Corso di Statistica Matematica
Allievi Meccanici II Anno, 1. semestre, sez. M-Sc

Lorenzo Valdettaro
e-mail: lorenzo.valdettaro@polimi.it

Libri consigliati:

1. M. Bramanti: Calcolo delle probabilità e statistica per il Corso di Diploma in
2. G. Cicchitelli: Probabilità e statistica. Maggioli Editore.
3. G. Cicchitelli: Complementi ed esercizi di statistica descrittiva ed inferenziale. Maggioli Editore. Ingegneria. Ed. Esculapio
4. D.C. Montgomery, G.C. Runger e N.F. Hubele: Engineering Statistics. Ed. John Wiley & Sons (libro in inglese).
5. A.M. Mood, F.A. Graybill e D.C. Boes: Introduzione alla statistica. Ed. McGraw-Hill

Questa copia è disponibile su Internet all'indirizzo

`http://www1.mate.polimi.it/didattica/statistica_matematica/lv`

Indice

1 Introduzione **5**

2 Statistica Descrittiva **7**
2.1 Tipi di Dati . . . 7
2.2 Metodi grafici . . . 9
2.3 Indici di Posizione, Dispersione e Forma . . . 10
2.3.1 Indici di Posizione . . . 11
2.3.2 Indici di dispersione . . . 13
2.3.3 Indici di forma . . . 16
2.4 Analisi comparative, correlazione tra variabili . . . 16
2.4.1 Frequenze congiunte per dati multivariati raggruppati in classi . . . 16
2.4.2 Covarianza, coefficiente di correlazione . . . 17
2.4.3 Scatterplot, o diagramma di dispersione . . . 18
2.4.4 Regressione lineare semplice . . . 19
2.4.5 Regressione lineare multipla . . . 21

3 Calcolo delle probabilità **23**
3.1 Esperimenti aleatori, spazio campionario, eventi . . . 23
3.2 Probabilità di eventi . . . 24
3.2.1 Definizione di probabilità per uno spazio campionario discreto . . . 24
3.2.2 Come si assegnano le probabilità . . . 25
3.3 Probabilità condizionata . . . 27
3.4 Indipendenza di eventi . . . 28

4 Variabili aleatorie **29**
4.1 Definizioni . . . 29
4.2 Indici di posizione di una variabile aleatoria . . . 32
4.3 Indici di dispersione di una variabile aleatoria discreta . . . 33
4.4 Analisi comparative tra variabili aleatorie discrete . . . 34

5 Modelli discreti di variabili aleatorie **37**
5.1 Processo di Bernoulli . . . 37
5.1.1 Media e varianza del processo di Bernoulli . . . 38
5.2 Processo di Poisson . . . 39

6 Legge dei grandi numeri **41**
6.1 Media campionaria di variabili aleatorie . . . 41
6.2 Disuguaglianza di Chebyshev . . . 41
6.3 Legge debole dei grandi numeri . . . 42

7 Variabili aleatorie continue **43**
7.1 Proprietà delle variabili aleatorie continue . . . 43
7.1.1 Valore atteso . . . 44

7.1.2 Varianza 44
7.2 Modelli continui di variabili aleatorie 44
7.2.1 Densità uniforme 44
7.2.2 Densità gaussiana (o normale) 45
7.2.3 La legge esponenziale 46
7.2.4 La legge gamma 47
7.3 Quantili 48
7.4 Teorema centrale del limite 50

8 Statistica inferenziale **53**
8.1 Modello statistico parametrico 53
8.2 Stima puntuale 54
8.2.1 Stima puntuale della media 54
8.2.2 Stima puntuale della varianza 54
8.3 Stima per intervalli 56
8.4 Campionamento da una popolazione normale 56
8.4.1 Legge chi-quadrato 56
8.4.2 Legge t di Student 58
8.5 Intervalli di confidenza 60
8.5.1 Intervalli di confidenza per la media 60
8.5.2 Intervalli di confidenza per la varianza 62

9 Test d'ipotesi **65**
9.1 Definizioni 65
9.1.1 Ipotesi statistica 65
9.1.2 Verifica d'ipotesi 66
9.1.3 Regione critica 67
9.1.4 Livello di significatività 67
9.1.5 p-value 68
9.2 Verifica di ipotesi sulla media (varianza nota) 69
9.3 Test su una frequenza (grandi campioni) 72
9.4 Verifica di ipotesi sulla media (varianza incognita) 73
9.5 Verifica d'ipotesi sulla varianza 74
9.6 Test chi-quadrato di buon adattamento 75
9.7 Test chi-quadrato di indipendenza 77
9.8 Verifica d'ipotesi sulla differenza tra due medie 78

Cap. 1. Introduzione

Scopo della statistica matematica: lo studio di popolazioni.

Esempi: presentazione di risultati elettorali, proiezioni di risultati elettorali, indagini statistiche, distribuzione degli errori nella produzione di dispositivi meccanici, ecc.

I dati devono essere **raccolti**, **presentati**, **analizzati**, **interpretati**.

Due approcci fondamentali: la **statistica descrittiva** e la **statistica inferenziale**

Statistica descrittiva: si propone di

1. raccogliere e presentare i dati in forma sintetica, grafica e/o tabulare;
2. caratterizzare alcuni aspetti in modo sintetico: indici di posizione (es. valore medio), di dispersione (es. varianza), e di forma (es. simmetria);
3. studiare le relazioni tra i dati riguardanti variabili diverse.

Esempio: studio della altezza e del peso di una popolazione: grafico della distribuzione dei valori, media e varianza, relazione tra peso e altezza, ecc.

Statistica inferenziale: si cerca di far rientrare la collezione dei dati in categorie (distribuzioni) matematiche prestabilite. Si cerca di determinare le distribuzioni e i parametri che meglio si adattano ai dati: test di ipotesi e stima dei parametri.

Esempio 1: sondaggio a campione riguardo alle intenzioni di voto: quale conclusione trarre sull'insieme della popolazione?

Esempio 2: si misurano i diametri di un campione di bulloni prodotti da una linea di produzione. Quale sarà il diametro medio e la variabilità nei diametri della produzione totale? Quanti bulloni risulteranno difettosi (diametri troppo larghi o troppo stretti)?

Si introduce il concetto di casualità: sondaggi diversi danno *probabilmente* risultati diversi.

Il corso si articola in 3 parti:

1. Statistica descrittiva
2. Calcolo delle probabilità e variabili aleatorie
3. Statistica inferenziale

Cap. 2. Statistica Descrittiva

Scopo: Introdurre gli strumenti basilari per l'analisi di un certo insieme di **dati**.

1. raccogliere e di presentare i dati in forma sintetica, grafica e/o tabulare: istogrammi, diagrammi a barre, grafici di frequenza cumulativa, boxplots, scatterplots
2. di caratterizzare alcuni aspetti in modo sintetico:
   (a) indici di posizione: valore medio, mediana, moda,
   (b) indici di dispersione: varianza, deviazione standard, quantile, quartile, differenza interquartile (IQR)
   (c) indice di forma: skewness, curtosi;
3. di studiare le relazioni tra i dati riguardanti variabili diverse: covarianza, coefficiente di correlazione, regressione lineare

2.1 Tipi di Dati

Possiamo dividere i dati in due categorie principali:

1. Dati di tipo **numerico**
   (a) Variabili numeriche **discrete**, se la grandezza osservata appartiene ad un insieme numerico finito o, al piú, numerabile (ad esempio ad $\mathbb{N}$).
   (b) Variabili numeriche **continue**, se la grandezza osservata appartiene ad un insieme non numerabile come, ad esempio $\mathbb{R}$ od un suo intervallo finito o meno.
2. Dati di tipo **categorico** se non sono numerici

Persona	Età	Altezza (metri)	Peso (Kg)	Genere musicale preferito
1	34	1.755	75.838	Lirica
2	43	1.752	77.713	Classica
3	35	1.747	76.448	Classica
4	33	1.831	85.514	Rap
5	51	1.748	74.241	Nessuna
6	29	1.754	78.706	Rap
7	47	1.752	77.295	Rock
8	51	1.696	65.507	Rock
9	59	1.784	85.392	Rock
10	24	1.743	80.905	Rap
...	...	...	...	...
	Var. num. **discreta**	Var. num. **continua**	Var. num. **continua**	Var. **cat.**

I dati presentati in una tabella così come sono raccolti sono detti **dati grezzi**.
Sono difficili da analizzare soprattutto se molto numerosi.

Un primo modo di analizzare i dati è quello di produrre dei **dati raggruppati** in **classi**. *Esempio*: consideriamo i dati relativi all'età degli individui appartenenti al campione della tabella che supponiamo essere composto da 200 persone e raggruppiamoli in **classi** di età:

Cl.	Freq. Ass.	Freq. Cum.	Freq. Rel.	Freq. Rel. Cum.	Freq. Perc.	Freq. Perc. Cum.
10-14	3.	3.	0.015	0.015	1.5	1.5
15-19	7.	10.	0.035	0.05	3.5	5.
20-24	17.	27.	0.085	0.135	8.5	13.5
25-29	19.	46.	0.095	0.23	9.5	23.
30-34	28.	74.	0.14	0.37	14.	37.
35-39	19.	93.	0.095	0.465	9.5	46.5
40-44	22.	115.	0.11	0.575	11.	57.5
45-49	21.	136.	0.105	0.68	10.5	68.
50-54	20.	156.	0.1	0.78	10.	78.
55-59	16.	172.	0.08	0.86	8.	86.
60-64	8.	180.	0.04	0.9	4.	90.
65-69	11.	191.	0.055	0.955	5.5	95.5
70-74	2.	193.	0.01	0.965	1.	96.5
75-79	1.	194.	0.005	0.97	0.5	97.
80-84	6.	200.	0.03	1.	3.	100.
85-90	0.	200.	0.	1.	0.	100.

Def.: la **frequenza assoluta** $f_a(k)$ relativa alla $k-esima$ classe è il numero di osservazioni che ricadono in quella classe.

$$f_a(k) = \#\{x_i | x_i = k\} \ (i = 1, \ldots, n)$$

essendo n il numero totale delle osservazioni (200 in questo caso).

Proprietà: $\sum_{k=1}^{N_c} f_a(k) = n$.

Def.: la **frequenza relativa** $f_r(k)$ della $k-esima$ classe è il rapporto $f_a(k)/n$

Proprietà: $\sum_{k=1}^{N_c} f_r(k) = 1$

Def.: la **frequenza percentuale** $f_p(k)$ è la quantità $f_p(k) = f_r(k) * 100$.

Proprietà: $\sum_{k=1}^{N_c} f_p(k) = 100$

Def.: la **frequenza cumulativa** $F_a(k)$ della $k-esima$ classe è il numero totale delle osservazioni che ricadono nelle classi fino a k compresa:

$$F_a(k) = \sum_{j=1}^{k} f_a(j)$$

Def.: la **frequenza relativa cumulativa** è il rapporto $F_r(k) = F_a(k)/n$, ed è sempre compresa fra 0 ed 1.

Def.: la **frequenza percentuale cumulativa** $F_p(k)$ è la quantità $F_p(k) = F_r(k) * 100$.

Il raggruppamento in classi costituite da intervalli contigui vale sia per variabili numeriche *discrete* che per variabili numeriche *continue*. Nel nostro esempio possiamo definire tanti intervalli di altezze in metri (es. [1.50-1.55], [1.55-1.60], [1.60-1.65], ...). Le frequenze sono definite nello stesso modo di prima.

Per le variabili categoriche le classi sono costituite naturalmente dalle categorie.

$$f_{\text{cat. } k} = \#\{x_i | x_i = \text{cat. } k\} \ (i = 1, \ldots, n)$$

e definizioni analoghe per le altre quantità f_r e f_p. Non ha senso invece definire la frequenza cumulativa.

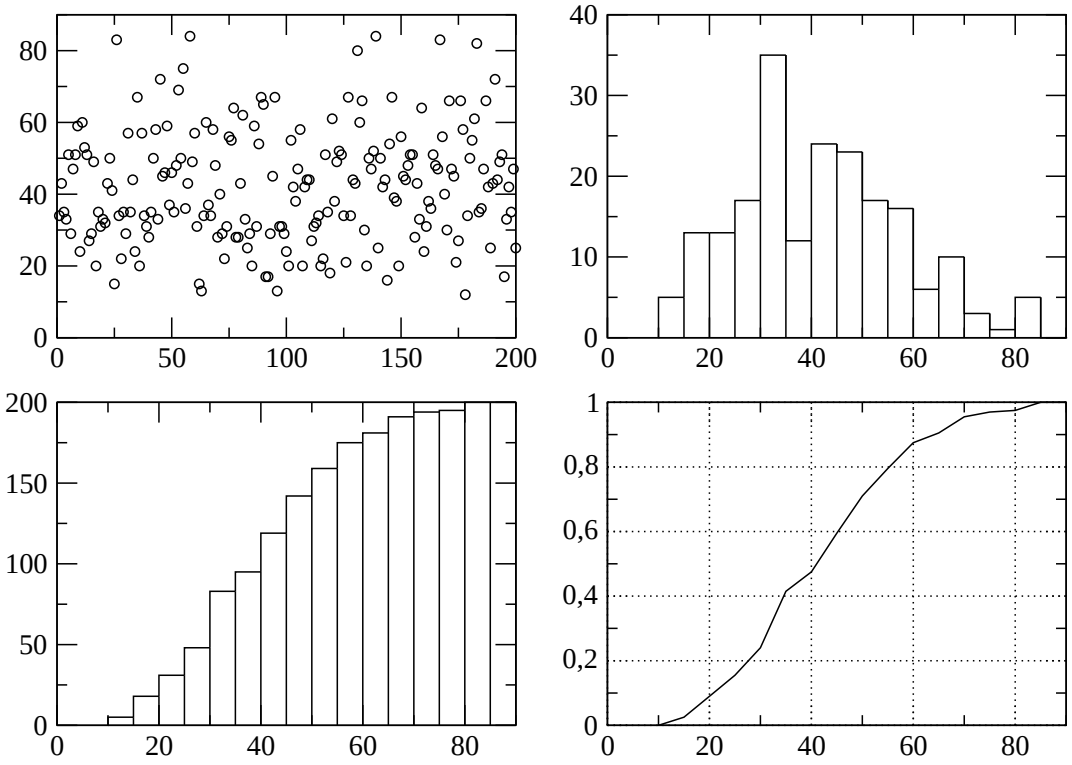
Nel nostro esempio:

$$f_{\text{Rock}} = \#\{x_i | x_i = \text{Rock}\} \ (i = 1, \ldots, n)$$

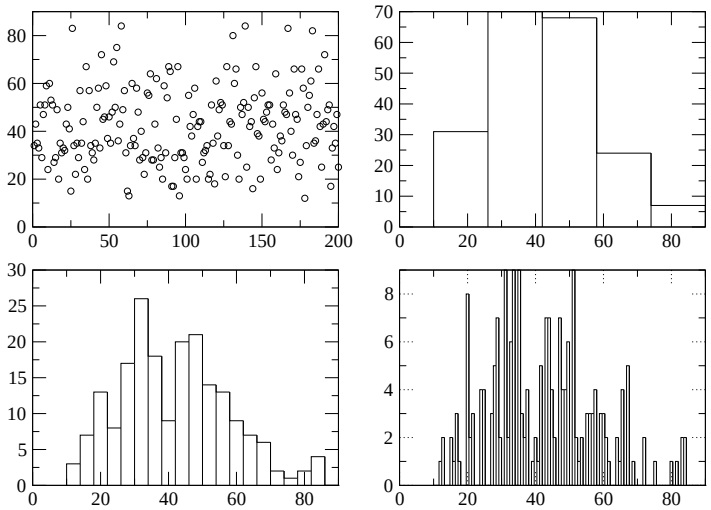
- Arbitrarietà nella scelta delle suddivisioni.
- Dalle frequenze non si può risalire alle osservazioni in quanto i dati sono stati raggruppati (*perdita di informazione*)

2.2 Metodi grafici

- **Istogramma**: grafico della distribuzione di frequenze per dati *numerici*. Le basi dei rettangoli adiacenti sono gli intervalli che definiscono le classi. Comandi Matlab *hist* e *histc*.
- **Diagramma a barre (o di Pareto)**: ad ogni classe corrisponde una barra la cui base non ha significato. Le barre non si disegnano adiacenti. Utili per rappresentare variabili di tipo *categorico*. Comandi Matlab *bar, pareto*.
- **Grafico di frequenza cumulativa**: si usa per dati *numerici*. in ascissa si riportano i valori osservati, oppure nella suddivisione in classi gli estremi degli intervalli di variabilità. In ordinata le frequenze cumulative corrispondenti. Comando Matlab *plot*.



Distribuzione delle età del campione di 200 persone: grafico dei dati grezzi, Istogramma della distribuzione delle frequenze assolute per le classi di età (comando Matlab *histc*), istogramma della distribuzione delle frequenze cumulative assolute, grafico della distribuzione delle frequenze cumulative relative (comando Matlab *plot*).



Istogrammi della distribuzione delle frequenze assolute per le classi di età. Sono state usate diverse scelte dei numeri di classi.

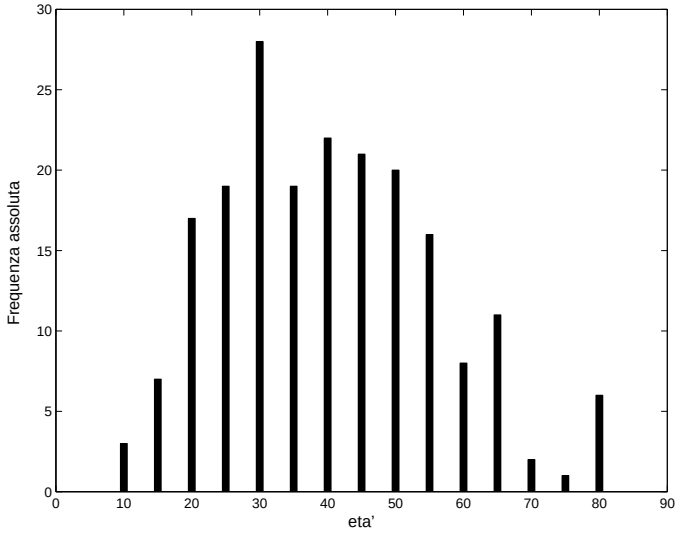


Diagramma a barre (o di Pareto) della distribuzione delle frequenze assolute per le classi di età.

2.3 Indici di Posizione, Dispersione e Forma

Si definiscono degli indici numerici che forniscono un'idea di massima di dove (indici di posizione) e come (indici di dispersione e di forma) i dati sono distribuiti.

2.3.1 Indici di Posizione

Gli indici di posizione più usati sono la **media**, la **mediana** e la **moda** associata al grafico della frequenza.

- **media** o **media campionaria** di n dati numerici $\{x_i,\ i = 1,\ldots,n\}$ (comando di Matlab *mean*):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  Esempio: supponiamo di aver misurato i seguenti 10 valori di una variabile discreta x:

$$x = [18 \quad 6 \quad 31 \quad 71 \quad 84 \quad 17 \quad 23 \quad 1 \quad 9 \quad 43]$$

  allora la media è:

$$\overline{x} = (18 + 6 + 31 + 71 + 84 + 17 + 23 + 1 + 9 + 43)/10 = 30.3$$

  Proprietà:

  - La media fornisce sempre un valore compreso fra il minimo ed il massimo valore dell'insieme dei dati. Supponiamo di avere ordinato i dati in ordine crescente: $x_1 \leq x_2 \cdots \leq x_n$. Allora:

$$nx_n - n\overline{x} = nx_n - \sum_{i=1}^{n} x_i = (n-1)x_n - \sum_{i=1}^{n-1} x_i = \sum_{i=1}^{n-1}(x_n - x_i) \geq 0$$

    e quindi $x_n - \overline{x} \geq 0$. Analogamente

$$n\overline{x} - nx_1 = \sum_{i=1}^{n} x_i - nx_1 = \sum_{i=2}^{n} x_i - (n-1)x_1 = \sum_{i=2}^{n}(x_i - x_1) \geq 0$$

    e quindi vale $x_1 \leq \overline{x} \leq x_n$

  - Media calcolata a partire dai dati raggruppati in classi.
    Dividiamo i dati in N_c classi indicando con x_{kl} il dato $l-esimo$ della $k-esima$ classe e con $f_a(k)$ la frequenza assoluta della $k-esima$ classe. Possiamo riorganizzare il calcolo della media nel seguente modo

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{k=1}^{N_c}\sum_{l=1}^{f_a(k)} x_{kl}$$

    ma

$$\sum_{l=1}^{f_a(k)} x_{kl} = f_a(k)\overline{x}_k$$

    essendo $\overline{x}_k$ la media dei dati della classe $k-esima$.
    Sostituendo si ha:

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{N_c} f_a(k)\overline{x}_k = \sum_{k=1}^{N_c} f_r(k)\overline{x}_k$$

    La media si ottiene dalle frequenze assolute o relative delle classi dei dati raggruppati se sono noti i valori medi dei dati in ciascuna classe Poichè di solito questi ultimi non sono noti, si sostituisce a ciascun $\overline{x}_k$ il valore centrale dell'intervallo associato alla classe k. In tal modo si ottiene un valore approssimato della media.

- Trasformazione lineare di dati.
  Abbiamo delle osservazioni $\{x_1, x_2, \ldots, x_n\}$ di cui abbiamo calcolato il valor medio $\overline{x}$. Ci interessa conoscere la media dei *dati trasformati linearmente* $y_i = ax_i + b$. Risulta
  $$\overline{y} = a\overline{x} + b$$
  Infatti
  $$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = b + a\frac{1}{n}\sum_{i=1}^{n} ax_i = a\overline{x} + b$$
  Esempio: $\{x_1, x_2, \ldots, x_n\}$ sono misure di temperatura in gradi Fahrenheit con valore medio $\overline{x}_F = 50$. Quale è la media in gradi centigradi?
  $$\overline{x}_C = \frac{100}{180}(\overline{x}_F - 32) = 10^o C$$
- Aggregazione di dati.
  Siano due campioni di osservazioni $\{x_1, x_2, \ldots, x_l\}$ e $y_1, y_2, \ldots, y_m\}$, di medie campionarie rispettive $\overline{x}$ e $\overline{y}$. Consideriamo il campione costituito dai dati aggregati $\{z_1, z_2, \ldots, z_n\}$= $\{x_1, x_2, \ldots, x_l, y_1, y_2, \ldots, y_m\}$, $n = l + m$. La media $\overline{z}$ di questo campione è:
  $$\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{l}{n}\overline{x} + \frac{m}{n}\overline{y}$$

- **mediana** di n dati numerici $\{x_i,\ i = 1 \ldots n\}$ (comando Matlab *median*):
  si dispongono i dati in ordine crescente (ad esempio con il comando Matlab *sort*). La mediana è il dato nella posizione centrale se n é dispari, oppure la media aritmetica dei due dati in posizione centrale, se n è pari.

  Nel nostro esempio:
  $$x = [1 \quad 6 \quad 9 \quad 17 \quad 18 \quad 23 \quad 31 \quad 43 \quad 71 \quad 84]$$
  $n = 10$ è pari e quindi la mediana è $\frac{(18+23)}{2} = 20.5$.

  Proprietà:

  - **media** e **mediana** non coincidono necessariamente; sono tanto più vicine quanto più i dati sono disposti regolarmente. Entrambi gli indici forniscono un valore più o meno centrato rispetto ai dati. La media è più facile da calcolare. La mediana è meno sensibile alla presenza di valori aberranti nei dati.
  - Mediana di dati raggruppati: si puó definire come quel valore che divide l'insieme dei dati raggruppati in due gruppi ugualmente numerosi.
    Infatti per definizione di mediana avremo che metà dei dati sarà minore (o uguale) e metà maggiore (o uguale) di essa.
    Per stimare la mediana basterá allora determinare il valore in corrispondenza del quale la frequenza cumulativa relativa o percentuale prendono il valore 0.5 o 50, rispettivamente.

- **moda** di n dati numerici raggruppati $\{x_i,\ i = 1 \ldots n\}$: punto di massimo assoluto nella distribuzione di frequenza. La moda è dunque il valore in corrispondenza del quale si ha la popolazione più numerosa. Se il valore massimo è raggiunto in più punti, allora la distribuzione delle frequenze si dice plurimodale, altrimenti è detta unimodale.

Esempio 1: campione di dati $\{1, 2, 2, 2, 4, 5, 6, 6, 8\}$.
Media campionaria: 4; mediana: 4; moda: 2

Esempio 2: campione di dati $\{1, 2, 2, 2, 4, 5, 6, 6, 53\}$.
Media campionaria: 9; mediana: 4; moda: 2

2.3.2 Indici di dispersione

Si vuole valutare come si disperdono i dati intorno alla media.
Osservazione. Definiamo lo scarto del dato $i-esimo$ come $s_i = x_i - \overline{x}$. La somma degli scarti **non** rappresenta un indice di dispersione poiché è identicamente nullo:

$$\sum_{i=1}^{n} s_i = \sum_{i=1}^{n} (x_i - \overline{x}) = 0$$

- **range** di un insieme di dati $\{x_i, i = 1, \dots\}$ non necessariamente finito:

  $$r = x_{\max} - x_{\min}$$

  dove $x_{\max}$ e $x_{\min}$ sono il valore massimo e minimo dell'insieme di dati.
  Il range fornisce un'informazione piuttosto grossolana, poiché non tiene conto della distribuzione dei dati all'interno dell'intervallo che li comprende.

