R - Esercitazione 3

Lorenzo Di Biagio
dibiagio@mat.uniroma3.it

Università Roma Tre

Venerdì 25 Ottobre 2013

Grafica (1)

I comandi grafici in R si distinguono in alto e basso livello.

I comandi di alto livello (`plot`, `curve`, `hist`, `boxplot`, `pie` ,...) creano un nuovo grafico sulla finestra grafica.

I comandi di basso livello (`lines` ,`points`, `abline`, `text`, `legend` ,...) aggiungono parti ad un grafico già esistente.

Con l'opzione `add=T` si forza un comando di alto livello ad agire come un comando di basso livello.

Alcuni parametri grafici:

- `bty`: tipo di box attorno al grafico ("n" - nessun box)
- `col`: colore (`colors()` dà lista dei nomi dei colori)
- `lty`: tipo di linea (intera, tratteggiata,...)
- `lwd`: spessore della linea
- `yaxt`, `xaxt`: tipo di asse ("n" - nessun asse)

Grafica (2)

RStudio ha un proprio dispositivo grafico.

Si possono utilizzare altri dispositivi grafici: ad esempio `quartz()` apre una finestra grafica in ambiente OS X.

Con `dev.cur()` si può conoscere il dispositivo grafico in uso.

Con `pdf(file="nome.pdf",paper="a4")` si apre il dispositivo grafico pdf. I grafici saranno salvati sul file nome.pdf nella directory di lavoro. Quando si vuole chiudere il file usare `dev.off()`.

Pacchetti aggiuntivi

Per installare pacchetti aggiuntivi utilizzare la finestra di RStudio, oppure direttamente la funzione `install.packages("nome pacchetto")`.

Per caricare pacchetti già installati utilizzare la funzione `library("nome pacchetto")`

Esempio: installare dal CRAN e caricare il pacchetto "actuar". Il pacchetto contiene, tra le altre cose, le funzioni generatrici dei momenti di alcune distribuzioni di probabilità. `mgfnorm` `mgfunif` sono le funzioni generatrici dei momenti della normale e dell'uniforme, rispettivamente.

Distribuzione Chi-Quadrato (1)

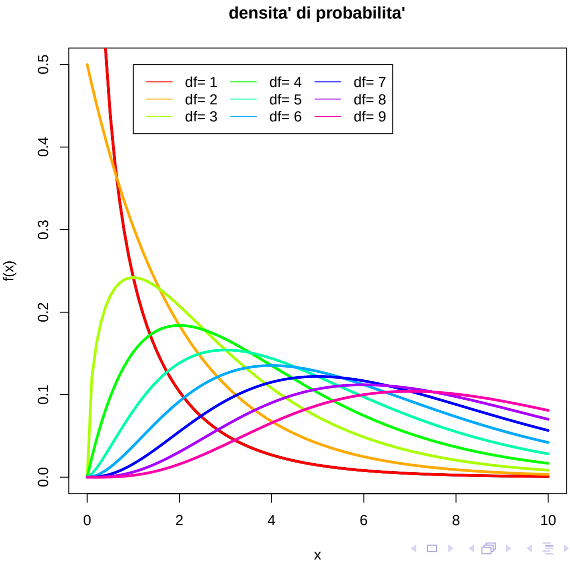
Siano $X_1, \ldots, X_n$ variabili aleatorie i.i.d. $\cong N(0,1)$. Allora la distribuzione di

$$U := \sum_{i=1}^{n} X_i^2$$

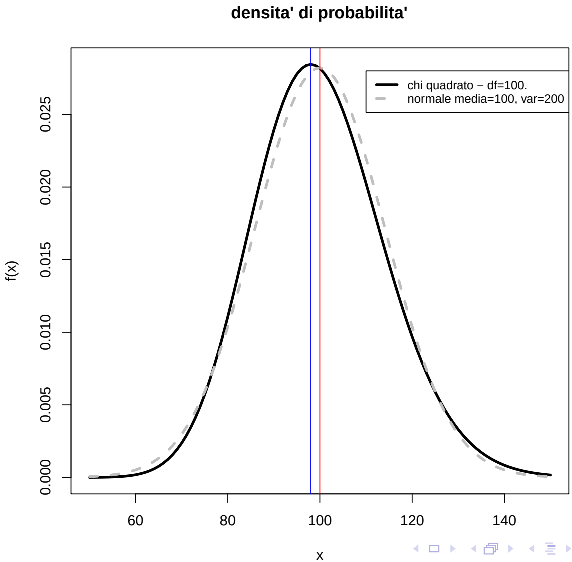
si dice distribuzione chi-quadrato con n gradi di libertà.

