Università Ca' Foscari di Venezia
Federica Giummolè

Informazioni generali

Probabilità e Statistica

A.A. 2013/2014

Informazioni generali

• Docente del corso: Federica Giummolè
email: giummole@unive.it
web: http://www.dst.unive.it/~giummole

• Ricevimento studenti: Martedì 12:00–13:00, via Torino stanza ospiti
http://www.dst.unive.it/~giummole/avvisi.html

• Lezioni: Martedì e Giovedì 8:45–10:15, via Torino aula 2

• Tutorato (Marco Fiorucci): Martedì 12:15–13:45, via Torino aula 2 o lab 5

• Moodle: http://moodle.unive.it/ per materiale didattico e test di autovalutazione

- Testi di riferimento:

Ross, S.M. (2007). *Calcolo delle probabilità*, seconda edizione, Apogeo.
Ross, S.M. (2009). *Probabilità e statistica per l'ingegneria e le scienze*, seconda edizione, Apogeo.

- L'esame consiste in una prova scritta con 10 domande a risposta multipla con sbarramento e alcuni esercizi teorici e pratici, anche sull'uso di R.

Laboratorio con R

R è un ambiente di sviluppo specifico per l'analisi statistica dei dati che utilizza un linguaggio di programmazione derivato e in larga parte compatibile con S.

- R è open-source e può essere scaricato gratuitamente dal sito `http://cran.r-project.org/`

- R funziona sotto UNIX, Windows e Mac

- R ha un *help* approfondito e dettagliato

- R ha eccellenti capacità grafiche

- R è un linguaggio di programmazione con molte funzioni predefinite e la possibilità di costruirne di nuove

- R è mantenuto e aggiornato da una squadra internazionale di esperti. Tutti possono contribuire con *packages* sempre nuovi

Università Ca' Foscari di Venezia
Federica Giummolè

Introduzione alla statistica

Probabilità e Statistica
A.A. 2013/2014

La statistica nella società dell'informazione

- Tutti dicono che viviamo nella società dell'informazione. Ma molti si lamentano che le informazioni sono troppe. E' facile raccoglierle, memorizzarle, distribuirle. E' difficile verificarle ed interpretarle.

- La statistica è la *tecnologia* necessaria per risolvere queste difficoltà.

- Uno statistico sa combinare informazioni di tipo differente, valutarne l'affidabilità, sintetizzare e presentare molti dati in maniera tale da evidenziare la storia che raccontano, costruire modelli (=visioni stilizzate di una parte di mondo) che facilitano l'interpretazione, e, ad esempio, permettono di calcolare previsioni o di formulare ipotesi di decisione.

Un po' di terminologia

- Un insieme (di individui o animali o oggetti o aziende o. . .) costituisce la parte del mondo che interessa, quella su cui dobbiamo produrre nuove conoscenze, quella che è coinvolta nelle decisioni da prendere. Questo insieme viene chiamato convenzionalmente la **popolazione di riferimento**. Gli elementi della popolazione sono chiamati genericamente **unità statistiche**.

- Alcune caratteristiche di tutte o di una parte delle unità statistiche vengono rilevate/misurate. Il risultato di questo rilevare/misurare costituisce quello che chiamiamo i **dati**. Le unità statistiche sono disomogenee rispetto ai fenomeni rilevati.

- L'obbiettivo è quello di trasformare i dati in nuove conoscenze o ipotesi di decisione. Ovvero, di trasformare i dati in affermazioni sul mondo (sulla popolazione di riferimento).

Un po' di terminologia

- Le caratteristiche rilevate sulle unità statistiche vengono chiamate **variabili**.

- I valori distinti assunti da una variabile sono chiamate le **modalità** della variabile stessa.

- Se le variabili di interesse non sono rilevate su tutte le unità statistiche, il sottoinsieme della popolazione oggetto della rilevazione è chiamato il **campione**.

Tipi di variabili

In statistica si parla di variabili:

- **qualitative** o **categoriali** quando le modalità utilizzate per descrivere il fenomeno analizzato prendono la forma di aggettivi o di altre espressioni verbali. Le variabili qualitative possono essere
  - **sconnesse** se non esiste nessun ordinamento naturale tra le modalità; esempi di variabili sconnesse sono: (i) il sesso, (ii) il tipo di servizio offerto da un albergo;
  - **ordinali** nel caso in cui un ordinamento naturale esiste; esempi di variabili qualitative ordinali sono: (i) il titolo di studio, (ii) il parere di un intervistato (ad es. classificato come "mediocre", "discreto", "buono").

  Quando le modalità sono solamente due (esempi (i) maschio vs. femmina, (ii) vivo vs. morto; (iii) buono vs. difettoso) si parla di variabili **dicotomiche** o **binarie**.

- **numeriche** quando le modalità sono espresse da numeri. Dal punto di vista dei modelli e delle tecniche utilizzate le variabili numeriche si suddividono a loro volta in

  - **discrete** o **intere** quando le modalità sono esprimibili da numeri interi; esempi sono: (i) il numero di clienti, (ii) il numero di pezzi prodotti;

  - **continue** o **reali** quando le modalità sono esprimibili da numeri reali; esempi sono: (i) il tempo d'attesa ad uno sportello, (ii) il peso di un manufatto.

Piccolo esempio (per fissare la terminologia)

Vogliamo avere un'idea sul numero di clienti e sul volume di vendite dei negozi di una città per tre categorie merceologiche ritenute simili. La popolazione di riferimento è l'insieme di tutti i negozi secondo le tre categorie merceologiche. Le unità statistiche sono i negozi. I dati si presentano in questa forma:

negozio	clienti	vendite	categoria
1	907	11.2	A
⋮	⋮	⋮	⋮
10	420	6.12	B
11	679	7.63	B
⋮	⋮	⋮	⋮
19	1010	11.77	C
20	621	7.41	A

Le variabili considerate nello studio sono tre:

clienti le cui *modalità* sono numeriche e discrete;

vendite (in migliaia di euro) le cui *modalità* sono numeriche e (con approssimazione) continue.

categoria le cui *modalità* sono sconnesse (A, B e C.)

Il modo in cui sono raccolti i dati può condizionare il loro tipo

Si consideri una macchina che deve forare delle lastre di metallo. Il diametro nominale dei fori è di 1mm con una tolleranza di 0,06mm. Ovvero un foro è *ben fatto* se il suo diametro è compreso tra 0,94mm e 1,06mm.

Allora, dati sulla *qualità* della produzione della macchina, potrebbero essere disponibili nella forma

1. "buono" vs. "difettoso" (dati dicotomici);

2. "troppo piccolo", "buono", "troppo grande" (dati qualitativi ordinali);

3. lunghezza del diametro (dati numerici continui).

Si osservi che le differenze non sono semplicemente dovute a come i dati vengono registrati ma possono anche essere dovute a come *i diametri vengono effettivamente misurati*. Ad esempio, raccogliere dati sui diametri nella forma (2) è più rapido e richiede strumenti meno costosi (bastano due bastoncini metallici di diametro rispettivamente uguale ai due estremi dell'intervallo di tolleranza) di quanto richiesto dalla forma (3).

Dati sperimentali vs dati osservazionali

Nell'analizzare dei dati è bene poi tenere presente il tipo di studio in cui sono stati rilevati. In particolare, è importante la distinzione tra:

- **studi sperimentali** ovvero situazioni in cui i dati sono stati raccolti in situazioni replicabili e controllate (esempio classico sono gli esperimenti di laboratorio);

- **studi osservazionali** ovvero situazioni in cui il ricercatore semplicemente rileva dei dati già esistenti (esempio: il numero di presenze alberghiere in una stagione, il prezzo di un'azione,...).

Il problema principale degli studi osservazionali è che non controllando i fattori che possono influenzare il fenomeno sotto indagine risulta difficile essere certi di averli individuati appropriatamente.

Metodi di raccolta dei dati

1. Esperimenti in laboratorio

2. Interviste telefoniche

3. Questionari inviati per posta

4. Social network

5. Carte fedeltà

6. ...

Il **campionamento** e il **disegno degli esperimenti** si occupano delle problematiche connesse con la raccolta dei dati.

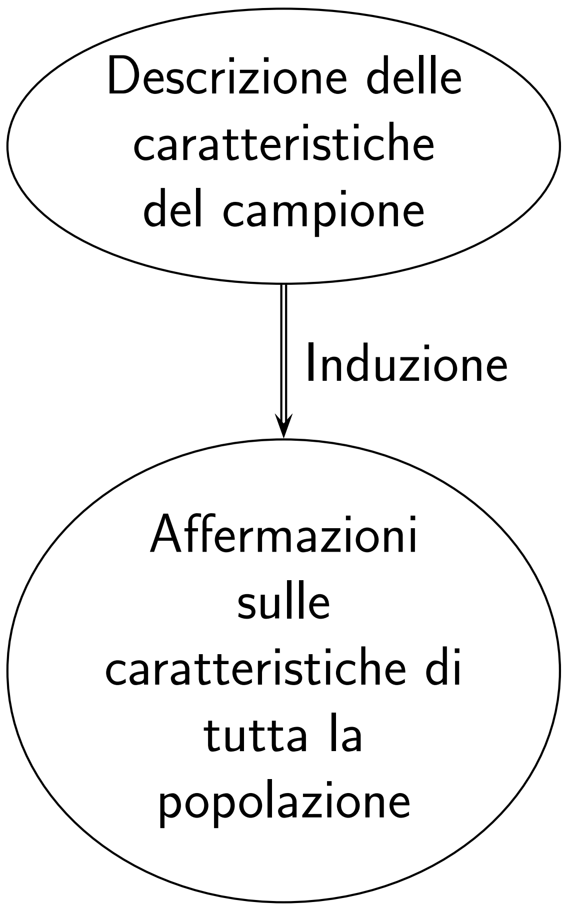
Statistica Descrittiva e Inferenza Statistica

Descrittiva: metodi per rappresentare, sintetizzare ed evidenziare le caratteristiche più significative di un insieme di dati. Usualmente si dispone di dati su tutta la popolazione di riferimento.

Inferenza: i dati disponibili sono stati rilevati solamente su una parte delle unità statistiche (il campione, da cui *indagini campionarie*). Si vogliono utilizzare le informazioni del campione per fare delle affermazioni sulle caratteristiche generali di tutta la popolazione.

Statistica Descrittiva e Inferenza Statistica

Statistica Descrittiva ed **Inferenza Statistica** nelle applicazioni non sono facilmente separabili. Infatti i problemi di *inferenza* vengono normalmente affrontati in accordo allo schema



La statistica descrittiva viene dunque utilizzata per un'analisi preliminare delle caratteristiche del campione.

Calcolo delle probabilità

Perché l'inferenza porti a risultati sensati, bisogna che sia noto il legame fra popolazione e campione. In particolare, il campione deve essere scelto in modo che rappresenti, in piccolo, la popolazione.

La relazione fra campione e popolazione si descrive attraverso il **calcolo delle probabilità**.

E' il calcolo delle probabilità che fornisce gli strumenti per l'inferenza e che permette di quantificare gli errori che commettiamo nel passaggio dal particolare (campione) al generale (popolazione).

Università Ca' Foscari di Venezia

Federica Giummolè

Statistica descrittiva

Probabilità e Statistica

A.A. 2013/2014

Variabili quantitative

In un reparto dove sono assemblati *lettori mp3* vengono provate in tre giorni diversi tre differenti organizzazioni delle linee di produzione. Le tre diverse organizzazioni sono chiamate nel seguito vecchia (quella in uso al momento dell'esperimento), nuova 1 e nuova 2.

Nei tre giorni, per ciascuno dei 288 addetti che lavorano nel reparto, viene rilevato

"il numero di operazioni completato"

che, ovviamente, può essere visto come una misura della produttività.

Domanda: qual è la migliore organizzazione del lavoro?

I dati

Vecchia organizzazione

725	724	710	724	700	724	713	692	683	712	684	707	703	691	709	702	705	715
704	705	697	725	692	719	694	717	696	707	726	703	705	712	710	697	698	694
701	715	701	707	706	701	687	708	719	713	699	702	694	708	712	704	703	687
709	693	715	707	710	700	718	702	718	705	723	718	701	698	692	684	716	710
708	707	695	726	710	709	692	707	717	709	710	718	708	720	705	714	687	707
707	723	695	676	705	684	717	719	715	710	711	696	696	715	686	702	708	713
701	692	713	700	704	726	702	706	706	700	700	687	696	694	699	709	704	704
715	706	688	724	713	686	697	710	704	724	721	717	690	707	713	685	706	699
687	702	701	708	704	705	702	701	699	699	685	712	678	706	706	695	707	718
706	716	703	721	714	704	697	693	711	697	710	713	702	715	714	716	698	714
704	717	700	692	718	699	698	690	710	703	702	719	710	725	721	713	699	703
698	712	714	707	691	711	712	718	702	711	709	700	719	692	716	700	707	714
717	714	703	709	711	704	689	712	714	711	692	720	697	698	700	689	693	707
699	704	696	708	713	714	712	708	704	720	705	703	712	719	713	716	712	703
717	695	711	697	693	701	699	697	724	713	706	705	704	707	704	719	711	700
694	706	705	698	697	697	700	705	722	712	703	688	694	708	703	690	706	704

Organizzazione ‘‘nuova 1’’

695	686	694	690	713	704	693	697	723	694	690	721	683	701	718	715	738	694
692	704	728	697	711	706	714	710	717	729	709	695	699	714	691	698	680	720
683	696	713	674	689	683	708	704	725	695	690	696	678	725	683	700	699	705
688	714	709	693	681	717	691	706	684	684	693	719	731	706	686	698	710	679
712	688	697	729	695	697	717	679	736	671	695	739	698	696	714	711	701	720
686	706	722	695	688	709	693	756	677	712	670	693	695	683	713	672	706	708
690	685	686	681	716	709	704	679	686	676	718	683	689	696	687	736	699	685
698	700	723	681	713	700	708	705	718	692	743	715	745	700	693	676	723	712
671	714	687	687	687	683	671	677	696	696	714	713	671	688	675	671	692	725
690	680	693	703	733	708	720	704	688	732	711	685	714	704	686	682	699	708
708	704	685	685	694	702	738	702	696	709	701	687	703	701	702	693	691	701
735	721	705	691	741	685	716	716	737	687	732	697	670	684	693	711	685	705
690	705	693	698	678	704	710	686	689	686	698	684	687	696	719	679	696	701
712	691	686	704	744	705	718	709	725	699	721	690	678	713	714	705	681	721
673	698	717	711	670	726	694	723	701	683	716	671	712	704	699	705	727	719
702	692	708	694	670	694	697	682	718	705	699	709	695	711	688	717	699	686

Organizzazione ‘‘nuova 2’’

698	715	675	710	731	721	705	718	693	702	713	730	707	710	744	725	724	701
737	715	704	723	705	702	698	729	698	723	716	698	732	724	721	722	728	740
727	709	724	746	704	740	729	708	721	714	739	713	752	732	713	692	734	727
725	690	749	706	758	722	697	722	705	723	748	730	706	688	709	739	709	744
704	716	748	713	744	721	723	733	707	723	702	734	690	715	711	705	718	702
706	742	742	736	740	712	722	731	713	704	704	735	700	717	746	735	717	718
691	696	720	735	716	745	714	698	709	704	704	684	749	747	715	717	731	700
747	709	705	749	704	697	694	715	737	734	705	726	710	716	740	731	714	733
726	752	714	710	714	753	749	728	696	733	731	728	686	706	710	729	729	730
722	707	716	702	728	716	743	750	715	735	710	734	712	706	719	709	702	712

710	729	728	720	721	752	715	712	717	692	724	720	739	719	712	713	734	734
710	711	722	743	707	729	712	681	739	699	721	706	703	708	719	708	724	730
726	731	734	739	727	759	718	716	715	719	693	729	738	710	730	726	719	726
733	717	701	723	720	744	730	698	729	696	717	713	705	700	715	710	735	726
732	701	707	724	708	730	721	720	706	700	735	706	725	725	735	695	709	705
702	737	688	727	717	708	720	724	731	706	730	714	703	721	712	748	734	724

Frequenze assolute

	vecchia	nuova 1	nuova 2
[670,675)	0	13	0
[675,680)	2	12	1
[680,685)	4	20	2
[685,690)	13	33	3
[690,695)	23	33	8
[695,700)	35	38	13
[700,705)	55	27	24
[705,710)	52	28	34
[710,715)	50	28	32
[715,720)	33	19	32
[720,725)	15	12	34
[725,730)	6	9	27
[730,735)	0	4	30
[735,740)	0	7	17
[740,745)	0	3	12
[745,750)	0	1	12
[750,755)	0	0	5
[755,760)	0	1	2
Totale	288	288	288

Frequenze relative

	vecchia	nuova 1	nuova 2
[670,675)	0,000	0,045	0,000
[675,680)	0,007	0,042	0,003
[680,685)	0,014	0,069	0,007
[685,690)	0,045	0,115	0,010
[690,695)	0,080	0,115	0,028
[695,700)	0,122	0,132	0,045
[700,705)	0,191	0,094	0,083
[705,710)	0,181	0,097	0,118
[710,715)	0,174	0,097	0,111
[715,720)	0,115	0,066	0,111
[720,725)	0,052	0,042	0,118
[725,730)	0,021	0,031	0,094
[730,735)	0,000	0,014	0,104
[735,740)	0,000	0,024	0,059
[740,745)	0,000	0,010	0,042
[745,750)	0,000	0,003	0,042
[750,755)	0,000	0,000	0,017
[755,760)	0,000	0,003	0,007
Totale	1,000	1,000	1,000

$$\text{frequenze relative} = \frac{\text{frequenze assolute}}{\text{numero totale di osservazioni}}$$

Tabelle di frequenza: notazioni

y_i : modalità/classe i del carattere y, $i = 1, 2, \ldots, k$ (k modalità/classi)

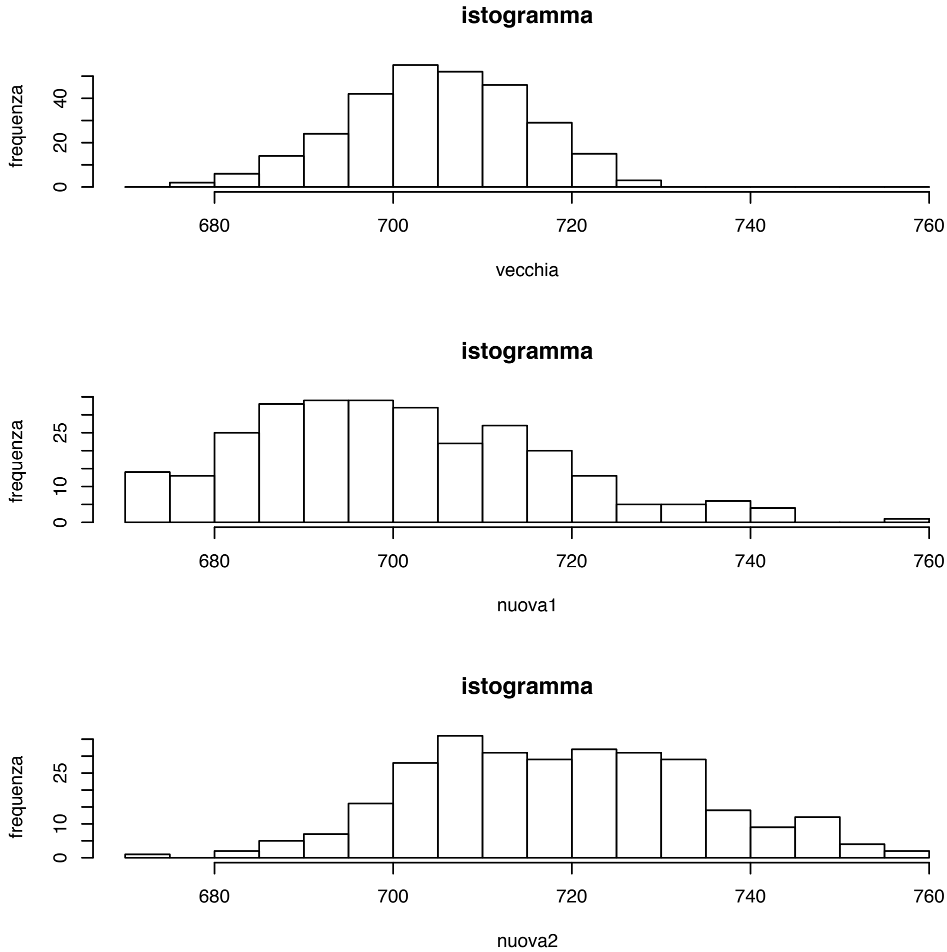
f_i : frequenza assoluta, numero di unità statistiche che possiedono la modalità/classe y_i

n : numero totale di osservazioni ($n = f_1 + f_2 + \cdots + f_k$)

p_i : frequenza relativa ($p_i = f_i/n$)

modalità/classe	freq. assolute	freq. relative
y_1	f_1	$p_1 = f_1/n$
y_2	f_2	$p_2 = f_2/n$
$\vdots$	$\vdots$	$\vdots$
y_k	f_k	$p_k = f_k/n$
Totale	n	1

Istogramma



Gli istogrammi in questo grafico sono stati costruiti ponendo:

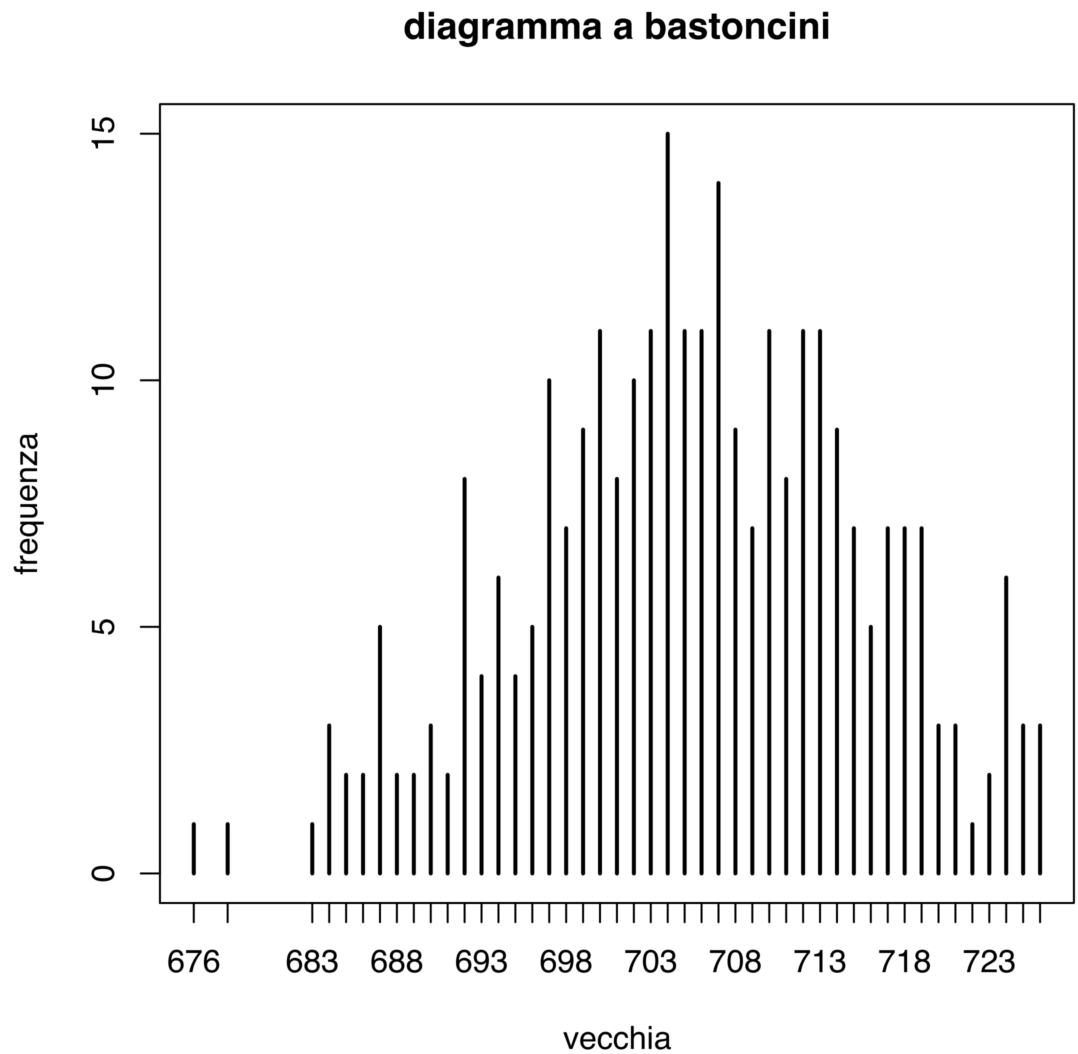
1. la base dei rettangoli pari agli intervalli riportati nella 1° colonna delle tabelle precedenti;

2. l'altezza dei rettangoli pari alle frequenze assolute.