- **Varianza campionaria** (comando di Matlab *var*):

  $$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

  Deviazione standard o **scarto quadratico medio** (comando di Matlab *std*). è la radice quadrata della varianza:

  $$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

  Esempio: i tempi per il taglio di una lastra in sei parti sono (espressi in minuti): $\{0.6, 1.2, 0.9, 1.0, 0.6, 0.8\}$. Calcoliamo s.

  $$\overline{x} = \frac{0.6 + 1.2 + 0.9 + 1.0 + 0.6 + 0.8}{6} = 0.85 \text{ (minuti)}$$

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
0.6	-0.25	0.0625
1.2	0.35	0.1225
0.9	0.05	0.0025
1.0	0.15	0.0225
0.6	-0.25	0.0625
0.8	-0.05	0.0025

  $$s = \sqrt{\frac{0.0625 + 0.1225 + 0.0025 + 0.0225 + 0.0625 + 0.0025}{5}}$$

  $$= \sqrt{\frac{0.2750}{5}} \approx 0.23 \text{(minuti)}$$

  Proprietà della varianza:

  – In alcuni testi si trova la definizione seguente di varianza:

  $$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{n-1}{n} s^2$$

  e analogamente per la deviazione standard. La definizione data sopra è da preferirsi a questa, per ragioni che verranno esposte nel capitolo di statistica inferenziale, quando parleremo di stimatori corretti. La maggior parte dei pacchetti software di analisi statistica usa la prima definizione. Per n grande la differenza è trascurabile.

- Modo alternativo di calcolare la varianza:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2$$

Con l'altra definizione di varianza si ottiene:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

In questo caso la varianza è pari alla differenza fra la media dei quadrati e il quadrato della media.
Varianza calcolata in base ai dati raggruppati:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\sum_{k=1}^{N_c}\sum_{l=1}^{f_a(k)}(x_{lk} - \overline{x})^2$$

sostituiamo x_{lk} con $\overline{x}_k$

$$s^2 \approx \frac{1}{n-1}\sum_{k=1}^{N_c}\sum_{l=1}^{f_a(k)}(\overline{x}_k - \overline{x})^2 = \frac{1}{n-1}\sum_{k=1}^{N_c} f_a(k)(\overline{x}_k - \overline{x})^2$$

- Trasformazione lineare di dati. Abbiamo delle osservazioni $\{x_1, x_2, \ldots, x_n\}$ di cui abbiamo calcolato la varianza s_x^2. Ci interessa conoscere la varianza dei *dati trasformati linearmente* $y_i = ax_i + b$. Risulta

$$s_y^2 = a^2 s_x^2$$

Infatti

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(ax_i - a\overline{x})^2$$

$$= a^2\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = a^2 s_x^2$$

- Dato il campione $\{x_1, \ldots, x_n\}$, definiamo la **variabile standardizzata**

$$y = \frac{x - \overline{x}}{s_x}$$

Il campione corrispondente $y_i = (x_i - \overline{x})/s_x$ ha media nulla e varianza 1. Infatti

$$\overline{y} = \frac{1}{s_x}(\overline{x} - \overline{x}) = 0$$

$$s_y^2 = \frac{1}{s_x^2}s_x^2 = 1$$

- **Quantili, percentili & C.**
sia $\{x_i, i = 1 \ldots n\}$ un campione di dati numerici *ordinato*.

  - Definizione di **p-esimo quantile** q_p $(0 < p < 1)$:
  se np non è intero, sia k l'intero tale che $k < np < k+1$: $q_p = x_{k+1}$.
  se $np = k$ con k intero, allora $q_p = (x_k + x_{k+1})/2$.
  - Il *p-esimo quantile* viene anche detto **100p-esimo percentile**.

- Il $p-esimo$ quantile o $100p-esimo$ percentile forniscono un valore che risulta maggiore o uguale del 100p% dei dati del campione.
- Il 25^o, 50^o e 75^o percentile vengono detti anche primo, secondo e tezo **quartile**, e indicati con Q_1, Q_2, Q_3. Q_1, Q_2, Q_3 sono tre numeri che dividono l'insieme di osservazioni in 4 gruppi contenenti ciascuno circa un quarto dei dati.
- Il secondo quartile Q_2 coincide con la mediana.
- Anche se é piu corretto annoverare i quantili tra gli indici di posizione, é possibile ricavare un indice di dispersione, chiamato **differenza interquartile**, o **IQR** (dall'inglese interquartile range) definito come la distanza fra il primo ed il terzo quartile:

$$IQR = Q_3 - Q_1$$

Esempio: Altezze di 20 persone di sesso maschile. Campione ordinato in modo crescente espresso in metri:
{ 1.58, 1.60, 1.66, 1.68, 1.70, 1.74, 1.74, 1.75, 1.75, 1.76, 1.78, 1.78, 1.78, 1.79, 1.80, 1.81, 1.82, 1.84, 1.88, 1.91}.

$$\overline{x} = 1.7575m, \quad s^2 = 0.0069m^2, \quad s = 8.3cm$$

Calcolo di $q_{0.5}$: $np = 20 * 0.5 = 10$ è intero. Pertanto:

$$q_{0.50} = \frac{x_{10} + x_{11}}{2} = \frac{1.76 + 1.78}{2} = 1.77m$$

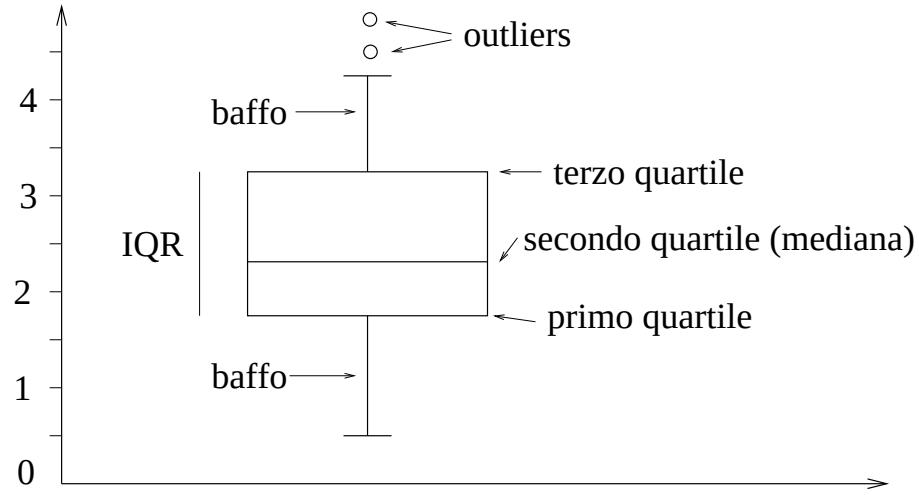
Analogamente:

$$q_{0.25} = \frac{x_5 + x_6}{2} = \frac{1.70 + 1.74}{2} = 1.72m$$

$$q_{0.75} = \frac{x_{15} + x_{16}}{2} = \frac{1.80 + 1.81}{2} = 1.805m$$

$IQR = q_{0.75} - q_{0.25} = 1.805 - 1.72 = 8.5cm$. range=$x_{20} - x_1 = 1.91 - 1.58 = 33cm$.

- Alcune informazioni contenute nella distribuzione di frequenza (e in particolare nei quartili) possono essere visualizzate graficamente con un **boxplot**.



Gli *outliers* sono dati che *giacciono fuori dai limiti*, la cui correttezza andrebbe accertata. Possono essere definiti in vari modi.
Ad esempio come quei dati che stanno sotto il 5^o percentile o sopra il 95^o percentile.
Altra definizione (usata da Matlab): si calcola un limite superiore $U = Q_3 + 1.5IQR$; il baffo superiore viene prolungato fino all'ultima osservazione che risulta minore o uguale di U. Gli outliers sono i dati che eccedono questo valore. Si segue analoga procedura per gli outliers inferiori.

2.3.3 Indici di forma

- La **skewness**
$$\gamma_3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma}\right)^3$$
È una grandezza *adimensionale*. Può assumere valori sia positivi che negativi.
Se è negativa denota una *coda* verso sinistra.
Se è positiva denota una *coda* verso destra.
se la distribuzione è simmetrica, allora la skewness è nulla, ma l'inverso non è vero.
Per trasformazioni lineari $y_i = ax_i + b$ la skewness non cambia: $\gamma_3^y = \gamma_3^x$.

- La **curtosi**
$$\gamma_4 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma}\right)^4$$
È una grandezza adimensionale e non negativa. Misura (in un certo senso) l'appiattimento della distribuzione delle frequenze, poiché assegna un peso elevato agli scarti grandi: valori elevati della curtosi segnalano distribuzioni significativamente diverse da 0 per grandi scarti, piccoli valori distribuzioni *appuntite* in corrispondenza di $\overline{x}$.
Per trasformazioni lineari $y_i = ax_i + b$ la curtosi non cambia: $\gamma_4^y = \gamma_4^x$.

- Momento centrato di ordine k:
$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$$
Se k è pari, allora $m_k > 0$: indice di dispersione ($m_2 = \sigma^2$) e di forma.
Se k è dispari, allora m_k può essere negativo: indice di simmetria.

2.4 Analisi comparative, correlazione tra variabili

Si effettuano osservazioni simultanee di più variabili su una medesima popolazione (ad esempio peso e altezza in un campione di persone). I dati in questo caso si dicono *multivariati*. Ci si domanda se esistono dei legami (associazione, dipendenza, correlazione) tra le variabili considerate.

2.4.1 Frequenze congiunte per dati multivariati raggruppati in classi

Come nel caso dei dati univariati è spesso utile raggruppare i dati in classi. Definiamo di seguito le frequenze congiunte per dati bivariati.

Def.: la **frequenza assoluta congiunta** $f_a(i,j)$ relativa alla $i-esima$ classe della prima variabile e alla $j-esima$ classe della seconda variabile è il numero delle osservazioni che ricadono in quelle classi.

$$f_a(i,j) = \#\{(x_k, y_k) | x_k = i, y_k = j\} \ (k = 1, \ldots, n)$$

essendo n il numero totale delle osservazioni.
Esempio: il numero di persone per le quali l'altezza è compresa tra 1.65 e 1.70 metri, e il peso è compreso tra 75 e 80 Kg.
Proprietà: $\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_a(i,j) = n$, dove N_1 è il numero di classi in cui è suddivisa la prima variabile e N_2 è il numero di classi in cui è suddivisa la seconda variabile.

Def.: la **frequenza relativa congiunta** $f_r(i,j)$ è il rapporto $f_a(i,j)/n$.
Proprietà: $\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_r(i,j) = 1$.

Def.: la **frequenza cumulativa congiunta** $F_a(i,j)$ è il numero totale delle osservazioni che ricadono fino a i compreso per la prima variabile e fino a j compreso per la seconda variabile:

$$F_a(i,j) = \sum_{k=1}^{i}\sum_{l=1}^{j} f_a(k,l)$$

Def.: la **frequenza marginale assoluta** $f_{ax}(i)$ relativa alla prima variabile x è:

$$f_{ax}(i) = \sum_{j=1}^{N_2} f_a(i,j) = \#\{x_k | x_k = i\} \ (k = 1,\dots,n)$$

f_{ax} è dunque la frequenza assoluta dei dati della prima variabile. Analogamente si definisce la frequenza marginale assoluta relativa alla seconda variabile y:

$$f_{ay}(j) = \sum_{i=1}^{N_1} f_a(i,j)$$

Nota: Dalle frequenze marginali si può ricavare la frequenza congiunta solo in casi molto specifici, ossia se le due variabili sono *indipendenti*, come vedremo più avanti nel corso.

Analogamente si definisce la **frequenza marginale relativa**: $f_{rx}(i) = f_{ax}(i)/n$, $f_{ry}(i) = f_{ay}(i)/n$

La media e la varianza si definiscono in modo naturale nel seguente modo:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \approx \frac{1}{n}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_a(i,j)\overline{x}_i = \frac{1}{n}\sum_{i=1}^{N_1} f_{ax}(i)\overline{x}_i = \sum_{i=1}^{N_1} f_{rx}(i)\overline{x}_i$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \approx \frac{1}{n-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_a(i,j)(\overline{x}_i-\overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{N_1} f_{ax}(i)(\overline{x}_i-\overline{x})^2$$

2.4.2 Covarianza, coefficiente di correlazione

Def. la **covarianza campionaria** delle variabili x e y è il numero

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \frac{1}{n-1}\sum_{i=1}^{n} x_i y_i - \frac{n}{n-1}\overline{x}\,\overline{y}$$

- Come per la varianza, anche nel caso della covarianza si trova in alcuni testi la definizione con n al denominatore al posto di $n-1$.

- Vale la proprietà: $s_{xy} = s_{yx}$.
- Vale la proprietà:

  $$s_{x+y}^2 = s_x^2 + s_y^2 + 2s_{xy}$$

  Dim.: mostriamo innanzitutto che $\overline{x+y} = \overline{x} + \overline{y}$:

  $$\overline{x+y} = \frac{1}{n}\sum_{i=1}^{n}(x_i+y_i) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{x} + \overline{y}$$

  Dunque:

  $$s_{x+y}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i+y_i-\overline{x}-\overline{y})^2 =$$

  $$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 + (y_i-\overline{y})^2 + 2(x_i-\overline{x})(y_i-\overline{y}) = s_x^2 + s_y^2 + 2s_{xy}$$

- Vale la proprietà:

$$s_{\alpha x,\beta y} = \alpha\beta s_{xy}$$

Infatti:

$$s_{\alpha x,\beta y} = \frac{1}{n-1}\sum_{i=1}^{n}(\alpha x_i - \alpha\overline{x})(\beta y_i - \beta\overline{y}) =$$

$$= \alpha\beta\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \alpha\beta s_{xy}$$

- Covarianza calcolata in base ai dati raggruppati:

$$s_{xy} \approx \frac{1}{n-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_a(i,j)(\overline{x}_i - \overline{x})(\overline{y}_j - \overline{y})$$

Def. il **coefficiente di correlazione campionario** di x e y è il numero

$$\rho_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{j=1}^{n}(x_j - \overline{x})^2 \sum_{k=1}^{n}(y_k - \overline{y})^2}}$$

- Il coefficiente di correlazione ha lo stesso segno della covarianza.

- Le variabili x e y si dicono
direttamente correlate se $s_{xy} > 0$ (e dunque se $\rho_{xy} \geq 0$),
inversamente correlate se $s_{xy} < 0$,
non correlate se $s_{xy} = 0$.

- Vale la proprietà:

$$-1 \leq \rho_{xy} \leq 1$$

. *Dim.*:

$$0 \leq s^2_{\frac{x}{s_x}+\frac{y}{s_y}} = s^2_{\frac{x}{s_x}} + s^2_{\frac{y}{s_y}} + 2s_{\frac{x}{s_x},\frac{y}{s_y}} = 2 + 2\frac{s_{xy}}{s_x s_y}$$

Ne deduciamo $\rho_{xy} \geq -1$. Ragionando in modo analogo su $s^2_{\frac{x}{s_x}-\frac{y}{s_y}}$ deduciamo che $\rho_{xy} \leq 1$.

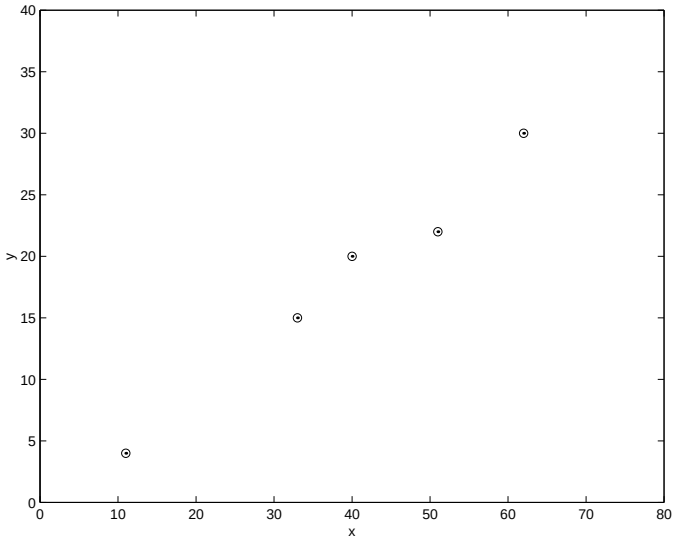
- $\rho_{xy} = \pm 1$ se e solo se esistono a e b t.c. $y_i = ax_i + b$. ρ_{xy} ha lo stesso segno di a.

- ρ_{xy} è invariante per trasformazione lineare: se $x'_i = ax_i + b$, $y'_i = cy_i + d$, si ha $\rho_{x'y'} = \rho_{xy}$

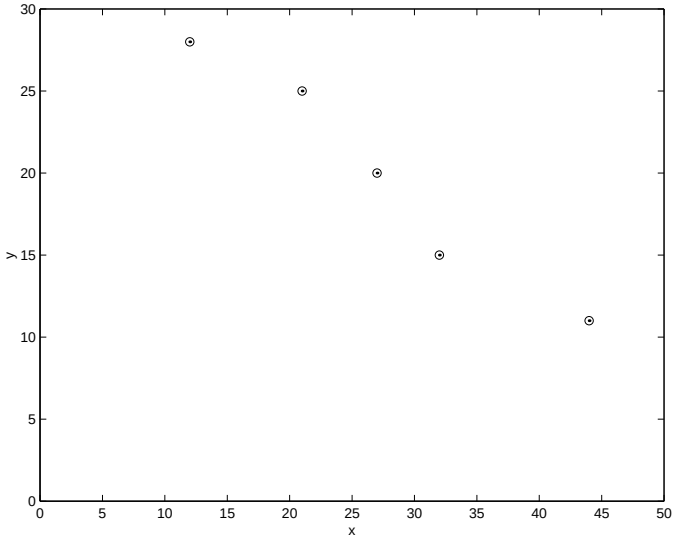
2.4.3 Scatterplot, o diagramma di dispersione

Lo **scatterplot** (comando di Matlab *plot*) è un metodo grafico utile per vedere se esistono delle correlazione tra due variabili. Si mette in ascissa una variabile, in ordinata un'altra, e si rappresentano le singole osservazioni con dei punti.
Se punti con ascissa piccola hanno ordinata piccola, e punti con ascissa grande hanno ordinata grande, allora esiste una correlazione diretta tra le due variabili ($\rho_{xy} > 0$).
Viceversa quando al crescere dell'una l'altra decresce si ha correlazione inversa ($\rho_{xy} > 0$).
Se i punti formano una nuvola indistinta i dati sono scorrelati.

Esempio 1: x_i, y_i = (11, 4), (51, 22), (40, 20), (62, 30), (33, 15).



I dati sono fortemente correlati. Infatti $\rho_{xy} = 0.9913$.
Esempio 2: $x_i, y_i = (21, 25), (12, 28), (32, 15), (44, 11), (27, 20)$.



I dati sono inversamente correlati. Infatti $\rho_{xy} = -0.9796$.

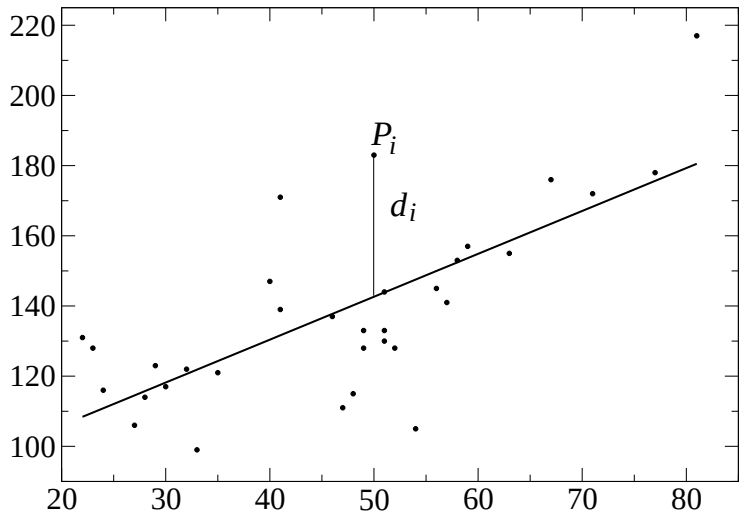
2.4.4 Regressione lineare semplice

Ricerca di una relazione lineare tra le variabili x e y. Stiamo supponendo di avere

$$y_i = ax_i + b + r_i \qquad (1)$$

dove r_i è un residuo che vogliamo quanto più piccolo possibile.

- Chiameremo x_i *predittore* e y_i *responso*.
- La retta che cerchiamo si chiama **retta di regressione semplice** (si dice semplice perché coinvolge un solo predittore), o anche **retta dei minimi quadrati**.
- Alla forma (1) si può essere arrivati dopo eventuali trasformazioni dei dati.



La retta di regressione è quella che rende minima la somma dei quadrati delle lunghezze d_i dei segmenti verticali congiungenti i punti osservati con la retta stessa.

Per stimare al meglio i coefficienti a e b utilizziamo il **principio dei minimi quadrati**: minimizziamo la quantità

$$f(a,b) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$$

Dal calcolo differenziale sappiamo che dobbiamo imporre:

$$\frac{\partial f(a,b)}{\partial a} = -\sum_{i=1}^{n} 2x_i [y_i - (ax_i + b)] = 0$$

$$\frac{\partial f(a,b)}{\partial b} = -\sum_{i=1}^{n} 2 [y_i - (ax_i + b)] = 0$$

Otteniamo:

$$a = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{s_{xy}}{s_x^2}$$

$$b = \overline{y} - a\overline{x} = \overline{y} - \frac{s_{xy}}{s_x^2}\overline{x}$$

Nota: $\sum_{i=1}^{n} r_i = 0$.

Per stimare la qualità di una regressione possiamo utilizzare i seguenti criteri:

- Il coefficiente di correlazione ρ_{xy} deve essere vicino a 1 o a -1.
- L'esame visivo dello scatterplot delle due variabili: i dati devono essere vicini alla retta di regressione.

- L'esame del grafico dei residui: in ascissa i valori previsti, in ordinata i valori dei residui. La nuvola dei punti deve avere un aspetto omogeneo, senza segni di curvatura, allargamenti o restringimenti.
  - Un grafico dei residui che presenti curvatura è un indizio che una dipendenza lineare non spiega bene i dati. Si può tentare di correggere questo difetto con trasformazioni di x e/o y, oppure si può provare a passare a una regressione multipla (che definiremo più avanti).
  - Un allargarsi/restringersi della nuvola di punti è un indizio che gli errori non sono tutti dello stesso tipo al variare di i. Si scelga quella combinazione di trasformazioni che danno la nuvola dei residui più omogenea possibile.

2.4.5 Regressione lineare multipla

Il responso y è *spiegato* da più predittori $(x^{(1)}, x^{(2)}, \ldots, x^{(d)})$. Ipotizziamo il modello teorico

$$y_i = a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)} + r_i \qquad (1)$$

I coefficienti $a_0, a_1, \ldots, a_d$ sono stimati usando il principio dei minimi quadrati: si rende minima la quantità

$$f(a_0, a_1, \ldots, a_d) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)}) \right]^2$$

Dobbiamo imporre:

$$\frac{\partial f(a_0, a_1, \ldots, a_d)}{\partial a_0} = -\sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)}) \right] = 0$$

$$\frac{\partial f}{\partial a_k} = -\sum_{i=1}^{n} 2 x_i^{(k)} \left[y_i - (a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)}) \right] = 0, \quad k = 1, \ldots, d$$

Questo sistema lineare di $d+1$ equazioni in $d+1$ incognite ammette una soluzione unica (supponendo che il determinante sia non nullo).

È comodo riscrivere il sistema di equazioni in forma matriciale: posto

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(d)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(d)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \ldots & x_n^{(d)} \end{bmatrix} \quad r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix} \qquad (2.1)$$

Il sistema da risolvere può essere scritto in forma matriciale come:

$$X^T X a = X^T y$$

La soluzione è

$$a = (X^T X)^{-1} X^T y$$

L'equazione $y_i = a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)}$ è l'equazione di un iperpiano. Esso rappresenta quell'iperpiano che rende minima la somma dei quadrati delle lunghezze d_i dei segmenti congiungenti i punti osservati all'iperpiano stesso

- Come per la regressione lineare semplice possiamo essere arrivati al modello lineare (1) dopo aver fatto trasformazioni sul responso e/o sui predittori.

- Tra i predittori possiamo inserire anche potenze e prodotti dei predittori fondamentali.
- Se i predittori sono tutti potenze di un unico predittore fondamentale, si parla di **regressione polinomiale**.
- Il grafico dei residui, ossia lo scatterplot dei punti $(a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)}, r_i)$, è anche in questo caso uno strumento di analisi grafica per controllare la bontà della regressione. Valgono le considerazioni già fatte nel caso della regressione semplice.
- Definiamo
  la *devianza totale* $DT = \sum_{i=1}^{n} (y_i - \overline{y})^2$,
  la *devianza spiegata* $DS = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$ (dove $\hat{y}_i$ sono i valori previsti $\hat{y}_i = a_0 + \sum_{k=1}^{d} a_k x_i^{(k)}$),
  la *devianza dei residui* $DR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.
  Vale la proprietà $DT = DS + DR$.
  Il coefficiente di *determinazione multipla* R^2 definito da

$$R^2 = \frac{DS}{DT} = 1 - \frac{DR}{DT}$$

  è sempre compreso tra 0 e 1 ed è un indice della frazione della variabilità di y spiegata dal modello. R^2 vicino a 1 è un buon indizio.