Ricordiamo che U ha media n e varianza $2n$. Per $n \geq 3$, la moda è $n-2$.

Distribuzione Chi-Quadrato (2)



Distribuzione Chi-Quadrato (3)



Distribuzione Chi-Quadrato (4)

Teorema: se $Z_1, \dots, Z_n$ è un campione casuale da una distribuzione normale standardizzata e $\overline{Z}$ è la sua media campionaria, allora

$$Y := \sum_{i=1}^{n} (Z_i - \overline{Z})^2$$

ha una distribuzione chi-quadrato con $n - 1$ gradi di libertà.

Esercizio 1

Simulare la variabile Y per $n = 5$ e ottenere 1000 realizzazioni campionarie. Confrontare la distribuzione dei dati osservati con la densità teorica di una v.a. chi-quadrato con 4 gradi di libertà (usare `hist` o `density`).

Distribuzione t di Student (1)

Sia Z una variabile normale standardizzata e sia U una variabile chi-quadrato con m gradi di libertà. Siano Z e U indipendenti. Allora la distribuzione di

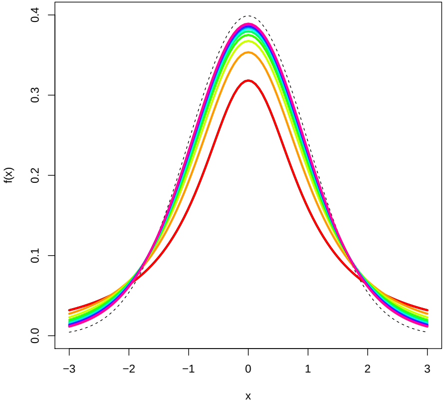
$$X := \frac{Z}{\sqrt{U/m}}$$

si dice distribuzione t (di Student) con m gradi di libertà.

Ricordiamo che X ha media 0 (se $m > 1$) e varianza $\frac{m}{m-2}$ (se $m > 2$). Se $m = 1$ X è una variabile di Cauchy.

Distribuzione t di Student (2)

Funzioni di densità di variabili t con $1, \ldots, 10$ gradi di libertà.
La curva tratteggiata è la normale.



Distribuzione t di Student (3)

Esercizio 2

Sia X_1, X_2 un campione casuale estratto da $N(0,1)$. Dire qual è la distribuzione di $Z = \frac{X_1+X_2}{\sqrt{(X_1-X_2)^2}}$. Simulare 300 realizzazioni campionarie di (X_1, X_2) e confrontare la distribuzione empirica di Z con quella teorica.

Esercizio 3

Aprire il database `statura.csv` che contiene informazioni sulle altezze (in cm) di un campione di 1802 individui divisi in maschi e femmine. Studiare la struttura del database, le medie e le varianze delle altezze per uomini e donne, anche attraverso strumenti grafici. Supponendo che l'altezza, per l'intera popolazione maschile, ubbidisca ad una legge normale di media μ e varianza σ^2 ignote, e che il campione degli uomini sia casuale, studiare l'affermazione "μ vale 174.5 cm" sul campione assegnato.

Distribuzione t di Student (4)

Esercizio 3:

Ricordiamo che se $X_1, \ldots, X_n$ è un campione casuale da una distribuzione normale con media μ e varianza σ^2 allora

$$Y := \frac{\sqrt{n(n-1)}(\overline{X} - \mu)}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2}}$$

ha una distribuzione t di Student con $n - 1$ gradi di libertà.

Con $n = 1001$, $\mu = 174.5$, Y sul nostro campione assume il valore di 2.361658.

Distribuzione t di Student (5)

Esercizio 3:

L'ipotesi $\mu = 174.5$ viene scartata quando il valore osservato della statistica t cade nella zona di rifiuto (in base al livello di significatività scelto).