Attenzione! questa regola è valida perché tutti gli intervalli hanno la stessa ampiezza. . .

Diagrammi a bastoncini

Il diagramma a bastoncini (da non confondere con l'istogramma!) è costruito disegnando in corrispondenza di ogni valore osservato un bastoncino di lunghezza uguale alla frequenza assoluta con cui quel valore è stato osservato.



Intervalli di differenti lunghezze

Può capitare o per scelta (si vuole fornire informazioni più dettagliate su parte della distribuzione) o per necessità (i dati sono già stati raggruppati in classi da qualcuno) di costruire degli istogrammi utilizzando intervalli di lunghezza differente.

In questo caso, le altezze dei rettangoli che compongono l'istogramma non devono essere proporzionali alle frequenze osservate ma alla **densità** delle osservazioni nelle singole classi:

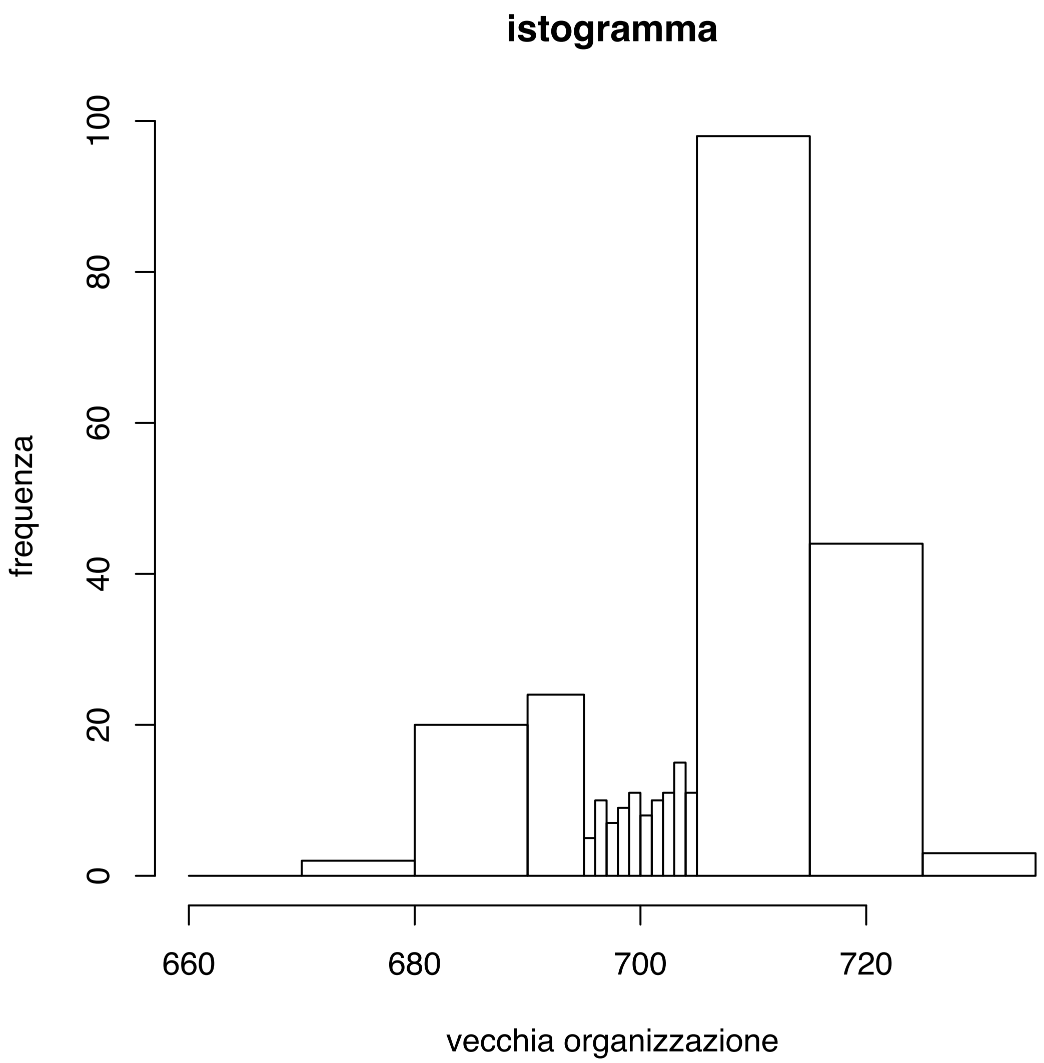
$$\begin{array}{c}\text{densità}\\ \text{di un intervallo}\end{array} = \frac{\text{frequenza dell'intervallo}}{\text{lunghezza dell'intervallo}}.$$

Per capire la definizione si pensi alla popolazione. E' la densità della popolazione non il numero totale di abitanti che ci dice quanto gli individui sono *addensati* in una certa regione geografica.

Istogramma per organizzazione "vecchia" costruito con:

1) intervalli più piccoli nella parte centrale;

2) altezze dei rettangoli proporzionali alle frequenze.

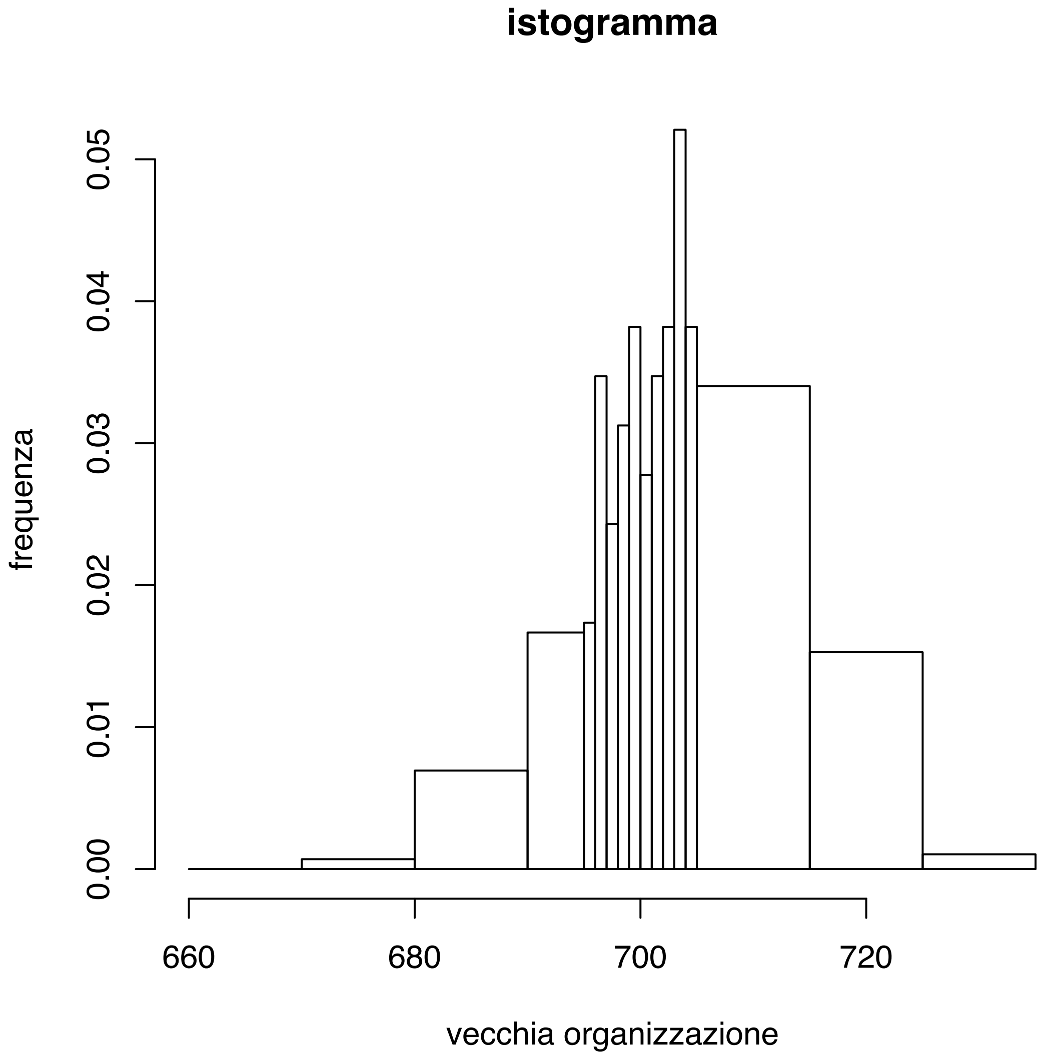


Sembra esserci un buco al centro, esattamente dove le osservazioni sono più *addensate*.

Istogramma per organizzazione “vecchia” costruito con:

1) intervalli più piccoli nella parte centrale;

2) altezze dei rettangoli proporzionali alle densità.



Il buco al centro è sparito. Il grafico correttamente ci dice che le osservazioni sono *addensate* intorno a 705.

Frequenze cumulate

Si ottengono "cumulando" progressivamente le frequenze.

Possono essere "assolute" o "relative".

Esempio di calcolo per organizzazione "nuova 1":

fine int.	freq. ass.	freq. cum. ass.	freq. cum. rel.
675	13	13	13/288=0.045
680	12	25=13+12	25/288=0.087
685	20	45=13+12+20	45/288=0.156
⋮	⋮	⋮	⋮
755	0	$287=13+12+\cdots+0$	287/288=0.997
760	1	$288=13+12+\cdots+0+1$	288/288=1

Funzione di ripartizione empirica

Osservazioni ordinate

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$$

Quindi

- la frazione $1/n$ di unità statistiche assumono valori della variabile Y inferiori o uguali ad $y_{(1)}$;
- la frazione $2/n$ di unità statistiche assumono valori della variabile Y inferiori o uguali ad $y_{(2)}$;
- ...
- la frazione i/n di unità statistiche assumono valori della variabile Y inferiori o uguali ad $y_{(i)}$;
- ...
- la frazione $n/n = 1$ di unità statistiche assumono valori della variabile Y inferiori o uguali ad $y_{(i)}$.

Funzione di ripartizione empirica:

$$\begin{aligned}\widehat{F}(y) &= \text{freq. rel. di unità che assumono valore} \leq y \\ &= \frac{\text{frequenza assoluta di unità che assumono valore} \leq y}{n}.\end{aligned}$$

Proprietà:

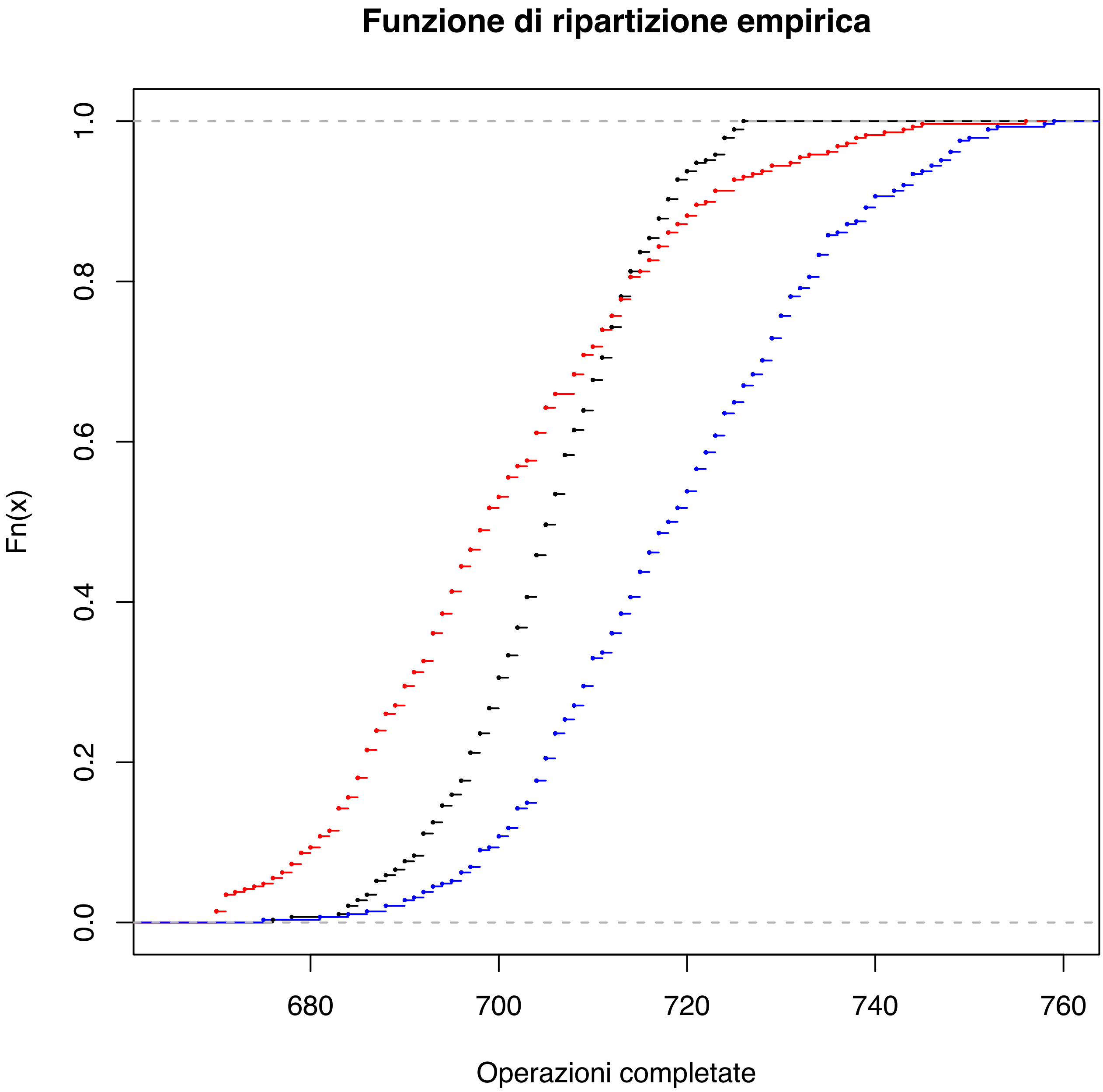
1. $0 \leq \widehat{F}(y) \leq 1$;

2. $\widehat{F}(-\infty) = 0$;

3. $\widehat{F}(\infty) = 1$;

4. $\widehat{F}(y)$ è una funzione ("a gradini") non decrescente;

5. $\widehat{F}(y)$ è continua da destra.

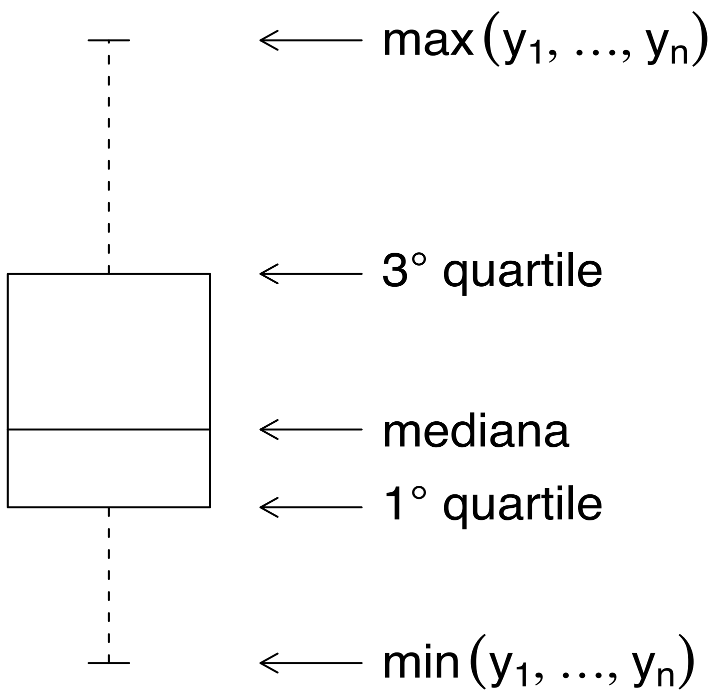

Funzione di ripartizione empirica
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
680
700
720
740
760
Operazioni completate

Diagrammi a scatola con baffi

Il nome deriva dall'inglese (box and whiskers plot, spesso abbreviato in **boxplot**).

Forniscono un'idea schematica di un insieme di dati basata sui quantili.

Sono costituiti, come dice il nome, da una scatola e da due baffi costruiti in accordo al disegno sottostante:



Una variante

Variante comunemente usata del boxplot:

1) la scatola è costruita come descritto precedentemente a partire dai tre quartili;

2) i baffi si estendono fino ai dati più lontani che siano però non più distanti di k volte lo scarto interquartile dalla scatola. Lo **scarto interquartile** è la differenza tra il terzo e il primo quartile (ossia l'ampiezza della scatola), k è una costante arbitraria tipicamente scelta uguale a 1.5. Ovvero, non accettiamo baffi esageratamente lunghi;

3) le osservazioni che sono oltre i baffi sono disegnate opportunamente sul grafico (ad esempio utilizzando un pallino).

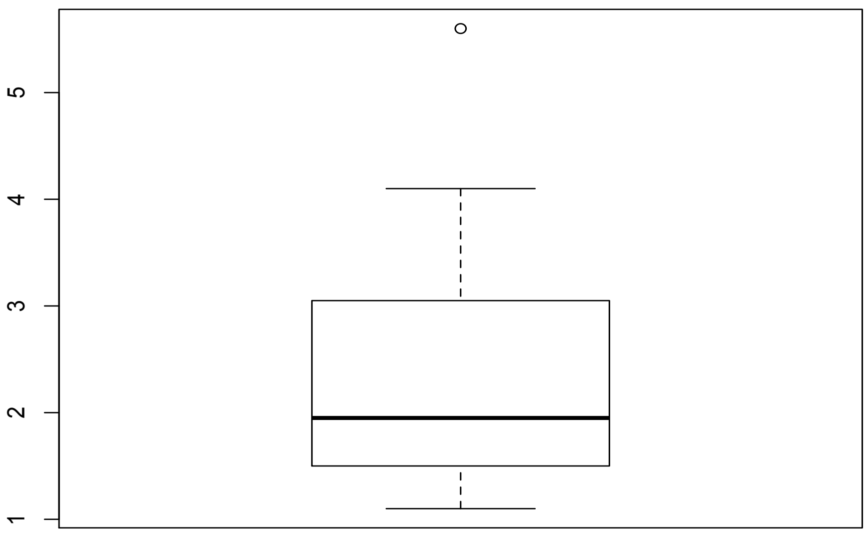
Esempio di costruzione di un boxplot

Dati (già ordinati):

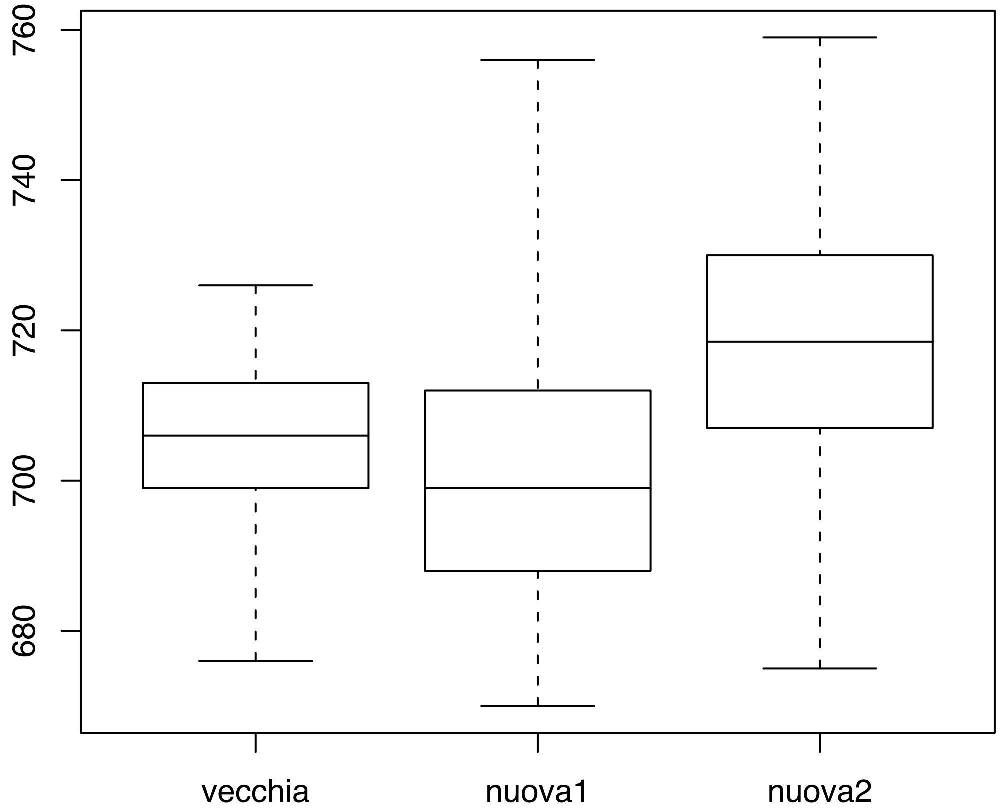
$$1.11.31.41.61.81.92.02.52.93.24.15.6$$

Perciò $y_{0.25} = 1.5$, $y_{0.5} = 1.95$ e $y_{0.75} = 3.05$. Quindi $1.5\times$(scarto interquartile)$= 1.5 \times 1.55 = 2.325$. Di conseguenza:

1. la scatola si estende da 1.5 a 3.05;
2. il baffo inferiore si estende fino all'osservazione più piccola tra quelle maggiori di $y_{0.25} - 2.325 = -0.825$, ovvero fino a 1.1;
3. il baffo superiore si estende fino all'osservazione più grande tra quelle minori di $y_{0.75} + 2.325 = 5.375$, ovvero fino a 4.1;
4. vanno disegnate esplicitamente nel diagramma le osservazioni più piccole di 1.1 o più grandi di 5.375; in questo caso l'osservazione pari a 5.6.