Modelli con retta di regressione per l'origine: Si ipotizza che il responso deve essere nullo quando i predittori sono nulli. In altre parole il coefficiente a_0 viene posto uguale a 0: $y_i = a_1 x_i^{(1)} + a_2 x_i^{(2)} + \cdots + a_d x_i^{(d)} + r_i$. Si procede come prima col principio dei minimi quadrati, ma si ottengono d equazioni nelle d incognite $a_1, \ldots, a_d$. La soluzione cambia.
Nota: non è piú vero che $DT = DS + DR$.

Cap. 3. Calcolo delle probabilità

Scopo: si vogliono ricavare dei modelli matematici per esperimenti aleatori.

3.1 Esperimenti aleatori, spazio campionario, eventi

Un **esperimento aleatorio** è un esperimento che a priori può avere diversi esiti possibili, e il cui esito effettivo dipende dal caso.

Esempi

1. (a) Si estraggono sei palline da un campione di 90 palline numerate progressivamente, e si guardano i numeri estratti.
   (b) Si entra in una classe di studenti e si conta il numero di assenti.
2. (a) Si lancia ripetutamente una moneta finché non esce testa; si conta il numero di lanci.
   (b) Si telefona ogni minuto a un determinato numero finché non lo si trova libero. Si conta il numero di tentativi.
3. (a) Si accende una lampadina e si misura il suo tempo di vita.
   (b) Si misura l'altezza di un individuo scelto a caso in un gruppo di persone.

Def. Chiamiamo **evento elementare** ω un possibile esito di un esperimento aleatorio.
Def. Lo **spazio campionario** Ω è l'insieme costituito da tutti gli eventi elementari.
Negli esempi 1a e 1b gli eventi elementari sono in numero finito.
Negli esempi 2a e 2b sono un'infinità numerabile ($\Omega = \mathbb{N}$)
Negli esempi 3a e 3b sono un'infinità non numerabile ($\Omega = \mathbb{R}$ o un intervallo di $\mathbb{R}$).

Lo spazio campionario è detto **discreto** se i suoi elementi sono in numero finito oppure un'infinità numerabile (es. 1 e 2). È detto **continuo** se è più numeroso, ad esempio $\mathbb{R}$ o un suo intervallo (es. 3).

Def. chiamiamo **evento** un qualsiasi sottoinsieme di Ω. La totalità degli eventi possibili è rappresentata dall' *insieme delle parti* di Ω.
Esempi di eventi.
1a: le palline estratte hanno numeri progressivi contigui.
1b: non vi sono più di 3 assenti.
2a: si ottiene testa dopo non meno di 10 lanci e non più di 20.
2b: non si aspetta più di 10 minuti.
3a: la lampadina dura almeno 300 ore.
3b: la persona misura meno di 1.80 metri.

Rappresentazione insiemistica degli eventi.

Linguaggio degli insiemi	*Linguaggio degli eventi*
Ω	evento certo
$\varnothing$	evento impossibile
insieme A	si verifica l'evento A
insieme $\overline{A}$	non si verifica A
$A \cup B$	si verificano A e/o B
$A \cap B$	si verificano sia A che B
$A \backslash B$	si verifica A e non si verifica B
$A \cap B = \varnothing$	A e B sono incompatibili
$B \subseteq A$	B implica A

Alcune proprietà degli insiemi; A, B e C sono sottoinsiemi qualsiasi di Ω:

$A \cup A = A$	idempotenza dell'unione
$A \cap A = A$	idempotenza dell'intersezione
$A \cup \varnothing = A$	
$A \cap \varnothing = \varnothing$	
$A \cup \Omega = \Omega$	
$A \cap \Omega = A$	
$A \cup \overline{A} = \Omega$	
$A \cap \overline{A} = \varnothing$	
$A \cup B = B \cup A$	commutatività dell'unione
$A \cap B = B \cap A$	commutatività dell'intersezione
$A \cup (B \cup C) = (A \cup B) \cup C$	associatività dell'unione
$A \cap (B \cap C) = (A \cap B) \cap C$	associatività dell'intersezione
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	distributività dell'unione risp. intersez.
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	distributività dell'intersez. risp. unione
$\overline{A \cup B} = \overline{A} \cap \overline{B}$	legge di De Morgan
$\overline{A \cap B} = \overline{A} \cup \overline{B}$	legge di De Morgan
$\overline{\overline{A}} = \mathrm{A}$	

3.2 Probabilità di eventi

In relazione a un evento siamo interessati a calcolarne la probabilità. Nei nostri esempi:
1a: prob. che le palline estratte abbiano numeri progressivi contigui.
1b: prob. che non vi siano più di 3 assenti.
2a: prob. che si ottenga testa dopo non meno di 10 lanci e non più di 20.
2b: prob. di aspettare non più di 10 minuti.
3a: prob. che la lampadina duri almeno 300 ore.
3b: prob. che la persona misuri meno di 1.80 metri.

3.2.1 Definizione di probabilità per uno spazio campionario discreto

Chiamiamo **probabilità** una qualsiasi funzione P con le proprietà seguenti:

- $P : \mathcal{P} \to [0,1]$, dove $\mathcal{P}$ è una famiglia di sottoinsiemi di Ω con le proprietà che $\Omega \in \mathcal{P}$, che se $A \in \mathcal{P}$ allora $\overline{A} \in \mathcal{P}$, e che se $A, B \in \mathcal{P}$, allora $A \cup B \in \mathcal{P}$. Ad esempio $\mathcal{P}$ potrebbe essere l'insieme delle parti di Ω
- $P(\Omega) = 1$
- condizione di *numerabile additività*: se $A_1, A_2, \ldots$ sono eventi disgiunti ($A_i \cap A_j = 0$ se

$i \neq j$), allora

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Si possono dimostrare le seguenti proprietà di uno spazio di probabilità:

1. $P(\varnothing) = 0$
2. $P(\overline{A}) = 1 - P(A)$.
3. se $A_1, A_2, \ldots, A_n$ sono elementi di $\mathcal{P}$ con n finito o infinito, allora $\bigcup_{i=1}^{n} A_i \in \mathcal{P}$ e $\bigcap_{i=1}^{n} A_i \in \mathcal{P}$.
4. $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ è vera anche per n finito.
5. se $A, B \in \mathcal{P}$ e $A \subset B$, allora $P(A) \leq P(B)$ e $P(B \backslash A) = P(B) - P(A)$.
6. se $A, B \in \mathcal{P}$, allora $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Esempio. Supponiamo che i pezzi prodotti da una certa macchina possano presentare due tipi di difetti, che chiameremo a e b. È stato stabilito che la probabilità che un pezzo presenti il difetto a è 0.1, la probabilità che non presenti il difetto b è 0.8, la probabilità che presenti entrambi i difetti è 0.01.
Qual'è la probabilità che un pezzo non presenti alcun difetto?
Indichiamo con A l'evento *il pezzo presenta il difetto a* e con B l'evento *il pezzo presenta il difetto b*. Le informazioni si traducono in: $P(A) = 0.1$, $P(\overline{B}) = 0.8$, $P(A \cap B) = 0.01$.
L'evento richiesto è l'evento $\overline{A} \cap \overline{B} = \overline{A \cup B}$. Pertanto:

$$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] =$$

$$= P(\overline{B}) + P(A \cap B) - P(A) = 0.8 + 0.01 - 0.1 = 0.71$$

3.2.2 Come si assegnano le probabilità

Sia Ω uno spazio discreto e $\{\omega_i\}$, $i = 1, \ldots,$ gli eventi elementari. Ogni evento A può essere visto come unione finita o infinita (numerabile) di eventi elementari (e perciò disgiunti). Allora

$$P(A) = P\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} P(\{\omega_i\})$$

Se conosciamo le probabilità $p_i = P(\{\omega_i\})$ degli eventi elementari, risulta completamente definita la funzione di probabilità su Ω.

Esempio.

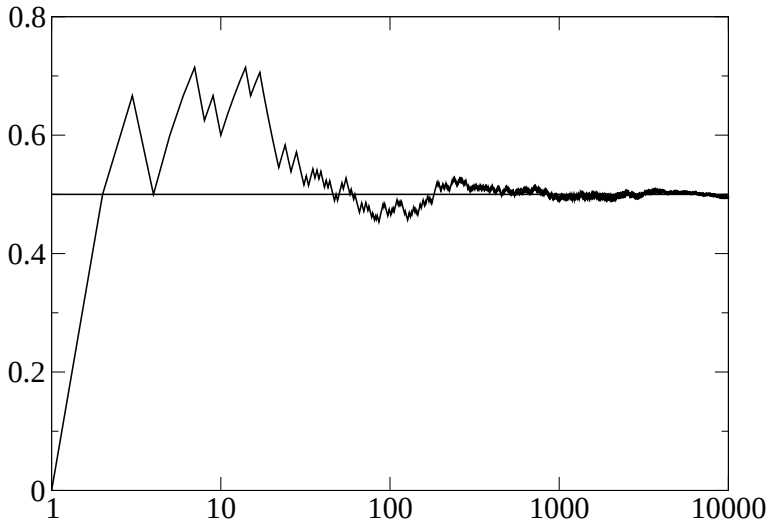
$$\Omega = \mathbb{N}, \quad p_i = \frac{1}{2^i}, \quad i = 1, 2, \ldots$$

Verifichiamo che p_i definisce una probabilità:

$$0 \leq p_i \leq 1, \quad P(\Omega) = \sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = \frac{1}{1 - 1/2} - 1 = 1$$

Calcoliamo ad esempio la probabilità dell'evento A *numero pari*:

$$P(A) = \sum_{i=1}^{\infty} p_{2i} = \sum_{i=1}^{\infty} p_{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^i = \frac{1}{1 - 1/4} - 1 = \frac{1}{3}$$



Frequenza relativa dell'evento testa in una successione di lanci

La probabilità classica. Consideriamo il caso in cui lo spazio campionario è *finito*. Facciamo l'ulteriore ipotesi che gli eventi elementari siano *equiprobabili*.

$$\Omega = \{\omega_1, \ldots, \omega_n\}, \quad p_k = \frac{1}{n}$$

La probabilità dell'evento A è

$$P(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{\#A}{n} = \frac{\#A}{\#\Omega}$$

dove $\#A$ rappresenta il numero degli eventi elementari che costituiscono l'evento A.
Dunque la probabilità *classica* di un evento è *il rapporto tra il numero dei casi favorevoli e il numero dei casi possibili*.

Esempio. Estraiamo due palline da un'urna che contiene 60 palline bianche e 40 palline nere. In questo caso $n = C_{100,2} = 100 * 99/2 = 4950$
($C_{n,k}$ è il *coefficiente binomiale* e rappresenta il numero di **combinazioni di** k **oggetti tra** n):

$$C_{n,k} = \binom{n}{k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$$

I modi possibili di estrarre 2 palline nere è $\#A = C_{40,2} = 40 * 39/2 = 780$.
La probabilità che le due palline estratte siano nere è: $p = \#A/n = 0.158$.

L'idea frequentista di probabilità. La probabilità dell'evento A è il limite della frequenza relativa con cui A si verifica in una lunga serie di prove ripetute sotto condizioni simili.
Da questo punto di vista la probabilità è dunque una frequenza relativa.

Esempio: si lancia una moneta n volte e si considera la frequenza relativa dell'evento *Testa* (numero di volte in cui si presenta T diviso per n). All'aumentare di n tale frequenza relativa tende a stabilizzarsi intorno al valore limite 0.5, che è la probabilità di T.

3.3 Probabilità condizionata

Ci chiediamo qual'è la probabilità di un evento A nell'ipotesi che l'evento B sia già verificato.

Def. Sia B un evento con $P(B) > 0$. Si chiama **probabilità di A condizionata a B** il numero

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

- $P(A|B)$, fissato B, è effettivamente una probabilità. Infatti $P(A|B) \leq 1$ in quanto $P(A \cap B) \leq P(B)$; inoltre $P(\Omega|B) = 1$ poiché $\Omega \cap B = B$; infine data una successione di eventi incompatibili A_i,

$$P\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \frac{P((\cup_{i=1}^{\infty} A_i) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots)}{P(A)}$$

  Essendo $A_i \cap B$ incompatibili, $P((A_1 \cap B) \cup (A_2 \cap B) \dots) = \sum_{i=1}^{\infty} P(A_i \cap B)$. Pertanto

$$P\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

- $P(A|B)$, fissato A, non è una probabilità.
  Ad esempio, mentre è vero che $P(\overline{A}|B) = 1 - P(A|B)$, in generale $P(A|\overline{B}) \neq 1 - P(A|B)$.

- Se lo spazio campionario è finito, la definizione data sopra trova una piena giustificazione: poiché l'evento B si è verificato, si tratta di determinare la probabilità dell'evento $A \cap B$ prendendo come nuovo spazio campionario l'insieme B. Agli eventi elementari di A si attribuiscono nuove probabilità $\pi_i = p_i/p(B)$. Si ha dunque:

$$P(A|B) = \sum_{A \cap B} \pi_i = \frac{\sum_{A \cap B} p_i}{\sum_B p_i} = \frac{P(A \cap B)}{P(B)}$$

- Nel caso della probabilità classica (ossia di n eventi elementari equiprobabili con $p_i = 1/n$), dato $B \neq \emptyset$ si ha:

$$P(A|B) = \frac{\frac{\#(A \cap B)}{\#\Omega}}{\frac{\#B}{\#\Omega}} = \frac{\#(A \cap B)}{\#B}$$

  Si considera B come nuovo spazio campionario e si fa riferimento solo agli eventi elementari che appartengono sia ad A che a B.

- *Esempio 1*. Una confezione contiene 25 transistors di buona qualità, 10 difettosi (cioè che si rompono dopo qualche ora), e 5 guasti. Un transistor viene scelto a caso e messo in funzione. Sapendo che non è guasto, qual'è la probabilità che sia di buona qualità?
  Evento A: il transistor scelto a caso è di buona qualità.
  Evento B: il transistor scelto a caso è difettoso.
  Evento C: il transistor scelto a caso è guasto.
  $P(A) = 25/40$, $P(B) = 5/40$, $P(C) = 10/40$.

$$P(A|\overline{C}) = \frac{P(A \cap \overline{C})}{P(\overline{C})} = \frac{P(A)}{1 - P(C)} = \frac{25/40}{35/40} = \frac{5}{7}$$

- *Esempio 2*. Problema delle tre carte: supponiamo di avere tre carte da gioco, una con faccia rossa e l'altra nera, una con entrambe le facce rosse e una con entrambe le facce nere. Si estrae una carta a caso e la si mette sul tavolo. Se la faccia visibile è rossa, qual'è la probabilità che la faccia coperta sia rossa?
  Sia A l'evento *la faccia coperta è rossa*.
  Sia B l'evento *la faccia visibile è rossa*.
  Dobbiamo calcolare $P(A|B)$.
  L'evento $A \cap B$ è l'evento *abbiamo scelto la carta con entrambe le facce rosse* la cui probabilità è pari a $1/3$.
  L'evento B ha probabilità $1/2$ poichè vi sono in totale tante facce rosse quante facce nere. Quindi
  $$P(B|A) = \frac{1/3}{1/2} = \frac{2}{3}$$

3.4 Indipendenza di eventi

Intuitivamente, due eventi A e B si dicono indipendenti se il verificarsi di uno dei due non modifica la probabilità che l'altro accada. Formalizzando:

$$A \text{ e } B \text{ indipendenti} \quad \Leftrightarrow \quad P(A|B) = P(A) \text{ (se } P(B) > 0)$$

o equivalentemente

$$P(B|A) = P(B) \text{ (se } P(A) > 0)$$

Dalla definizione si ricava che $P(A \cap B) = P(A)P(B)$.

Def. due eventi A e B si dicono **indipendenti** se

$$P(A \cap B) = P(A)P(B)$$

Esempio. Un'urna contiene 6 palline rosse e 4 palline bianche. Ne estraggo una, ne guardo il colore, la reintroduco e ne estraggo una seconda. Qual'è la probabilità che entrambe siano bianche?

Gli eventi B_i = *la i-esima pallina estratta è bianca* sono indipendenti. Pertanto $P(B_1 \cap B_2) = P(B_1)P(B_2) = \frac{4}{10} x \frac{4}{10} = .16$.

Nota: se l'estrazione avveniva senza reimmissione, i due eventi B_1 e B_2 non erano più indipendenti.

Def. Gli eventi $A_1, \ldots, A_n$ si dicono indipendenti se per ogni $k \leq n$ e per ogni scelta degli indici $i_1, \ldots, i_k$ tutti distinti, vale

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

Nota: eventi indipendenti sono sicuramente indipendenti a due a due. *Non è vero però il viceversa: controesempio di Bernstein*. Consideriamo un tetraedro con le facce di questi colori: 1 blu, 1 rossa, 1 gialla, 1 rossa blu e gialla. Lanciamo il tetraedro e osserviamo qual'è il colore che compare sulla faccia appoggiata. Consideriamo i tre eventi:
B = esce il colore blu.
R = esce il colore rosso.
G = esce il colore giallo.
Chiaramente $P(B) = P(R) = P(G) = 1/2$.
$P(B \cap R) = P(R \cap G) = P(B \cap G) = 1/4 = P(B)P(R) = P(R)P(G) = P(B)P(G)$: gli eventi B, R e G sono a due a due indipendenti.
Però $P(B \cap R \cap G) = 1/4 \neq P(B)P(R)P(G) = 1/8$: B, R e G non sono indipendenti.

Cap. 4. Variabili aleatorie

4.1 Definizioni

Def. Si chiama **variabile aleatoria** X una funzione definita sullo spazio campionario Ω che ad ogni evento elementare ω associa un numero reale:

$$X: \quad \Omega \longrightarrow \mathbb{R}$$

$$\omega \longrightarrow X(\omega)$$

Esempio. Sia Ω lo spazio campionario generato dal lancio di due dadi: $\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$. Definiamo X la somma dei numeri che si verificano:

$$(i,j) \longrightarrow i+j$$

In genere quello che ci interessa di una variabile aleatoria è di *calcolare la probabilità che essa assuma determinati valori*.

Nell'esempio precedente del lancio di due dadi ci può interessare di conoscere la probabilità che la somma dei numeri sia pari a 5, oppure che sia inferiore a 7, ecc..

Useremo in seguito le seguenti abbreviazioni:

$\{X = a\}$	indica	$\{\omega : X(\omega) = a\}$
$\{a < X \leq b\}$	indica	$\{\omega : a < X(\omega) \leq b\}$
$\{X \in I\}$	indica	$\{\omega : X(\omega) \in I\}$

Def. Una variabile aleatoria si dice **discreta** se lo spazio campionario Ω su cui è definita è discreto. Le possibili determinazioni di X possono essere indicate mediante la successione $x_1, x_2, \ldots, x_n$.

Def. Chiamiamo **legge, o distribuzione di probabilità** di una v.a. X l'applicazione

$$I \longrightarrow P(X \in I) \quad \forall I \subseteq \mathbb{R}$$

Per una v.a. discreta la legge è assegnata quando si fornisca la tabella delle probabilità p_i che la variabile aleatoria assuma il valore x_i.

La funzione

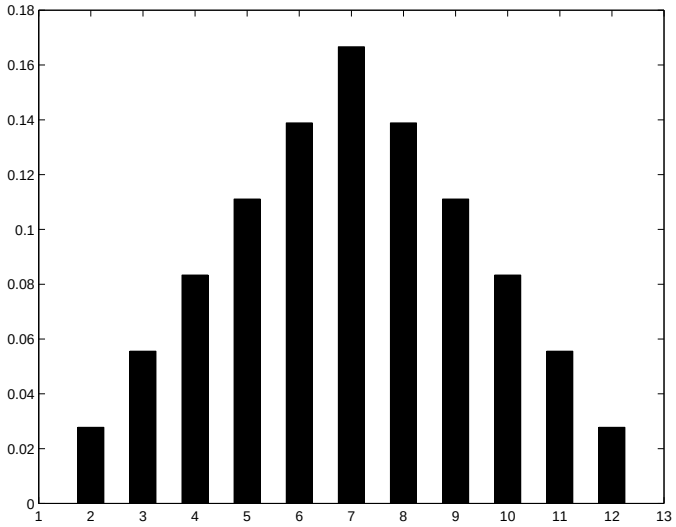
$$x_i \longrightarrow p_X(x_i) = P(X = x_i)$$

viene chiamata **densità (discreta) di X, o funzione di probabilità**.
Precisamente, la legge è data da:

$$P(X \in I) = \sum_{x_i \in I} p_X(x_i)$$

Nell'esempio precedente del lancio di due dadi, la v.a. assume valori interi compresi tra 2 e 12. La densità di X è data dalla seguente tabella:

X	2	3	4	5	6	7	8	9	10	11	12
p_X	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Densità discreta della v.a. *somma dei punti di due dadi*

Def. Chiamiamo **funzione di ripartizione** della v.a. X la funzione che dà, per ogni x, la probabilità che la variabile aleatoria assuma valori minori o uguali a x. Per una v.a. discreta:

$$x \longrightarrow F_X(x) = \sum_{i \text{ t.c. } x_i \leq x} p_X(x_i)$$

La funzione di ripartizione di una v.a. discreta è costante a tratti: nell'intervallo $[x_i, x_{i+1})$ è costante, mentre in x_{i+1} cresce della quantità p_{i+1}.

Nel nostro esempio del lancio di due dadi la funzione di ripartizione è data dalla seguente tabella:

X	$x < 2$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$
F_X	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$

X	$[7,8)$	$[8,9)$	$[9,10)$	$[10,11)$	$[11,12)$	$x \geq 12$
F_X	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

Def. Due v.a. X, Y si dicono **indipendenti** se per ogni coppia di intervalli $I, J \subseteq \mathbb{R}$, risulta

$$P(X \in I, Y \in J) = P(X \in I)P(Y \in J)$$

(la virgola a primo membro tra eventi espressi tramite variabili aleatorie sta per l'intersezione).

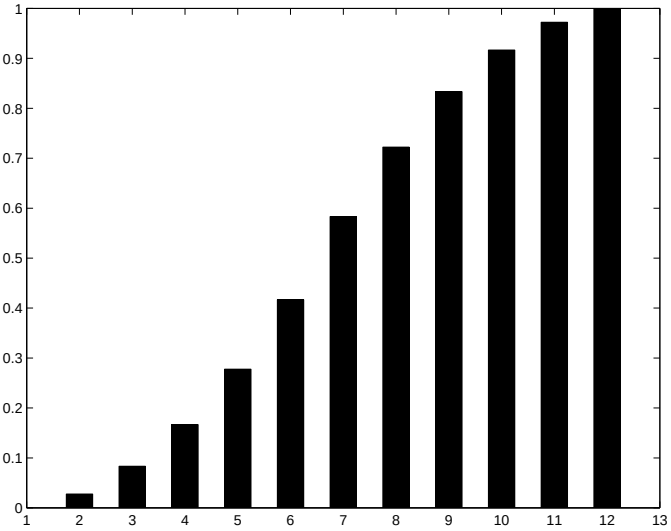
Tradotto nel linguaggio degli eventi ciò significa

$$P(\{(\omega : X(\omega) \in I) \cap (\omega : Y(\omega) \in J)\}) = P(\{(\omega : X(\omega) \in I\})P(\{\omega : Y(\omega) \in J)\})$$

Si definisce allo stesso modo l'indipendenza di n v.a.: $X_1, X_2, \ldots, X_n$ sono indipendenti se scelti comunque n intervalli $I_1, I_2, \ldots, I_n \subseteq \mathbb{R}$ si ha

$$P(X_1 \in I_1, X_2 \in I_2, \ldots, X_n \in I_n) = P(X_1 \in I_1)P(X_2 \in I_2) \ldots P(X_n \in I_n)$$

Nota: la definizione di indipendenza di n v.a. richiede una sola condizione, mentre quella di indipendenza di n eventi richiedeva che $P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$ per ogni $k \leq n$



Funzione di ripartizione della v.a. *somma dei punti di due dadi*

e per ogni scelta degli indici $i_1, \ldots, i_k$ tutti distinti. Le due definizioni sembrerebbero diverse, tuttavia, per l'arbitrarietà della scelta degli intervalli I_i, esse in realtà sono analoghe. Infatti se scegliamo $I_i = \mathbb{R}$ per $i > k$, allora $P(X_i) = 1$ per $i > k$, e

$$P(X_1 \in I_1, X_2 \in I_2, \ldots, X_k \in I_k) = P(X_1 \in I_1)P(X_2 \in I_2) \ldots P(X_k \in I_k)$$

Def. La funzione $p_{X_1 \ldots X_n}$ definita da

$$p_{X_1 \ldots X_n}(y_1, \ldots, y_n) = P(X_1 = y_1, \ldots, X_n = y_n)$$

è detta **densità congiunta** di $X_1, \ldots, X_n$.

Def. Si definisce la **funzione di ripartizione congiunta** delle v.a. X e Y la funzione che dà, per ogni x e y, la probabilità che X assuma valori minori o uguali a x e Y assuma valori minori o uguali a y:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} p_{XY}(u, v)$$

Def. Si definisce la **funzione di probabilità marginale** di X (risp. Y) la funzione di probabilità della singola v.a. X (risp. Y):

$$p_X(x) = P(X = x) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_x p_{XY}(x, y)$$

L'indipendenza di n variabili aleatorie si traduce nella seguente:

$$p_{X_1 \ldots X_n}(y_1, \ldots, y_n) = p_{X_1}(y_1) \times \cdots \times p_{X_n}(y_n)$$

dove p_{X_i} indica la densità di probabilità della v.a. X_i.

