussiana.
Calcolare (più precisamente "stimare") direttamente $A$ dai dati non è naturale.
Invece possiamo stimare $Q$ e $\mu$, usando gli usuali stimatori empirici di media e covarianza.
Con R, per stimare $Q$, basta usare il comando $cov(B)$, dove $B$ è il nome della tabella.
In questo senso dicevamo sopra che negli esempi di solito si conosce $Q$ (in genere empirica), non $A$.

# Generazione di punti aleatori nota Q

Sopra abbiamo visto il modo banale di generare punti aleatori nota $A$ (basta generare punti standard ed applicare $A$).
Se è nota $Q$, bisogna risolvere l'equazione $AA^T = Q$. Essa ha infinite soluzioni. Quella "canonica" è

$$A = \sqrt{Q}$$

definita come quell'unica matrice simmetrica e definita positiva che al quadrato dà $Q$. Come la si calcola?
Dopo averla calcolata, basta generare punti standard $z$ ed applicarci $\sqrt{Q}$.

# Decomposizione spettrale

- Essendo simmetrica, $Q$ è diagonalizzabile: esiste una base ortonormale $\mathbf{e}_1, ..., \mathbf{e}_n$ di $\mathbb{R}^n$ fatta di autovettori di $Q$,

$$Q\mathbf{e}_i = \lambda_i \mathbf{e}_i$$

e, posto

$$Q_e = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}, \qquad U = \begin{pmatrix} \mathbf{e}_1 & \dots & \mathbf{e}_n \\ & & \end{pmatrix}$$

vale

$$Q = U Q_e U^T.$$

- Essendo $Q$ positiva, gli autovalori $\lambda_i$ risultano positivi.

# Definizione di radice quadrata

Ricordando $Q = UQ_eU^T$, basta ora porre

$$\sqrt{Q} = U\sqrt{Q_e}U^T.$$

Si verifica subito che $\sqrt{Q}\sqrt{Q} = Q$ e che $\sqrt{Q}$ è simmetrica e definita positiva.
Ecco i comandi di R:

```
e = eigen(Q)
U = e$vectors
B = U %*% diag(sqrt(e$values)) %*% t(U)
```

# Esercizio riassuntivo

Svolgiamo insieme un esercizio riassuntivo.

1. Creiamo un vettore di punti gaussiani generati con R, che corrispondano ad una gaussiana dilatata in una direzione e ruotata di $\pi/10$
2. Di tali punti calcoliamo la matrice di covarianza empirica, che chiameremo $Q$
3. Di $Q$ calcoliamo la radice quadrata $\sqrt{Q}$
4. Usando $\sqrt{Q}$, simuleremo punti gaussiani aventi covarianza $Q$
5. osserveremo ad occhio la somiglianza coi punti creati all'inizio (volendo ci sarebbero dei test statistici)

L'esercizio si trova svolto in una scheda a parte.

# Verso PCA

- Supponiamo di avere dei punti nello spazio 3D.
- Ad esempio, supponiamo che siano stati generati da una gaussiana tridimensionale.
- Se vogliamo vederli nel modo più "aperto" ("sparpagliato") possibile, meno "sovrapposto", come dobbiamo ruotare il disegno?
- Immaginate che i punti siano una nuvola ellissoidale; l'ellissoide avrà un asse più lungo, un secondo asse più lungo ed un terzo asse più corto. Se ci mettiamo nel piano dei due assi più lunghi abbiamo la visuale più sparpagliata possibile.
- Come si trovano gli assi dell'ellisoide?

# Le superfici di livello di una gaussiana qualsiasi

**Theorem**

*Supponiamo* $\det Q \neq 0$. *Allora le superfici di livello della densità*

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det Q}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T Q^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

*ovvero gli insieli della forma*

$$\left\{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}-\boldsymbol{\mu})^T Q^{-1}(\mathbf{x}-\boldsymbol{\mu}) = r_a^2\right\}$$

*sono ellissoidi aventi come assi gli autovettori* $\mathbf{e}_1, ..., \mathbf{e}_n$ *di* $Q$, *e lunghezze degli assi i numeri* $\sqrt{\lambda_1}, ..., \sqrt{\lambda_n}$, *dove* $\lambda_1, ..., \lambda_n$ *sono gli autovalori corrispondenti a* $\mathbf{e}_1, ..., \mathbf{e}_n$. *Useremo la convenzione che sia* $\lambda_1 \geq ... \geq \lambda_n$.

# Verso PCA

Dal precedente teorema è chiaro come agire:

- dato un insieme di puntui nello spazio $\mathbb{R}^n$
- si calcola la loro matrice di covarianza empirica $Q$
- si trova il suo spettro $\mathbf{e}_1, ..., \mathbf{e}_n$, $\lambda_1, ..., \lambda_n$
- si decide che il piano individuato da $\mathbf{e}_1, \mathbf{e}_2$ è quello che fornisce la miglior visione dei punti.
- L'ipotesi di gaussianità vedremo che non serve, strettamente parlando. Però ci ha fornito in modo facile e geometricamente intuitivo il risultato.

# Il metodo PCA

- Data una tabella (si veda l'esempio nella scheda a parte)
- le righe vengono interpretate come punti di uno spazio $\mathbb{R}^n$
- il metodo PCA calcola la matrice di covarianza empirica $Q$ della tabella (cioè di tali punti)
- e la decomposizione spettrale di $Q$. Tutto questo col comando `princomp(B)`
- col comando `biplot` si ottiene il piano individuato da $\mathbf{e}_1, \mathbf{e}_2$ con disegnati i punti di partenza, corrispondenti alle righe ("gli individui") della tabella.
- Nelle lezioni successive vedremo un po' di teoria e di pratica in più su PCA e la sua versione funzionale, fPCA.