Le tre organizzazioni della produzione



Variabili qualitative

I dati si riferiscono ad un'indagine ISTAT condotta nel 2001 sugli esercizi ricettivi, ovvero alberghi, campeggi e villaggi turistici, alloggi agro-turistici ed altri esercizi (ostelli, case per ferie, rifugi alpini, .etc.), divisi per area geografica.

I dati prendono la forma di una lunga tabella di questo tipo:

esercizio	tipo	area geografica
1	albergo	Nord
2	camp. e vill. tur.	Sud
⋮	⋮	⋮

Per ogni esercizio (*unità statistica*) sono state rilevate due variabili: il *tipo* di esercizio e l'*area geografica* dell'esercizio.

Tabelle di frequenza

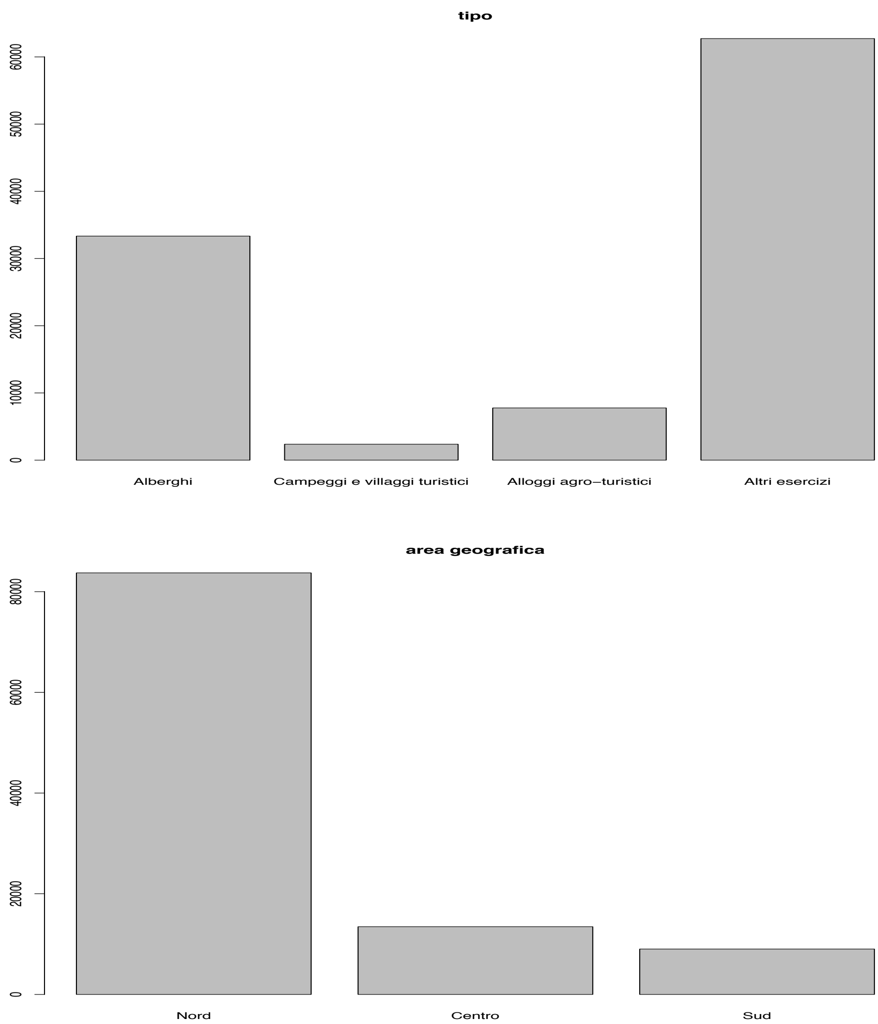
La variabile *tipo* ha la seguente distribuzione di frequenze

tipo	freq.	freq. rel.
Alberghi	33.338	0,314
Campeggi e villaggi turistici	2.371	0,022
Alloggi agro-turistici	7.769	0,073
Altri esercizi	62.727	0,591
TOTALE	106.205	1,00

La variabile *area geografica* ha invece la seguente distribuzione di frequenze

area geografica	freq.	freq. rel.
Nord	83.732	0,788
Centro	13.454	0,127
Sud	9.019	0,085
TOTALE	106.205	1,00

Diagramma a barre: frequenze assolute



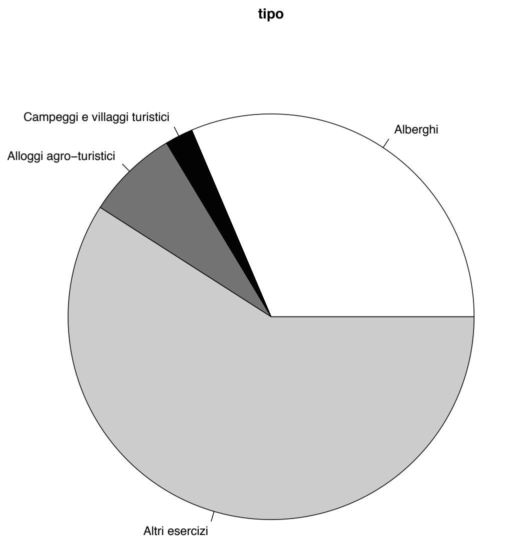
La rappresentazione grafica più utilizzata è il diagramma a barre, in cui ogni modalità è rappresentata da una barra di altezza pari alla frequenza (assoluta o relativa) della modalità. Si osservi che i rettangoli, contrariamente al caso di un istogramma, sono disegnati *staccati*.

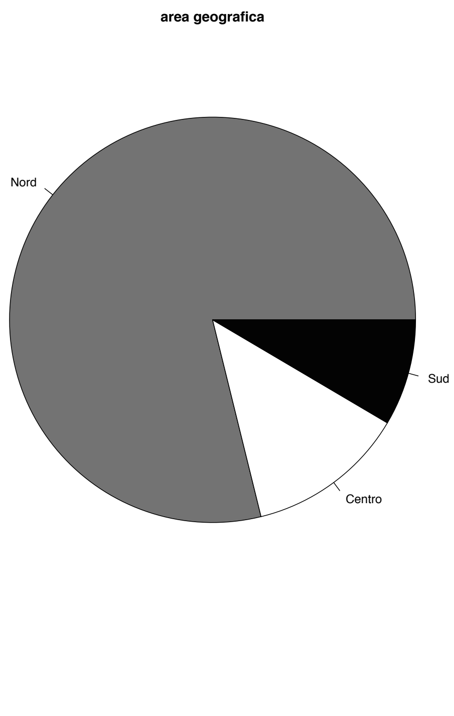
Notiamo che, se la variabile non è ordinale, l'ordine delle modalità nell'asse delle ascisse del grafico è arbitrario.

Diagramma a torte: frequenze relative

Una diversa rappresentazione grafica per variabili qualitative è data dal diagramma a torta, in cui ogni modalità è rappresentata da una fetta di torta proporzionale alla sua frequenza relativa:

$$angolo = 360 \cdot frequenza\ relativa$$



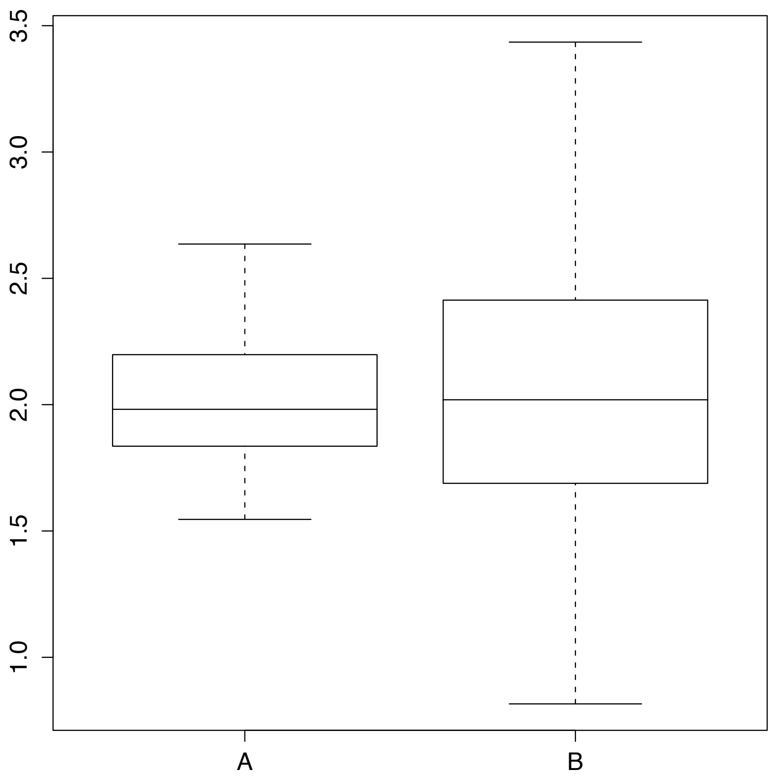


La variabilità

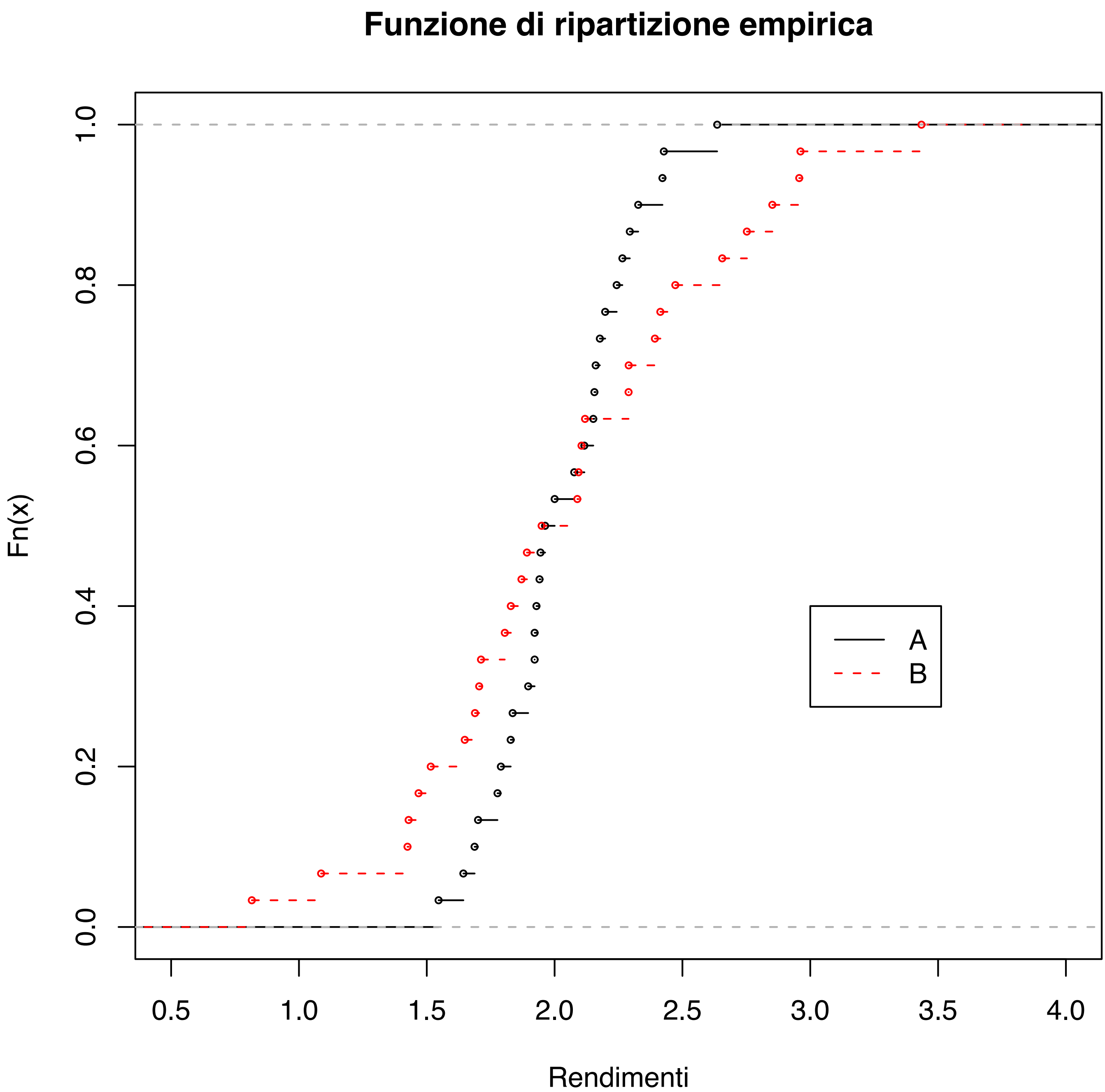
Per confrontare le *performance* di due tipologie di fondi, etichettate come A e B abbiamo preso in considerazione i rendimenti di 30 fondi per ciascuna tipologia. Riportiamo di seguito i diagrammi a scatola dei rendimenti.

Gruppo A
1.643 2.117 1.897 1.836 2.294 1.929 2.243 1.777 1.922 1.945
2.156 2.265 2.177 1.941 2.198 1.922 1.828 2.422 2.151 1.790
2.427 1.687 2.000 2.327 1.700 2.160 1.963 2.636 1.546 2.077

Gruppo B
2.752 1.805 2.290 2.105 2.472 1.087 3.435 0.816 1.705 1.516
2.094 2.957 1.689 1.468 1.829 1.949 2.289 2.414 2.656 2.089
2.852 1.712 1.649 1.870 2.962 1.892 1.429 2.392 1.424 2.119



e le rispettive funzioni di ripartizione

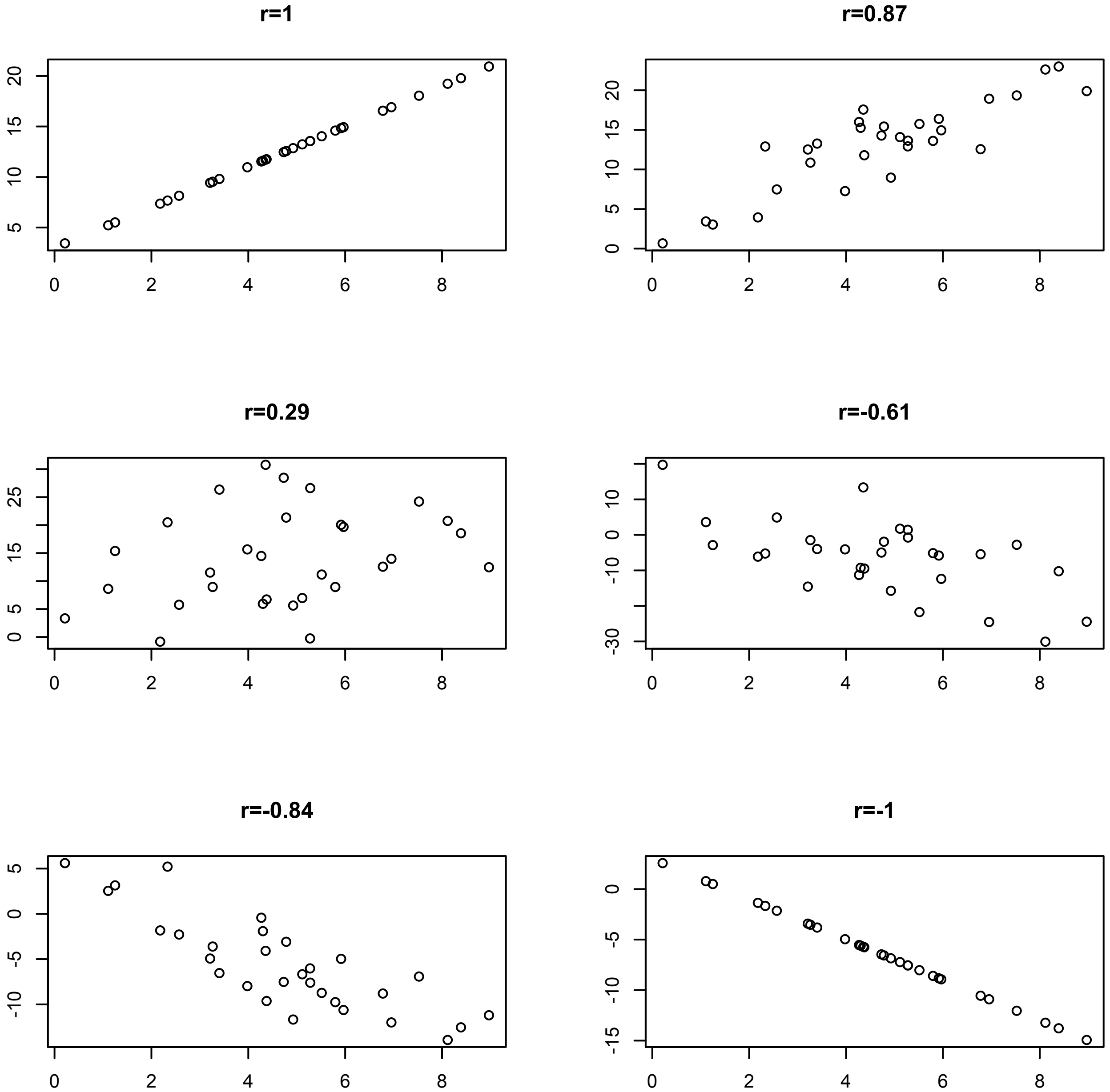


Indici di variabilità

	A	B
varianza	0,06	0,34
scarto quadratico medio	0,25	0,58
campo di variazione	1,09	2,62
scarto interquartile	0,34	0,72
MAD	0,31	0,58

La tabella mostra chiaramente come tutti gli indici considerati evidenzino la maggiore variabilità dei rendimenti (leggi 'rischio') dei fondi di tipo B.

La correlazione

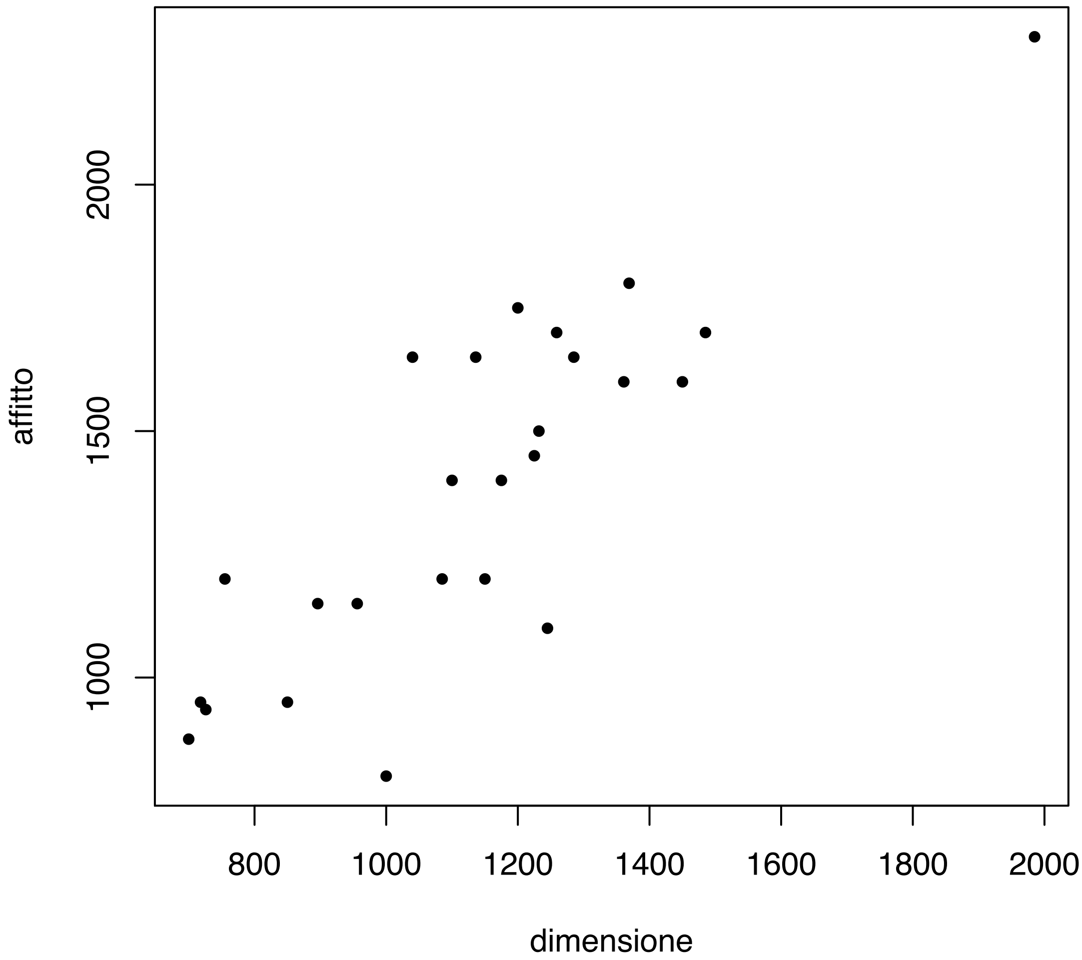


I dati

Un agente immobiliare intende prevedere gli affitti mensili degli appartamenti sulla base della loro dimensione. Per questo conduce un'indagine e reperisce i dati su 25 appartamenti in una zona residenziale. La seguente tabella mostra i dati ottenuti per i 25 appartamenti. L'affitto è l'affitto mensile in dollari e la dimensione è espressa in piedi al quadrato.

	affitto	dimensione
1	950	850
2	1600	1450
3	1200	1085
4	1500	1232
5	950	718
6	1700	1485
7	1650	1136
8	935	726
9	875	700
10	1150	956
11	1400	1100
12	1650	1285
13	2300	1985
14	1800	1369
15	1400	1175
16	1450	1225
17	1100	1245
18	1700	1259
19	1200	1150
20	1150	896
21	1600	1361
22	1650	1040
23	1200	755
24	800	1000
25	1750	1200

Diagramma di dispersione



Abbiamo semplicemente disegnato i punti osservati sul piano. E' evidente una forte relazione, certamente crescente come ci si poteva attendere.