Esempio 1. Un'urna contiene 5 palline numerate da 1 a 5. Estraiamo due palline, con reimmissione della prima pallina. Siano X_1 e X_2 i risultati della prima e della seconda estrazione rispettivamente.

$$P(X_1 = i, X_2 = j) = \frac{1}{25} = P(X_1 = i)P(X_2 = j)$$

pertanto X_1 e X_2 sono eventi indipendenti.

Esempio 2. Un'urna contiene 5 palline numerate da 1 a 5. Estraiamo due palline senza reimmissione. Siano Y_1 e Y_2 i risultati della prima e della seconda estrazione rispettivamente. Per $i \neq j$ $P(Y_1 = i, Y_2 = j) = \frac{1}{20}$ mentre $P(Y_1 = i)P(Y_2 = j) = \frac{1}{25}$ pertanto Y_1 e Y_2 non sono eventi indipendenti.

4.2 Indici di posizione di una variabile aleatoria

Sia X una v.a. discreta che assume i valori $x_1, \ldots, x_n$, e sia p_X la sua densità di probabilità. Si chiama **valore atteso, o media, o speranza matematica** di X, e la si denota con $\mathbb{E}(X)$, la quantità

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_X(x_i)$$

Se i valori della v.a. sono un'infinità numerabile, la somma diventa una serie; si dice che X ammette valor medio se $\sum_{i=1}^{\infty} |x_i| p_X(x_i) < \infty$. In questo caso il valor medio è definito da

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_X(x_i)$$

e il valore atteso $\mathbb{E}(X)$ è definito a condizione che la serie converga assolutamente.

Esempio. Consideriamo la variabile aleatoria associata al lancio di due dadi. Il valore atteso è

$$\mathbb{E}(X) = 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} +$$
$$+8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} = 7$$

Se X assume n valori distinti $x_1, \ldots, x_n$ con uguale probabilità $p_X(x_i) = 1/n$, allora $\mathbb{E}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$ è la media aritmetica dei valori assunti.

Proprietà del valore atteso

1. Per una trasformazione lineare della v.a. il valore atteso si trasforma linearmente:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

$$\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$$

2. Valore atteso di una funzione di v.a.: sia X una v.a. e f una funzione continua su $\mathbb{R}$. Allora il valore atteso di $f(X)$ si può calcolare così:

$$\mathbb{E}[f(X)] = \sum_{k} f(x_k) p_X(x_k)$$

3. Valore atteso di una funzione di più v.a.: siano $X_1, \ldots, X_n$ n v.a. e f una funzione continua su $\mathbb{R}$. Allora il valore atteso di $f(X_1, \ldots, X_n)$ si può calcolare così:

$$\mathbb{E}[f(X_1, \ldots, X_n)] = \sum_{k_1, \ldots, k_n} f(x_{1k_1}, \ldots, x_{nk_n}) p_{X_1 \ldots X_n}(x_{1k_1}, \ldots, x_{nk_n})$$

   Dove x_{ik_i} indica il k_i-esimo valore della v.a. X_i

4. Se $X_1, \ldots, X_n$ sono v.a. indipendenti con valore atteso finito, allora

$$\mathbb{E}(X_1 X_2 \ldots X_n) = \mathbb{E}(X_1)\mathbb{E}(X_2) \ldots \mathbb{E}(X_n)$$

Esempio. Nella trasmissione di un'immagine il colore di ogni pixel è descritto da un vettore a 8 bits $(a_1, \ldots, a_8)$, dove gli a_i possono valere 0 oppure 1. Durante la trasmissione di ogni bit si può avere un errore con probabilità $p_b = 2 \times 10^{-4}$, indipendentemente da un bit all'altro.

1. Qual'è la probabilità che un singolo pixel venga trasmesso correttamente?

2. Per un'immagine composta da $512 \times 256 = 131072$ pixels quale sarà il numero medio di pixels distorti?

Soluzione.

1. Consideriamo l'evento A_i = lo i-esimo bit non è stato distorto ($i = 1, \ldots, 8$). La probabilità p_p che un singolo pixel venga trasmesso correttamente è

$$p_p = P(A_1 \cap \cdots \cap A_8) = P(A_1) \ldots P(A_8) = (1 - p_b)^8 \approx 0.9984$$

2. Definiamo la v.a. X_i che vale 1 se lo i-esimo pixel è stato distorto, 0 altrimenti. Si ha $P(X_i = 1) = 1 - 0.9984 = 1.6 \times 10^{-3}$. Chiamiamo $S_n = \sum_{i=1}^{n} X_i$. Il valor medio che cerchiamo è $\mathbb{E}(S_n)$ con $n = 131072$:

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\mathbb{E}(X_1) \approx 131072 \times 1.6 \times 10^{-3} = 209.7$$

4.3 Indici di dispersione di una variabile aleatoria discreta

Definiamo ora gli indici caratteristici di *dispersione* (la varianza, la deviazione standard) di una variabile aleatoria discreta.

Def. Sia X una v.a. discreta avente valore atteso finito. Si definisce **varianza** di X la quantità

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

purché questo valore sia finito. In caso contrario X non ha varianza finita.

La varianza di una v.a. dunque rappresenta una misura della sua dispersione rispetto al valore atteso $\mathbb{E}(X)$.

Proprietà della varianza:

1. Per calcolare la varianza possiamo ricorrere alla formula:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} [x_i - \mathbb{E}(X)]^2 p_X(x_i)$$

   che si ottiene sfruttando la proprietà del valore atteso di una funzione di v.a..

2. $\mathrm{Var}(X) > 0$

3. $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

4. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ per qualsiasi valore di $a, b \in \mathbb{R}$

5. Se $X_1, X_2, \ldots, X_n$ sono v.a. *indipendenti*, allora

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)$$

   Se le v.a. non sono indipendenti questa proprietà non vale. Ad esempio $\mathrm{Var}(X + X) = \mathrm{Var}(2X) = 4\mathrm{Var}(X) \neq \mathrm{Var}(X) + \mathrm{Var}(X)$.

Def. La **deviazione standard, o scarto quadratico medio**, è definito come

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

Def. Sia X una v.a. con valore atteso μ_X e varianza σ_X^2 finite. Si dice **standardizzata** di X la v.a.

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Z ha valore atteso nullo e varianza pari a 1.

4.4 Analisi comparative tra variabili aleatorie discrete

Def. Siano X e Y due v.a. aventi varianza finita; si definisce la **covarianza** di X e Y come:

$$\text{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

La covarianza viene anche denotata con il simbolo σ_{XY}.
Proprietà della covarianza:

1. $\text{cov}(X,Y) = \text{cov}(Y,X)$
2. $\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
3. $\text{cov}(X,X) = \text{Var}(X)$
4. $\text{cov}(aX+b, cY+d) = ac\,\text{cov}(X,Y)$
5. $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X,Y)$

   Da questa proprietà discende che condizione necessaria e sufficiente affinché due v.a. X e Y siano indipendenti è che $\text{cov}(X,Y) = 0$.

Def. Siano X e Y due v.a. aventi varianza finita; si definisce **coefficiente di correlazione** di X, Y, la quantità

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Il coefficiente di correlazione è sempre compreso tra -1 e 1. Due variabili aleatorie si dicono *incorrelate* se $\text{cov}(X,Y) = 0$ e dunque se $\rho_{XY} = 0$.

Se due v.a. X, Y sono indipendenti, allora sono incorrelate (ma non vale il viceversa!). Infatti, se sono indipendenti allora $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Perciò $\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.

Nel caso in cui $\rho_{XY} = \pm 1$, si dimostra che le v.a. sono linearmente dipendenti: $Y = aX + b$. La costante a ha lo stesso segno di ρ_{XY}.

Esercizio. Un'urna contiene k palline nere e $n-k$ palline bianche. Se ne estraggono due senza rimpiazzo. Definiamo le v.a. X_1 e X_2:

$X_1 = 1$ se la 1^a pallina estratta è *nera*,
$X_1 = 0$ se la 1^a pallina estratta è *bianca*,
$X_2 = 1$ se la 2^a pallina estratta è *nera*,
$X_2 = 0$ se la 2^a pallina estratta è *bianca*.

Calcolare il coefficiente di correlazione tra X_1 e X_2.

Soluzione.

$$P(X_1 = 1) = P(X_2 = 1) = \frac{k}{n}; \quad P(X_1 = 0) = P(X_2 = 0) = \frac{n-k}{n}$$

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = \frac{k}{n}1 + \frac{n-k}{n}0 = \frac{k}{n}; \quad \mathbb{E}(X_1^2) = \mathbb{E}(X_2^2) = \frac{k}{n}1 + \frac{n-k}{n}0 = \frac{k}{n}$$

$$\text{Var}(X_1) = \text{Var}(X_2) = \mathbb{E}(X_1^2) - [\mathbb{E}(X_1)]^2 = \frac{k(n-k)}{n^2}$$

$$\text{cov}(X_1, X_2) = \left(1 - \frac{k}{n}\right)\left(1 - \frac{k}{n}\right) P(X_1 = 1, X_2 = 1) +$$

$$+ \left(1 - \frac{k}{n}\right)\left(-\frac{k}{n}\right) P(X_1 = 1, X_2 = 0) + \left(-\frac{k}{n}\right)\left(1 - \frac{k}{n}\right) P(X_1 = 0, X_2 = 1) +$$

$$+ \left(-\frac{k}{n}\right)\left(-\frac{k}{n}\right) P(X_1 = 0, X_2 = 0)$$

$$= \left(1 - \frac{k}{n}\right)\left(1 - \frac{k}{n}\right)\frac{k}{n}\frac{(k-1)}{(n-1)} + 2\left(-\frac{k}{n}\right)\left(1 - \frac{k}{n}\right)\frac{k}{n}\frac{(n-k)}{(n-1)} +$$
$$+ \left(-\frac{k}{n}\right)\left(-\frac{k}{n}\right)\frac{n-k}{n}\frac{(n-k-1)}{(n-1)}$$
$$= \frac{-k(n-k)}{n^2(n-1)}$$
$$\rho_{X_1X_2} = \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} = -\frac{1}{n-1}$$

Analogie tra variabile aleatoria e insiemi di dati numerici:

dati numerici	*variabile aleatoria*
n-upla di numeri $(x_1, \ldots, x_n)$	v.a. X
$f_r(k) = \#\{x_i \mid x_i = k\}/n \ (i = 1, \ldots, n)$	$x_i \longrightarrow p_X(x_i) = P(X = x_i)$
$\overline{x} = \sum_{i=1}^{N_c} \overline{x}_i f_r(i)$	$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_X(x_i)$
$\sigma^2 = \sum_{i=1}^{N_c} f_r(i)(\overline{x}_i - \overline{x})^2$	$\mathrm{Var}(X) = \sum_{i=1}^{n}(x_i - \mathbb{E}(X))^2 p_X(x_i)$
$\sigma_{xy} = \sum_{i=1}^{N_c} f_r(i)(\overline{x}_i - \overline{x})(\overline{y}_i - \overline{y})$	$\mathrm{cov}(X, Y) = \sum_{i=1}^{n}(x_i - \mathbb{E}(X))\times$ $\times((y_i - \mathbb{E}(Y))p_{XY}(x_i, y_i)$

Cap. 5. Modelli discreti di variabili aleatorie

5.1 Processo di Bernoulli

Def. Si dice **esperimento di Bernoulli** o **prova di Bernoulli** un esperimento aleatorio in cui sono possibili solo due esiti. Le rispettive probabilità sono p e $1-p$, dove p è un numero reale compreso tra 0 e 1. p viene chiamato *parametro* della prova di Bernoulli. Convenzionalmente l'evento con probabilità p viene chiamato *successo* mentre quello con probabilità $1-p$ viene chiamato *insuccesso*.

Esempio 1. Il lancio di una moneta è un esperimento di Bernoulli. Se la moneta non è truccata il parametro p vale $1/2$.

Esempio 2. Lancio due dadi e considero *successo* l'evento *la somma dei punti dei due dadi è 7*, e insuccesso l'evento complementare. Il parametro p vale $1/6$.

La variabile aleatoria associata alla prova di Bernoulli si indica con $X \sim B(p)$ e prende i valori 1 in caso di successo e 0 in caso di insuccesso:

$$p_X(1) = p \quad p_X(0) = 1 - p$$

In modo compatto:

$$p_X(a) = p^a(1-p)^{1-a}$$

Def. Si dice **processo di Bernoulli** una sequenza di esperimenti di Bernoulli di uguale parametro p, tra loro indipendenti.

La sequenza può essere finita oppure infinita (numerabile). Nel secondo caso si parla di processo di Bernoulli *illimitato*.

Esempio. Si controllano 100 pezzi prodotti e si registra il numero di pezzi difettosi.

Def. Consideriamo un processo di Bernoulli di parametro p, di n prove. Si definisce **binomiale** di parametri n e p, e la si scrive $X \sim B(n,p)$, la v.a. che conta il numero complessivo di successi ottenuti nelle n prove. Dunque $B(n,p)$ è la somma di n v.a. Bernoulliane di parametro p, indipendenti tra loro:

$$X = \sum_{i=1}^{n} X_i, \quad X \sim B(n,p), \quad X_i \sim B(p)$$

Teorema. La v.a. binomiale di parametri n e p può assumere valori interi compresi tra 0 e n. La sua densità discreta è:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

Dim. La probabilità $P_{X_1 \ldots X_n}(a_1, \ldots, a_n)$, per l'indipendenza degli eventi, vale:

$$P_{X_1 \ldots X_n}(a_1, \ldots, a_n) = p_{X_1}(a1) \ldots p_{X_n}(a_n) =$$

$$= p^{a_1}(1-p)^{1-a_1} \dots p^{a_n}(1-p)^{1-a_n} = p^{\sum_{i=1}^n a_i}(1-p)^{n-\sum_{i=1}^n a_i}$$

L'evento *k successi nelle n prove* richiede che $\sum_{i=1}^n a_i = k$:

$$P_{X_1 \dots X_n}(a_1, \dots, a_n) = p^k(1-p)^{n-k}$$

L'evento può essere ottenuto con $\binom{n}{k}$ scelte diverse dell'insieme degli a_i. Pertanto:

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Osservazione. Dire che n v.a. $X_1, \dots, X_n$ sono identicamente distribuite vuol dire che

$$p_{X_1}(x) = p_{X_2}(x) = \dots = p_{X_n}(x) \quad \forall x$$

Questo non vuol dire che le variabili siano identiche!

Ad esempio, sia $S_n = X_1 + \dots + X_n$ con X_i bernoulliane indipendenti tra di loro: S_n è una binomiale $S_n \sim B(n,p)$. Invece, se $X_1 = X_2 = \dots = X_n$, le v.a. X_i non sono indipendenti, e $S_n = X_1 + \dots + X_n = nX_1$. S_n assume soltanto i due valori 0 e n, e la densità di probabilità associata è $p(n) = p$, $p(0) = 1-p$.

Esercizio. Una compagnia aerea sa che il 10% dei passeggeri che hanno prenotato non si presenta alla partenza. In base a questa considerazione accetta 32 prenotazioni su 30 posti liberi. Supponendo che i comportamenti dei passeggeri siano indipendenti, qual'è la probabilità che almeno uno rimanga a terra?

Soluzione. Sia X la v.a. che vale 0 se il passeggero con prenotazione non si presenta, 1 se si presenta. $X \sim B(0.9)$. Gli eventi sono 32, e cerchiamo la probabilità che la binomiale di parametri $n = 32$ e $p = 0.9$ abbia valore maggiore di 30:

$$p(S_{32} > 30) = \binom{32}{31} 0.9^{31} 0.1^1 + \binom{32}{32} 0.9^{32} 0.1^0 \approx 0.122 + 0.034 = 0.156$$

Def. Consideriamo una successione di prove Bernoulliane indipendenti di parametro p. Si chiama **geometrica** la v.a. che rappresenta il numero di prove necessario affinché si presenti per la prima volta l'evento *successo*. La indicheremo $X \sim \text{Geo}(p)$. La funzione di probabilità associata è:

$$p_X(k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \dots$$

Infatti questa è la probabilità che in k prove Bernoulliane le prime $k-1$ siano insuccessi e l'ultimo sia un successo.

5.1.1 Media e varianza del processo di Bernoulli

Sia $X \sim B(P)$. Allora

$$\mathbb{E}(X) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p$$

$$\text{Var}(X) = (0-p)^2 p_X(0) + (1-p)^2 p_X(1) = p^2(1-p) + (1-p)^2 p = p(1-p)$$

Sia ora $X \sim B(n,p)$. Siccome X è somma di n v.a. Bernoulliane *indipendenti*, si avrà:

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) = np, \quad \text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p)$$

5.2 Processo di Poisson

Un processo di Poisson è un processo di Bernoulli in cui il parametro n è molto grande, mentre il valore atteso $\mathbb{E}[B(n,p)] = np = \lambda$ è un numero finito noto.

Esempio 1. Sia X il numero di utenti che chiamano un centralino telefonico in un giorno. Si vuole conoscere la distribuzione di probabilità di X, sapendo che il numero n delle persone che *potrebbero* chiamare il centralino è molto grande, che le azioni di questi utenti sono indipendenti, e che in media si verificano λ chiamate al giorno.

Allora X è un processo di Bernoulli $X \sim B(n,p)$, e la sua densità di probabilità è

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Sapendo che $\mathbb{E}[B(n,p)] = np = \lambda$, ossia che $p = \lambda/n$:

$$p_X(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, \ldots, n$$

Esempio 2. Un filo di rame di lunghezza L metri possiede delle imperfezioni. Sappiamo che in media si verificano λ imperfezioni ogni L metri, e che le posizioni delle imperfezioni sul filo sono variabili casuali indipendenti. Vogliamo sapere la funzione di probabilità della v.a. *numero di imperfezioni* del filo di lunghezza L.

Sezioniamo il filo in n intervalli di lunghezza L/n. Se n è abbastanza grande la probabilità che vi siano più di una imperfezione in ogni intervallo è trascurabile. X è un processo di Bernoulli $X \sim B(n,p)$ con valore atteso $\mathbb{E}(X) = np = \lambda$. La distribuzione di probabilità è pertanto la stessa dell'esempio precedente.

Altri esempi che danno luogo alla stessa distribuzione di probabilità sono:

1. il numero di automobili che passa per un determinato incrocio in un determinato intervallo di tempo;
2. il numero di persone che si reca in un negozio in un giorno feriale;
3. il numero di guasti che si verificano in un impianto in un giorno lavorativo;
4. il numero di pixels difettosi di uno schermo a cristalli liquidi;

Def. Si chiama **processo di Poisson** un esperimento aleatorio schematizzabile come la registrazione del numero di successi di eventi casuali in un determinato intervallo di tempo con le proprietà seguenti:

- la probabilità che in un intervallino di tempo ci sia più di un successo è trascurabile,
- la probabilità di successo è la stessa per tutti gli intervallini di tempo e proporzionale alla lunghezza dell'intervallino,
- il successo o meno in un intervallino di tempo è indipendente dal fatto che ve ne sia uno in un altro intervallino di tempo.

Allora se λ è il numero medio di successi, la v.a. associata al numero di successi ha una distribuzione data da

$$p_X(k) = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, \ldots, n$$

Calcoliamo esplicitamente questo limite:

$$p_X(k) = \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Questa distribuzione viene chiamata **legge di Poisson** di parametro λ, e si scrive $X \sim P(\lambda)$.
Proprietà.

- Nel processo di Poisson il numero medio di successi è proporzionale all'intervallo di tempo considerato. Pertanto possiamo scrivere $\lambda = \nu t$. Il parametro ν viene chiamato *intensità* del processo di Poisson. Rappresenta il numero medio di successi nell'unità di tempo. La legge della variabile aleatoria è $X \sim P(\lambda) = P(\nu t)$.
- Verifichiamo che il valore atteso della legge di Poisson è pari a λ:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=0}^{\infty} \frac{k e^{-\lambda} \lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda$$

- Calcoliamo la varianza: la varianza della binomiale $X \sim B(n,p)$ è $\mathrm{Var}(X) = np(1-p) = \lambda(1-\lambda/n)$. Il limite per $n \to \infty$ è dunque $\mathrm{Var}(X) = \lambda$.
- Siano $X_1, \ldots, X_n$ n v.a. indipendenti con legge di Poisson $X_i \sim P(\lambda_i)$ (λ_i possono essere diversi tra loro). Allora

$$X_1 + \cdots + X_n \sim P(\lambda_1 + \cdots + \lambda_n)$$

Cap. 6. Legge dei grandi numeri

6.1 Media campionaria di variabili aleatorie

Date n v.a. X_i $(i = 1, \ldots, n)$, definiamo la v.a. **media campionaria** $\overline{X}_n$ nel modo seguente:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Il valore atteso della v.a. media campionaria è

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i)$$

Siano ora $X_1, X_2, \ldots$ v.a. indipendenti identicamente distribuite con media finita μ e varianza finita σ^2. Il valore atteso della v.a. media campionaria è

$$\mathbb{E}(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \mu$$

Mentre la varianza vale:

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

Ossia la varianza diminuisce al crescere del campione di v.a..

6.2 Disuguaglianza di Chebyshev

Sia X una v.a. con valore atteso μ e varianza σ^2. Sia δ un numero reale positivo prefissato. Vale la seguente disuguaglianza:

$$P(|X - \mu| \geq \delta) \leq \frac{\sigma^2}{\delta^2}$$

Diamo la dimostrazione per una v.a. discreta, ma notiamo che la disuguaglianza vale anche per le v.a. continue.

Sia A l'intervallo dei valori di X compresi tra $\mu - \delta$ e $\mu + \delta$, e sia $\overline{A}$ il complementare di A rispetto ad $\mathbb{R}$:

$$A = \{x : |x - \mu| \leq \delta\}, \qquad \overline{A} = \{x : |x - \mu| > \delta\}$$

Scriviamo la varianza come:

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_X(x_i) = \sum_{i : x_i \in A} (x_i - \mu)^2 p_X(x_i) + \sum_{j : x_j \in \overline{A}} (x_j - \mu)^2 p_X(x_j)$$

Considerato che $\sum_{i:x_i \in A} (x_i - \mu)^2 p_X(x_i)$ è una quantità certamente positiva, e poi che in $\overline{A}$ vale $(x - \mu)^2 \geq \delta^2$, si ha:

$$\sigma^2 \geq \sum_{j:x_j \in \overline{A}} (x_j - \mu)^2 p_X(x_j) \geq \delta^2 \sum_{j:x_j \in \overline{A}} p_X(x_j)$$

Ma ovviamente $\sum_{j:x_j \in \overline{A}} p_X(x_j) = P(|X - \mu| \geq \delta)$ e da ciò discende direttamente la disuguaglianza di Chebyshev.

La disuguaglianza di Chebyshev può essere scritta nelle seguenti forme alternative:

$$P(|X - \mu| \geq \delta\sigma) \leq \frac{1}{\delta^2}$$

$$P(|X - \mu| < \delta\sigma) \geq 1 - \frac{1}{\delta^2}$$

6.3 Legge debole dei grandi numeri

Siano $X_1, X_2, \ldots$ v.a. indipendenti identicamente distribuite con media finita μ. Allora per ogni $\epsilon > 0$ vale

$$P(|\overline{X}_n - \mu| > \epsilon) \to 0 \quad \text{per} \quad n \to \infty$$

Questo fatto si esprime dicendo che la successione di v.a. $\{\overline{X}_n\}$ tende in probabilità a μ per $n \to \infty$.

Dim. Dimostriamo la legge nell'ipotesi che esista finita la varianza $\mathrm{Var}(X_i) = \sigma^2$ (la legge però è vera anche se non esiste la varianza).
Applichiamo la disuguaglianza di Chebyshev alla v.a. $\overline{X}_n$:

$$P\left(|\overline{X}_n - \mu| \geq \delta \frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{\delta^2} \quad \text{per ogni } \delta > 0$$

Scegliamo $\delta = \epsilon\sqrt{n}/\sigma$:

$$P\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2 n} \quad \text{per ogni } \epsilon > 0$$

Per $n \to \infty$ si ottiene la tesi.

Cap. 7. Variabili aleatorie continue

Assumiamo che lo spazio campionario degli eventi sia uno spazio continuo, e che i valori che può assumere la v.a. X formino un insieme continuo su $\mathbb{R}$.

La **legge** della v.a., come nel caso della v.a. discreta, è l'applicazione

$$I \to P(X \in I) \quad \text{per ogni intervallo } I \subseteq \mathbb{R}$$

Def. La **densità (continua)** di una v.a. continua X è la funzione $f_X : \mathbb{R} \to \mathbb{R}$ che determina la legge della v.a. X nel seguente modo:

$$P(X \in I) = \int_I f_X(t)dt \quad \forall I \subseteq \mathbb{R}$$

Naturalmente f_X deve soddisfare le segueti proprietà:

$$f_X(t) \geq 0 \quad \forall t \in \mathbb{R}, \qquad \int_{\mathbb{R}} f_X(t)dt = 1$$

- Attenzione: $f_X(x)$ non è una funzione di probabilità. In particolare $f_X(x) \neq P(X = x)$, e non necessariamente $f_X(x) \leq 1$.

- La legge della v.a. continua è non nulla solo su intervalli di lunghezza finita, mentre è nulla in singoli punti o insiemi numerabili di punti. Pertanto $P(X = a) = 0$ per ogni $a \in \mathbb{R}$.

Def. Sia X una v.a. continua. Si definisce la **funzione di ripartizione** di X la funzione $F_X : \mathbb{R} \to [0, 1]$ definita da

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)dx$$

Proprietà.

- $F_X(t)$ è monotona non decrescente e continua.
- $F_X(t) \to 0$ per $t \to -\infty$
  $F_X(t) \to 1$ per $t \to +\infty$.
- $P(a < X < b) = F_X(b) - F_X(a) \quad \forall a, b \in \mathbb{R}, \quad a < b$
- Nei punti in cui la densità $f_X(t)$ è continua, $F_X(t)$ è derivabile, e $F'_X(t) = f_X(t)$.

7.1 Proprietà delle variabili aleatorie continue

Diamo nel seguito definizioni e proprietà per le v.a. continue di quantità analoghe a quelle già definite per le v.a. discrete.

7.1.1 Valore atteso

Def. Si chiama **valore atteso, o media, o speranza matematica** di una v.a. continua X, il numero

$$\mathbb{E}(X) = \int_{\mathbb{R}} t f_X(t) dt$$

a condizione che l'integrale esista finito.

Proprietà

- $\mathbb{E}(aX_1 + b) = a\mathbb{E}(X_1) + b$
- $\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$
- $X_1, \ldots, X_n$ indipendenti: $\mathbb{E}(X_1 X_2 \ldots X_n) = \mathbb{E}(X_1)\mathbb{E}(X_2) \ldots \mathbb{E}(X_n)$.
- valore atteso di una funzione di v.a.:

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(t) f_X(t) dt$$

7.1.2 Varianza

Def. Si definisce **varianza** di una v.a. continua X il numero

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{\mathbb{R}} [t - \mathbb{E}(X)]^2 f_X(t) dt$$

a condizione che l'integrale esista finito.

Proprietà

- $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \int_{\mathbb{R}} t^2 f_X(t) dt - \left(\int_{\mathbb{R}} t f_X(t) dt\right)^2$
- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$
- X_1,...,X_n indipendenti: $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$

7.2 Modelli continui di variabili aleatorie

7.2.1 Densità uniforme

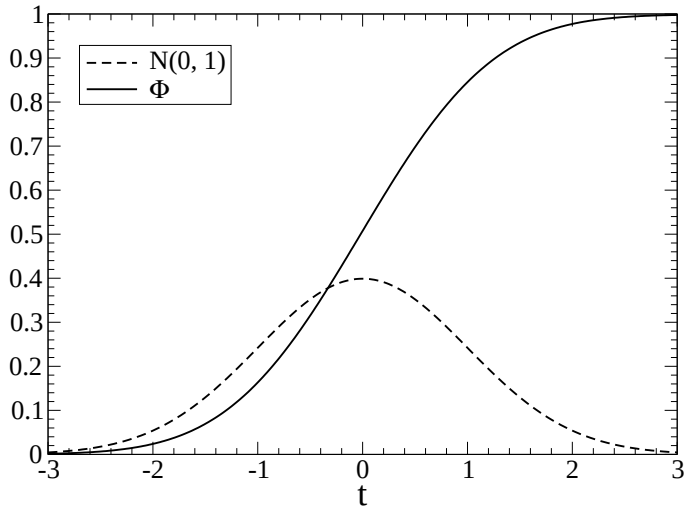
Definiamo la v.a. $X \sim U(a, b)$ che ha densità uniforme sull'intervallo $[a, b]$:

$$f_X(t) = \frac{1}{b-a} I_{[a,b]}(t)$$

dove $I_{[a,b]}$ è la *funzione indicatrice* dell'intervallo $[a, b]$ che vale 1 all'interno dell'intervallo e 0 fuori.

$$P(t_1 < X < t_2) = \int_{t_1}^{t_2} \frac{1}{b-a} I_{[a,b]}(t) dt = \frac{|[a, b] \cap [t_1, t_2]|}{b-a}$$

Nota: la funzione $f_X(t)$ in questo caso è discontinua.



Densità normale standard $\mathcal{N}(0,1)$ e sua funzione di ripartizione Φ

7.2.2 Densità gaussiana (o normale)

- Densità gaussiana standard
  La v.a. normale standard viene chiamata $X \sim \mathcal{N}(0,1)$. La sua densità è definita come

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

  Si dimostra che questa funzione è integrabile su $\mathbb{R}$ con integrale pari a 1.

$$P(a < X < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt$$

  La funzione di ripartizione vale

$$F_X(t) \equiv \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx$$

  Calcoliamo valore atteso e varianza della normale standard.

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t e^{-t^2/2} dt = 0$$

  poiché l'integrando è una funzione dispari.

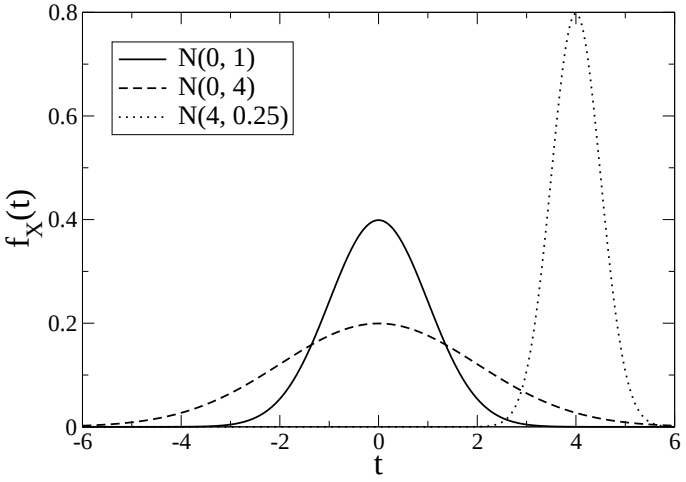
$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^2 e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t \left[t e^{-t^2/2} \right] dt =$$

$$= \frac{-t e^{-t^2/2}}{\sqrt{2\pi}} \Bigg|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2} dt = 1$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 1$$

- Densità gaussiana.
  La v.a. gaussiana $Y \sim \mathcal{N}(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ e $\sigma > 0$, è definita da:

$$Y = \sigma X + \mu \quad \text{con } X \sim \mathcal{N}(0,1)$$



Alcuni esempi di densità gaussiana

Si ha pertanto

$$P(a < Y < b) = P(a < \sigma X + \mu < b) = P\left(\frac{a-\mu}{\sigma} < X < \frac{b-\mu}{\sigma}\right) =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-t^2/2} dt = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

La densità della gaussiana è dunque

$$f_Y(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

La funzione di ripartizione vale

$$F_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Calcoliamo valore atteso e varianza della gaussiana $Y \sim \mathcal{N}(\mu, \sigma^2)$. Sia X la v.a. normale standard: $X \sim \mathcal{N}(0, 1)$.

$$\mathbb{E}(Y) = \mathbb{E}(\sigma X + \mu) = \sigma\mathbb{E}(X) + \mu = \mu$$

$$\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \sigma^2 \mathrm{Var}(X) = \sigma^2$$

Si dimostra che se $X_1, \ldots, X_n$ sono n v.a. *indipendenti* con distribuzione gaussiana $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, allora

$$X_1 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$$

7.2.3 La legge esponenziale

Sia X un processo di Poisson di intensità ν (ossia $X \sim P(\nu t)$). Si dice *v.a. esponenziale* di parametro ν, e si denota con $Y \sim \mathrm{Esp}(\nu)$ la v.a. Y che misura l'istante del primo successo del processo di Poisson. Y è funzione del tempo, dunque è una v.a. continua.

Ad esempio se il processo di Poisson è quello che descrive il numero di guasti nel tempo in un apparecchio meccanico, la v.a. che descrive il tempo di attesa del suo primo guasto è una v.a. esponenziale.

Troviamo l'espressione della densità esponenziale. Dalla definizione:

$$P(Y > t) = P(X = 0) = e^{-\nu t} \quad \text{per } t > 0, \qquad P(Y > t) = 1 \quad \text{per } t \le 0$$

Quindi la funzione di ripartizione è

$$F_Y(t) = P(Y \le t) = 1 - P(Y > t) = 1 - e^{-\nu t} \ \text{ per } t > 0, \quad F_Y(t) = 0 \ \text{ per } t \le 0$$

La densità di Y è la derivata della f.d.r.:

$$f_Y(t) = \nu e^{-\nu t} \quad \text{per } t > 0, \qquad f_Y(t) = 0 \quad \text{per } t \le 0$$

Calcoliamo valore atteso e varianza della legge esponenziale:

$$\mathbb{E}(Y) = \int_{\mathbb{R}} t f_Y(t) dt = \int_0^{+\infty} t\nu e^{-\nu t} dt = -te^{-\nu t}\big|_0^{+\infty} + \int_0^{+\infty} e^{-\nu t} dt = \frac{1}{\nu}$$

$$\mathbb{E}(Y^2) = \int_{\mathbb{R}} t^2 f_Y(t) dt = -t^2 e^{-\nu t}\Big|_0^{+\infty} + \int_0^{+\infty} 2te^{-\nu t} dt = \frac{2}{\nu^2}$$

Perciò

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = \frac{1}{\nu^2}$$

La legge esponenziale descrive anche il tempo medio di attesa tra due successi successivi in un processo di Poisson.

Nell'esempio dei guasti di un apparecchio meccanico essa fornisce la probabilità del tempo di attesa tra due guasti successivi. Il valor medio $1/\nu$ viene anche detto *tempo medio tra due guasti*.

Una proprietà importante della legge esponenziale è la sua **assenza di memoria**: se aspettiamo un successo nel processo di Poisson, e dopo un tempo T non siè verificato, la probabilità di dover aspettare ancora per un tempo t è uguale a quella che avevamo in partenza. Formalmente:

$$P(Y \ge T + t | Y \ge T) = P(Y \ge t)$$

Infatti

$$P(Y \ge T + t | Y \ge T) = \frac{P(Y \ge T + t, Y \ge T)}{P(Y \ge T)} = \frac{P(Y \ge T + t)}{P(Y \ge T)} =$$
$$= \frac{e^{-\nu(T+t)}}{e^{-\nu T}} = e^{-\nu t} = P(Y \ge t)$$

7.2.4 La legge gamma

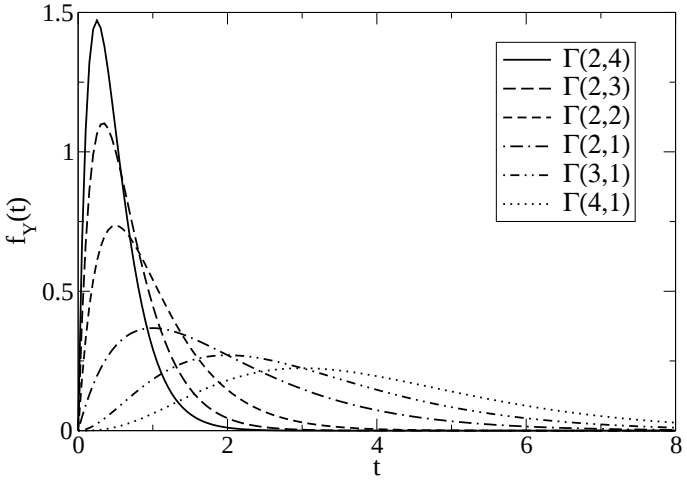
Sia X un processo di Poisson di intensità ν. Si dice **gamma** la v.a. che misura l'istante dell'n-esimo successo del processo di Poisson. Si scrive $Y \sim \Gamma(n, \nu)$.

Poiché il tempo di attesa tra due successi successivi è una v.a. di legge $\mathrm{Esp}(\nu)$, e poiché gli eventi sono indipendenti, la legge gamma è somma di n v.a. i.i.d. di legge $\mathrm{Esp}(\nu)$. La funzione di ripartizione $F_Y(t) = 1 - P(Y > t)$ si ottiene notando che l'evento $Y > t$ significa *l'n-esimo successo avviene dopo l'istante t* e coincide con l'evento *il numero di successi fino all'istante t è* $\le n - 1$. Perciò

$$F_Y(t) = 1 - P(Y > t) = 1 - P(X \le n - 1) = 1 - \sum_{k=0}^{n-1} e^{-\nu t} \frac{(\nu t)^k}{k!}$$

La densità continua è la derivata della f.d.r.:

$$f_Y(t) = \frac{d}{dt} F_Y(t) = \nu \sum_{k=0}^{n-1} e^{-\nu t} \frac{(\nu t)^k}{k!} - \nu \sum_{k=0}^{n-1} e^{-\nu t} \frac{(\nu t)^{k-1}}{(k-1)!} = \nu e^{-\nu t} \frac{(\nu t)^{n-1}}{(n-1)!}$$



Densità delle leggi $\Gamma(n, \nu)$

- Nel caso particolare $n = 1$ la legge gamma coincide con la legge esponenziale: $\Gamma(1, \nu) = \text{Esp}(\nu)$, e $f_Y(t) = \nu e^{-\nu t}$
- Il valore atteso e la varianza della legge gamma sono semplici da ottenere poiché $Y \sim \Gamma(n, \nu)$ è somma di n v.a. i.i.d. $X \sim \text{Esp}(\nu)$:

$$\mathbb{E}(Y) = n\mathbb{E}(X) = \frac{n}{\nu}, \qquad \text{Var}(Y) = n\text{Var}(X) = \frac{n}{\nu^2}$$

7.3 Quantili

In molti problemi statistici occorre ragionare in direzione inversa, ossia assegnato $\alpha \in [0, 1]$, determinare x tale che $P(X < x) = \alpha$.

Def. Si definisce **quantile di ordine** α (o quantile α-esimo) di una v.a. continua con funzione di ripartizione strettamente crescente il numero q_α dato da:

$$P(X < q_\alpha) = \alpha$$

Il quantile è dunque l'inversa della funzione di ripartizione:

$$P(X < q_\alpha) \equiv F(q_\alpha) = \alpha, \quad q_\alpha = F^{-1}(\alpha)$$

Ovviamente vale la proprietà:

$$P(X > q_{1-\alpha}) = \alpha$$

Esempio: i quantili della normale standard. Consideriamo la v.a. normale standard $X \sim \mathcal{N}(0, 1)$ di densità $f_X(t) = \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ e f.d.r. $F_X(t) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du$

Il quantile q_α è l'unico numero che soddisfa

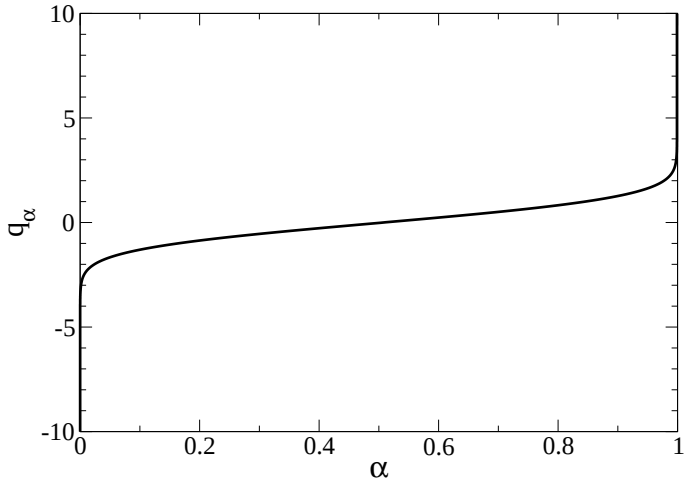
$$\Phi(q_\alpha) = P(X < q_\alpha) = \alpha$$

Per la simmetria della gaussiana rispetto a $x = 0$ si ha:

$$q_{1-\alpha} = -q_\alpha$$

Infatti $\Phi(-x) = 1 - \Phi(x)$. Applichiamo questa proprietà a $x = q_\alpha$:

$$\Phi(-q_\alpha) = 1 - \Phi(q_\alpha) = 1 - \alpha$$



Quantili della normale standard

$$\Phi^{-1}(\Phi(-q_\alpha)) = \Phi^{-1}(1-\alpha)$$

$$-q_\alpha = q_{1-\alpha}$$

Sempre per la proprietà di simmetria si ha anche:

$$P(|X| < q_{\frac{1+\alpha}{2}}) = \alpha \qquad P(|X| > q_{1-\frac{\alpha}{2}}) = \alpha$$

Alcuni quantili della normale standard sono riassunti nella seguente tabella:

α	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
q_α	1.2816	1.6449	1.96	2.3263	2.7578	3.0902	3.2905

Esempio 1. Calcoliamo il valore atteso della normale standard $X \sim \mathcal{N}(0,1)$.

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t e^{-t^2/2} dt = 0$$

poiché l'integrando è una funzione dispari.

Esempio 2. Calcoliamo il valore atteso della gaussiana $Y \sim \mathcal{N}(\mu, \sigma^2)$.

$$\mathbb{E}(Y) = \mathbb{E}(\sigma X + \mu) = \sigma \mathbb{E}(X) + \mu = \mu$$

Esempio 1. Calcoliamo la varianza della normale standard $X \sim \mathcal{N}(0,1)$.

$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^2 e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t \left[t e^{-t^2/2} \right] dt =$$

$$= \frac{-t e^{-t^2/2}}{\sqrt{2\pi}} \Bigg|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2} dt = 1$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 1$$

Esempio 2. Calcoliamo la varianza della gaussiana $Y \sim \mathcal{N}(\mu, \sigma^2)$.

$$\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \sigma^2 \mathrm{Var}(X) = \sigma^2$$

7.4 Teorema centrale del limite

Sia $\{X_i\}$, $i \geq 1$ una successione di v.a. indipendenti identicamente distribuite con valore atteso $\mathbb{E}(X_i) = \mu$ e varianza finita $\mathrm{Var}(X_i) = \sigma^2$. Sia $\overline{X}_n$ la media campionaria

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Allora posto S_N^* la *media campionaria standardizzata*

$$S_n^* = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

vale, per ogni t fissato e per $n \to \infty$:

$$S_n^* \to \mathcal{N}(0,1), \quad P(S_n^* \leq t) \to \Phi(t) = \frac{1}{2\pi}\int_{-\infty}^{t} e^{-x^2/2}dx$$

Di conseguenza possiamo affermare che

$$\overline{X}_n \to \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad P(\overline{X}_n \leq t) \to \Phi\left(\frac{\sqrt{n}(t-\mu)}{\sigma}\right)$$

$$S_n = X_1 + \cdots + X_n \to \mathcal{N}(n\mu, n\sigma^2), \quad P(S_n \leq t) \to \Phi\left(\frac{t-n\mu}{\sqrt{n}\sigma}\right)$$

L'interpretazione statistica di questo teorema è la seguente: sia $X_1, \ldots, X_n$ un campione casuale di ampiezza n estratto da una popolazione di distribuzione qualsiasi, avente valore atteso μ e varianza σ^2. Allora, al crescere di n, la media campionaria standardizzata S_n^* tende a distribuirsi con legge normale standard.

Quanto debba essere grande n affinché l'approssimazione della media campionaria con la legge normale standard sia accettabile dipende dalla legge di partenza. Una regola empirica è di richiedere che $n \geq 30$. Questo valore va aumentato se la legge di partenza è fortemente asimmetrica, e può essere diminuito se essa è fortemente simmetrica: ad es. per la legge uniforme l'approssimazione è già buona per $n = 10$.

Applicazione. Approssimazione della Binomiale. Siano $X_i \sim B(p)$ n v.a. Bernoulliane i.i.d., e sia $S_n = X_1 + \cdots + X_n \sim B(n,p)$. La media di una prova di Bernoulli è $\mu = p$ e la sua varianza $\sigma^2 = p(1-p)$.

Dal teorema centrale del limite possiamo affermare che, per n grande,

$$S_n \to \mathcal{N}(n\mu, n\sigma^2) = \mathcal{N}[np, np(1-p)]$$

$$P(S_n \leq t) \to \Phi\left(\frac{t-np}{\sqrt{np(1-p)}}\right)$$

Una buona norma è di applicare l'approssimazione normale della binomiale quando sono verificate le condizioni $np > 5$, $n(1-p) > 5$. Ricordiamo che se n è grande e p è piccolo, la binomiale può essere approssimata dalla legge di Poisson $P(np)$.

Correzione di continuità. Se la variabile aleatoria che vogliamo approssimare con la legge normale è discreta, conviene correggere la legge normale in modo da tenere conto del fatto che la funzione di probabilità è costante a tratti.

Precisamente, supponiamo che la v.a. assuma valori interi, come nel caso della binomiale. Allora $P(X \leq t)$ è costante per $t \in [k, k+1)$ con k intero. Conviene allora usare l'approssimazione

$$P(X \leq k) \simeq \Phi\left(\frac{\sqrt{n}(k+0.5-\mu)}{\sigma}\right)$$

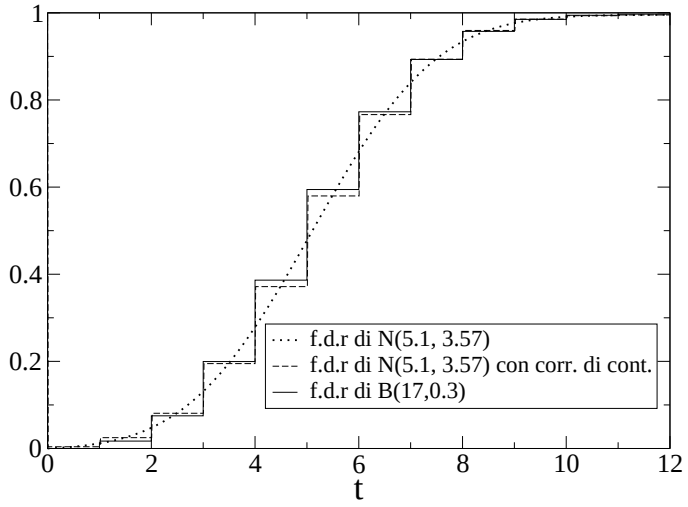


Grafico della f.d.r. della legge $B(17, 0.3)$, della sua approssimazione normale e della sua approssimazione normale con correzione di continuità.

Cap. 8. Statistica inferenziale

La distribuzione di probabilità di una v.a., dipendente da uno o più parametri θ, permette di assegnare una probabilità a qualsiasi campione. Scopo della statistica inferenziale è di procedere all'inverso, ossia a partire dai dati di un campione di una popolazione, si vuole determinare il parametro incognito θ.

Esempio. Una macchina produce componenti meccanici di dimensioni specificate con un livello di tolleranza dato. Al di fuori dei limiti di tolleranza il pezzo viene giudicato difettoso. Il produttore vuole garantire che la percentuale dei pezzi difettosi non superi il 5%. Il modello di produzione è ben rappresentato mediante un processo di Bernoulli di v.a. X_i che vale 1 se lo i-esimo pezzo è difettoso e 0 altrimenti: $X_i \sim B(p)$. Il parametro incognito è p, e si vuole stimarlo sulla base di osservazioni a campione.

8.1 Modello statistico parametrico

Def. Un **modello statistico parametrico** è una famiglia di leggi di v.a., dipendenti da uno o più parametri θ: $\{p_X(x,\theta)\}$, $\theta \in I$, $I \subseteq \mathbb{R}^n$, $n \geq 1$. La legge è nota a meno dei parametri θ.

Def. Un **campione casuale** di dimensione n estratto da una popolazione di densità $p_X(x,\theta)$ è una n-upla di v.a. $X_1, \ldots, X_n$ i.i.d. ciascuna con densità $p_X(x,\theta)$.

Def. Consideriamo una v.a. X avente densità di probabilità $p_X(x,\theta)$, dove θ rapresenta uno o più parametri. Si dice **statistica** una v.a. $T = T(X_1, X_2, \ldots, X_n)$ funzione del campione casuale $(X_1, X_2, \ldots, X_n)$.
N.B.: T **NON** deve dipendere da θ.

Ad esempio, la media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ di n v.a. è una statistica.

Def. Si dice **stimatore** del parametro θ una statistica usata per stimare θ o una sua funzione $g(\theta)$. Assegnata la funzione $T = T(X_1, X_2, \ldots, X_n)$, una volta estratto un particolare campione $(x_1, x_2, \ldots, x_n)$, il valore $\tau = T(x_1, x_2, \ldots, x_n)$ si dice **stima** di $g(\theta)$.
N.B.: lo stimatore è una variabile aleatoria, mentre la stima è un numero reale.

Proprietà degli stimatori.

- *Correttezza.* Uno stimatore T di $g(\theta)$ si dice *corretto*, o *non distorto*, se $\mathbb{E}(T) = g(\theta)$ per ogni θ.

  Uno stimatore non corretto si dice *distorto* e la quantità $E(T) - g(\theta)$ si dice *distorsione* dello stimatore.

- *Correttezza Asintotica.* Uno stimatore $T_n = T_n(X_1, X_2, \ldots, X_n)$ si dice *asintoticamente corretto* se la distorsione si annulla al crescere dell'ampiezza del campione:

$$\lim_{n\to\infty} \mathbb{E}(T_n) - g(\theta) = 0, \quad \forall\, \theta$$

- *Consistenza* Uno stimatore corretto, o asintoticamente corretto, si dice consistente se

$$\lim_{n\to\infty} P(|T_n - g(\theta)| \leq \epsilon) = 1 \quad \forall \epsilon > 0, \quad \forall\, \theta$$

o, alternativamente, se

$$Var(T_n) \to 0 \quad \text{per } n \to \infty$$

8.2 Stima puntuale

Scopo della *stima puntuale* è di utilizzare opportune statistiche per stimare i valori dei parametri incogniti della distribuzione di partenza. Vedremo nei prossimi paragrafi esempi di statistiche per stimare il valore atteso e la varianza di una distibuzione. Nella sezione successiva ci proporremo di fornire degli intervalli entro cui riteniamo che tali parametri appartengano. Questa parte della statistica inferenziale viene chiamata *stima per intervalli*.