Covarianza e correlazione

Calcoliamo la covarianza e la correlazione dei nostri dati, ponendo x =dimensione e y =affitto:

$$\begin{aligned}
\text{cov}(x,y) &= \frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y} \\
&= \frac{1}{25} \cdot 41480210 - 1135.32 \cdot 1386.4 \\
&= 85200.75
\end{aligned}$$

$$\begin{aligned}
\text{cor}(x,y) &= \frac{\text{cov}(x,y)}{\text{sqm}(x) \cdot \text{sqm}(y)} \\
&= \frac{85200.75}{282.8249 \cdot 354.3854} \\
&= 0.85
\end{aligned}$$

Tabelle di contingenza: il Titanic

Dopo il disastro, una commissione d'inchiesta del *British Board of Trade* ha compilato una lista di tutti i 1316 passeggeri con alcune informazioni aggiuntive riguardanti: se è stato salvato (SI, NO), la classe (I, II, III) in cui viaggiavano, il sesso, l'età,. . . .

Ci limitiamo a considerare le informazioni sull'esito e la classe. Quindi dal nostro punto di vista i dati sono costituiti da una lunga lista del tipo

Passeggero	Classe	Salvato
nome 1	II	SI
nome 2	III	NO
nome 3	I	NO
⋮	⋮	⋮
nome 1316	III	SI

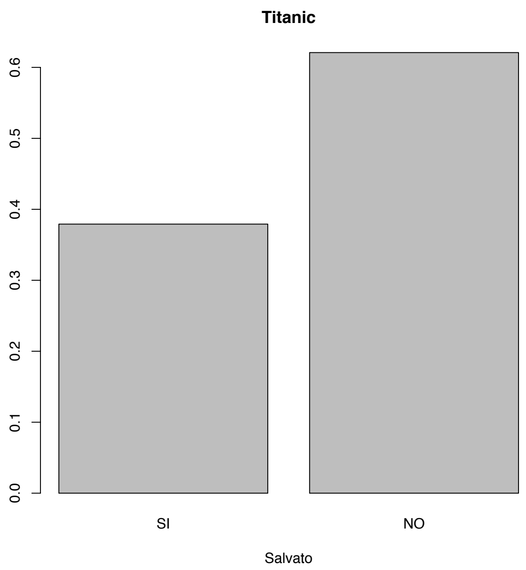
Una variabile alla volta

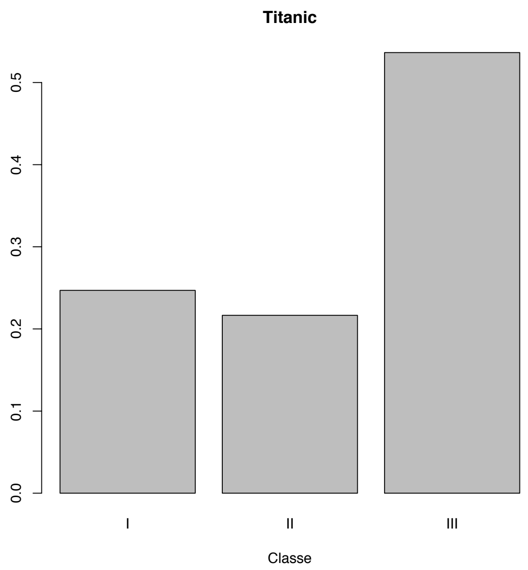
La variabile Salvato ha la seguente distribuzione di frequenze

Salvato	Freq. assolute	Freq. relative
SI	499	0,379
NO	817	0,621
	1316	1,000

La variabile Classe ha invece la seguente distribuzione

Classe	Freq. assolute	Freq. relative
I	325	0,247
II	285	0,216
III	706	0,537
	1316	1,00





Le due variabili assieme: frequenze congiunte

La prima sintesi che possiamo operare consiste nel costruire una tabella del tipo

	Classe			
Salvato	I	II	III	totale
SI	203	118	178	499
NO	122	167	528	817
totale	325	285	706	1316

dove consideriamo tutti i possibili incroci di modalità delle due variabili ($2 \times 3 = 6$).

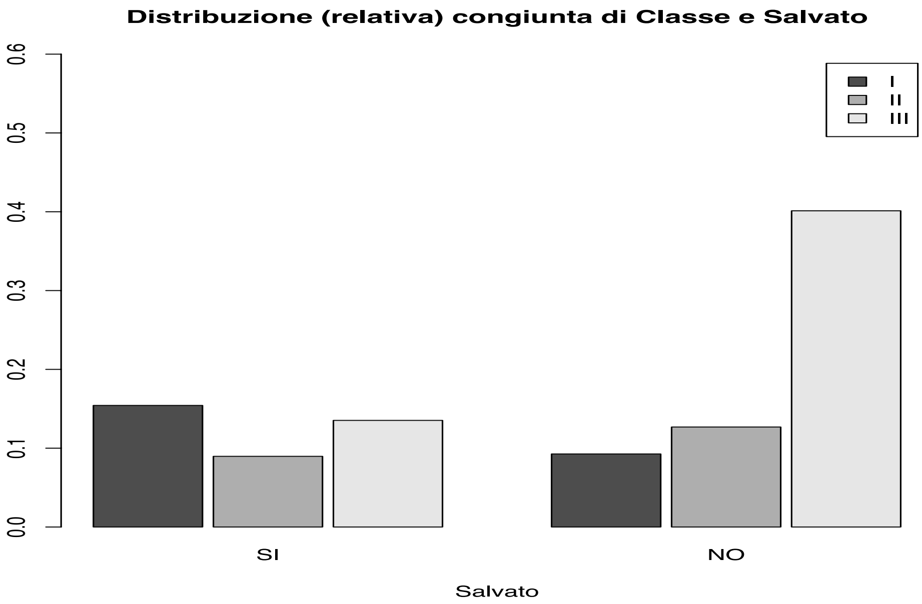
Possiamo anche considerare le frequenze relative, ottenute semplicemente dividendo le frequenze assolute per il numero totale $n = 1316$ di unità

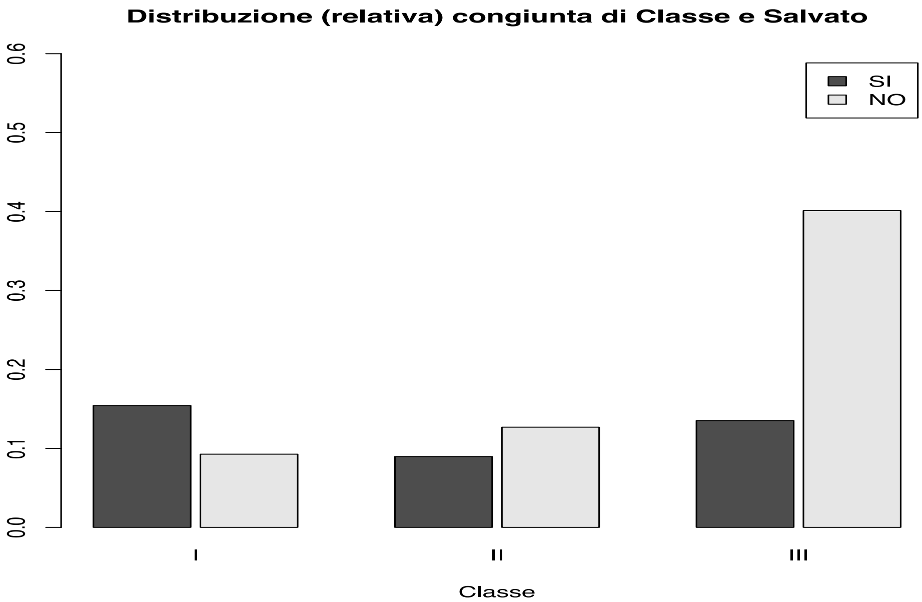
	Classe			
Salvato	I	II	III	totale
SI	0,154	0,090	0,135	0,38
NO	0,093	0,127	0,401	0,62
totale	0,247	0,217	0,536	1,000

Frequenze congiunte: rappresentazione grafica

Possiamo rappresentare le frequenze (sia assolute che relative) della tabella attraverso un appropriato diagramma a barre.

La stessa informazione può essere rappresentata in due modi diversi ("per riga" o "per colonna"):



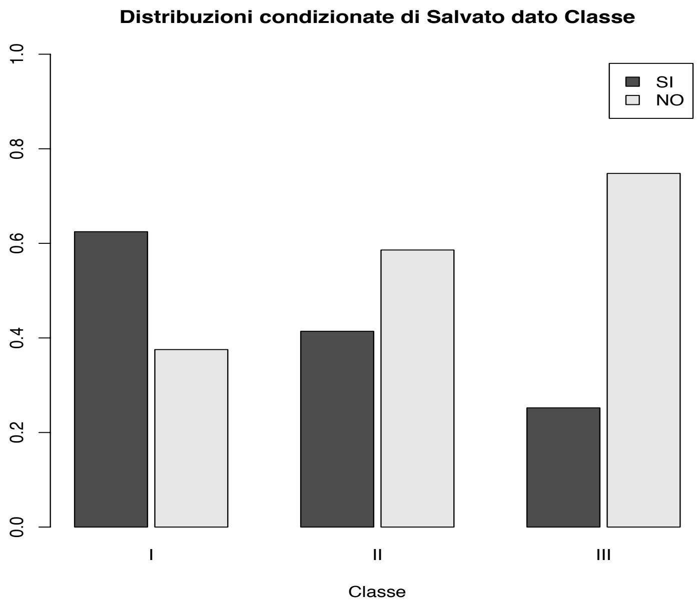


Distribuzioni condizionate di Salvato dato Classe

Ci sono tre distribuzioni condizionate di Salvato dato Classe (le tre colonne), una per ogni modalità di Classe (I, II, II).
Le distribuzioni condizionate relative si ottengono dividendo ogni colonna per il totale di colonna

Salvato	Classe I	Classe II	Classe III
SI	203	118	178
NO	122	167	528
totale	325	285	706

Salvato	Classe I	Classe II	Classe III
SI	0,62	0,41	0,25
NO	0,38	0,59	0,75
totale	1,00	1,00	1,00

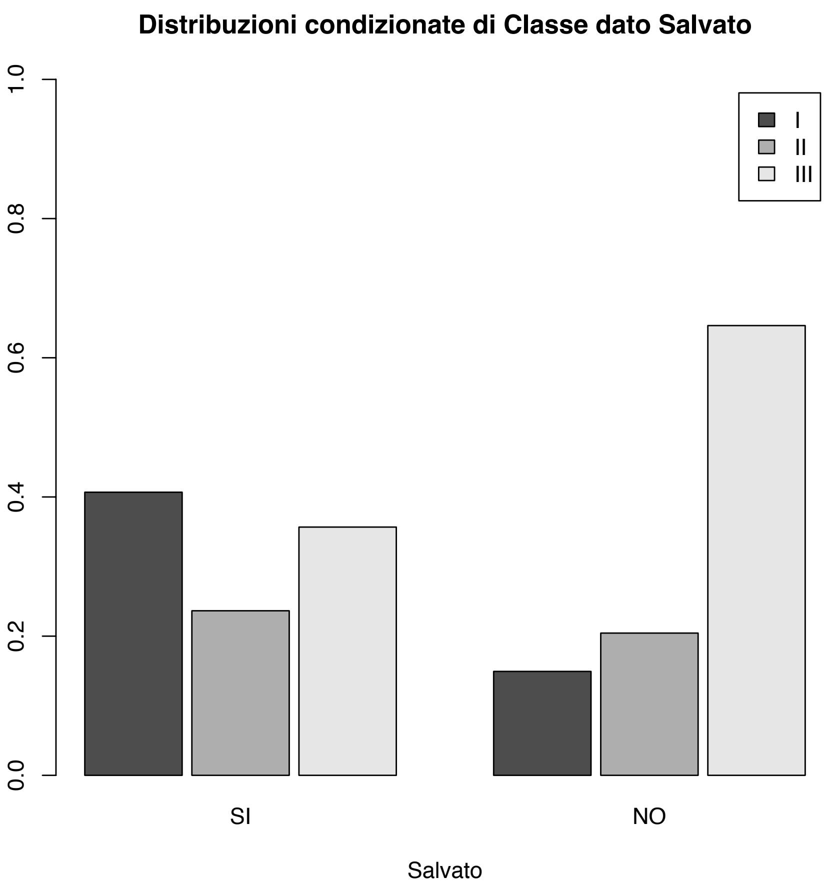


Distribuzioni condizionate di Classe dato Salvato

Ci sono due distribuzioni condizionate di Classe dato Salvato (le due righe), una per ogni modalità di Salvato (SI, NO).
Le distribuzioni condizionate relative si ottengono dividendo ogni riga per il totale di riga

	Classe			
Salvato	I	II	III	totale
SI	203	118	178	499
NO	122	167	528	817

	Classe			
Salvato	I	II	III	totale
SI	0,41	0,24	0,36	1,00
NO	0,15	0,20	0,65	1,00



Frequenze attese

La tabella delle frequenze attese è quella che si osserverebbe se fra le due variabili non ci fosse nessun tipo di dipendenza:

	classe			
salvato	I	II	III	totale
SI	123,2	108,1	267,7	499
NO	201,8	176,9	438,3	817
totale	325	285	706	1316

Il confronto con le frequenze osservate è particolarmente istruttivo.

	Classe			
Salvato	I	II	III	totale
SI	203	118	178	499
NO	122	167	528	817
totale	325	285	706	1316

Ad esempio, ci indica che, senza la preferenza accordata ai passeggeri di I classe, si sarebbero salvati un centinaio di passeggeri di III classe in più.

Quindi, sembra esserci evidenza contro l'ipotesi di indipendenza tra le due variabili.

L'indice X^2 di Pearson

E' una *misura della distanza* fra le frequenze osservate e le frequenze attese.

$$
\begin{aligned}
X^2 &= \frac{(203-123,2)^2}{123,2} + \frac{(118-108,1)^2}{108,1} + \ldots \\
&\quad \cdots + \frac{(528-438,3)^2}{438,3} \\
&= 133,05
\end{aligned}
$$

$$
\tilde{X}^2 = \frac{133,05}{1316 \cdot \min(1,2)} = 0,1011\,.
$$

Purtroppo, per sapere se il valore che abbiamo ottenuto è grande o piccolo, abbiamo bisogno del calcolo delle probabilità...

Università Ca' Foscari di Venezia
Federica Giummolè

Possibilità

Probabilità e Statistica
A.A. 2013/2014

Contare le possibilità

Principio fondamentale del conteggio: se una scelta può essere fatta in m_1 modi diversi e un'altra scelta può essere fatta in m_2 modi diversi, allora esistono in totale $m_1 \times m_2$ possibilità di scelta.

• Esempio: 10 cavalieri e 12 dame partecipano a un ballo. Ci sono $10 \times 12 = 120$ possibili coppie danzanti.

Principio fondamentale del conteggio generalizzato: se ciascuna di r scelte successive può essere fatta in m_i modi rispettivamente, allora esistono in totale

$$\prod_{i=1}^{r} m_i = m_1 \times \ldots \times m_r$$

possibilità di scelta.

• Esempio: una commissione parlamentare deve essere composta da un membro del partito A, che conta 10 rappresentanti, da un membro del partito B, che conta 15 rappresentanti, e da un membro del partito C, che conta 2 rappresentanti. Ci sono in totale $10 \times 15 \times 2 = 300$ possibili commissioni parlamentari.

Disposizioni

Consideriamo un insieme di n elementi. Una disposizione di r di essi è una scelta ordinata di r elementi tra quegli n. Si distinguono le disposizioni con ripetizione da quelle semplici (senza ripetizione), a seconda o meno che uno stesso elemento possa essere scelto più di una volta.

Le *disposizioni con ripetizione* di n elementi presi r alla volta sono in numero di

$$\prod_{i=1}^{r} n = n^r,$$

per il principio fondamentale del conteggio generalizzato.

• Esempio: le parole lunghe due lettere che si possono comporre con le lettere I, L, A sono $3^2 = 9$: $II, IL, IA, LI, LL, LA, AI, AL, AA$.

• Esempio: un bit può assumere i valori 0 o 1. Un byte è una fila di otto bit. Quanti byte ci sono?

Disposizioni semplici

Le *disposizioni semplici* di n elementi presi r alla volta sono in numero di

$$n \times (n-1) \ldots \times (n-r+1),$$

per il principio fondamentale del conteggio generalizzato.

• Esempio: le parole di due lettere diverse che si possono comporre con le lettere I, L, A sono $3 \times 2 = 6$: IL, IA, LI, LA, AI, AL.

• Esempio: di 10 concorrenti in una gara ciclistica vengono classificati solo i primi 3 arrivati. Quante possibili classifiche ci sono?

Campionamento da un'urna

Il *campionamento casuale da un'urna* è una estrazione di palle da un'urna. Può essere fatto con o senza reintroduzione.

Per *casuale* si intende dire che prima di ogni estrazione l'urna viene 'mescolata' appropriatamente per essere riportata a una condizione di irriconoscibilità e di dislocazione casuale delle palle. Un'operazione del genere viene fatta per le estrazioni del lotto.

La *reintroduzione* fa invece riferimento al fatto di riimmettere nell'urna ciascuna palla subito dopo averla estratta e averne registrate le caratteristiche di interesse, per esempio il suo numero o il suo colore.

Dunque:

- se un'urna contiene n palle distinguibili (per esempio numerate da 1 a n) e r palle vengono estratte con reintroduzione, le estrazioni possibili sono in numero di n^r .
- se un'urna contiene n palle distinguibili e r palle vengono estratte senza reintroduzione, le estrazioni possibili sono in numero di

$$n \times (n-1) \ldots \times (n-r+1)$$

.

Permutazioni

Le disposizioni semplici di n elementi presi n alla volta si chiamano anche permutazioni perché rappresentano tutti i modi in cui n elementi possono essere messi in fila. Esse sono in numero di

$$n \times (n-1) \ldots \times 2 \times 1,$$

una quantità per cui esiste il simbolo speciale $n!$ che si legge *n fattoriale*.

• Esempio: le permutazioni delle lettere I, L, A sono $3 \times 2 \times 1 = 3! = 6$:

$$ILA, IAL, LIA, LAI, AIL, ALI.$$

• Esempio: le possibili file che si possono fare con 10 bambini dell'asilo sono $10! = ?$

• Esempio (più difficile): supponiamo di fare due file, maschietti a destra e femminucce a sinistra. Ci sono $6!$ possibili file di maschietti e $4!$ file di femminucce possibili. In tutto ci sono quindi $6! \times 4! = 17280$ possibili file, per un'altra applicazione del principio del conteggio.

Combinazioni

Quanti sono i sottoinsiemi di 3 lettere dell'insieme di 5 lettere $\{A, B, C, D, E\}$? Finora sappiamo che ci sono $5 \times 4 \times 3 = 60$ parole di tre lettere diverse. Ma, per esempio, le parole ABC e BCA rappresentano lo stesso sottoinsieme, perché nella definizione di sottoinsieme l'ordine non conta. Ci sono $3! = 6$ parole equivalenti per ogni scelta, e ci sono quindi $60/6 = 10$ sottoinsiemi cercati. Essi sono ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE.

In generale, un sottoinsieme di ampiezza r da n elementi si chiama *combinazione di n elementi r alla volta*. Il numero di combinazioni di n elementi r alla volta è

$$\frac{n \times (n-1) \dots (n-r+1)}{r!} =: \binom{n}{r}$$

e si chiama anche coefficiente binomiale n su r.

• Esempio: la professoressa Tremendi interroga ogni Lunedí 10 studenti da una classe di 25. Esistono $\binom{25}{10}$ possibilità.

Binomio di Newton

Il nome coefficiente binomiale deriva dalla seguente espressione:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

detta formula del binomio di Newton.

- Esempio:

$$\begin{aligned}(a+b)^2 &= \binom{2}{0} a^2 b^0 + \binom{2}{1} ab + \binom{2}{2} a^0 b^2 \\ &= a^2 + 2ab + b^2.\end{aligned}$$

Fenomeni aleatori

La logica del certo è la logica della teoria degli insiemi e del calcolo su proposizioni (o eventi) che possono assumere il valore di vero o falso.

Il *calcolo delle probabilità* è invece la logica dell'incerto.

La probabilità si usa per ragionare sui possibili risultati di un *fenomeno (o esperimento) aleatorio*, del quale cioè non si può prevedere con certezza l'esito.

Aleatorio = opposto di deterministico.

Esempi di fenomeni aleatori

1. Il lancio di un dado.

2. Il lancio di una stessa moneta 4 volte.

3. La classificazione di 10 pezzi prodotti sequenzialmente da una macchina in conformi o non conformi, secondo che siano o non siano entro le specifiche di progetto.

4. L'estrazione di una mano di poker, cioè un insieme di cinque carte, da un mazzo di 52.

5. L'osservazione del tempo di guasto [min] di un circuito elettrico formato da tre resistenze in serie.

6. La registrazione giornaliera dei livelli massimi di polveri totali [mcg/mc] nell'aria alla centralina del Parco di San Giuliano nel Gennaio 2013.

Spazio campionario, risultati, eventi

Ω = *spazio campionario* = insieme dei possibili *risultati* di un fenomeno aleatorio. Un generico risultato si può indicare con $\omega \in \Omega$.

Di un *evento*, si può dire se sia vero o falso una volta che il fenomeno aleatorio di interesse è stato osservato.
Formalmente, evento = sottoinsieme $A \subset \Omega$.

I possibili risultati $\{\omega\}$, visti come singoletti, cioè insiemi contenenti un solo elemento, sono anch'essi eventi, detti *eventi elementari*.

Ω viene anche chiamato l'*evento certo*, perché sicuramente si verificherà.

Esempi di spazi campionari

1. $\Omega = \{1, 2, 3, 4, 5, 6\}$.

2. $\Omega =$ le sedici possibili sequenze di quattro dei simboli T e C, dove T indica 'testa' e C indica 'croce' (un possibile risultato è per esempio $\omega = TCCC$, cioè una testa seguita da tre croci).

3. $\Omega =$ le 2^{10} possibili sequenze di dieci dei simboli C e N, dove C indica 'conforme' e N indica 'non conforme' (un possibile risultato è per esempio $\omega = CCCNCCCNCC$).