8.2.1 Stima puntuale della media

La media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ di n v.a. i.i.d. è una statistica; in virtù della legge dei grandi numeri si può affermare che è uno stimatore corretto e consistente del valore atteso $\theta = \mathbb{E}(X_i)$.

Esempi. Nell'esempio visto in precedenza della produzione di componenti meccanici, il modello statistico parametrico è la famiglia di leggi di Bernoulli $B(p)$, p è il parametro da determinare; La media campionaria $T_n = \overline{X}_n$ è uno stimatore non distorto e consistente di p.

Per una v.a. normale $\mathcal{N}(\mu, \sigma^2)$ i parametri sono $\theta = (\mu, \sigma^2)$. La media campionaria è uno stimatore non distorto e consistente di μ.

Lo stesso discorso si applica alle altre v.a. notevoli:
Per la binomiale $X \sim B(n,p)$, $\theta = (n,p)$, la media campionaria è uno stimatore di $\mathbb{E}(X) = np$.
Per l'esponenziale $X \sim Exp(\lambda)$, $\theta = \lambda$, la media campionaria è uno stimatore di $\mathbb{E}(X) = 1/\lambda$.
Per il modello Gamma $X \sim \Gamma(n,\lambda)$, $\theta = (n,\lambda)$, la media campionaria è uno stimatore di $\mathbb{E}(X) = n/\lambda$.

Date n v.a. i.i.d. X_i, la v.a. $X_n = \sum_{i=1}^n \lambda_i X_i$ con $\sum_{i=1}^n \lambda_i = 1$ è uno stimatore corretto del valore atteso $\theta = \mathbb{E}(X_i)$. È anche consistente se $\sum_{i=1}^n \lambda_i^2 \to 0$ per $n \to \infty$.

8.2.2 Stima puntuale della varianza

Siano n variabili aleatorie i.i.d. X_i aventi ciascuna media μ e varianza σ^2.

Consideriamo la statistica

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$$

Calcoliamo il valor medio di $\hat{S}_n^2$:

$$E(\hat{S}_n^2) = \frac{1}{n}\sum_{i=1}^n E[(X_i - \overline{X}_n)^2]$$

Per completare il calcolo riscriviamo $\hat{S_n}^2$ come

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n [(X_i - \mu) - (\overline{X}_n - \mu)]^2$$

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 + (\overline{X}_n - \mu)^2 - 2(\overline{X}_n - \mu)\frac{1}{n}\sum_{i=1}^n (X_i - \mu)$$

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - (\overline{X}_n - \mu)^2$$

$$E(\hat{S}_n^2) = E\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}_n-\mu)^2\right\}$$

$$E(\hat{S}_n^2) = \frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2] - E[(\overline{X}_n-\mu)^2]$$

$$E(\hat{S}_n^2) = \frac{1}{n}n\sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2$$

La statistica $\hat{S}_n$ è dunque uno stimatore distorto della varianza σ^2; la distorsione è $E(\hat{S}_n^2) - \sigma^2 = -\sigma^2/n$ e quindi $\hat{S}_n^2$ è uno stimatore asintoticamente corretto.

È evidente però, che il calcolo precedente porta a definire facilmente uno stimatore corretto di σ^2 come:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$$

Per questo nuovo stimatore infatti vale

$$E(S_n^2) = \sigma^2$$

La v.a. S_n^2 prende il nome di **varianza campionaria**. Si può dimostrare con calcoli facili ma noiosi che S_n^2 è uno stimatore consistente :

$$\lim_{n\to\infty}\mathrm{Var}(S_n^2) = 0$$

purché $E[(X_i-\mu)^4] < \infty$

In pratica, una volta estratto un particolare campione $(x_1, x_2, \ldots, x_n)$, si ottiene il valore corrispondente di s_n^2:

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

ossia s_n^2 è la varianza campionaria dei dati $x_1, \ldots, x_n$.

Osservazione. Se è noto il valore atteso $\mathbb{E}(X_i) = \mu$ della v.a. X_i, allora per stimare la varianza si può usare la statistica seguente:

$$T_n = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

Attenzione: T_n è una statistica solo se il valore atteso $\mathbb{E}(X_i) = \mu$ è noto; altrimenti μ è un parametro incognito e T_n non è più una statistica.

Dimostriamo che T_n è uno stimatore di σ^2:

$$\mathbb{E}(T_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i-\mu)^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2 - 2\mu X_i + \mu^2) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}(X_i^2) - 2\mu\mathbb{E}(X_i) + \mu^2\right\} = \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}(X_i^2) - \mu^2\right\} = \sigma^2$$

Esempio. Vogliamo stimare i parametri r e λ di una popolazione con distribuzione $\Gamma(r,\lambda)$. Effettuiamo un campionamento e consideriamo gli stimatori $\overline{X}_n$ e S_n^2. I loro valori attesi sono rispettivamente $\mathbb{E}(\overline{X}_n) = r/\lambda$ e $\mathbb{E}(S_n^2) = r/\lambda^2$. Dunque $\overline{X}_n$ e S_n^2 sono stimatori non distorti dei parametri r/λ e r/λ^2. Nella pratica, a campionamento effettuato otteniamo i valori $\overline{x}_n$ e s_n^2. Risolvendo per r e λ otteniamo le stime $\hat{\lambda} = \overline{x}_n/s_n^2$ e $\hat{r} = \overline{x}_n^2/s_n^2$.
Si può anche dire che $\overline{X}_n/S_n^2$ e $\overline{X}_n^2/S_n^2$ sono stimatori (distorti!) rispettivamente di λ ed r.

8.3 Stima per intervalli

Uno stimatore T, come ad esempio $\overline{X}_n$, fornisce, *a campionamento eseguito*, una stima del valore di θ del quale è però ignota l'accuratezza. Descriviamo questa proprietà degli stimatori dicendo che forniscono una stima puntuale del/dei parametro(i) incogniti. Se lo stimatore è asintoticamente corretto e consistente $E(T)$ darà una stima sempre più accurata al crescere dell'ampiezza del campione; tuttavia non sempre è possibile aumentare n. È necessario quindi un metodo per ottenere dal campione stesso anche una stima dell'accuratezza della stima puntuale. Questo metodo consiste nella costruzione di un intervallo, detto **intervallo di confidenza o intervallo fiduciario**, che verosimilmente contiene il valore vero del parametro incognito. In tale ottica, parliamo di stima per *intervalli* di θ.

8.4 Campionamento da una popolazione normale

Per stimare i parametri di una distribuzione normale, è utile definire alcune distribuzioni continue.

8.4.1 Legge chi-quadrato

Def. Si dice **legge chi-quadrato con n gradi di libertà**, la legge della variabile aleatoria

$$Y = \sum_{i=1}^{n} X_i^2,$$

dove X_i sono n v.a. indipendenti, ciascuna di legge $\mathcal{N}(0,1)$. Si scrive $Y \sim \chi^2(n)$.

Come vedremo, la legge chi-quadrato è utile per stimare la varianza di una popolazione normale.

Proprietà. Si dimostra che la legge $\chi^2(n)$ coincide con la legge gamma di parametri $n/2$, $1/2$: $\chi^2(n) = \Gamma(n/2, 1/2)$. Da questa proprietà si possono ricavare molte informazioni sulla legge chi-quadrato:

- La funzione densità è:

  $$f_Y(t) = c_n t^{n/2-1} e^{-t/2} \quad \text{per } t > 0, \quad f_Y(t) = 0 \quad \text{per } t \leq 0$$

  Il valore di c_n viene ottenuto imponendo che $\int_0^{+\infty} c_n t^{n/2-1} e^{-t/2} = 1$.

- il valore atteso: ricordando la proprietà che il valore atteso della $\Gamma(r, \nu)$ è pari a r/ν,

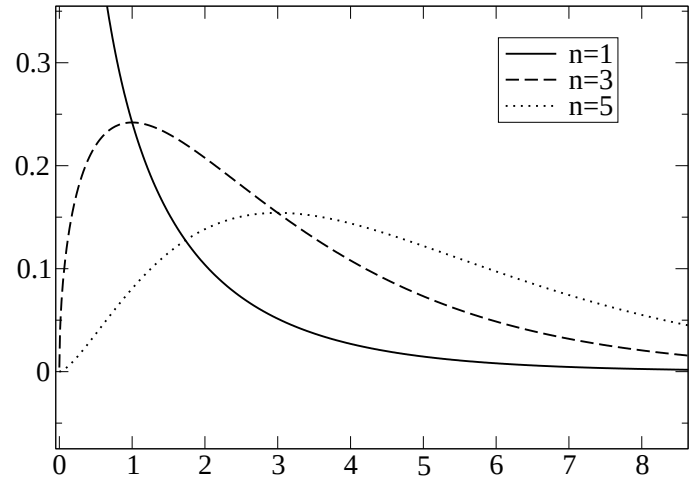
  $$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^{n} X_i^2\right) = \frac{n/2}{1/2} = n$$

- la varianza: ricordando la proprietà che la varianza della $\Gamma(r, \nu)$ è pari a r/ν^2,vale:

  $$\mathrm{Var}(Y) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i^2\right) = \frac{n/2}{1/4} = 2n$$

- Se Y_1 e Y_2 sono v.a. indipendenti con leggi rispettive $Y_1 \sim \chi^2(n)$, $Y_2 \sim \chi^2(m)$, allora

  $$Y_1 + Y_2 \sim \Gamma\left(\frac{n}{2} + \frac{m}{2}, \frac{1}{2}\right) = \chi^2(n+m)$$



Funzione densità della v.a. chi-quadrato $\chi^2(n)$ per alcuni valori di n.

- Per n grande, il teorema centrale del limite porta a concludere che

$$Y \sim \chi^2(n) \to \mathcal{N}(n, 2n), \quad P(Y \le t) \simeq \Phi\left(\frac{t-n}{\sqrt{2n}}\right)$$

- Indichiamo con $\chi^2_\alpha(n)$ i quantili della legge chi-quadrato:

$$P(Y \le \chi^2_\alpha(n)) = \alpha$$

  I valori dei quantili sono tabulati per i primi valori di n. Per n grande ($n > 30$) si possono determinare i quantili da quelli della normale, sfruttando l'approssimazione normale:

$$\Phi\left(\frac{t-n}{\sqrt{2n}}\right) = \alpha, \quad q_\alpha \simeq \frac{\chi^2_\alpha(n) - n}{\sqrt{2n}}, \quad \chi^2_\alpha(n) \simeq q_\alpha\sqrt{2n} + n$$

  Un'approssimazione leggermente migliore di questa, valida sempre per $n > 30$, è

$$\chi^2_\alpha(n) \simeq \frac{1}{2}\left(q_\alpha + \sqrt{2n-1}\right)^2$$

L'importanza della legge chi-quadrato è dovuta alle seguenti proprietà:
Siano $X_1, X_2, \ldots, X_n$, n v.a. normali i.i.d. di legge $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Allora

- La somma delle standardizzate vale

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

  Questa proprietà discende direttamente dalla definizione della legge chi-quadrato come somma di v.a. normali standard indipendenti.

- Se $\overline{X}_n$ è la media campionaria,

$$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}_n}{\sigma}\right)^2 \sim \chi^2(n-1)$$

  la media μ viene sostituita con la media campionaria $\overline{X}_n$, e la v.a. trovata ha legge chi-quadrato con un grado di libertà in meno.

Non dimostreremo questa proprietà. Intuitivamente si può capire che le n v.a. $X_i - \overline{X}_n$ non sono più indipendenti, poiché la loro somma è nulla. Questa relazione sottrae un grado di libertà alla somma dei loro quadrati.

In termini della varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, la formula precedente si può riscrivere come

$$\frac{(n-1)}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

- Si dimostra infine che la varianza campionaria S_n^2 e la media campionaria $\overline{X}_n$ sono v.a. tra loro indipendenti. Notiamo che questa proprietà non è semplice da dimostrare e non vale in generale per una legge qualsiasi.

Esempio. Una ditta produce bulloni del diametro medio di $2cm$. Dall'esperienza passata è noto che la deviazione standard del loro diametro è di $0.1cm$. Si può supporre inoltre che il diametro effettivo di un bullone abbia una distribuzione normale. Una seconda ditta intende comprare una partita di bulloni ma non crede ai parametri forniti dalla prima ditta sul valor medio e sulla varianza, e pone come requisito che la varianza campionaria di 20 pezzi scelti a caso non superi $(0.12cm)^2$. Qual'è la probabilità che la partita venga scartata?

Risposta. Applichiamo la formula $(n-1)S_n^2/\sigma^2 \sim \chi^2(n-1)$ con $n = 20$, $\sigma = 0.1cm$. Poniamo $Y \sim \chi^2(19)$.

$$P\left(S_n^2 > (0.12cm)^2\right) = P\left(Y > \frac{19 * 0.12^2}{0.1^2}\right) = P(Y > 27.36) \simeq 0.1$$

Il valore di $P(Y > 27.36)$ è stato ricavato dalle tavole.

8.4.2 Legge t di Student

La legge t di Student è utile per stimare il valor medio di una popolazione normale quando non sia nota la varianza.

Def. Si dice **Legge t di Student con n gradi di libertà**, la legge di una v.a.

$$T = \frac{Z}{\sqrt{Y/n}}, \quad \text{dove } Z \sim \mathcal{N}(0,1),\ Y \sim \chi^2(n)$$

e si richiede che Z e Y siano indipendenti. Si usa scrivere $T \sim t(n)$.

Si può calcolare esplicitamente la densità della $t(n)$:

$$f_T(t) = c_n \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

il coefficiente di normalizzazione c_n si ricava imponendo che $\int_{-\infty}^{+\infty} f_T(t)dt = 1$.

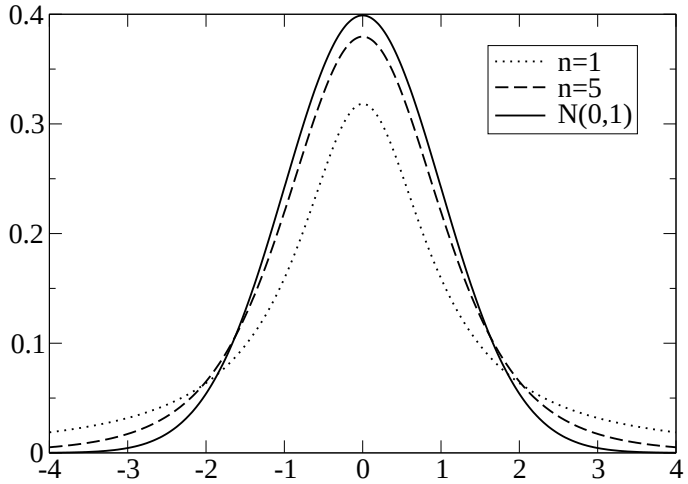
Proprietà.

- Per $n \to \infty$, la legge $t(n)$ tende alla normale standard $\mathcal{N}(0,1)$. Infatti è facile stabilire che

$$\lim_{n\to\infty} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} = e^{-t^2/2}$$

- La densità della $t(n)$ è una funzione simmetrica pari, perciò il valore atteso è nullo, tranne che per $n = 1$, per il quale non esiste.

  La varianza vale $n/(n-2)$ per $n > 2$: è sempre maggiore di uno, e tende a 1 per $n \to \infty$.



Funzione densità della v.a. t di Student $t(n)$ e confronto con la normale standard $\mathcal{N}(0,1)$

- Indichiamo con $t_\alpha(n)$ i quantili della legge $t(n)$:

  $$P(T \le t_\alpha(n)) = \alpha$$

  Per la simmetria della funzione densità, valgono le seguenti proprietà, del tutto simili a quelle relative ai quantili della normale:

  $$P(T \ge t_{1-\alpha}(n)) = \alpha, \quad P(|T| \ge t_{1-\frac{\alpha}{2}}(n)) = \alpha, \quad P(|T| \le t_{\frac{1+\alpha}{2}}(n)) = \alpha$$

  Per valori di n maggiori di 120 si possono approssimare i quantili della $t(n)$ con quelli della normale standard. Per n minore di 120 i valori si ricavano dalle tavole.

L'importanza della legge t di Student è dovuta alla seguente proprietà: siano X_1, X_2, ..., X_n, n v.a. normali i.i.d. di legge $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Allora

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

Infatti, sappiamo che $\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, e dunque

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

D'altra parte, abbiamo visto in precedenza che

$$\frac{(n-1)}{\sigma^2} S_n^2 \sim \chi^2(n-1)$$

Essendo S_n^2 e $\overline{X}_n$ indipendenti, otteniamo

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{S_n^2/\sigma^2}} \sim t(n-1)$$

Esempio. La ditta che vuole decidere se comprare la partita di bulloni dell'esempio precedente, procede a una misurazione a campione di 50 bulloni, e trova che il diametro medio del campione è di $2.04cm$ con una deviazione standard campionaria di $0.15cm$. Supponendo

ancora che il diametro dei bulloni segua una legge normale, calcolare la probabilità che il valore medio differisca di meno di $0.1cm$ dal valore dichiarato di $2cm$.

Soluzione. Si considera la v.a.

$$T_n = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$$

dove S_n è la deviazione standard campionaria. Per quanto visto prima T_n ha distribuzione t di Student con $n-1$ gradi di libertà. Dunque

$$P\left(t_\alpha < \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} < t_\beta\right) = \beta - \alpha$$

$$P(\overline{X}_n - t_\beta S_n/\sqrt{n} < \mu < \overline{X}_n - t_\alpha S_n/\sqrt{n}) = \beta - \alpha$$

Imponiamo

$$\overline{X}_n - t_\beta S_n/\sqrt{n} = 1.99, \quad \overline{X}_n - t_\alpha S_n/\sqrt{n} = 2.01$$

$$t_\beta = (\overline{X}_n - 1.99)\sqrt{n}/S_n \simeq 2.357, \quad t_\alpha = (\overline{X}_n - 2.01)\sqrt{n}/S_n \simeq 1.414$$

Dalle tavole risulta $\beta \simeq 0.9888$, $\alpha \simeq 0.9182$. Pertanto la probabilità cercata è $\beta - \alpha \simeq 0.07$.

8.5 Intervalli di confidenza

Sia $\{X_1, \ldots, X_n\}$ un campione aleatorio estratto da una popolazione di densità $p_X(x, \theta)$. Siano $T_1 = t_1(X_1, \ldots, X_n)$ e $T_2 = t_2(X_1, \ldots, X_n)$ due statistiche, e sia $g(\theta)$ una funzione del (dei) parametro(i) θ. Fissato $\alpha \in [0, 1]$, l'intervallo aleatorio (T_1, T_2) si dice **intervallo di confidenza per $g(\theta)$, al livello del** $100\alpha\%$ se

$$p(T_1 < g(\theta) < T_2) = \alpha$$

A campionamento eseguito, l'intervallo ottenuto $[t_1(x_1, \ldots, x_n), t_2(x_1, \ldots, x_n)]$ si chiama intervallo di confidenza per $g(\theta)$, al livello del $100\alpha\%$, calcolato dal campione. Questo intervallo perde il significato di probabilità: **non** è vero che la probabilità che $g(\theta)$ sia compresa tra $t_1(x_1, \ldots, x_n)$ e $t_2(x_1, \ldots, x_n)$ è pari ad α. Per questo motivo si parla di confidenza e non di probabilità.
È vero invece che se effettuassimo numerosi campionamenti e calcolassimo per ciascuno di questi l'intervallo di confidenza allo stesso livello, ci aspettiamo che una proporzione del $100\alpha\%$ degli intervalli contenga il valore di $g(\theta)$.

8.5.1 Intervalli di confidenza per la media

Utilizzando i risultati descritti sopra ci proponiamo ora di costruire gli *intervalli fiduciari per la media* nei due casi in cui rispettivamente la varianza sia nota e la varianza sia incognita.

- *Intervallo fiduciario per la media di una popolazione con varianza nota.*

  Consideriamo un campione casuale $(X_1, X_2, \ldots, X_n)$ di ampiezza n estratto da una popolazione avente valor medio μ incognito e varianza σ^2 nota. Lo stimatore per μ è la media campionaria $\overline{X}_n$ per la quale supporremo che:

  $$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

  Questa relazione è esatta se la legge della popolazione è normale, mentre vale solo asintoticamente per $n \to \infty$ altrimenti (teorema centrale del limite).

La standardizzata di $\overline{X}_n$

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

è distribuita secondo la normale standard

$$Z_n \sim \mathcal{N}(0,1).$$

Fissato il livello di confidenza α, possiamo affermare che

$$P\left(\frac{|\overline{X}_n - \mu|}{\sigma/\sqrt{n}} \leq q_{\frac{1+\alpha}{2}}\right) = \alpha$$

ovvero

$$P(|\overline{X}_n - \mu| \leq q_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = \alpha$$

I due valori estremi di μ sono

$$\mu_{\pm} = \overline{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{\frac{1+\alpha}{2}}$$

Abbiamo dunque costruito un intervallo *casuale*, centrato su $\overline{X}_n$, di ampiezza nota $2q_{(1+\alpha)/2}\sigma/\sqrt{n}$. Tale intervallo casuale ha probabilità α di contenere il valore vero μ. Una volta eseguito il campionamento ed ottenuta la stima del valor medio $\overline{x}_n$ si ottiene l'intervallo di confidenza

$$[\overline{x}_n - q_{\frac{1+\alpha}{2}}\sigma/\sqrt{n}, \overline{x}_n + q_{\frac{1+\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$$

N.B. Non si può però affermare che esso contiene μ con probabilità α.

Osservazione. Si noti che l'ampiezza dell'intervallo fiduciario, fissati σ ed n, è tanto più grande quanto maggiore è il livello di fiducia poiché $q_{(1+\alpha)/2} \to \infty$ per $\alpha \to 1$. Pertanto innalzare il livello fiduciario aumenta il margine di errore su μ. Se si vuole mantenere un margine di errore prefissato, e nel contempo un livello fiduciario elevato, è necessario aumentare n; si noti come l'ampiezza $2q_{(1+\alpha)/2}\sigma/\sqrt{n}$ decresca come $1/\sqrt{n}$; quindi per diminuire l'errore di un ordine di grandezza è necessario aumentare l'ampiezza del campione di due ordini di grandezza.

- *Intervallo fiduciario per la media di una popolazione con varianza incognita.*

  Consideriamo come nel caso precedente un campione casuale $(X_1, X_2, \ldots, X_n)$ di ampiezza n estratto da una popolazione $f_X(x,\theta)$ con legge normale. Costruiamo la v.a.

$$T_n = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

  Fissato il livello fiduciario α abbiamo:

$$P(|T_n| \leq t_{\frac{1+\alpha}{2}}(n-1)) = \alpha$$

$$P\left(\frac{|\overline{X}_n - \mu|}{S_n/\sqrt{n}} \leq t_{\frac{1+\alpha}{2}}(n-1)\right) = \alpha$$

$$P\left(|\overline{X}_n - \mu| \leq t_{\frac{1+\alpha}{2}}(n-1)\frac{S_n}{\sqrt{n}}\right) = \alpha$$

  I due valori estremi di μ sono

$$\mu_{\pm} = \overline{X}_n \pm \frac{S_n}{\sqrt{n}} t_{\frac{1+\alpha}{2}}(n-1)$$

  Osservazioni

- L'intervallo fiduciario anche in questo caso è centrato su $\overline{X}_n$; la sua ampiezza $2t_{\frac{1+\alpha}{2}}(n-1)S_n/\sqrt{n}$, però, non è più nota a priori, ma è a sua volta una v.a..
- Fissato un livello fiduciario α, l'ampiezza dell'intervallo e quindi l'errore nella stima di μ tende a zero per $n \to \infty$, poiché S_n^2 è uno stimatore consistente di σ^2.

Esempio. Un laboratorio di analisi controlla il quantitativo medio di catrame contenuto in una certa marca di sigarette. In un campione di 30 sigarette si trovano i seguenti valori per la media campionaria $\overline{x}_n$ e la deviazione standard campionaria s_n:

$$\overline{x}_n = 10.92mg, \qquad s_n = 0.51mg$$

Si determini l'intervallo fiduciario per il quantitativo medio di catrame al livello del 99%

Soluzione.

$$\alpha = 0.99 \qquad \frac{1+\alpha}{2} = 0.995 \qquad t_{0.995}(29) \simeq 2.756$$

Gli estremi dell'intervallo sono

$$10.92 - 2.756\frac{0.51}{\sqrt{30}}, \qquad 10.92 + 2.756\frac{0.51}{\sqrt{30}}$$

$$10.92 - 0.25, \qquad 10.92 + 0.25$$

Si noti che se avessimo considerato la deviazione standard campionaria come il valore vero ed avessimo considerato il quantile $q_{0.995} \simeq 2.33$ avremmo trovato un intervallo fiduciario leggermente più stretto:

$$10.92 \pm \frac{0.51}{\sqrt{50}}2.33 \simeq 10.92 \pm 0.17$$

8.5.2 Intervalli di confidenza per la varianza

Ci proponiamo di costruire gli *intervalli fiduciari per la varianza* nei due casi in cui rispettivamente il valor medio sia noto e il valor medio sia incognito.

- *Intervallo fiduciario per la varianza di una popolazione con media nota.*

  Partiamo come prima da un campione casuale $(X_1, X_2, \ldots, X_n)$ di ampiezza n estratto da una popolazione con legge normale $\mathcal{N}(\mu, \sigma^2)$.

  Essendo μ nota, la v.a.

  $$T_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

  è una statistica. Si ha poi

  $$\frac{nT_n^2}{\sigma^2} \sim \chi^2(n)$$

  Nei casi visti in precedenza del calcolo di intervalli di confidenza per il valor medio si procedeva a costruire un intervallo centrato sull'origine e tale che la probabilità che la v.a. standardizzata appartenga a detto intervallo sia α. Qui il procedimento va modificato perché la densità della legge chi-quadrato non è simmetrica rispetto all'origine.

  Si possono adottare due punti di vista: quello di maggiorare la varianza (è il caso ad esempio in cui si voglia avere una stima degli errori sperimentali), e quello in cui si voglia costringere la varianza all'interno di un intervallo (è il caso in cui si vuole determinare il valore esatto della varianza).

Nel primo caso otteniamo una maggiorazione sul valore della varianza imponendo che

$$P\left(\frac{nT_n^2}{\sigma^2} \le \chi^2_\alpha(n)\right) = \alpha \quad \Rightarrow \quad P\left(\frac{nT_n^2}{\sigma^2} \ge \chi^2_{1-\alpha}(n)\right) = \alpha$$

$$P\left(\sigma^2 \le \frac{nT_n^2}{\chi^2_{1-\alpha}(n)}\right) = \alpha$$

Otteniamo in definitiva che

$$\sigma^2 \in \left[0, \frac{nT_n^2}{\chi^2_{1-\alpha}(n)}\right]$$

Nel secondo caso, poniamo $Y \sim \chi^2(n)$. È ragionevole considerare un intervallo $[a,b]$, $0 < a < b$, tale che $P(a < Y < b) = \alpha$, e inoltre che $P(Y < a) = P(Y > b)$ (ossia le code hanno uguale probabilità):

$$P(Y < a) = P(Y > b) = \frac{1-\alpha}{2}$$

Si ricava:

$$a = \chi^2_{\frac{1-\alpha}{2}}(n), \qquad b = \chi^2_{\frac{1+\alpha}{2}}(n)$$

Con questa scelta dell'intervallo possiamo scrivere

$$P\left(\chi^2_{\frac{1-\alpha}{2}}(n) \le \frac{nT_n^2}{\sigma^2} \le \chi^2_{\frac{1+\alpha}{2}}(n)\right) = \alpha$$

ovvero

$$P\left(\frac{nT_n^2}{\chi^2_{\frac{1+\alpha}{2}}(n)} \le \sigma^2 \le \frac{nT_n^2}{\chi^2_{\frac{1-\alpha}{2}}(n)}\right) = \alpha$$

Il valore esatto σ^2 della varianza ha probabilitá α di essere contenuto nell'intervallo aleatorio

$$\left[\frac{nT_n^2}{\chi^2_{\frac{1+\alpha}{2}}(n)}, \frac{nT_n^2}{\chi^2_{\frac{1-\alpha}{2}}(n)}\right]$$

- *Intervallo fiduciario per la varianza di una popolazione con media incognita.*

  Partiamo sempre da un campione casuale $(X_1, X_2, \ldots, X_n)$ di ampiezza n estratto da una popolazione con legge normale $\mathcal{N}(\mu, \sigma^2)$.

  La v.a. T_n^2 che abbiamo usato prima *non* è una statistica poiché è funzione della media che non è nota.
  La v.a. varianza campionaria invece è una statistica:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

  Come nel caso precedente, distinguiamo i due casi in cui vogliamo rispettivamente maggiorare la varianza e costringerla in un intervallo opportuno.

Nel primo caso imponiamo che

$$P\left(\frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_\alpha(n-1)\right) = \alpha \quad \Rightarrow \quad P\left(\frac{(n-1)S_n^2}{\sigma^2} \geq \chi^2_{1-\alpha}(n-1)\right) = \alpha$$

$$P\left(\sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{1-\alpha}(n-1)}\right) = \alpha$$

L'intervallo di confidenza della varianza con livello α è dunque

$$\left[0, \frac{(n-1)S_n^2}{\chi^2_{1-\alpha}(n-1)}\right]$$

Nel secondo caso poniamo $Y \sim \chi^2(n-1)$ e consideriamo un intervallo di confidenza $[a,b]$ tale che

$$P(Y < a) = P(Y > b) = \frac{1-\alpha}{2}$$

$$P\left(\chi^2_{\frac{1-\alpha}{2}}(n-1) \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{\frac{1+\alpha}{2}}(n-1)\right) = \alpha$$

ovvero

$$P\left(\frac{(n-1)S_n^2}{\chi^2_{\frac{1+\alpha}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{\frac{1-\alpha}{2}}(n-1)}\right) = \alpha$$

Il valore esatto σ^2 della varianza ha probabilitá α di essere contenuto nell'intervallo aleatorio

$$\left[\frac{(n-1)S_n^2}{\chi^2_{\frac{1+\alpha}{2}}(n-1)}, \frac{(n-1)S_n^2}{\chi^2_{\frac{1-\alpha}{2}}(n-1)}\right]$$

Cap. 9. Test d'ipotesi

9.1 Definizioni

9.1.1 Ipotesi statistica

Sia $f_X(x, \theta)$ la densità di probabilità di una certa popolazione statistica, θ uno o più parametri tutti o in parte incogniti.
Si dice **ipotesi statistica** un'asserzione sul valore vero dei parametri θ incogniti.
Un'ipotesi statistica si dice **semplice** se specifica completamente la legge $f_X(x, \theta)$, altrimenti si dice **composta**.

Esempi. Supponiamo che una certa grandezza sia distribuita normalmente secondo la legge $\mathcal{N}(\mu, 4)$, allora

- L'ipotesi $\mu = 5$ *è semplice* perché specifica completamente la normale.
- L'ipotesi $\mu \leq 3$ *è composta* perché non specifica completamente la normale.

In generale, il valore di θ varia all'interno di un insieme Θ detto **spazio dei parametri** e un'ipotesi su θ ha la forma

$$\theta \in \Theta_0$$

dove Θ_0 è un sottoinsieme di Θ. Se l'ipotesi è semplice, allora Θ_0 si riduce ad un punto.

L'ipotesi che intendiamo sottoporre a verifica si dice **ipotesi nulla** e si indica con H_0. H_0 viene ritenuta *vera fino a prova contraria*. Nella costruzione di un test, si sceglie come H_0 l'ipotesi alla quale si è disposti a rinunciare solo in caso di forte evidenza del contrario.

Esempi.

- $H_0 : \mu = 4$
- $H_0 : \mu > 5$
- $H_0 : 0.1 \leq p \leq 0.7$

L'ipotesi complementare ad H_0, $\theta \in \overline{\Theta}_0$, si dice **ipotesi alternativa** e si indica con H_1. H_1 è vera se e solo se H_0 è falsa.

Esempi.

- $H_0 : \mu = 4, H_1 : \mu \neq 4$
- $H_0 : \mu > 5, H_1 : \mu \leq 5$
- $H_0 : 0.1 \leq p \leq 0.7, \quad H_1 : p < 0.1, p > 0.7$

9.1.2 Verifica d'ipotesi

Si dice **verifica**, o *test* il procedimento con cui si decide, sulla base di una stima ottenuta dai dati campionari, se accettare o meno l'ipotesi.

Ad esempio, se l'ipotesi nulla fosse $H_0 : \mu = 4$ per una distribuzione normale $\mathcal{N}(\mu, 1)$, si potrebbe pensare di usare come stimatore la media campionaria di un campione. Non sarebbe però ragionevole richiedere che il valore della media campionaria ottenuto sia esattamente uguale a 4, perché entrano in gioco le fluttuazioni statistiche. È più sensato richiedere che il valore medio si situi in un intorno opportunamente piccolo del valore 4.

Nella esecuzione di un test si possono avere i seguenti esiti

- H_0 *è vera* ed il test la *accetta*. La decisione è corretta.
- H_0 *è vera* ed il test la *rifiuta*. In questo caso si commette un errore di **I tipo**.
- H_0 *è falsa* ed il test la *accetta*. In questo caso si commette un errore di **II tipo**.
- H_0 *è falsa* ed il test la *rifiuta*. La decisione è corretta.

Riassumendo:

	H_0 è vera	H_0 è falsa
Rifiutiamo H_0	*Errore del I tipo*	Decisione corretta
Accettiamo H_0	Decisione corretta	*Errore del II tipo*

L'errore del I tipo è considerato più grave di quello del II tipo.

Esempio 1. Consideriamo il processo ad un imputato. Formuliamo prima l'ipotesi H_0 *l'imputato è colpevole*. Otteniamo la tabella seguente:

	H_0: L'imputato è colpevole	H_1: L'imputato è innocente
Viene assolto	*Errore del I tipo*	Decisione corretta
Viene condannato	Decisione corretta	*Errore del II tipo*

Mentre se assumessimo come ipotesi H_0: *l'imputato è innocente* otterremmo:

	H_0: L'imputato è innocente	H_1: L'imputato è colpevole
Viene condannato	*Errore del I tipo*	Decisione corretta
Viene assolto	Decisione corretta	*Errore del II tipo*

Ritenendo che sia più grave condannare un innocente rispetto a lasciare un colpevole in libertà, dobbiamo scegliere come ipotesi nulla H_0 la seconda: *l'imputato è innocente*.

Esempio 2. Due persone giocano con un dado. Una delle due persone ha il sospetto che il dado sia truccato. Decide di effettuare un gran numero di lanci e di registrare il numero di volte in cui esce il 6. L'ipotesi nulla è $H_0 : p(6) = 1/6$ (ipotesi *innocentista*). Il test d'ipotesi sarà del tipo: *rifiuto* H_0 *se* $|\overline{X}_n - 1/6| > k$ (per un valore opportuno di k), dove $\overline{X}_n$ è la media campionaria delle v.a. Bernoulliane $X_i \sim B(1/6)$ che valgono 1 se all'i-esimo lancio del dado è venuto il 6.

Esempio 3. Una ditta produce bicchieri con spessore medio alla base dichiarato di $4mm$. Una seconda ditta prima di decidere se comprarne una grossa partita vuole effettuare delle misurazioni su un campione. Essendo importante che lo spessore abbia una valore minimo garantito (per ragioni di robustezza dei bicchieri), formula l'ipotesi H_0 : *lo spessore medio della base è almeno pari a quello dichiarato*. il test d'ipotesi sarà: *rifiuto* H_0 *se* $\overline{X}_n < k$ per un opportuno valore di k.

9.1.3 Regione critica

Fissata l'ipotesi nulla H_0, il test d'ipotesi consiste nello scegliere una statistica appropriata $T(X_1, \ldots, X_n)$, e nello stabilire una regola di decisione per accettare o rifiutare l'ipotesi. Precisamente, adottiamo la seguente *regola di decisione*: *si rifiuti* H_0 *se* $T(X_1, \ldots, X_n) \in I$, dove I è un intervallo o un insieme numerico. Allora l'insieme $\mathcal{R}$ delle realizzazioni campionarie che portano a rifiutare H_0, cioè

$$\mathcal{R} = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) \in I\}$$

è detta *regione critica*, o *di rifiuto* del test.

Esempio. H_0 è l'ipotesi dell'esempio 3: *lo spessore medio della base del bicchiere è almeno pari a quello dichiarato*. La statistica è la media campionaria $\overline{X}_n$; fissiamo k e stabiliamo la regola di decisione si rifiuti H_0 se $\overline{X}_n < k$; l'insieme $\mathcal{R}$ è l'insieme dei possibili risultati campionari che forniscono una media campionaria nella regione critica:

$$\mathcal{R} = \{(x_1, \ldots, x_n) : \overline{x}_n < k\}$$

Una volta definita la regione critica, si può pensare di calcolare, per ogni valore possibile del parametro incognito $\theta \in \Theta_0$, la probabilità che l'ipotesi venga rifiutata:

$$\pi(\theta) = P_\theta \left(T(X_1, \ldots, X_n) \in I\right), \quad \theta \in \Theta_0$$

Si può anche scrivere

$$\pi(\theta) = P_\theta \left(T(X_1, \ldots, X_n) \in I \,|\, \theta \in \Theta_0\right)$$

Nota. Se l'ipotesi H_0 è semplice, questo calcolo si riduce a un valore (θ può assumere un solo valore θ_0).

Nell'esempio precedente, supponiamo che lo spessore della base dei bicchieri segua una legge normale con varianza σ^2 nota; il parametro incognito è μ, e

$$\pi(\mu) = P_\mu(\overline{X}_n < k)$$

Si può calcolare esplicitamente $\pi(\mu)$, ricordando che $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$:

$$P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < q_{\alpha(\mu)}\right) = \alpha(\mu)$$

$$P\left(\overline{X}_n < \mu + q_{\alpha(\mu)} \frac{\sigma}{\sqrt{n}}\right) = \alpha(\mu)$$

Poniamo $\mu + q_{\alpha(\mu)} \frac{\sigma}{\sqrt{n}} = k$:

$$q_{\alpha(\mu)} = \frac{(k-\mu)\sqrt{n}}{\sigma}, \qquad \pi(\mu) = \alpha(\mu) = \Phi\left(\frac{(k-\mu)\sqrt{n}}{\sigma}\right)$$

9.1.4 Livello di significatività

Definiamo ora il **livello di significatività**, anche detto **ampiezza del test**. Consideriamo il problema di verifica dell'ipotesi

$$H_0 : \theta \in \Theta_0$$

contro

$$H_1 : \theta \notin \Theta_0$$

L'*ampiezza*, o *livello di significatività* α del test basato su un campione di dimensione n con regione critica

$$\mathcal{R} = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) \in I\}$$

è definito come

$$\alpha = \sup_{\theta \in \Theta_0} \pi(\theta) = \sup_{\theta \in \Theta_0} P(T(X_1, \ldots, X_n) \in I)$$

α rappresenta la massima probabilità di rifiutare l'ipotesi nulla, quando questa è vera; è cioè la massima probabilità di fare un errore di tipo I. Più α è piccolo, più siamo tranquilli di non sbagliare, se la regola di decisione ci porta a rifiutare l'ipotesi nulla.

Nella pratica il valore di α viene stabilito a priori, prima di eseguire il campionamento, e il valore di k viene ottenuto di conseguenza. Valori tipici per il livello di significatività α sono 0.1, 0.05, 0.01.

Nell'esempio precedente scegliamo un livello di significatività α e cerchiamo il valore di k corrispondente.

$$\sup_{\mu \geq 4} P(\overline{X}_n < k) = \sup_{\mu \geq 4} \Phi\left(\frac{k-\mu}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{k-4}{\sigma/\sqrt{n}}\right)$$

Scegliendo

$$k = 4 + \frac{q_\alpha \sigma}{\sqrt{n}} = 4 - \frac{q_{1-\alpha}\sigma}{\sqrt{n}}$$

otteniamo proprio

$$\sup_{\mu \geq 4} P(\overline{X}_n < k) = \Phi(q_\alpha) = \alpha$$

Dunque al livello di significatività α, la regola di decisione dell'ipotesi nulla *lo spessore alla base è almeno pari a quello dichiarato* è si rifiuti H_0 se $\overline{x}_n < 4 - \frac{q_{1-\alpha}\sigma}{\sqrt{n}}$.

Riassumiamo i passi di un test statistico:

1. Si scelgono l'ipotesi nulla H_0 e la sua alternativa H_1. Nella scelta va condotto un giudizio su quale delle due ipotesi sia la più importante.
2. Si sceglie una statistica per stimare il parametro su cui effettuare il test, e si stabilisce la forma che costituisce la regione critica (ad esempio: si rifiuti H_0 se $\overline{X}_n < k$; k è ancora indeterminato).
3. Si sceglie il livello di significatività α a cui si vuole eseguire il test. Più α è piccolo e più difficilmente rifiuteremo l'ipotesi nulla, e più certi saremo di non sbagliare quando la rifiutiamo.
4. Si determina la regione del rifiuto in funzione del valore α scelto (ad esempio si rifiuti H_0 se $\overline{X}_n < 4 - \frac{q_{1-\alpha}\sigma}{\sqrt{n}}$).
5. Si esegue il campionamento, si calcola la statistica definita nel punto 2 e si vede se il risultato appartiene o meno alla regione di rifiuto: in caso positivo si rifiuta l'ipotesi nulla, in caso negativo la si accetta.

9.1.5 p-value

Poiché k dipende dal livello di significatività impostato, una stessa ipotesi che è stata rifiutata diciamo al livello dell' 1% può essere invece accettata ad un livello inferiore.

Esiste un livello di significatività limite, detto $p - value$, pari al più basso livello di significatività a cui i dati campionari consentono di rifiutare l'ipotesi nulla.

Lo si ottiene ponendo l'uguaglianza al posto della disuguaglianza che definisce la regione critica del test.

Nell'esempio precedente, il $p - value$ è soluzione dell'equazione

$$\overline{x}_n = 4 - \frac{q_{1-\overline{\alpha}}\sigma}{\sqrt{n}}$$

ossia

$$\overline{\alpha} = \Phi\left(\frac{\overline{x}_n - 4}{\sigma/\sqrt{n}}\right)$$

- Un $p-value$ molto piccolo significa che H_0 può venire rifiutata con tranquillità.
- Un $p-value$ basso ma non piccolissimo, dell'ordine dei consueti livelli di significatività (cioè 0.01, 0.05, ...) vuol dire che la decisione di rifiutare H_0 dipende dal livello di significatività impostato.
- Un $p-value$ alto vuol dire che H_0 si può plausibilmente accettare come vera.

9.2 Verifica di ipotesi sulla media (varianza nota)

Affrontiamo in modo sistematico il problema del test di ipotesi sulla media. Iniziamo dal caso in cui la varianza σ^2 della popolazione sia nota.

- Supponiamo per cominciare che l'ipotesi nulla sia

  $$H_0 : \mu \geq \mu_0$$

  mentre

  $$H_1 : \mu < \mu_0$$

  è l'ipotesi alternativa.

  - Il primo passo nella costruzione del test è la scelta di una statistica, *detta statistica test*, mediante la quale si stima il parametro incognito a partire dai dati campionari. Nel caso della media la statistica è naturalmente la media campionaria

    $$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

    per la quale vale (esattamente o asintoticamente)

    $$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

  - Osserviamo che stiamo verificando l'ipotesi $\mu \geq \mu_0$ che rigetteremo solo nel caso che la stima ottenuta dal campione sia "nettamente" al di sotto di μ_0.
    Fissiamo allora un valore $k < \mu_0$ e decidiamo di accettare H_0 se risulterà $\overline{x}_n > k$ e di rifiutarla nel caso opposto.
  - La regione critica del test è l'insieme dei valori campionari

    $$\mathcal{R} = \{(x_1, \ldots, x_n) : \overline{x}_n < k\}$$

  - *Come scegliere k?*
    Fissando il livello di significatività α del test, viene fissato di conseguenza il valore di k. Per legare k ad α partiamo dalla seguente espressione

    $$P\left(\overline{X}_n < k \,\middle|\, H_0\right) = P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{k-\mu}{\sigma/\sqrt{n}} \,\middle|\, H_0\right) = \alpha(\mu)$$

    Poiché $X = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ (vale per un campione numeroso in virtù del teorema centrale del limite, o per un campione qualsiasi estratto da una popolazione gaussiana), otteniamo

    $$P\left(X < \frac{k-\mu}{\sigma/\sqrt{n}} \,\middle|\, H_0\right) = \alpha(\mu)$$

Il valore massimo di questa probabilità, al variare di μ in H_0 (cioè per $\mu \geq \mu_0$), viene assunto in $\mu = \mu_0$. Pertanto

$$\frac{k - \mu_0}{\sigma/\sqrt{n}} = q_\alpha$$

$$k = \mu_0 - \frac{\sigma q_{1-\alpha}}{\sqrt{n}}$$

L'ipotesi $H_0 : \mu \geq \mu_0$ sarà accettata (al livello α) se la media stimata $\overline{x}_n$ risulta *maggiore* di k, altrimenti sarà rigettata:
rifiuto H_0 se

$$\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} < -q_{1-\alpha}$$

Poiché k dipende dal livello di significatività impostato, una stessa ipotesi che è stata rifiutata diciamo al livello dell' 1% può essere invece accettata ad un livello inferiore.

– Il $p - value$ si ottiene ponendo l'uguaglianza al posto della diseguaglianza precedente:

$$\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} = -q_{1-\overline{\alpha}}$$

$$\overline{\alpha} = \Phi\left(\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}\right)$$

- Il caso in cui l'ipotesi nulla e quella alternativa siano

$$H_0 : \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0$$

si tratta in modo del tutto analogo al precedente. Diamo solo i risultati.
Si rifiuta H_0 se

$$\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} > q_{1-\alpha}$$

Il $p - value$ è

$$\overline{\alpha} = \Phi\left(\frac{\mu_0 - \overline{x}_n}{\sigma/\sqrt{n}}\right)$$

- Vediamo infine il caso in cui l'ipotesi nulla e quella alternativa siano

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0$$

In questo caso il test deve essere costruito in modo da rifiutare uno scostamento, di qualunque segno, maggiore di un certo k da determinarsi a partire da α.
Per determinare la relazione tra k ed α scriviamo

$$P\left(|\overline{X}_n - \mu_0| > k\right) = P\left(\frac{|\overline{X}_n - \mu_0|}{\sigma/\sqrt{n}} > \frac{k}{\sigma/\sqrt{n}}\right) = \alpha$$

da cui, ricordando che per una v.a. $X \sim \mathcal{N}(0,1)$ si ha

$$P(|X| > q_{1-\frac{\alpha}{2}}) = \alpha$$

otteniamo

$$\frac{k}{\sigma/\sqrt{n}} = q_{1-\frac{\alpha}{2}}$$

$$k = q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Fissato α si rifiuterà H_0 se

$$|\overline{x}_n - \mu_0| > q_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Il p-value viene ottenuto ponendo l'uguaglianza

$$|\overline{x}_n - \mu_0| = q_{1-\frac{\overline{\alpha}}{2}} \frac{\sigma}{\sqrt{n}}$$

da cui si ricava

$$\overline{\alpha} = 2 - 2\Phi\left(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\right)$$

Esempio. Da una popolazione normale di media incognita e deviazione standard $\sigma = 3$ si estrae un campione di ampiezza 20, e si sottopone a test l'ipotesi nulla $H_0 : \mu = 100$.

a) Troviamo la regione critica ai livelli dell'1%, del 5% e del 10%.

   Per quanto visto sopra la regione critica del test è data da quei valori di $\overline{x}_n$ per cui si ha

$$\frac{|\overline{x}_n - 100|}{\sigma/\sqrt{n}} > q_{1-\frac{\alpha}{2}}$$

   con

α	0.01	0.05	0.1
$q_{1-\frac{\alpha}{2}}$	2.578	1.96	1.6449

b) Supponendo di avere estratto un campione per cui $\overline{x}_n = 98.5$, si tragga una conclusione, per ciascuno dei tre livelli di significatività.

   Sostituiamo nella formula precedente $\overline{x}_n$ con 98.5:

$$\frac{|\overline{x}_n - 100|}{\sigma/\sqrt{n}} = \frac{|98.5 - 100|}{3/\sqrt{20}} \approx 2.2361$$

   al livello dell'1% il test viene accettato, mentre ai livelli del 5% e del 10% viene rifiutato.

c) Calcoliamo infine il p-value:

$$\overline{\alpha} = 2 - 2\Phi\left(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\right) = 2 - 2\Phi\left(\frac{|98.5 - 100|}{3/\sqrt{20}}\right) \approx 0.0253$$

   Tutti i test con livello di significatività inferiore al 2.53% sono accettati, mentre quelli con livello maggiore sono rifiutati.

Riassumiamo i risultati ottenuti in questa sezione nella tabella seguente:

H_0	H_1	Rifiutare H_0 se	p-value
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert z\rvert > q_{1-\alpha/2}$	$2 - 2\Phi(\lvert z\rvert)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$z > q_{1-\alpha}$	$\Phi(-z)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$z < -q_{1-\alpha}$	$\Phi(z)$

$$\text{dove } \; z = \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

9.3 Test su una frequenza (grandi campioni)

Vogliamo sottoporre a verifica d'ipotesi un campione tratto da una popolazione Bernoulliana $X_i \sim B(p)$.

Consideriamo le ipotesi nulle

$$H_0 : p = p_0; \quad H_0 : p \leq p_0; \quad H_0 : p \geq p_0$$

e le loro rispettive alternative

$$H_1 : p \neq p_0; \quad H_1 : p > p_0; \quad H_1 : p < p_0$$

Utilizziamo la proprietà che, se il campione è sufficientemente numeroso, la media campionaria tende a una v.a. normale:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Perciò

$$\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim \mathcal{N}(0,1)$$

Possiamo ragionare come nella sezione precedente. L'unica differenza è che la deviazione standard σ viene sostituita con $\sqrt{p_0(1-p_0)}$. Otteniamo in definitiva la tabella seguente

H_0	H_1	Rifiutare H_0 se	p-value
$p = p_0$	$p \neq p_0$	$\lvert z\rvert > q_{1-\alpha/2}$	$2 - 2\Phi(\lvert z\rvert)$
$p \leq p_0$	$p > p_0$	$z > q_{1-\alpha}$	$\Phi(-z)$
$p \geq p_0$	$p < p_0$	$z < -q_{1-\alpha}$	$\Phi(z)$

$$\text{dove } \; z = \frac{\overline{x}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Esempio. Un partito politico ha ricevuto nelle ultime elezioni il 35% dei voti. Quattro anni dopo, da un sondaggio d'opinione basato su 300 interviste si è trovato che il 32% degli intervistati ha dichiarato di essere disposto a votare per quel partito. Ci si chiede se, rispetto al risultato elettorale, la situazione del partito sia peggiorata.