4. $\Omega =$ i $\binom{52}{5}$ possibili sottoinsiemi delle 52 carte (uno dei quali è per esempio il poker d'assi più il tre di picche).

5. $\Omega = \mathcal{R}^+ := [0, \infty)$, cioè i numeri non negativi, visto che il tempo di guasto è un numero non negativo.

6. $\Omega =$ tutte le possibili sequenze di 31 numeri non negativi (la maggior parte contenuti tra 10 e 350).

Esempi di eventi

1. Il dado dà un punteggio superiore a quattro: A $= \{5, 6\}$.

2. Otteniamo almeno tre teste sui quattro lanci:
A $= \{TTTC, TTCT, TCTT, CTTT, TTTT\}$.

3. Tutti i pezzi sono conformi:
A $= \{CCCCCCCCCC\}$ (questo è anche un singoletto).

4. Si ottiene un poker: l'evento di interesse è dato da tutte le possibili mani contenenti un poker, che sono in numero di 13x48 perché 13 sono i possibili poker e 48 sono, per ogni dato poker, i modi di scegliere la quinta carta.

5. Il circuito ha una durata di meno di 50 ore : A $= [0, 50)$.

6. In nessun giorno si è superato il limite di 300 [mcg/mc]:
A $= \{(x_1, \ldots, x_{31}) : 0 \leq x_i \leq 300, i = 1, \ldots, 31\}$.

Operazioni logiche sugli eventi

La *negazione* o *complemento* di un evento A, indicata con $\bar{A}$, è l'evento che è vero quando A è falso ed è falso quando A è vero.

La negazione dell'evento certo è l'evento impossibile: $\bar{\Omega} = \emptyset$ (evento impossibile = insieme vuoto).

L'*intersezione* di due eventi A e B, indicata con $A \cap B$, è l'evento che è vero quando sia A che B sono veri e altrimenti è falso.

L'*unione* di due eventi A e B, indicata con $A \cup B$, è l'evento che è vero quando o A oppure B oppure entrambi sono veri, altrimenti falso.

L'evento A *implica* l'evento B, in simboli $A \subset B$, se il verificarsi di A implica il verificarsi di B.

Partizioni

Due eventi A e B si dicono *incompatibili*, o *disgiunti*, se non è possibile che siano entrambi veri, cioè se $A \cap B = \emptyset$.

Una famiglia di eventi si dice una *partizione* dell'evento certo se ogni coppia di insiemi della famiglia ha intersezione vuota e l'unione di tutti i componenti della famiglia è Ω.

Partizione numerabile $C_1, C_2, \ldots$:

$$C_i \cap C_j = \emptyset \quad \forall i, j$$
$$\bigcup_{i=1}^{\infty} C_i = \Omega,$$

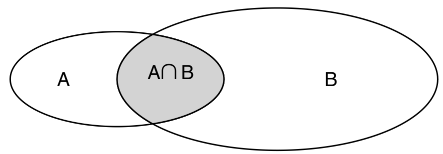
Una partizione finita $C_1, \ldots, C_n$, si può pensare come a una 'piastrellatura' di Ω come illustrato in figura.

Un qualsiasi evento A si può scrivere come unione delle sue intersezioni con gli elementi di una partizione:

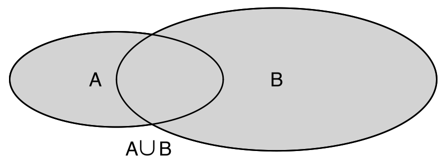
$$A = \bigcup_{i=1}^{\infty} (A \cap C_i)$$

Diagrammi di Venn

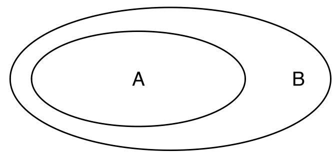
intersezione



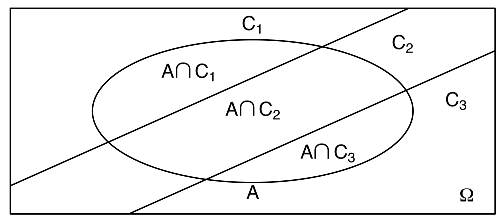
unione



inclusione



partizione



Esempio

Fenomeno aleatorio: lancio di un dado.
Eventi:

$$\text{A} = \{5, 6\} = \text{il risultato del lancio è superiore a 4}$$
$$\text{B} = \{2, 4, 6\} = \text{il risultato del lancio è pari.}$$

Allora

$$A \cap B = \{6\} = \begin{array}{l}\text{il risultato del lancio è pari} \\ \text{e superiore a 4}\end{array}$$
$$A \cup B = \{2, 4, 5, 6\} = \begin{array}{l}\text{il risultato del lancio è pari} \\ \text{oppure superiore a 4}\end{array}$$

Partizione dei numeri divisibili per 3 e non:

$$C_1 = \{3, 6\}$$
$$C_2 = \{1, 2, 4, 5\}$$

Abbiamo quindi

$$A = (A \cap C_1) \cup (A \cap C_2) = \{6\} \cup \{5\}$$

Università Ca' Foscari di Venezia
Federica Giummolè

Probabilità

Probabilità e Statistica
A.A. 2013/2014

Probabilità

La probabilità è una funzione degli eventi di uno spazio campionario, a valori nell'intervallo $[0, 1]$, definita tramite i seguenti assiomi:

1. $0 \leq \mathsf{P}(A) \leq 1$

2. $\mathsf{P}(\Omega) = 1$

3. Se A_1, $A_2, \ldots$ sono una sequenza di eventi incompatibili, cioè se $A_i \cap A_j = \emptyset \quad \forall i \neq j$, allora

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{P}(A_n)$$

Commenti

La probabilità dell'evento A, $\mathsf{P}(A)$, è un numero tra 0 e 1 che indica il grado di fiducia del ricercatore nell'avverarsi dell'evento A. Più $\mathsf{P}(A)$ è vicina a 1, più ci aspettiamo che l'evento si avveri. Una volta osservato il fenomeno aleatorio, sappiamo se A si è verificato o meno, e la sua probabilità non serve più.

Si può pensare alla probabilità come a una massa unitaria (in virtú della condizione di normalizzazione 2) da spargere sullo spazio campionario.

La massa che va a finire su eventi disgiunti è la somma delle masse sui singoli eventi.

Proprietà della probabilità

Probabilità del complemento: dato un evento A,

$$\mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A).$$

Probabilità dell'evento impossibile:

$$\mathsf{P}(\emptyset) = \mathsf{P}(\bar{\Omega}) = 1 - \mathsf{P}(\Omega) = 0.$$

Probabilità dell'unione: dati due eventi A e B,

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Probabilità di una partizione: se $C_1, C_2, \ldots$ sono una partizione, allora

$$\mathsf{P}(\bigcup_{i=1}^{\infty} C_i) = \mathsf{P}(\Omega) = 1.$$

Spazi campionari finiti

Se lo spazio campionario costituisce un insieme finito, $\Omega = \{\omega_1, \dots, \omega_n\}$, allora un'assegnazione di probabilità è data da n valori $p_1, \dots, p_n$ tali che:

1. $p_i \in [0, 1]$, $\forall i = 1, \dots, n$;

2. $\sum_{i=1}^{n} p_i = 1$;

3. $p_i = \mathsf{P}(\omega_i)$, $\forall i = 1, \dots, n$.

Dato che ogni evento $A \subset \Omega$ si può scrivere come unione (finita) degli eventi elementari (disgiunti) che lo costituiscono,

$$A = \{\omega_{i_1}, \dots, \omega_{i_r}\} = \bigcup_{k=1}^{r} \{\omega_{i_k}\},$$

si ha che

$$\mathsf{P}(A) = \sum_{k=1}^{r} \mathsf{P}(\{\omega_{i_k}\}) = \sum_{k=1}^{r} p_{i_k}.$$

Eventi elementari equiprobabili

In particolare, se possiamo supporre (per ragioni di simmetria) che tutti gli eventi elementari abbiano la stessa probabilità, allora

$$p_i = \mathsf{P}(\omega_i) = \frac{1}{n}, \quad \forall i = 1, \dots, n.$$

Per ogni evento $A = \{\omega_{i_1}, \dots, \omega_{i_r}\}$ si può dunque scrivere

$$\mathsf{P}(A) = \frac{r}{n} = \frac{\#A}{\#\Omega} = \frac{\text{numero casi favorevoli}}{\text{numero casi possibili}}.$$

Attenzione! Questa formula vale solo se gli eventi elementari sono *equiprobabili*.

• Esempio: Qual è la probabilità che il risultato del lancio di un dado equilibrato sia un numero divisibile per 3? Dato che il dado non è truccato, si può assumere che ognuno dei 6 possibili risultati abbia la stessa probabilità pari a 1/6. I casi favorevoli al nostro evento sono 2 ({3} e {6}) mentre quelli possibili sono 6. Il risultato è dunque $2/6 = 1/3$. Se il dado fosse truccato questo procedimento di calcolo non sarebbe corretto.

Esempio

Si consideri un'urna composta da quattro palle bianche numerate da 1 a 4 e tre palle nere numerate da 1 a 3. Si campioni casualmente una palla dall'urna.

È ragionevole assumere che ciascuna palla abbia probabilità 1/7 di essere estratta.

Consideriamo gli eventi

$$
\begin{aligned}
B &= \text{“viene estratta una palla bianca”};\\
N &= \text{“viene estratta una palla nera”}\\
&\qquad (\text{nota: } N = \bar{B});\\
C_i &= \text{“viene estratto il numero } i\text{”}, i = 1, 2, 3, 4;\\
D &= \text{“viene estratto un numero dispari”}.
\end{aligned}
$$

Probabilità nell'esempio

$$
\begin{aligned}
\mathsf{P}(B) &= 4/7 \\
\mathsf{P}(N) &= 1 - \mathsf{P}(B) = 3/7 \\
\mathsf{P}(C_i) &= 2/7, \quad i = 1, 2, 3 \\
\mathsf{P}(C_4) &= 1/7, \\
\mathsf{P}(D) &= 4/7 \\
\mathsf{P}(B \cap D) &= 2/7 \\
\mathsf{P}(B \cup D) &= \mathsf{P}(B) + \mathsf{P}(D) - \mathsf{P}(B \cap D) = 6/7 \\
\mathsf{P}(B \cap C_4) &= \mathsf{P}(C_4) = 1/7 \quad (C_4 \text{ implica } B) \\
\mathsf{P}(B \cap C_2) &= 1/7 \quad (C_2 \text{ non implica } B)
\end{aligned}
$$

Popolazioni e sottopopolazioni

Consideriamo una popolazione con N elementi suddivisi, a seconda che possiedano o meno una certa caratteristica, in due sottopopolazioni rispettivamente di K e $N-K$ elementi.

Qual è la probabilità che su n elementi estratti casualmente esattamente k abbiano quella caratteristica (e i rimanenti $n-k$ no)?

$\Omega = \{(x_1, \dots, x_n), x_i \in \text{ popolazione } \forall i\}$,
dove ogni n-upla ha la stessa probabilità di essere estratta (estrazioni casuali).
La cardinalità di Ω cambia a seconda che le estrazioni avvengano con o senza reinserimento.

$A_k =$"k elementi su n hanno la caratteristica richiesta".
Anche la cardinalità di A_k dipende dalla modalità di campionamento.

...continua

- Con reinserimento

$$\#\Omega = N^n$$

$$\#A_k = \binom{n}{k} K^k (N-K)^{n-k}$$

$$\Rightarrow \mathsf{P}(A_k) = \binom{n}{k} \left(\frac{K}{N}\right)^k \left(\frac{N-K}{N}\right)^{n-k}$$

- Senza reinserimento ($n < N$)

$$\#\Omega = N(N-1)(N-2)\dots(N-n+1)$$

$$\begin{aligned}\#A_k = &\binom{n}{k} K(K-1)\dots(K-k+1) \\ &(N-K)\dots(N-K-(n-k)+1)\end{aligned}$$

$$\Rightarrow \mathsf{P}(A_k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Università Ca' Foscari di Venezia
Federica Giummolè

Probabilità condizionata

Probabilità e Statistica
A.A. 2013/2014

Definizione

Sia B un evento di probabilità positiva.
La probabilità *condizionata* dell'evento A dato l'evento B è

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}, \quad \mathsf{P}(B) > 0.$$

Altre espressioni equivalenti: $\mathsf{P}(A|B)$ è la probabilità subordinata (a volte anche condizionale) di A subordinatamente a B. Da notare l'uso della sbarra verticale $|$.

$\mathsf{P}(A|B)$ rappresenta la probabilità di A valutata in presenza dell'informazione aggiuntiva che B si verifichi.

Intuitivamente, si restringe il campo delle possibilità non alla totalità dei possibili risultati Ω ma ad un suo sottoinsieme proprio $B \subset \Omega$.

Esempio (urna)

Si consideri l'urna dell'esempio al lucido 22.

Si valutino le probabilità condizionate di estrarre 1 dato che la palla è bianca, di estrarre 1 dato che la palla è nera e di estrarre una palla nera dato che si estrae 1.

Formalmente:

$$\mathsf{P}(C_1|B) = \frac{\mathsf{P}(C_1 \cap B)}{\mathsf{P}(B)} = 1/4$$
$$\mathsf{P}(C_1|N) = \frac{\mathsf{P}(C_1 \cap N)}{\mathsf{P}(N)} = 1/3$$
$$\mathsf{P}(N|C_1) = \frac{\mathsf{P}(N \cap C_1)}{\mathsf{P}(C_1)} = 1/2.$$

Nota: $\mathsf{P}(N|C_1)$ e $\mathsf{P}(C_1|N)$ significano cose molto diverse e non sono in relazione diretta. Invece, per esempio, $\mathsf{P}(N|C_1) = 1-\mathsf{P}(B|C_1) = 1/2$, perché le probabilità condizionate allo stesso evento obbediscono alle leggi della probabilità (fare per esercizio!).

La formula delle probabilità composte

La definizione di probabilità condizionata si può anche usare come formula pratica per la fattorizzazione della probabilità di un'intersezione:

$$\mathsf{P}(A \cap B) = \mathsf{P}(A|B)\,\mathsf{P}(B),$$

sempre che $\mathsf{P}(A|B)$ sia ben definita.

Questa formula si generalizza ad un qualsiasi numero di eventi $A_1, \ldots, A_n$ e viene anche chiamata la *formula delle probabilità composte*:

$$\begin{gathered}\mathsf{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathsf{P}(A_n|A_1 \cap \cdots \cap A_{n-1}) \ldots \\ \ldots \mathsf{P}(A_3|A_1 \cap A_2)\,\mathsf{P}(A_2|A_1)\,\mathsf{P}(A_1).\end{gathered}$$

Esempio (urna)

Si consideri ora l'esperimento che consiste nell'estrazione di 3 palline senza reinserimento dalla solita urna. Qual è la probabilità che le prime due siano bianche e la terza nera?

Siano

$B_i =$ "pallina bianca all'i-esima estrazione"
$N_i =$ "pallina nera all'i-esima estrazione"

Si ha che (probabilità composte)

$$\begin{aligned} \mathsf{P}(B_1 \cap B_2 \cap N_3) &= \mathsf{P}(N_3 | B_1 \cap B_2)\, \mathsf{P}(B_2 | B_1)\, \mathsf{P}(B_1) \\ &= \frac{3}{5}\frac{3}{6}\frac{4}{7} = \frac{6}{35}. \end{aligned}$$

Qual è la probabilità che siano tutte tre nere?

Eventi indipendenti

Nel caso particolare in cui

$$\mathsf{P}(A|B) = \mathsf{P}(A)$$

si dice che A e B sono *indipendenti*.

Si ha allora

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)$$

che può anche essere presa come definizione di eventi indipendenti.

La definizione si estende così: gli eventi $A_1, \ldots, A_n$ si dicono indipendenti se, comunque si prendono $k > 1$ di essi, si ha

$$\mathsf{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathsf{P}(A_{i_1}) \ldots \mathsf{P}(A_{i_k}).$$

• Esempio (urna): se le estrazioni dell'esempio al lucido precedente si effettuano con reinserimento, allora i tre eventi B_1, B_2 e N_3 sono indipendenti e si ha:

$$\mathsf{P}(B_1 \cap B_2 \cap N_3) = \mathsf{P}(B_1)\,\mathsf{P}(B_2)\,\mathsf{P}(N_3) = \frac{4}{7}\frac{4}{7}\frac{3}{7}.$$

Nota: eventi indipendenti e eventi disgiunti sono cose molto diverse. Due eventi sono disgiunti o meno a prescindere dalle loro probabilità.

Esempio

Si consideri l'esperimento di lanciare un dado equo due volte.

Si definiscano i seguenti eventi:

$A =$ "la somma dei dadi è 6"
$B =$ "la somma dei dadi è 7"
$C =$ "il primo dado dà 4"

Si ha

$$\mathrm{P}(A) = \frac{5}{36}\ , \quad \mathrm{P}(B) = \frac{1}{6}\ , \quad \mathrm{P}(C) = \frac{1}{6}\ .$$

Poiché

$$\mathrm{P}(A \cap C) = \mathrm{P}((4,2)) = \frac{1}{36}\ ,$$

allora A e C non sono indipendenti.

B e C sono invece indipendenti, ma non disgiunti. Infatti,

$$\mathrm{P}(B \cap C) = \frac{1}{36} = \mathrm{P}(B)\,\mathrm{P}(C).$$

Infine, A e B sono disgiunti ma non indipendenti:

$$\mathrm{P}(A \cap B) = \mathrm{P}(\emptyset) = 0.$$

Popolazioni e sottopopolazioni (bis)

Alla luce di quanto detto, possiamo reinterpretare i risultati al lucido 25.

Indichiamo con

$B_i =$ "l'i-esimo elemento estratto ha la proprietà richiesta"

• Con reinserimento: i risultati delle estrazioni successive sono indipendenti

$$\Rightarrow \quad \mathsf{P}(B_i) = \frac{K}{N} \quad \mathsf{P}(\bar{B}_i) = \frac{N-K}{N} \quad \forall i.$$

$$\begin{aligned}
\mathsf{P}(A_k) &= \binom{n}{k} \mathsf{P}(B_1 \cap \dots \cap B_k \cap \bar{B}_{k+1} \cap \dots \cap \bar{B}_n) \\
&= \binom{n}{k} \mathsf{P}(B_1) \dots \mathsf{P}(B_k) \mathsf{P}(\bar{B}_{k+1}) \dots \mathsf{P}(\bar{B}_n) \\
&= \binom{n}{k} \left(\frac{K}{N}\right)^k \left(\frac{N-K}{N}\right)^{n-k}
\end{aligned}$$

...continua

- Senza reinserimento: i risultati delle estrazioni successive NON sono indipendenti

$$
\begin{aligned}
\mathsf{P}(A_k) &= \binom{n}{k} \mathsf{P}(B_1 \cap \cdots \cap B_k \cap \bar{B}_{k+1} \cap \cdots \cap \bar{B}_n) \\
&= \binom{n}{k} \mathsf{P}(B_1) \ldots \mathsf{P}(B_k | B_1 \cap \cdots \cap B_{k-1}) \\
&\quad \mathsf{P}(\bar{B}_{k+1} | B_1 \cap \cdots \cap B_k) \ldots \mathsf{P}(\bar{B}_n | B_1 \cap \cdots \cap \bar{B}_{n-1}) \\
&= \binom{n}{k} \frac{K}{N} \frac{K-1}{N-1} \cdots \frac{K-(k-1)}{N-(k-1)} \\
&\quad \frac{N-K}{N-k} \cdots \frac{N-K-(n-k-1)}{N-(n-1)}
\end{aligned}
$$

Test diagnostici

La frazione dei soggetti affetti da una certa malattia (per esempio, la sieropositività HIV oppure la tubercolosi) in una popolazione si chiama *prevalenza*.

Si consideri un test diagnostico per la malattia.

La *sensitività* di un test è la probabilità che il test, somministrato a un malato, sia positivo.

La *specificità* di un test è la probabilità che il test, somministrato a un non malato, sia negativo.

Situazione ideale: sensitività = specificità = 1. La situazione ideale è spesso non raggiungibile, e i test reali sono imperfetti, cioè con sensitività < 1 e specificità < 1.

Falsi positivi e falsi negativi

Si immagini di somministrare un test diagnostico non perfetto a una persona estratta a caso dalla popolazione e si considerino gli eventi:

$$\begin{aligned} M &= \text{la persona estratta è malata} \\ + &= \text{il test dà risultato positivo} \\ - &= \text{il test dà risultato negativo} \\ \bar{M} \cap + &= \text{il test dà un falso positivo} \\ M \cap - &= \text{il test dà un falso negativo} \end{aligned}$$

Si ha allora

$$\begin{aligned} \mathsf{P}(M) &= \text{prevalenza} \\ \mathsf{P}(+|M) &= \text{sensitività} \\ \mathsf{P}(-|\bar{M}) &= \text{specificità} \end{aligned}$$

Probabilità di un falso positivo:

$$\begin{aligned} \mathsf{P}(\bar{M} \cap +) &= \mathsf{P}(\bar{M})\,\mathsf{P}(+|\bar{M}) \\ &= (1 - \text{prevalenza}) \times (1 - \text{specificità}). \end{aligned}$$

Probabilità di un falso negativo:

$$\begin{aligned} \mathsf{P}(M \cap -) &= \mathsf{P}(M)\,\mathsf{P}(-|M) \\ &= \text{prevalenza} \times (1 - \text{sensitività}). \end{aligned}$$

Esempio

Si studi un nuovo test per l'HIV.

Sia

$$\text{prevalenza} = \mathsf{P}(HIV) = 0.001$$

la proporzione di HIV nella popolazione studiata.

Sia inoltre:

$$\begin{aligned} \mathsf{P}(+|HIV) &= .95 \quad \text{sensitività} \\ \mathsf{P}(-|\overline{HIV}) &= .98 \quad \text{specificità} \end{aligned}$$

La probabilità di falso positivo è

$$\begin{aligned} \mathsf{P}(\overline{HIV} \cap +) &= \mathsf{P}(\overline{HIV})\,\mathsf{P}(+|\overline{HIV}) \\ &= (1-0.001)(1-0.98) = 0.01998. \end{aligned}$$

La probabilità di falso negativo è

$$\begin{aligned} \mathsf{P}(HIV \cap -) &= \mathsf{P}(HIV)\,\mathsf{P}(-|HIV) \\ &= 0.001(1-0.95) = 0.00005. \end{aligned}$$

La legge della probabilità totale

Se $C_1, C_2, \ldots$ sono una partizione dell'evento certo, la probabilità di un qualsiasi evento A può essere scritta come

$$\mathsf{P}(A) = \sum_i \mathsf{P}(A \cap C_i) = \sum_i \mathsf{P}(C_i)\,\mathsf{P}(A|C_i)$$

che si dice *legge della probabilità totale* o formula della partizione.

La prima uguaglianza viene dal fatto che

$$A = \bigcup_i (A \cap C_i),$$

che sono eventi a due a due disgiunti. La seconda uguaglianza segue dalla definizione di probabilità condizionata, sempre che $\mathsf{P}(C_i) > 0$, $\forall i = 1, 2, \ldots$.

Esempio HIV

Nell'esempio di prima, HIV e $\overline{HIV}$ costituiscono una semplice partizione formata da due 'piastrelle'.

Calcoliamo la probabilità che il test, somministrato ad una persona campionata a caso dalla popolazione, sia positivo:

$$
\begin{aligned}
\mathsf{P}(+) &= \mathsf{P}(+|HIV)\,\mathsf{P}(HIV) + \mathsf{P}(+|\overline{HIV})\,\mathsf{P}(\overline{HIV}) \\
&= 0.95 \times 0.001 + (1 - 0.98) \times (1 - 0.001) \\
&= 0.02093
\end{aligned}
$$

cioè in pratica avremo, a lungo andare, il 2 per cento di positivi, siano essi veri positivi o falsi positivi.

La formula di Bayes (1702-1761)

Sia data la partizione $C_1, C_2, \ldots$ e tutti i suoi elementi abbiano probabilità positiva. Sia A un ulteriore evento, anch'esso con probabilità positiva.

Fissiamo l'attenzione su uno specifico elemento C_m della partizione. Abbiamo allora

$$\mathsf{P}(C_m|A) = \frac{\mathsf{P}(C_m \cap A)}{\mathsf{P}(A)}$$

(definizione di probabilità condizionata)

$$= \frac{\mathsf{P}(A|C_m)\,\mathsf{P}(C_m)}{\sum_i \mathsf{P}(A|C_i)\,\mathsf{P}(C_i)}$$

(legge della probabilità totale).

Esempio del test diagnostico

Perchè dovremmo essere interessati alla probabilità condizionata di un singolo elemento C_m?

Riconsideriamo l'esempio HIV. È di primario interesse l'evento che una persona risultata positiva sia effettivamente malata.

Usando i numeri di prima:

$$\mathsf{P}(HIV|+) =$$

$$= \frac{\mathsf{P}(HIV)\,\mathsf{P}(+|HIV)}{\mathsf{P}(HIV)\,\mathsf{P}(+|HIV) + \mathsf{P}(\overline{HIV})\,\mathsf{P}(+|\overline{HIV})}$$

$$= ?$$

Nel campo della diagnostica, P(malattia$|+$) viene chiamata a volte valore predittivo positivo.

Università Ca' Foscari di Venezia
Federica Giummolè

Variabili aleatorie

Probabilità e Statistica
A.A. 2013/2014

Abbiamo già visto vari esempi di fenomeni aleatori e spazi campionari ad essi collegati.

Uno spazio campionario relativo ad un esperimento o ad un fenomeno casuale può essere di varia natura. In particolare, non è detto che sia un insieme numerico.

In molte situazioni, anziché essere interessati allo specifico risultato di un esperimento, siamo interessati ad una sua funzione numerica.

• Esempio: si fa una scommessa in cui si vincono 1000 lire se lanciando una moneta equilibrata si ottiene testa e se ne perdono 1500 se il risultato è croce. Chiaramente ciò che ci preme di più è l'importo della vincita (negativo se si tratta della perdita) e non il risultato esatto del lancio della moneta.

Variabili aleatorie

Una *variabile aleatoria* X è una funzione che assume valori numerici determinati dall'esito di un certo fenomeno aleatorio. Formalmente, se Ω è lo spazio campionario relativo al fenomeno di interesse, X è una particolare funzione

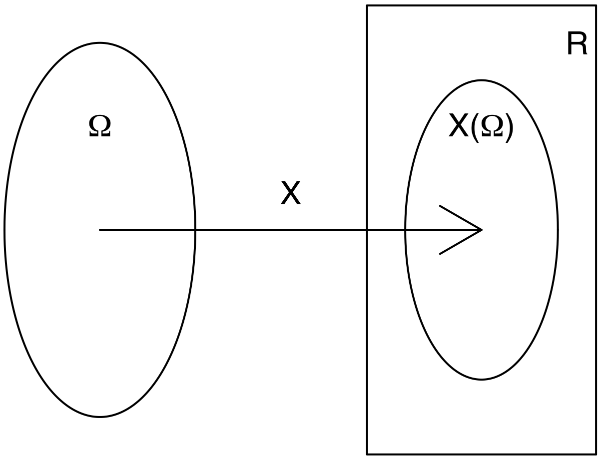
$$X : \Omega \rightarrow \mathrm{R} \; .$$

Le variabili aleatorie si indicano di solito con la lettera maiuscola.

• Esempi:
1. $S_4 =$ numero di teste in 4 lanci consecutivi di una moneta; i possibili valori di S_4 sono: 0, 1, 2, 3 e 4.
2. $X =$ vincita nel gioco descritto nel lucido precedente. I possibili valori di X sono 1000 e -1500.
3. $T =$ tempo di vita di un componente elettronico prodotto da una ditta. I possibili valori di T sono tutti i numeri reali maggiori di 0.

Spazio campionario indotto

Una variabile aleatoria associata ad un esperimento definisce dunque un nuovo spazio campionario numerico, costituito da tutti i possibili valori assunti dalla variabile stessa.



In questo modo si passa da un generico spazio campionario ad un sottoinsieme di R.

Bisogna ora assegnare le probabilità agli eventi del nuovo spazio campionario. Vedremo due modi differenti di fare quest'assegnazione, a seconda che lo spazio campionario indotto sia un sottoinsieme discreto (come nei precedenti esempi 1 e 2) o continuo (esempio 3) di R.

Variabili aleatorie discrete

Una variabile aleatoria *discreta* X assume valori in un insieme finito o numerabile di punti, $\{x_1, x_2, \ldots, x_i, \ldots\}$.

Un modello probabilistico per X è un'assegnazione di probabilità ad ogni suo possibile valore:

$$\mathsf{P}(X = x_i) = p_i, \quad \forall i = 1, 2, \ldots \ .$$

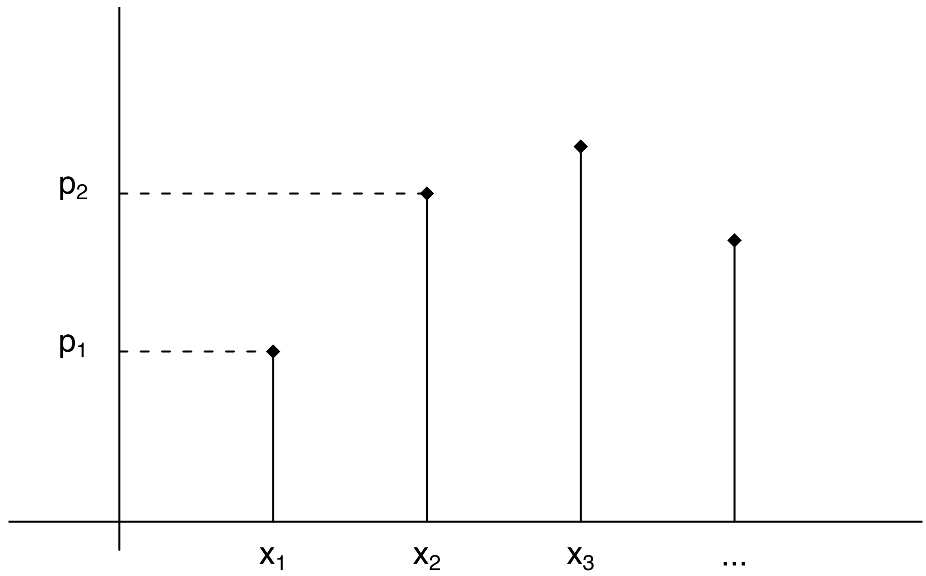
Le probabilità p_i sono tali che:
1. $0 \leq p_i \leq 1, \quad \forall i = 1, 2, \ldots$;
2. $\sum_i p_i = 1$.

Per calcolare $\mathsf{P}(X \in A)$, si sommano le probabilità dei singoli valori che appartengono ad A:

$$\mathsf{P}(X \in A) = \sum_{i: x_i \in A} p_i \ .$$

Funzioni di probabilità

Un'assegnazione di probabilità per X viene chiamata *funzione di probabilità* e può essere rappresentata graficamente tramite un diagramma a bastoncini:



Esempio (urna)

Consideriamo la solita urna.

• X = numero estratto dall'urna.
$X \in \{1, 2, 3, 4\}$
$\mathsf{P}(X = 1) = 2/7$, $\mathsf{P}(X = 2) = 2/7$,
$\mathsf{P}(X = 3) = 2/7$, $\mathsf{P}(X = 4) = 1/7$.
Si scrive anche

$$X = \begin{cases} 1 & 2 & 3 & 4 \\ 2/7 & 2/7 & 2/7 & 1/7 \end{cases}$$

Se $A = \{1, 2\}$, allora
$\mathsf{P}(X \in A) = \mathsf{P}(X \leq 2) = \frac{2}{7} + \frac{2}{7} = \frac{4}{7}$.

• Supponiamo di vincere 3000 lire se la pallina estratta è bianca e di perderne 2000 se è nera.
Y = importo vinto nel gioco.

$$Y = \begin{cases} 3000 & -2000 \\ 4/7 & 3/7 \end{cases}$$

...continua

• Consideriamo ora un'estrazione di due palline senza reinserimento.

$S =$ somma dei due numeri estratti.

$$S = \begin{cases} 2 & 3 & 4 & 5 & 6 & 7 \\ 2/42 & 8/42 & 10/42 & 12/42 & 6/42 & 4/42 \end{cases}$$

Infatti, ad esempio,

$$\begin{aligned} \mathsf{P}(S=2) &= \mathsf{P}(X_1 = 1 \cap X_2 = 1) \\ &= \mathsf{P}(X_1 = 1)\,\mathsf{P}(X_2 = 1 | X_1 = 1) \\ &= \frac{2}{7} \times \frac{1}{6} \end{aligned}$$

Cosa cambia se l'estrazione avviene con reinserimento? ($\rightarrow$ esercizio)

Variabili aleatorie continue

Una variabile aleatoria *continua* X assume valori in un insieme continuo di punti (un sottoinsieme di R non numerabile).

Un modello probabilistico per X è un'assegnazione di probabilità ad ogni sottoinsieme di suoi possibili valori:

$$\mathrm{P}(X \in A) = \text{ area su A sottesa da una curva.}$$

La curva è il grafico di una funzione $f(x)$ tale che:

1. $f(x) \geq 0$, $\quad \forall x \in \mathrm{R}$;
2. $\int_{\mathrm{R}} f(x)\, dx = 1$, cioè l'area totale sotto il grafico di $f(x)$ è 1.

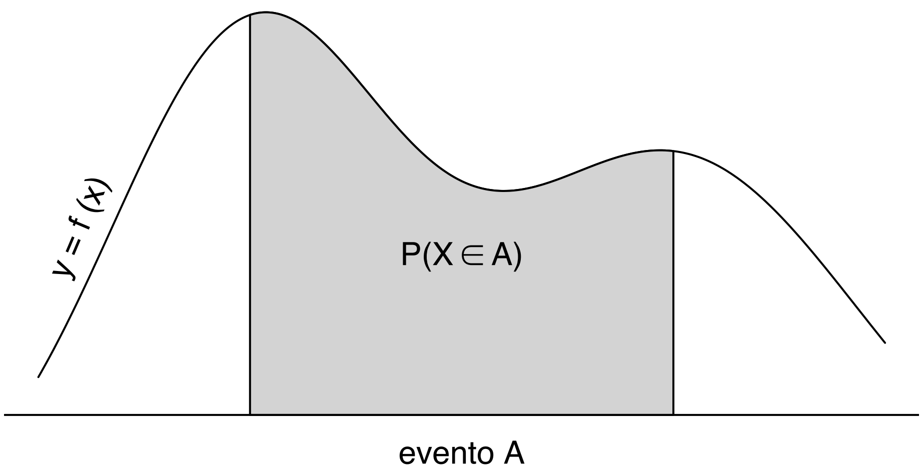
Densità di probabilità

Una funzione $f(x)$ con le proprietà precedenti viene chiamata *densità di probabilità*.

Una volta assegnata una densità di probabilità per la variabile X, si può scrivere:

$$\mathrm{P}(X \in A) = \int_A f(x)\, dx \ ,$$

per ogni evento A di R.



Importante: $\mathrm{P}(X = x) = 0, \ \forall x \in \mathrm{R}$.

Esempio

Sia X una variabile aleatoria continua con densità

$$f(x) = 2\mathrm{e}^{-2x}\mathbb{1}_{(0,+\infty)}(x).$$

$f(x)$ è davvero una densità:
1. $f(x) \geq 0, \ \forall x \in \mathrm{R}$,
2. $\int_{\mathrm{R}} f(x)\,dx = \int_0^{+\infty} 2\mathrm{e}^{-2x}\,dx = 1$.

Inoltre:

$$\mathrm{P}(X \in (1,2)) = \int_1^2 2\mathrm{e}^{-2x}\,dx = \mathrm{e}^{-2} - \mathrm{e}^{-4}$$

e

$$\mathrm{P}(X \in (-1,1)) = \int_{-1}^0 0\,dx + \int_0^1 2\mathrm{e}^{-2x}\,dx = 1 - \mathrm{e}^{-2}.$$

Funzione di ripartizione

Si dice *funzione di ripartizione* (o di distribuzione cumulativa) di una variabile aleatoria X la funzione $F : \mathsf{R} \to [0, 1]$ così definita:

$$F(x) = \mathsf{P}(X \leq x), \quad \forall x \in \mathsf{R}.$$

La funzione di ripartizione in un dato punto x è semplicemente la probabilità (per questo il suo codominio è $[0, 1]$) che la variabile X assuma valori minori o al più uguali a x.

La funzione di ripartizione F ha le seguenti proprietà:
1. F è non decrescente,
2. F è continua a destra,
3. $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to+\infty} F(x) = 1$.

F.r. di una v.a. discreta

Se

$$X = \begin{cases} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{cases}$$

è una v.a. discreta, allora

$$F(x) = \sum_{i:x_i \leq x} \mathsf{P}(X = x_i) = \sum_{i:x_i \leq x} p_i.$$

La funzione di ripartizione di una v.a. discreta è una funzione costante a tratti, con salti in corrispondenza dei punti massa $x_1, x_2, \ldots$.

Dalla funzione di ripartizione si può risalire alla funzione di probabilità della variabile così:

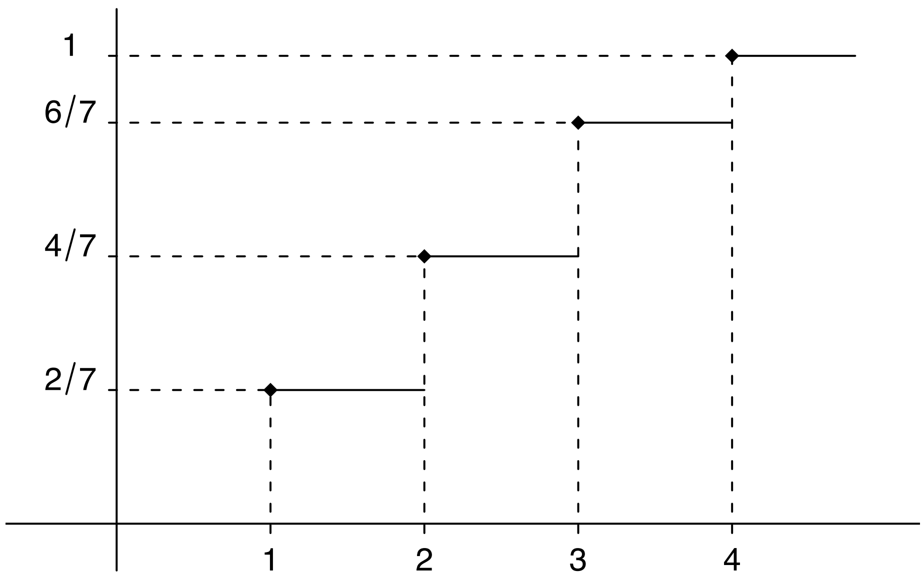
$$\mathsf{P}(X = x) = F(x) - F(x^-), \quad \forall x \in \mathsf{R},$$

cioè la probabilità che X assuma il valore x è uguale al salto della funzione di ripartizione nel punto x.

Esempio

Per la variabile X dell'esempio nel lucido 6 si ha:

$$F(x) = \begin{cases} 0 & x < 1 \\ 2/7 & 1 \leq x < 2 \\ 4/7 & 2 \leq x < 3 \\ 6/7 & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$



Esercizio: trovare la funzione di probabilità di X a partire dalla funzione di ripartizione.

F.r. di una v.a. continua

Se X è una v.a. continua con densità $f(x)$, allora

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

La funzione di ripartizione di una v.a. continua è una funzione continua.

Dalla funzione di ripartizione si può risalire alla densità di probabilità della variabile così:

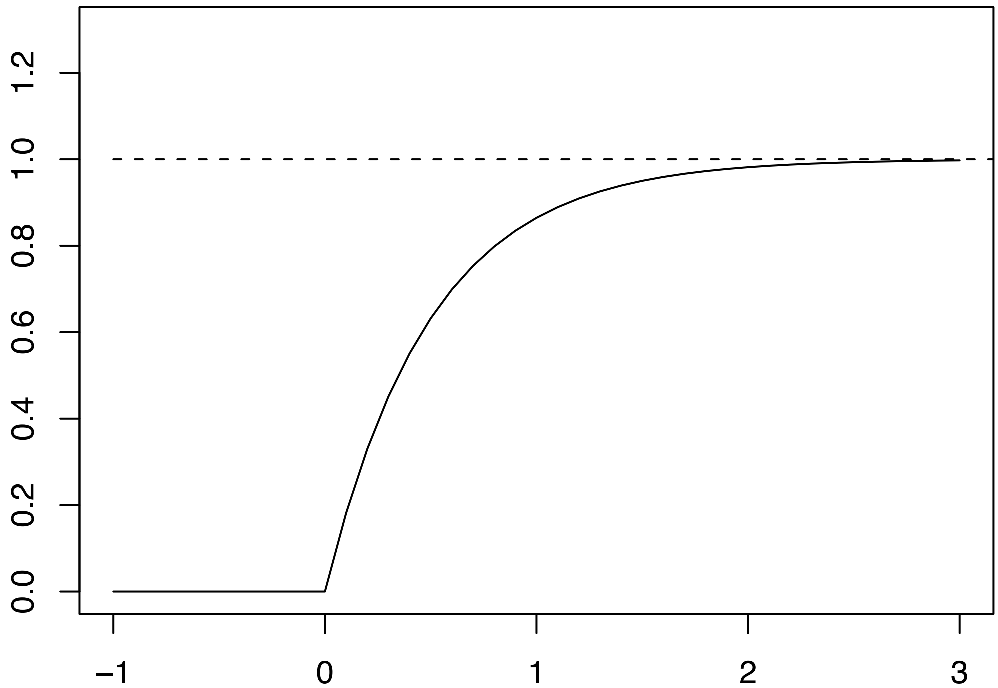
$$f(x) = \frac{dF(x)}{dx},$$

in tutti i punti in cui $F(x)$ è derivabile, cioè la densità di X in un punto x è uguale alla derivata della funzione di ripartizione nel punto stesso (se questa esiste).

Esempio

Per la variabile X dell'esempio nel lucido 10 si ha:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \mathrm{e}^{-2x} & x \geq 0 \end{cases}$$



Esercizio: trovare la densità di X a partire dalla funzione di ripartizione.

Costanti caratteristiche

Una costante caratteristica o indice è un numero associato ad una variabile aleatoria o alla sua distribuzione di probabilità. Conoscere il valore di un indice significa avere informazione sulla distribuzione stessa.

Definiremo ora alcune importanti costanti caratteristiche: il valore atteso (che è un indice di posizione), la varianza (che è un indice di dispersione) e i quantili di una variabile aleatoria.

Valore atteso di una variabile aleatoria discreta

Sia

$$X = \begin{cases} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{cases}$$

Allora il *valore atteso* (o media) di X è

$$\begin{aligned} \mathsf{E}(X) &= \sum_i x_i p_i \\ &= x_1 p_1 + x_2 p_2 + \dots \ . \end{aligned}$$

• Esempio: per l'esempio nei lucidi 6 e 7 abbiamo:

$$\mathsf{E}(X) = 1\frac{2}{7} + 2\frac{2}{7} + 3\frac{2}{7} + 4\frac{1}{7} = \frac{16}{7}$$

$$\mathsf{E}(Y) = 3000\frac{4}{7} - 2000\frac{3}{7} = \frac{6000}{7}$$

$$\mathsf{E}(S) = ?$$

• $\mathbb{1}_A$ è la variabile aleatoria che vale 1 con probabilità $\mathsf{P}(A)$ e 0 con probabilità $\mathsf{P}(\bar{A})$. Il suo valore atteso è $\mathsf{P}(A)$.

La scommessa

Facciamo un gioco: tu scegli un prezzo p da pagare per giocare, io lo pago e poi guadagno 1 euro se l'evento A si verifica o 0 euro se A non si verifica. Qual è il prezzo giusto della scommessa? Giusto significa che per me (e anche per te) è indifferente essere quello che punta o quello che paga.

Se X è il mio guadagno aleatorio nella scommessa, allora

$$X = 1 \times \mathbb{1}_A + 0 \times \mathbb{1}_{\bar{A}} - p.$$

Un prezzo ragionevole per giocare è dunque quello che mi dà un guadagno atteso nullo (altrimenti nessuno accetterebbe la scommessa):

$$0 = \mathsf{E}\,(X)\,,$$

cioè

$$p = 1 \times \mathsf{P}(A) + 0 \times \mathsf{P}(\bar{A}) = \mathsf{P}(A).$$

Valore atteso di una variabile aleatoria continua

Sia X una v.a. continua con densità $f(x)$. Allora il *valore atteso* (o media) di X è

$$\mathsf{E}\,(X) \;=\; \int_{\mathsf{R}} x f(x)\, dx.$$

- Esempio: per l'esempio nel lucido 10 abbiamo:

$$\mathsf{E}\,(X) = \int_0^{+\infty} 2x\mathrm{e}^{-2x}\, dx = \frac{1}{2}.$$

Proprietà del valore atteso

Il valore atteso ha le seguenti proprietà:

1. $\mathsf{E}(a) = a$, dove a è una costante;

2. $\mathsf{E}(aX + b) = a\,\mathsf{E}(X) + b$, dove a e b sono costanti;

3. $\mathsf{E}(X_1 + X_2) = \mathsf{E}(X_1) + \mathsf{E}(X_2)$, dove X_1 e X_2 sono v.a. .