Si tratta di un test d'ipotesi sul parametro p di una popolazione Bernoulliana $B(p)$. L'ipotesi da verificare (ipotesi nulla) è

$$H_0 : p \geq 0.35$$

mentre l'ipotesi alternativa è

$$H_0 : p < 0.35$$

La standardizzata vale

$$z = \frac{\overline{x}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{.32 - .35}{\sqrt{.35 * .65/300}} \approx -1.0894$$

Il $p - value$ corrispondente al dato campionario è

$$\overline{\alpha} = \Phi(z) \approx \Phi(-1.0894) \approx 0.1380$$

L'ipotesi H_0 viene accettata da ogni test il cui livello di significatività sia inferiore al $p - value$, cioè al 13.8%.

9.4 Verifica di ipotesi sulla media (varianza incognita)

Consideriamo ora il caso in cui la varianza σ^2 della popolazione sia incognita.

Riprendiamo per cominciare la verifica dell'ipotesi

$$H_0 : \mu \geq \mu_0, \qquad H_1 : \mu < \mu_0$$

Come nel caso a varianza nota, fissiamo un valore $k < \mu_0$ e decidiamo di rifiutare H_0 se $\overline{x}_n$ dovesse risultare inferiore a k. La probabilità che $\overline{X}_n < k$, *supponendo vera* H_0 è

$$P\left(\overline{X}_n < k | H_0\right) = P\left(\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} < \frac{k-\mu}{S_n/\sqrt{n}} \,\middle|\, H_0\right) =$$

$$= P(T_n < t_{\alpha(\mu)}) = \alpha(\mu)$$

Nella espressione precedente si è sostituito σ con il suo stimatore S_n e siamo passati dalla standardizzata $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ alla v.a.

$$T_n = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

Seguendo lo stesso ragionamento del caso a varianza nota possiamo dire che

$$\alpha = \sup_{\mu \geq \mu_0} \alpha(\mu) = \alpha(\mu_0)$$

e quindi fissato il livello di significatività α abbiamo

$$P(T_n < -t_{1-\alpha}) = \alpha$$

ovvero

$$\frac{k - \mu_0}{S_n/\sqrt{n}} = -t_{1-\alpha}$$

Una volta effettuato il campionamento ed ottenute le stime $\overline{x}_n$ ed s_n rifiuteremo H_0 se

$$\frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}} < -t_{1-\alpha}$$

Osservazione. Il procedimento seguito è esatto se la popolazione da cui si estrae il campione è normale; è ancora approssimativamente valido per popolazioni non normali, se il campione è sufficientemente numeroso.

Il p-value viene ottenuto ponendo

$$t_{1-\overline{\alpha}} = -\frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}}$$

Ragionando in modo del tutto analogo si ottengono facilmente le regioni critiche e i p-value nei casi $H_0 : \mu \leq \mu_0$ e $H_0 : \mu = \mu_0$.

I risultati sono riassunti nella tabella seguente.

H_0	H_1	Rifiutare H_0 se	p-value
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert t\rvert > t_{1-\alpha/2}(n-1)$	$t_{1-\overline{\alpha}/2}(n-1) = \lvert t\rvert$
$\mu \leq \mu_0$	$\mu > \mu_0$	$t > t_{1-\alpha}(n-1)$	$t_{1-\overline{\alpha}}(n-1) = t$
$\mu \geq \mu_0$	$\mu < \mu_0$	$t < -t_{1-\alpha}(n-1)$	$t_{1-\overline{\alpha}}(n-1) = -t$

$$\text{dove } \; t = \frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}}$$

9.5 Verifica d'ipotesi sulla varianza

Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da una popolazione normale $\mathcal{N}(\mu, \sigma^2)$. Ci proponiamo di sottoporre a verifica l'ipotesi H_0 riguardante la varianza σ^2.

Esempio. In un processo di produzione di wafers al silicio si richiede che la varianza dello spessore del singolo wafer sia al più di 0.5 micron. Avendo riscontrato una varianza campionaria di 0.64 micron su un campione di 50 wafers, si vuole sottoporre a verifica con livello di significatività $\alpha = 0.05$ l'ipotesi H_0: *la deviazione standard dello spessore dei wafers è minore o uguale a 0.5 micron.*

L'analisi è diversa a seconda che il valore medio μ sia noto o incognito.

- Trattiamo innanzitutto il caso in cui μ sia ignoto. Una statistica test appropriata per la varianza della popolazione è la varianza campionaria

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Sottoponiamo a verifica l'ipotesi

$$H_0 : \sigma^2 \leq \sigma_0^2$$

Il test d'ipotesi sarà del tipo:

$$\textit{rifiuto } H_0 \textit{ se } S_n^2 > k$$

per un valore opportuno di k da stabilire in funzione del livello di significatività α scelto.

Precisamente k viene ottenuto imponendo che

$$\sup_{\sigma} P(S_n^2 > k | \sigma^2 \leq \sigma_0^2) = \alpha$$

La v.a. $(n-1)S_n^2/\sigma^2$ ha una legge chi-quadrato con $n-1$ gradi di libertà.

Il sup della probabilità viene assunto per $\sigma = \sigma_0$, infatti:

$$P(S_n^2 > k) = P\left(\frac{(n-1)S_n^2}{\sigma^2} > \frac{k(n-1)}{\sigma^2}\right) = P\left(\chi^2 > \frac{k(n-1)}{\sigma^2}\right)$$

Quando σ cresce il secondo membro della disuguaglianza decresce, e pertanto la probabilità cresce. Dunque il sup viene assunto per σ massimo, cioè in σ_0.

Otteniamo dunque:

$$P\left(\frac{(n-1)S_n^2}{\sigma^2} > \chi^2_{1-\alpha}(n-1)\right) = \alpha$$

dove $\chi^2_{1-\alpha}(n-1)$ è il quantile $1-\alpha$ della legge $\chi^2(n-1)$.

$$P\left(S_n^2 > \frac{\sigma_0^2}{n-1}\chi^2_{1-\alpha}(n-1)\right) = \alpha$$

il valore di k è perciò

$$k = \frac{\sigma_0^2}{n-1}\chi^2_{1-\alpha}(n-1)$$

e la regola di decisione del test è in definitiva:

$$\textit{rifiuto } H_0 \textit{ se } (n-1)S_n^2/\sigma_0^2 > \chi^2_{1-\alpha}(n-1).$$

Il p-value $\overline{\alpha}$ viene ottenuto ponendo l'uguaglianza

$$\chi^2_{1-\overline{\alpha}} = \frac{(n-1)S_n^2}{\sigma_0^2}$$

Nell'esempio precedente sui wafers di silicio, abbiamo

$$\frac{(n-1)S_n^2}{\sigma_0^2} = \frac{49 * 0.64}{0.5} = 62.72, \quad \chi^2_{0.95}(49) \simeq 66.34,$$

pertanto l'ipotesi H_0 viene accettata.

Il p-value è dato da:

$$\chi^2_{1-\overline{\alpha}} = 62.72 \quad \Rightarrow \quad \overline{\alpha} \simeq 0.09$$

Ogni test con livello di significatività inferiore al 9% viene accettato, mentre ogni test con livello di significatività superiore al 9% viene rifiutato.

Ragionando in modo analogo per le ipotesi nulle di altro tipo (H_0: $\sigma^2 = \sigma_0^2$, e H_0: $\sigma^2 \geq \sigma_0^2$), si ottiene la seguente tabella:

H_0	H_1	Rifiutare H_0 se	p-value
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi^2_{1-\frac{\alpha}{2}}(n-1)$ o $\chi^2 < \chi^2_{\frac{\alpha}{2}}(n-1)$	
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{1-\alpha}(n-1)$	$\chi^2_{1-\overline{\alpha}}(n-1) = \chi^2$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{\alpha}(n-1)$	$\chi^2_{\overline{\alpha}}(n-1) = \chi^2$

$$\text{dove } \chi^2 = \frac{(n-1)s_n^2}{\sigma_0^2}$$

- Nel caso in cui il valore medio μ sia noto, la statistica test appropriata è

$$T_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Si ha poi

$$\frac{nT_n^2}{\sigma^2} \sim \chi^2(n)$$

Le formule della tabella precedente rimangono invariate, a patto di sostituire χ^2 con la quantità nT_n^2/σ_0^2, e di sostituire $n-1$ con n nel numero dei gradi di libertà della legge chi-quadrato:

H_0	H_1	Rifiutare H_0 se	p-value
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi^2_{1-\frac{\alpha}{2}}(n)$ o $\chi^2 < \chi^2_{\frac{\alpha}{2}}(n)$	
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{1-\alpha}(n)$	$\chi^2_{1-\overline{\alpha}}(n) = \chi^2$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{\alpha}(n)$	$\chi^2_{\overline{\alpha}}(n) = \chi^2$

$$\text{dove } \chi^2 = \frac{nt_n^2}{\sigma_0^2}, \quad t_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

9.6 Test chi-quadrato di buon adattamento

Supponiamo di avere n osservazioni di una variabile X raggruppate in N_c classi. Le classi possono rappresentare

- valori assunti da una variabile discreta: ogni classe raggruppa le osservazioni che assumono un determinato valore o un gruppo di valori.
- Intervalli di valori assunti da una variabile continua.

- caratteristiche qualitative assunte da una variabile categorica (colori, pariti votati, ecc.).

Sia $f_r(k)$ la frequenza relativa della $k-esima$ classe.

Supponiamo di possedere una stima teorica dei valori che dovrebbe assumere la frequenza relativa. Ci poniamo il problema di valutare la bontà di adattamento delle frequenze osservate alle frequenze ipotizzate.

Esempio. La legge ipotizzata per il tempo di vita in mesi di una lampadina è una legge esponenziale $X \sim Exp(0.33)$. Su un campione di 100 lampadine sono state riscontrate le seguenti durate:

	freq. oss. f_r	freq. ipotizzata p
$X \leq 1$	0.39	$1 - e^{-0.33} \simeq 0.281$
$1 < X \leq 2$	0.24	$e^{-0.33} - e^{-0.66} \simeq 0.202$
$2 < X \leq 3$	0.12	$e^{-0.66} - e^{-0.99} \simeq 0.145$
$3 < X \leq 5$	0.16	$e^{-0.99} - e^{-1.65} \simeq 0.180$
$5 < X \leq 10$	0.09	$e^{-1.65} - e^{-3.3} \simeq 0.155$

Ci chiediamo se la legge esponenziale è adeguata a descrivere il fenomeno osservato.

Per risolvere questo tipo di problema si considera la seguente statistica test:

$$Q = \sum_{i=1}^{N_c} \frac{(np_i - N_i)^2}{np_i} = n \sum_{i=1}^{N_c} \frac{(p_i - f_r(i))^2}{p_i}$$

dove N_i sono le frequenze assolute ($N_i = nf_r(i)$), cioè il numero di osservazioni del campione appartenente alla $i-esima$ classe. Si noti che np_i sono le frequenze assolute ipotizzate.

La statistica Q viene detta **chi-quadrato calcolato dal campione**. Q è tanto più piccola quanto migliore è l'adattamento delle frequnze osservate a quelle ipotizzate. Si può allora pensare di utilizzare Q per fare un **test di adattamento**, nel modo seguente:

L'ipotesi nulla è H_0: *le osservazioni provengono da una popolazione distribuita secondo le frequenze relative attese* $p_1, p_2, \ldots, p_{N_c}$.

Il test è del tipo:
si rifiuti H_0 *se* $Q > k$, per un valore di k opportuno.

Il teorema seguente permette di determinare la costante k una volta fissato il livello di significatività α:

Teorema: Estraiamo un campione casuale di ampiezza n da una popolazione ripartita in N_c classi di frequenze relative $p_1, p_2, \ldots, p_{N_c}$. Si noti che deve essere $p_i \geq 0$ e $\sum_{i=1}^{N_c} p_i = 1$. Sia N_i la frequenza assoluta osservata relativa alla classe $i-esima$. Allora la statistica

$$Q = \sum_{i=1}^{N_c} \frac{(np_i - N_i)^2}{np_i}$$

è una v.a. la cui legge tende alla legge chi-quadrato $\chi^2(N_c - 1)$ per $n \to \infty$.
Se le frequenze relative attese p_i, invece di essere assegnate a priori, sono calcolate dopo aver stimato r parametri incogniti *dai dati del campione*, allora $Q \sim \chi^2(N_c - 1 - r)$.

Il teorema è applicabile a condizione che le frequenze assolute attese siano $np_i \geq 5 \ \forall i$, altrimenti la legge per Q non sarebbe ben approssimabile con la legge chi-quadrato. Se risultasse che $np_i < 5$ per qualche valore di i, allora bisognerebbe accorpare opportunamente alcune classi contigue, finché la condizione non sia verificata.

Dal teorema si ricava facilmente che il test d'ipotesi è:
si rifiuti H_0 *se* $Q > \chi^2_{1-\alpha}(N_c - 1)$ (oppure $Q > \chi^2_{1-\alpha}(N_c - 1 - r)$ se r è il numero di parametri stimati dai dati).

Nell'esempio di prima del tempo di vita di una lampadina, effettuiamo il test di adattamento per un livello di significatività del 10%:

$$Q = 100 \left[\frac{(0.39 - 0.281)^2}{0.281} + \frac{(0.24 - 0.202)^2}{0.202} + \frac{(0.12 - 0.145)^2}{0.145} + \right.$$

$$+\frac{(0.16-0.180)^2}{0.180}+\frac{(0.09-0.155)^2}{0.155}\Bigg] \simeq 8.3219$$

mentre

$$\chi^2_{1-\alpha}(N_c-1)=\chi^2_{0.9}(4)\simeq 7.7794$$

Risulta $Q > 7.7794$: il test viene rifiutato.
Il valore del p-value per il test è dato da:

$$\chi^2_{1-\overline{\alpha}}(N_c-1)=Q \quad \Rightarrow \quad \overline{\alpha}\simeq 8.1\%$$

I test con livelli di significatività minori dell'8.1% vengono accettati.

Supponiamo adesso che la durata di vita delle lampadine segua una legge esponenziale di parametro incognito, e ricaviamo il valore di ν dal campione: $\nu = 1/\overline{x} \simeq 0.46$. Il nuovo valore di Q è pari a 0.6834. Lo dobbiamo confrontare con $\chi^2_{1-\alpha}(N_c-1-1) = \chi^2_{0.9}(3) \simeq 6.2514$. Questa volta il test viene accettato. Si noti che la legge chi-quadrato possiede solo tre gradi di libertà, in quanto ν è stato stimato a partire dal campione e ciò toglie un grado di libertà. Il p-value è pari a 0.877, valore estremamente elevato: l'approssimazione esponenziale risulta ottima.

9.7 Test chi-quadrato di indipendenza

Questo test viene applicato al seguente problema: date n osservazioni congiunte di due variabili, ci si chiede se le due variabili sono indipendenti tra loro.
Il problema era stato affrontato per variabli numeriche nel primo capitolo mediante il calcolo del coefficiente di correlazione. Il metodo che esponiamo ora è alternativo e può essere applicato anche a variabili di tipo categorico.

Consideriamo il caso di due variabili X e Y associate alla medesima popolazione; effettuiamo un campionamento e raggruppiamo i dati in classi. Se le due variabili sono indipendenti, allora

$$P(X\in A_i, Y\in B_j)=P(X\in A_i)P(Y\in B_j)$$

dove A_i sono le classi relative alla variabile X, e B_j quelle relative a Y.

La probabilità $P(X\in A_i)$ può essere stimata con la frequenza marginale relativa $f_{rX}(i)$, e analogamente $P(Y\in B_j)\simeq f_{rY}(j)$. L'ipotesi di indipendenza si traduce nella

$$f_r(i,j)=f_{rX}(i)f_{rY}(j)$$

Mentre per le frequenze assolute:

$$f_a(i,j)=f_{aX}(i)f_{aY}(j)/n$$

Costruiamo la statistica chi-quadrato calcolata dai dati:

$$Q=\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\frac{(f_a(i,j)-f_{aX}(i)f_{aY}(j)/n)^2}{f_{aX}(i)f_{aY}(j)/n}$$

Per quanto abbiamo visto prima, per n sufficientemente grande e per una suddivisione in classi con almeno 5 elementi in ogni classe, la statistica Q è approssimabile con una legge chi-quadrato.

I gradi di libertà della legge chi-quadrato si calcolano in questo modo: Il numero di classi è in totale $N_c = N_1N_2$. I valori di $f_{rX}(i)$ per $i = 1,\dots,N_1-1$ sono stimati dal campione (l'ultimo valore, per $i = N_1$ viene ricavato dal fatto che deve essere $\sum_{i=1}^{N_1} f_{rX}(i) = 1$). Analogamente anche i valori di $f_{rY}(j)$ per $j = 1,\dots,N_2-1$ sono stimati dal campione. In totale i parametri

stimati dal campione sono in numero $N_1 - 1 + N_2 - 1$. Dunque il numero dei gradi di libertà della legge chi-quadrato è

$$N_c - 1 - (N_1 - 1 + N_2 - 1) = N_1 N_2 - N_1 - N_2 + 1 = (N_1 - 1)(N_2 - 1)$$

L'ipotesi nulla è

H_0: *le variabili X e Y sono indipendenti tra loro.*

Il test sull'ipotesi di indipendenza è:

si rifiuti H_0 se $Q > \chi^2_{1-\alpha}((N_1 - 1)(N_2 - 1))$.

Esempio. A un campione di 150 persone è stato chiesto il colore e l'animale preferiti. I risultati sono presentati nella seguente tabella:

	rosso	blu	verde	giallo	totale
gatto	7	17	16	13	53
cane	8	28	22	9	67
cavallo	5	10	9	6	30
totale	20	55	47	28	150

Ci chiediamo se il colore preferito è indipendente dall'animale preferito, per un livello di significatività del 10%.

Applichiamo la formula trovata sopra: si rifiuta il test d'indipendenza se $Q > \chi^2_{1-\alpha}((N_1 - 1)(N_2 - 1))$.

$Q \simeq 3.2983$, mentre $\chi^2_{1-\alpha}((N_1 - 1)(N_2 - 1)) = \chi^2_{0.9}(6) \simeq 10.6446$: il test viene accettato, ossia si può concludere che le due variabili colore e animale sono indipendenti. Il p-value è dato da $\chi^2_{1-\overline{\alpha}}(6) = 3.2983$: $\overline{\alpha} \simeq 77.1\%$.

9.8 Verifica d'ipotesi sulla differenza tra due medie

Vogliamo confrontare le medie di due popolazioni diverse, estraendo un campione casuale da ciascuna.

Consideriamo il caso di due popolazioni normali indipendenti: $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Estraiamo dalla prima un campione casuale di ampiezza n_X, e dalla seconda una campione di ampiezza n_Y; n_X e n_Y non sono necessariamente uguali.

Vogliamo confrontare le medie delle due popolazioni; formuliamo a tale fine una delle seguenti ipotesi nulle:

$$H_0 : \mu_x - \mu_Y \geq \delta, \quad H_0 : \mu_X - \mu_Y = \delta, \quad H_0 : \mu_x - \mu_Y \leq \delta$$

dove δ è un parametro dato.

I test di verifica delle ipotesi cambiano a seconda che le varianze siano note oppure incognite. Tratteremo due casi: quello in cui σ_X^2 e σ_Y^2 sono entrambe note, e quello in cui σ_X^2 e σ_Y^2 sono incognite ma uguali. Nella pratica, il caso in cui σ_X^2 e σ_Y^2 sono entrambe incognite si può ricondurre a quello in cui esse siano note, purché i campioni siano sufficientemente grandi ($n_X, n_Y > 30$), usando le varianze campionarie come se fossero i valori esatti delle varianze.

- Le varianze σ_X^2 e σ_Y^2 sono entrambe note. Si costruisce la statistica test

$$\overline{X}_{n_X} - \overline{Y}_{n_Y} - \delta \sim \mathcal{N}\left(\mu_X - \mu_Y - \delta, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)$$

  ossia

$$\frac{\overline{X}_{n_X} - \overline{Y}_{n_Y} - \delta - (\mu_X - \mu_Y - \delta)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim \mathcal{N}(0, 1)$$

  Le regole di decisione dei test si ricavano allo stesso modo di quelle trovate nel precedente capitolo sulla media di una popolazione gaussiana in cui la varianza sia nota.

  Riassumiamo i risultati nella tabella seguente:

H_0	H_1	Rifiutare H_0 se	p-value
$\mu_X - \mu_Y = \delta$	$\mu_X - \mu_Y = \delta$	$\lvert z\rvert > q_{1-\alpha/2}$	$2 - 2\Phi(\lvert z\rvert)$
$\mu_X - \mu_Y \leq \delta$	$\mu_X - \mu_Y > \delta$	$z > q_{1-\alpha}$	$\Phi(-z)$
$\mu_X - \mu_Y \geq \delta$	$\mu_X - \mu_Y < \delta$	$z < -q_{1-\alpha}$	$\Phi(z)$

$$\text{dove } z = \frac{\overline{x}_{n_X} - \overline{y}_{n_Y} - \delta}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

- Il secondo caso è quello in cui le varianze sono entrambe incognite ma uguali: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Si considera la seguente statistica:

$$T = \frac{\overline{X}_{n_X} - \overline{Y}_{n_Y} - \delta}{\sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$$

Dove S_X^2 e S_Y^2 sono le varianze campionarie dei due campioni. T ha una legge t di Student con n_X+n_Y-2 gradi di libertà. Infatti:

$$\frac{(n_X-1)S_X^2}{\sigma^2} \sim \chi^2(n_X-1), \quad \frac{(n_Y-1)S_Y^2}{\sigma^2} \sim \chi^2(n_Y-1)$$

per la proprietà della legge χ^2 la loro somma è anch'essa una legge chi-quadrato, con n_X+n_Y-2 gradi di libertà:

$$S^2 = \frac{(n_X-1)S_X^2}{\sigma^2} + \frac{(n_Y-1)S_Y^2}{\sigma^2} \sim \chi^2(n_X+n_Y-2)$$

La variabile aleatoria

$$\frac{\overline{X}_{n_X} - \overline{Y}_{n_Y} - \delta}{\sqrt{\left(\frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}\right)}}$$

ha legge $\mathcal{N}(0,1)$. Pertanto per definizione della legge t di Student:

$$\frac{\overline{X}_{n_X} - \overline{Y}_{n_Y} - \delta}{\sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} = \frac{\frac{\overline{X}_{n_X} - \overline{Y}_{n_Y} - \delta}{\sqrt{\left(\frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}\right)}}}{\sqrt{S^2/(n_X+n_Y-2)}} \sim t(n_X+n_Y-2)$$

Le regole di decisione dei test si ricavano allo stesso modo di quelle trovate nel precedente capitolo sulla media di una popolazione gaussiana in cui la varianza sia incognita.

H_0	H_1	Rifiutare H_0 se	p-value
$\mu_X - \mu_Y = \delta$	$\mu_X - \mu_Y \neq \delta$	$\lvert t\rvert > t_{1-\alpha/2}(n)$	$t_{1-\overline{\alpha}/2}(n) = \lvert t\rvert$
$\mu_X - \mu_Y \leq \delta$	$\mu_X - \mu_Y > \delta$	$t > t_{1-\alpha}(n)$	$t_{1-\overline{\alpha}}(n) = t$
$\mu_X - \mu_Y \geq \delta$	$\mu_X - \mu_Y < \delta$	$t < -t_{1-\alpha}(n)$	$t_{1-\overline{\alpha}}(n) = -t$

$$\text{dove } t = \frac{\overline{x}_{n_X} - \overline{y}_{n_Y} - \delta}{\sqrt{\frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X+n_Y-2}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}, \quad n = n_X + n_Y - 2$$

Esempio. L'osservazione dei tempi di cui hanno bisogno i clienti di un ufficio postale per effettuare le loro operazioni ha dato i seguenti risultati: su 150 persone il tempo medio ad operazione allo sportello A è risultato pari a 85 secondi, con una deviazione standard campionaria di 15 secondi, mentre allo sportello B su 200 persone la media è stata di 81 secondi e deviazione standard 20 secondi.

Al livello di confidenza del 5% ci domandiamo se è plausibile che i clienti passino meno tempo al primo sportello che al secondo.

L'ipotesi nulla è $H_0\colon \mu_A \leq \mu_B$

dove μ_A e μ_B sono i tempi medi passati rispettivamente agli sportelli A e B. Siamo nel caso di varianze incognite, non necessariamente uguali. Le approssimiamo allora con i valori delle varianze campionarie: $\sigma_A^2 = 15^2$, $\sigma_B^2 = 20^2$.

La regola di decisione del test è: *si rifiuti* H_0 *se* $z > q_{0.95}$, con

$$z = \frac{\overline{x}_A - \overline{y}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{85 - 81}{\sqrt{\frac{225}{150} + \frac{400}{200}}} \simeq 2.1381$$

Il quantile vale $q_{0.95} = 1.6449$. Pertanto il test viene rifiutato.
Calcoliamo infine il p-value: $\overline{\alpha} = \Phi(-2.1381) \simeq 1.6\%$.