Varianza di una variabile aleatoria discreta

Sia

$$X = \begin{cases} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{cases}$$

Allora la *varianza* di X è

$$\mathsf{Var}\,(X) = \sum_i (x_i - \mathsf{E}\,(X))^2 p_i \ .$$

Una formula pratica per il calcolo è

$$\mathsf{Var}\,(X) = \sum_i x_i^2 p_i - [\mathsf{E}\,(X)]^2 \ .$$

• Esempio: per l'esempio nei lucidi 6 e 7 abbiamo:

$$\mathsf{Var}\,(X) = 1\frac{2}{7} + 2^2\frac{2}{7} + 3^2\frac{2}{7} + 4^2\frac{1}{7} - \left(\frac{16}{7}\right)^2 .$$

Esercizio: calcolare la varianza per le altre due variabili dell'esempio.

Varianza di una variabile aleatoria continua

Sia X una v.a. continua con densità $f(x)$. Allora la *varianza* di X è

$$\mathrm{Var}\,(X) = \int_{\mathrm{R}} (x - \mathrm{E}\,(X))^2 f(x)\,dx.$$

Una formula pratica per il calcolo è

$$\mathrm{Var}\,(X) = \int_{\mathrm{R}} x^2 f(x)\,dx - [\mathrm{E}\,(X)]^2 \ .$$

- Esempio: per l'esempio nel lucido 10 abbiamo:

$$\begin{aligned} \mathrm{Var}\,(X) &= \int_0^{+\infty} 2x^2 \mathrm{e}^{-2x}\,dx - \left(\frac{1}{2}\right)^2 \\ &= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

Proprietà della varianza

La varianza ha le seguenti proprietà:

1. $\text{Var}(a) = 0$, dove a è una costante;

2. $\text{Var}(aX + b) = a^2 \text{Var}(X)$, dove a e b sono costanti.

Valore atteso di $g(X)$

Sia $Y = g(X)$ una v.a. ottenuta trasformando la v.a. X tramite la funzione $g : \mathsf{R} \to \mathsf{R}$. Il valore atteso di Y si può calcolare anche senza conoscere direttamente la distribuzione di probabilità di Y:

- X discreta: $\mathsf{E}(Y) = \sum_i g(x_i)p_i$.

- X continua: $\mathsf{E}(Y) = \int_{\mathsf{R}} g(x)f(x)\,dx$.

Anche la varianza di X è il valore atteso di una trasformata della X tramite la funzione $g(x) = (x - \mathsf{E}(X))^2$ e si può scrivere come

$$\mathsf{Var}(X) = \mathsf{E}\left(X^2\right) - [\mathsf{E}(X)]^2 .$$

Le proprietà enunciate nei lucidi 20 e 23 si dimostrano utilizzando le formule precedenti.

Moda di una v.a.

La *moda* di una variabile aleatoria X è il punto (o i punti) in cui la funzione di probabilità (o di densità) assume valore massimo.

• Esempi:
1. La variabile X dell'esempio nei lucidi 6 e 7 ha tre mode in 1, 2 e 3, mentre la Y ne ha una in 3000. E la S?
2. La variabile X dell'esempio nel lucido 10 ha moda in 0.

La moda, così come il valore atteso di una v.a., è un indice di posizione. Altri importanti indici di posizione sono la mediana e, in generale, i quantili di una v.a..

Mediana di una v.a.

La *mediana* di una variabile aleatoria X è il minimo valore m per cui

$$F(m) = \mathsf{P}(X \leq m) \geq \frac{1}{2}.$$

Per una v.a. continua (cioè con f.r. continua) la mediana è l'unico punto m in cui $F(m) = \mathsf{P}(X \leq m) = \mathsf{P}(X \geq m) = 1/2$.

• Esempi:
1. La variabile X dell'esempio nei lucidi 6 e 7 ha mediana in 2, mentre la Y ce l'ha in 3000. E la S? Disegnate i grafici delle relative f.r. e verificate quanto appena detto.
2. Per la variabile X dell'esempio nel lucido 10 la mediana è tale che

$$1 - \mathrm{e}^{-2m} = \mathrm{e}^{-2m} = \frac{1}{2},$$

cioè $m = \log(2)/2$.

Quantili di una v.a.

I *quantili* sono indici di posizione che generalizzano il concetto di mediana di una distribuzione.

Fissato un valore $\alpha \in (0,1)$, il *quantile di livello* α di una variabile aleatoria X è il minimo valore q_α per cui

$$F(q_\alpha) = \mathsf{P}(X \leq q_\alpha) \geq \alpha.$$

Per una v.a. continua (con f.r. continua) il quantile di livello α è l'unico punto q_α in cui $F(q_\alpha) = \mathsf{P}(X \leq q_\alpha) = \alpha$.

La mediana non è altro che il quantile di livello 1/2.

I *percentili* sono i quantili di livello $k/100$, con $k = 1, 2, \ldots, 99$.
I *decili* sono i quantili di livello $k/10$, con $k = 1, 2, \ldots, 9$.
I *quartili* sono i quantili di livello 0.25 (primo quartile), 0.5 (mediana) e 0.75 (terzo quartile).

• Esercizio: Calcolare i quartili delle variabili definite nei lucidi 6, 7 e 10.

Università Ca' Foscari di Venezia

Federica Giummolè

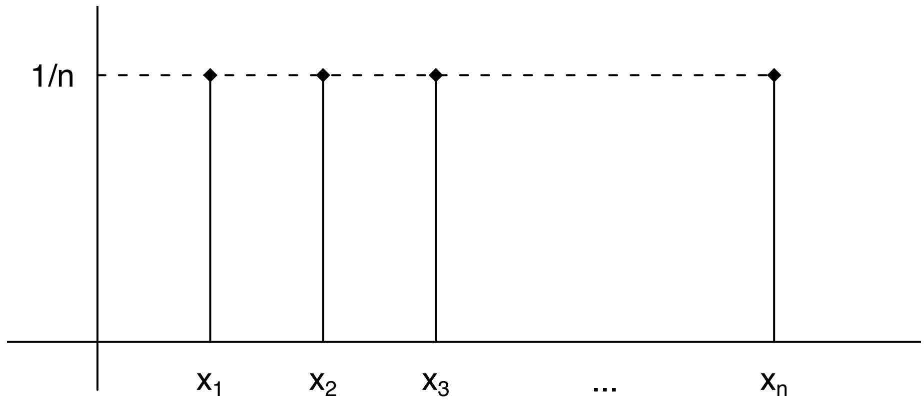
Variabili aleatorie discrete

Probabilità e Statistica

A.A. 2013/2014

Distribuzione uniforme

Consideriamo una variabile aleatoria X che assume un numero finito di valori, $\{x_1, \ldots, x_n\}$, tutti con la stessa probabilità $p_i = 1/n$, $i = 1, \ldots, n$.



Si dice allora che X ha una *distribuzione uniforme* e si scrive $X \sim U\{x_1, \ldots, x_n\}$.

Esempi:

1. X descrive il risultato del lancio di un dado equilibrato. I valori possibili per X sono gli interi fra 1 e 6, ciascuno con probabilità 1/6 di verificarsi (dado non truccato).
2. Si fa una scommessa in cui si guadagnano 1000 lire se lanciando una moneta equilibrata si ottiene testa e se ne perdono 1500 se il risultato è croce. La variabile X che indica il guadagno ottenuto, ha distribuzione uniforme sull'insieme $\{x_1 = 1000, x_2 = -1500\}$.

Popolazioni e sottopopolazioni 3

Riconsideriamo il campionamento da una popolazione divisa in due sottopopolazioni.

Questo schema costituisce un modello applicabile a molte situazioni differenti:
difettoso/non difettoso in controllo della qualità;
sotto soglia/sopra soglia in rilevazioni di inquinamento;
destra/sinistra in *polls* elettorali.

Conoscendo la composizione della popolazione (proporzione di individui con la proprietà di interesse) e il tipo di campionamento (con o senza reinserimento), abbiamo imparato a calcolare la probabilità di eventi del tipo:
A_k="k elementi su n hanno la proprietà richiesta".

Spesso si indica come "successo" l'estrazione di un individuo con la proprietà richiesta e l'evento A_k diventa "k successi su n estrazioni".

Possiamo ora interpretare questi risultati utilizzando le variabili aleatorie.

Distribuzione ipergeometrica

Sia X la variabile aleatoria che conta il numero di successi su n estrazioni senza reinserimento da una popolazione con N elementi dei quali K sono considerati successo.

Si dice allora che X ha una *distribuzione ipergeometrica* di parametri N, K e n e si scrive $X \sim Ip(N, K, n)$.

Si ha che

$$\mathsf{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \ ,$$

per $k = \max\{0, n-(N-K)\}, \dots, \min\{n, K\}$, con $n \le N$.

Non dimostreremo che

$$\mathsf{E}\,(X) = n\frac{K}{N}$$

e

$$\mathsf{Var}\,(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1} \ .$$

Esempio

Da un mazzo di 52 carte se ne scelgono 13. Qual è la probabilità che fra quelle 13 carte ci siano almeno 2 figure?

Sia X la variabile che conta il numero di figure in 13 carte estratte dal mazzo.

Allora $X \sim Ip(N = 52, K = 12, n = 13)$ e la probabilità richiesta è

$$\begin{aligned} \mathsf{P}(X \geq 2) &= 1 - \mathsf{P}(X \leq 1) \\ &= 1 - \mathsf{P}(X = 0) - \mathsf{P}(X = 1) \\ &= 1 - \frac{\binom{12}{0}\binom{52-12}{13-0}}{\binom{52}{13}} - \frac{\binom{12}{1}\binom{52-12}{13-1}}{\binom{52}{13}} \ . \end{aligned}$$

Il numero atteso di figure fra le 13 carte estratte è

$$\mathsf{E}\,(X) = 13\frac{12}{52} = 3 \ .$$

Distribuzione binomiale

Sia X la variabile aleatoria che conta il numero di successi su n estrazioni con reinserimento da una popolazione con N elementi dei quali K sono considerati successo.

La probabilità di successo rimane invariata, uguale a $p = K/N$, ad ogni estrazione successiva.

Si dice allora che X ha una *distribuzione binomiale* di parametri n e $p \in (0, 1)$ e si scrive $X \sim Bi(n, p)$.

Si ha che

$$\mathsf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n \ .$$

Si osservi che si tratta effettivamente di una distribuzione di probabilità:

$$\begin{aligned} \sum_{k=0}^{n} \mathsf{P}(X = k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= (p + 1 - p)^n = 1, \end{aligned}$$

per la formula del binomio di Newton.

Esempio (urna)

Si consideri ancora la solita urna.

Sia S_3 la variabile che conta il numero di palline bianche ottenute in tre estrazioni con reinserimento.

Qual è la $\mathsf{P}(S_3 = 2)$?

Considerando l'estrazione di pallina bianca come successo e visto che $\mathsf{P}(\text{bianca}) = 4/7$, si può affermare che $S_3 \sim Bi(3, 4/7)$.

Allora,

$$\mathsf{P}(S_3 = 2) = \binom{3}{2} \left(\frac{4}{7}\right)^2 \left(\frac{3}{7}\right)^{(3-2)} = 0.4198 \ .$$

Campionamento da popolazioni infinite

La distribuzione binomiale si utilizza anche nel caso di campionamento da popolazioni infinite ($N \to \infty$), quando si conosce ad ogni estrazione la probabilità di successo p.

Se la popolazione viene considerata infinita, non si distingue nemmeno fra estrazioni con o senza reinserimento perchè si assume che in ogni caso la probabilità di successo in estrazioni successive non cambi, e si usa sempre la binomiale.

Esempio

Dalla popolazione (finita, ma così numerosa da considerarsi infinita) dei pezzi prodotti da una macchina in una fabbrica, si sceglie un lotto di $n = 20$ pezzi (senza reinserimento, ma questo non ha più importanza). Si conosce la probabilità $p = 0.05$ che un pezzo prodotto dalla macchina sia difettoso.

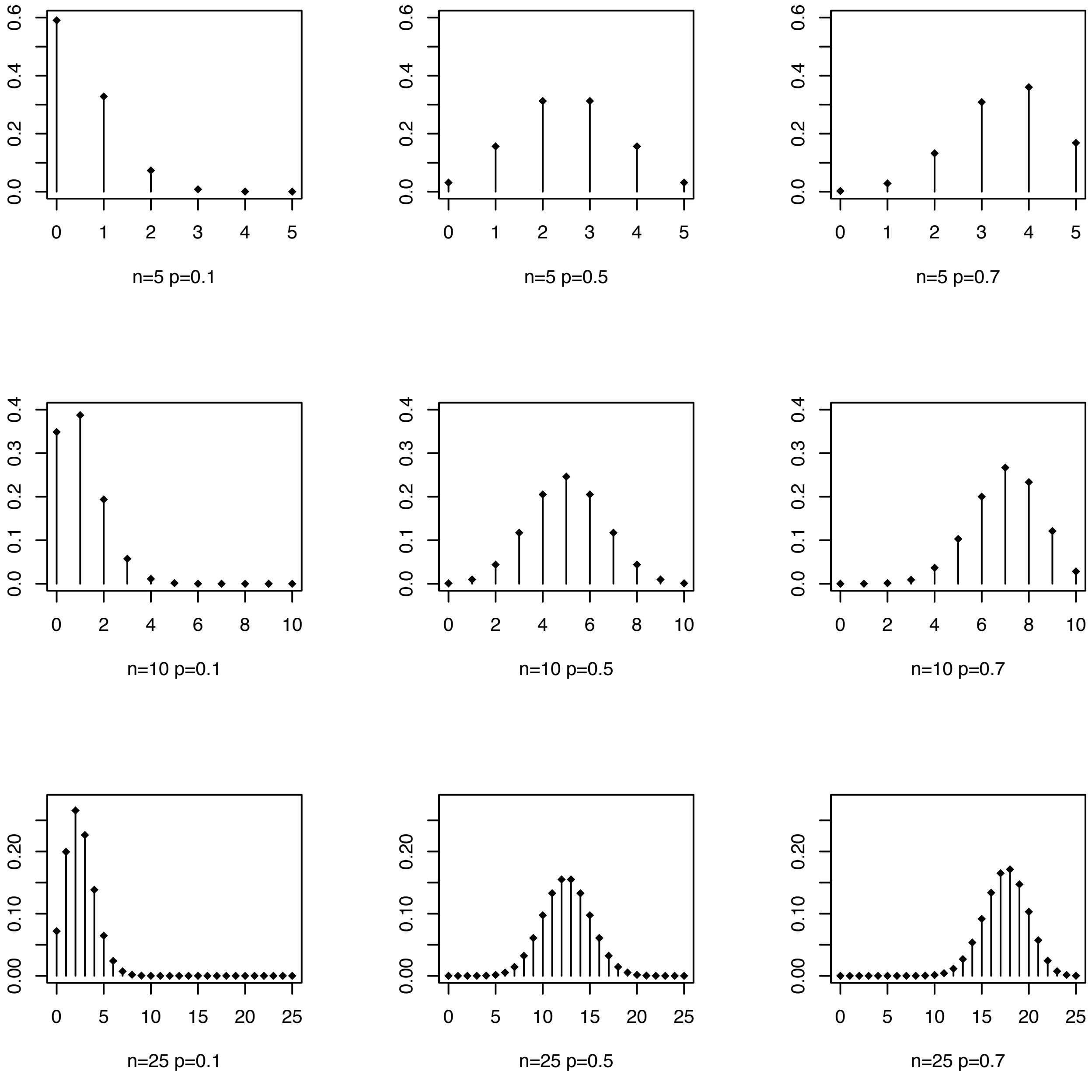
Se X è la v.a. che conta il numero di pezzi difettosi del lotto, allora $X \sim Bi(n = 20, p = 0.05)$ e, ad esempio,

$$\begin{aligned} \text{P(2 dif. su 20)} &= \text{P}(X = 2) \\ &= \binom{20}{2} 0.05^2 0.95^{18} = 0.1887 \ , \end{aligned}$$

$$\text{P(al piu' 1 dif. su 20)} = \text{P}(X \leq 1)$$
$$= \binom{20}{0} 0.05^0 0.95^{20} + \binom{20}{1} 0.05^1 0.95^{19} = 0.7358 \ .$$

Varie distribuzioni binomiali

Ecco come varia la distribuzione binomiale al variare dei suoi due parametri:



Un caso particolare è la *distribuzione di Bernoulli* che assume solo due possibili valori, 0 e 1, con probabilità $1-p$ e p. Si ottiene dalla binomiale con $n=1$.

...ancora binomiale

Il valore atteso di $X \sim Bi(n,p)$ è

$$\begin{aligned}
\mathsf{E}\,(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k} \\
&= np\sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!}p^{k}(1-p)^{n-1-k} = np \ .
\end{aligned}$$

La varianza di X si calcola tenendo conto che

$$\begin{aligned}
\mathsf{E}\left(X^2\right) &= \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n} (k^2-k)\binom{n}{k}p^k(1-p)^{n-k} + \mathsf{E}\,(X) \\
&= n(n-1)p^2 + np \ .
\end{aligned}$$

Si ha dunque

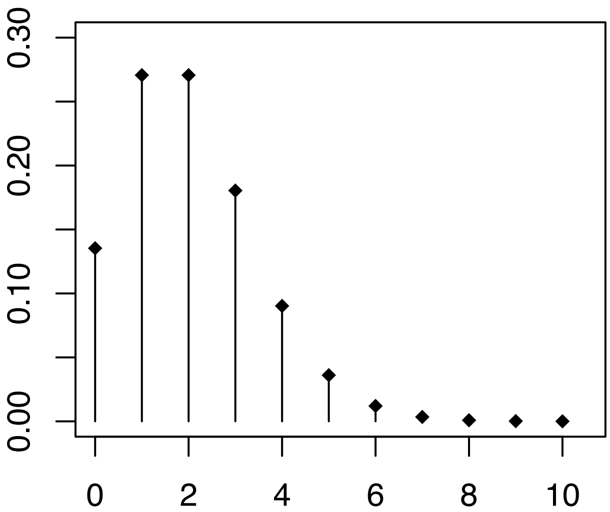
$$\mathsf{Var}\,(X) = \mathsf{E}\left(X^2\right) - [\mathsf{E}\,(X)]^2 = np(1-p) \ .$$

Distribuzione di Poisson

Una variabile X che assume valori nell'insieme dei numeri naturali N, ha *distribuzione di Poisson* di parametro $\lambda > 0$ se

$$\mathsf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots \ .$$

Scriveremo allora $X \sim \mathcal{P}(\lambda)$.



E' facile vedere che $\sum_i p_i = 1$:

$$\sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1 \ .$$

...ancora Poisson

La variabile di Poisson viene utilizzata come modello per il conteggio di manifestazioni di un certo fenomeno di interesse:
chiamate in arrivo ad un centralino in un certo intervallo di tempo;
macchine transitanti ad un casello autostradale in un certo periodo del giorno;
difetti rilevati in un pezzo di filo d'acciaio prodotto da una ditta;
terremoti manifestatisi in una data area nell'arco degli ultimi 10 anni.

Il valore atteso di $X \sim \mathcal{P}(\lambda)$ è

$$\begin{aligned} \mathsf{E}\,(X) &= \sum_{i=0}^{+\infty} i \frac{\lambda^i}{i!} e^{-\lambda} \\ &= \lambda \sum_{i=1}^{+\infty} \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda \ . \end{aligned}$$

Allo stesso modo si dimostra che

$$\mathsf{Var}\,(X) = \lambda \ .$$

Esempio

Agli sportelli di un ufficio postale arrivano clienti con una media di 10 ogni mezz'ora. Si può inoltre supporre che il numero di clienti in arrivo segua una distribuzione di Poisson. Qual è la probabilità che nella prossima mezz'ora entrino nell'ufficio non più di 3 clienti? E almeno 12?

$X =$ "n. clienti ogni mezz'ora" $\sim \mathcal{P}(\lambda = 10)$.

$$\begin{aligned}
\mathsf{P}(X \leq 3) = & \, \mathsf{P}(X = 0) + \mathsf{P}(X = 1) \\
& + \mathsf{P}(X = 2) + \mathsf{P}(X = 3) \\
= & \frac{10^0}{0!} e^{-10} + \frac{10^1}{1!} e^{-10} \\
& + \frac{10^2}{2!} e^{-10} + \frac{10^3}{3!} e^{-10} \\
= & 0.0103 \ ,
\end{aligned}$$

$$\mathsf{P}(X \geq 12) = 1 - \mathsf{P}(X \leq 11) = \dots \ .$$

Tavole per il calcolo

Utilizzando *R* (o un qualsiasi pacchetto statistico) si può oggi lavorare con le principali funzioni di probabilità, di densità e di ripartizione.

Una volta (e ancora adesso negli esami...) si usavano invece delle tabelle preconfezionate, le tavole. Nelle tavole sono registrati i valori di alcune funzioni di ripartizione relativi a diversi valori dei parametri caratterizzanti la distribuzione.

Alcune tavole sono nell'appendice del vostro libro: imparate ad usarle!

• Esempio: Se $X \sim \mathcal{P}(10)$, allora $\mathrm{P}(X \geq 12) = 1 - 0.697 = 0.303$.

Approssimazione binomiale/Poisson

Quando $n \to \infty$ e $p \to 0$ in modo tale che $np \to \lambda$ con λ costante, allora la funzione di probabilità di una v.a. binomiale di parametri n e p si può approssimare con la funzione di probabilità di una Poisson di parametro λ:

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda},$$

per $k = 0, 1, \ldots$.

L'approssimazione viene utilizzata nella pratica quando $n \geq 100$ e $p \leq 0.05$. Si sceglie allora $\lambda = np$.

Esempio

Una fabbrica di materiale elettrico fornisce il 3% delle lampadine vendute in un grande magazzino. Qual è la probabilità che su 100 lampadine scelte a caso dalle scorte del grande magazzino ve ne siano al massimo 3 provenienti da quella fabbrica?

$$\sum_{k=0}^{3} \binom{100}{k} 0.03^k 0.97^{100-k} = 0.64724921$$

è difficile da calcolare senza computer! Dato che $n = 100$ è grande e $p = 0.03$ è piccolo, si può utilizzare l'approssimazione con la Poisson di parametro $\lambda = np = 3$,

$$\sum_{k=0}^{3} \frac{3^k}{k!} e^{-3} = 0.64723189,$$

che è più facile da calcolare e molto precisa.

Distribuzione geometrica

Sia X una variabile aleatoria che conta il numero di ripetizioni indipendenti necessarie per osservare il primo successo in un esperimento che ha probabilità di successo p.

Si dice allora che X ha una *distribuzione geometrica* di parametro $p \in (0,1)$ e si scrive $X \sim Ge(p)$.

Si ha che

$$\mathsf{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \dots \ .$$

Si tratta effettivamente di una distribuzione di probabilità:

$$\begin{aligned}\sum_{k=1}^{\infty} \mathsf{P}(X = k) &= p \sum_{k=1}^{\infty} (1-p)^{k-1} \\ &= p \, \frac{1}{1-(1-p)} = 1.\end{aligned}$$

• Esercizio: Si dimostri che $\mathsf{E}\,(X) = 1/p$ e $\mathsf{Var}\,(X) = (1-p)/p^2$.

Mancanza di memoria

La distribuzione geometrica è l'unica distribuzione discreta con la proprietà di *mancanza di memoria*. Ciò significa che, se $X \sim Ge(p)$, allora

$$\begin{aligned} \mathsf{P}(X > m+n | X > m) &= \frac{\mathsf{P}(X > m+n)}{\mathsf{P}(X > m)} \\ &= \mathsf{P}(X > n) \ . \end{aligned}$$

Insomma, se sappiamo che ci vogliono più di m ripetizioni dell'esperimento per ottenere un successo, allora la probabilità che ce ne vogliano almeno altre n è uguale alla probabilità che ce ne vogliano più di n, come se si iniziasse a contare dall'$(m+1)$-esima ripetizione.

Questa proprietà si dimostra tenendo conto che

$$\mathsf{P}(X > k) = (1-p)^k \ , \quad k = 0, 1, 2, \dots \ .$$

Esempio (urna)

Dalla solita urna si estrae una pallina, se ne legge il numero e la si rimette nell'urna. Qual è la probabilità di trovare il numero 4 per la prima volta alla quinta estrazione? Se sappiamo che il 4 non è uscito fino alla terza estrazione, qual è la probabilità che si debbano fare in tutto più di 7 estrazioni?

Supponiamo che X conti il numero di estrazioni (con reinserimento) che bisogna fare per trovare per la prima volta il numero 4.
Allora $X \sim Ge(1/7)$, dove $1/7$ è la probabilità di estrarre il numero 4.

$$\mathrm{P}(X = 5) = \frac{1}{7} \times \left(\frac{6}{7}\right)^4 = 0.0771 \ ,$$

$$\mathrm{P}(X > 7 | X > 3) = \mathrm{P}(X > 4) = \left(\frac{6}{7}\right)^4 = 0.7347 \ .$$

Università Ca' Foscari di Venezia

Federica Giummolè

Variabili aleatorie continue

Probabilità e Statistica

A.A. 2013/2014

Distribuzione uniforme

Immaginiamo che X sia una variabile che può assumere un qualsiasi valore nell'intervallo (a, b), indifferentemente.

La sua densità di probabilità è allora costante nell'intervallo (a, b) e nulla al di fuori:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{altrove} \end{cases}$$

X è una variabile aleatoria *uniforme* e si scrive $X \sim U(a, b)$.

E' facile verificare che

$$\mathrm{E}\,(X) = \frac{a+b}{2} \text{ e } \mathrm{Var}\,(X) = \frac{(b-a)^2}{12} .$$

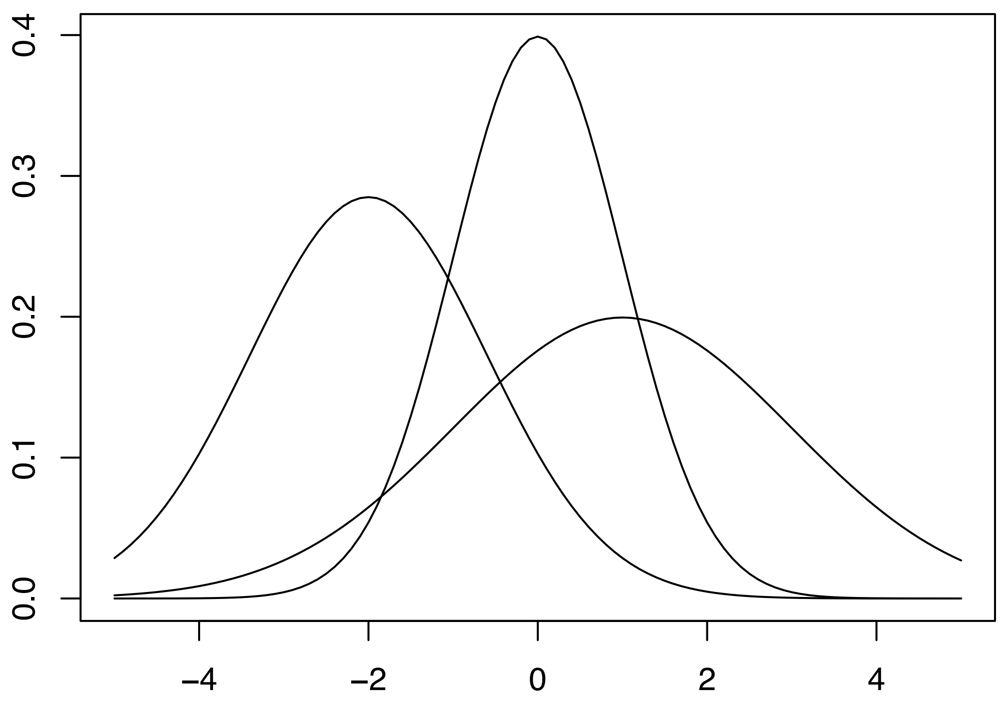
Distribuzione normale

Una variabile aleatoria X ha distribuzione *normale* se la sua funzione di densità ha la forma

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \mathrm{R} \, ,$$

con $\mu \in \mathrm{R}$ e $\sigma > 0$.

Si scrive allora che $X \sim N(\mu, \sigma^2)$.

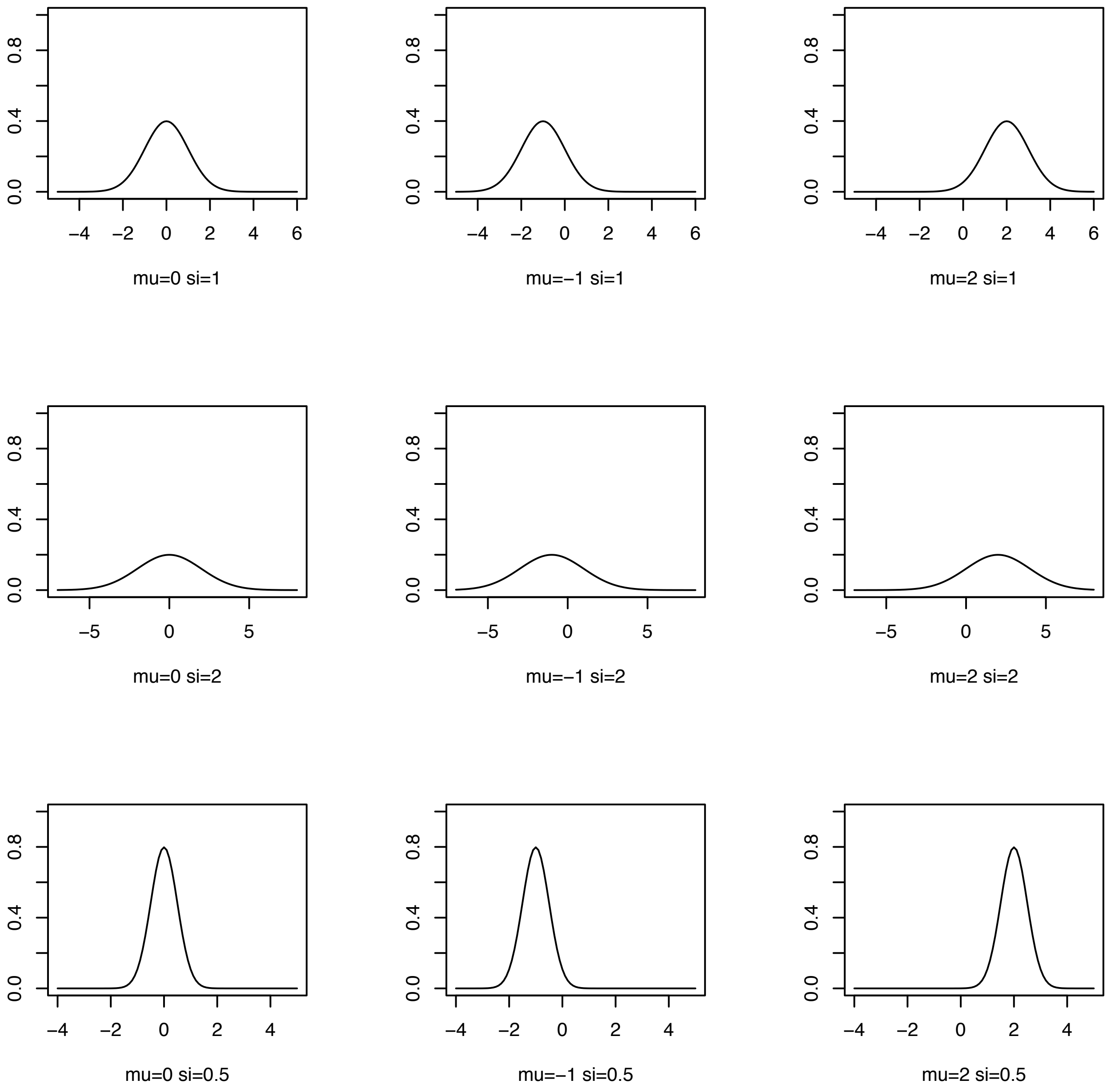


...continua

Se $X \sim N(\mu, \sigma^2)$, si può dimostrare che

$$\mathsf{E}\,(X) = \mu \text{ e } \mathsf{Var}\,(X) = \sigma^2 \ .$$

Alcuni grafici della densità gaussiana in funzione dei parametri μ e σ^2:



La standardizzazione

La funzione di ripartizione di una v.a. normale non si può calcolare in forma esplicita. Alcuni suoi valori si trovano tabulati nelle tabelle, ma soltanto per il caso $\mu = 0$ e $\sigma = 1$.

La variabile $Z \sim N(0, 1)$ viene anche chiamata *normale standard* e ad essa si riducono tutte le altre normali mediante *standardizzazione*:

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Allora,

$$\mathsf{P}(a < X < b) = \mathsf{P}(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}) \ .$$

Esempio

Una macchina produce tubi di diametro X, in mm. Il diametro X potrebbe essere modellizzato con una distribuzione normale di media μ mm e varianza σ^2 mm^2.

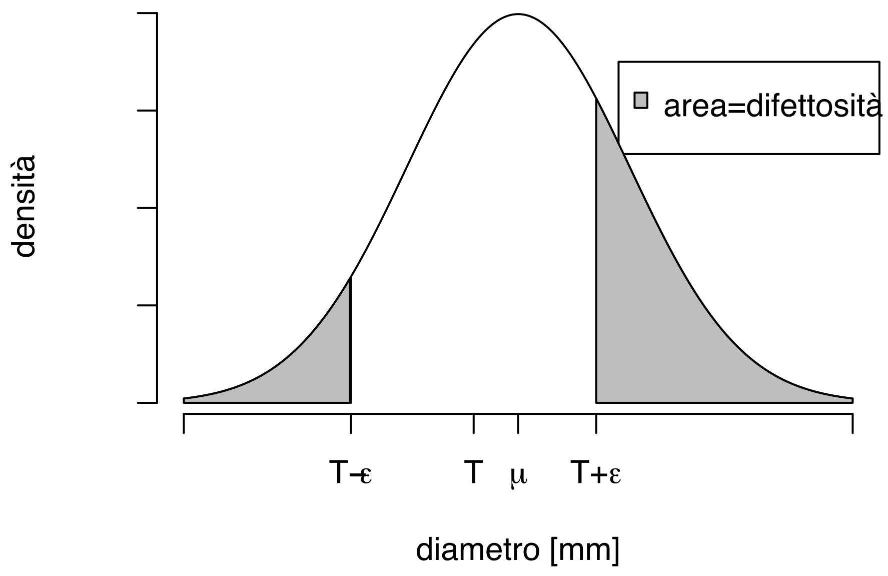
Supponiamo che per contratto i tubi debbano essere di diametro T mm più o meno un margine di specifica di ϵ mm.
Si noti che T non è necessariamente uguale a μ perché la macchina potrebbe essere centrata su valori leggermente diversi.

• Calcolare la probabilità che un generico tubo sia difettoso (cioè fuori specifica, o non conforme).
• Calcolare la probabilità che due tubi su 10 siano difettosi.

Soluzione

La prima probabilità cercata è

$$
\begin{aligned}
p &= \mathsf{P}(\text{“X è difettoso”}) \\
&= \mathsf{P}(X < T-\epsilon \cup T+\epsilon < X) \\
&= \mathsf{P}(X < T-\epsilon) + \mathsf{P}(T+\epsilon < X) \\
&= \int_{-\infty}^{T-\epsilon} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx \\
&\quad + \int_{T+\epsilon}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx.
\end{aligned}
$$



Dati μ, σ, T e ϵ siamo in grado di calcolare p (in R, per esempio, con il comando `pnorm`).

... continua

Ora, la variabile

$$S_{10} = \text{numero di tubi difettosi su 10}$$

conta il numero di successi in 10 prove indipendenti in cui ogni prova può risultare un successo (difettoso o non conforme) con probabilità

$$p = \mathsf{P}(\text{“X è difettoso”})$$

o un insuccesso (conforme) con probabilità $1 - p$.

Allora la probabilità che due tubi su 10 siano difettosi è data dalla formula della densità binomiale:

$$\mathsf{P}(S_{10} = 2) = \binom{10}{2} p^2 (1-p)^8.$$

• Esercizio: Fare i calcoli con $\mu = 3$, $\sigma = 0.9$, $T = 2.84$ e $\epsilon = 0.18$.
($p = 0.156$, $\mathsf{P}(S_{10} = 2) = 0.282$)

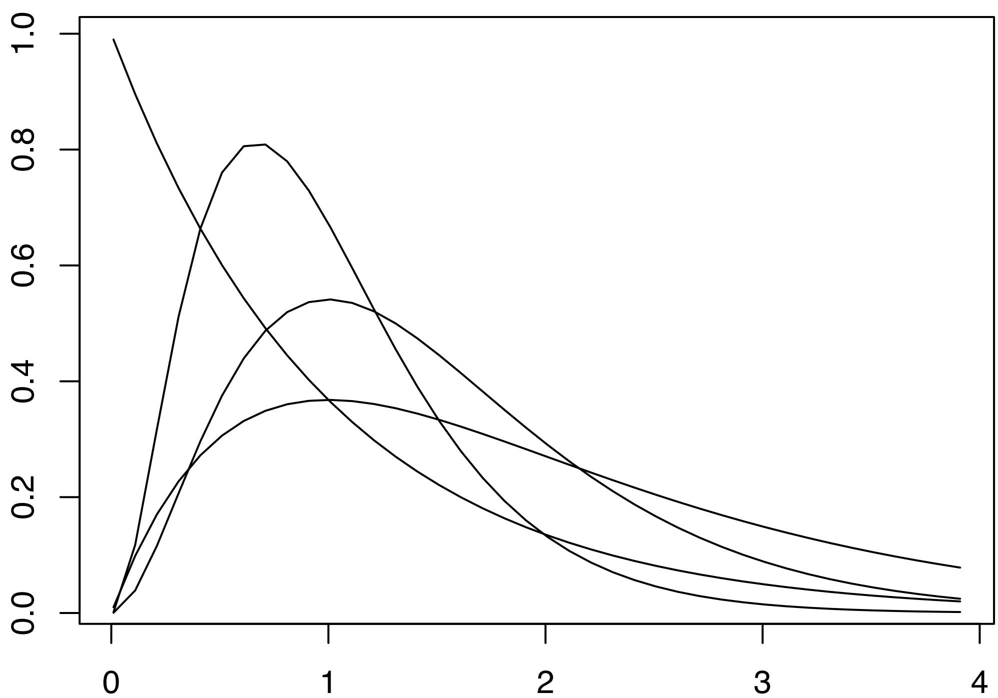
Distribuzione gamma

Si dice che X ha distribuzione *gamma* di parametri $\alpha > 0$ e $\lambda > 0$ se la sua densità di probabilità è della forma

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \in (0, \infty) \\ 0 & \text{altrove} \end{cases}$$

dove $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$.

Si scrive allora $X \sim Ga(\alpha, \lambda)$. Alcuni grafici della densità gamma al variare dei parametri:



Ancora gamma

Si ha che

$$\mathsf{E}\,(X) = \frac{\alpha}{\lambda} \quad \text{e} \quad \mathsf{Var}\,(X) = \frac{\alpha}{\lambda^2} \ .$$

• Esempio: se X è una v.a. gamma di media 4 e varianza 4, si vede facilmente che

$$\alpha = 4 \quad \text{e} \quad \lambda = 1 \ .$$

Si può allora calcolare (integrando ripetutamente per parti)

$$\begin{aligned} \mathsf{P}(X < 4) &= \int_0^4 \frac{e^{-x}x^3}{6}\,dx \\ &= \frac{1}{6}\left[-(x^3 + 3x^2 + 6x + 6)e^{-x}\right]_0^4 = \\ &= 0.567 \ . \end{aligned}$$

In questo caso il calcolo esplicito è stato possibile poiché α è un numero intero e

$$\Gamma(4) = \int_0^\infty x^{4-1}e^{-x}\,dx = (4-1)! = 6 \ .$$

Distribuzione esponenziale

Un caso particolare di distribuzione gamma con $\alpha = 1$ è la distribuzione *esponenziale*, $X \sim Exp(\lambda)$, già incontrata in vari esempi. La sua densità è

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \in (0, \infty) \\ 0 & \text{altrove} \end{cases}$$

e la sua funzione di ripartizione si può calcolare in forma esplicita:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Evidentemente,

$$\mathrm{E}\,(X) = \frac{1}{\lambda} \quad \text{e} \quad \mathrm{Var}\,(X) = \frac{1}{\lambda^2}\ .$$

La distribuzione esponenziale si usa per modellizzare tempi di attesa:
-il tempo che passa fra l'arrivo di un treno in stazione e il successivo;
-la vita di un certo componente elettronico di un'automobile;
-la lunghezza fra due difetti consecutivi di una fibra ottica.

Mancanza di memoria

La distribuzione esponenziale è l'unica distribuzione continua con la proprietà di *mancanza di memoria* (come la geometrica fra le distribuzioni discrete):

$$\mathsf{P}(X > m + n | X > m) = \frac{\mathsf{P}(X > m + n)}{\mathsf{P}(X > m)}$$
$$= \mathsf{P}(X > n) \ .$$

Insomma, se sappiamo che ci vogliono più di m ore prima che avvenga un certo evento, allora la probabilità che ce ne vogliano almeno altre n è uguale alla probabilità che ce ne vogliano più di n, come se si iniziasse a misurare il tempo dall'm-esima ora.

Questa proprietà si dimostra tenendo conto che

$$\mathsf{P}(X > x) = e^{-\lambda x} \ , \quad x > 0 \ .$$

Esempio

I clienti che si presentano agli sportelli della filiale di una banca sono una media di 30 ogni ora. Assumiamo che i tempi di attesa fra due arrivi consecutivi dei clienti abbiano distribuzione esponenziale. Siamo interessati al calcolo della probabilità che il tempo di attesa fra l'arrivo di un cliente e il successivo sia più di 5 minuti.

X=tempo di attesa in minuti $\sim Exp(1/2)$

$$\Rightarrow \mathsf{P}(X > 5) = e^{-5/2}$$

X=tempo di attesa in ore $\sim Exp(30)$

$$\Rightarrow \mathsf{P}(X > 1/12) = e^{-30/12}$$

... ancora sulla distribuzione di Poisson

La variabile di Poisson viene spesso utilizzata per i conteggi.

• Esempio: le telefonate in arrivo ad una centralina sono in media 12 ogni 10 minuti. Inoltre si può supporre che la variabile che conta le telefonate in arrivo in un intervallo di 10 minuti abbia distribuzione di Poisson:

$X_{10} =$ "numero telefonate in 10 min." $\sim \mathcal{P}(12)$.

Possiamo allora calcolare probabilità del tipo:

$$\text{P(meno di 6 telefonate fra le 11 e le 11,10)} = \\ = \text{P}(X_{10} < 6) = 0.02 \ .$$

E se fossimo interessati a intervalli di tempo di durata diversa da 10 minuti?

Il processo di Poisson

Il processo di Poisson è una successione di variabili aleatorie $\{X_t\}_{t \geq 0}$, con distribuzione di Poisson il cui parametro dipende dall'indice t:

$$X_t \sim \mathcal{P}(\alpha \times t).$$

Si usa per contare il numero di manifestazioni di un fenomeno di interesse in un qualsiasi intervallo di tempo di ampiezza t, quando il numero medio di queste manifestazioni dipende soltanto dall'ampiezza (e non dalla posizione) dell'intervallo considerato.

Possiamo allora scrivere:

$X_t =$ "n. di eventi in un intervallo di tempo t"

$\alpha =$ "n. medio di eventi nell'unità di tempo".

Esempio

Il numero di telefonate in arrivo ad una centralina segue un processo di Poisson nel tempo, con una media di 12 telefonate ogni 10 minuti.

Scegliamo il minuto come unità di misura del tempo. Abbiamo dunque:

$\alpha =$ "n. medio di telefonate in 1 min." $= 12/10$
e
$X_t =$ "n. di telefonate in t min." $\sim \mathcal{P}(12/10 \times t)$.

Possiamo allora calcolare, ad esempio,

$$\begin{aligned} &\mathsf{P}(\text{più di 20 telefonate in mezz'ora}) = \\ &\qquad = \mathsf{P}(X_{30} > 20) = 1 - \mathsf{P}(X_{30} \leq 20), \end{aligned}$$

con $X_{30} \sim \mathcal{P}(12/10 \times 30) = \mathcal{P}(36)$.

...ancora sulla distribuzione esponenziale

Associata ad ogni processo di Poisson c'è una varabile aleatoria esponenziale che misura il tempo fra due manifestazioni successive del fenomeno in questione:

$X_t =$ "n. di eventi in un intervallo di tempo t"
$\sim \mathcal{P}(\alpha \times t)$,

$T =$ "tempo trascorso fra due eventi successivi"
$\sim Exp(\alpha)$.

Per l'esempio del lucido precedente,

$$\begin{aligned} &\text{P(nessuna telefonata in 15 minuti)} = \\ &= \text{P}(X_{15} = 0) = \frac{(12/10 \times 15)^0}{0!} e^{-12/10 \times 15} = e^{-18} \\ &= \text{P}(T > 15) = e^{-12/10 \times 15} = e^{-18}, \end{aligned}$$

dove $T \sim Exp(12/10)$ misura il tempo che passa fra l'arrivo di due telefonate successive.