



Università degli Studi di Catania
Laurea Informatica Corso Triennale - L31

Metodi Matematici e Statistici

Prof. Orazio Muscato

Dipartimento di Matematica e Informatica

16 marzo 2017

Indice

1	Statistica descrittiva	6
1.1	Indici di tendenza centrale	10
1.2	Indici di variabilità	15
1.2.1	Dati raggruppati per classi	18
1.3	Indici di forma	21
1.4	La correlazione tra due serie di dati	23
1.5	Analisi di regressione per una serie di dati	25
1.5.1	Metodo dei minimi quadrati	25
1.5.2	Regressione lineare	25
1.5.3	Parabola dei minimi quadrati	30
2	Elementi di Probabilità	33
2.1	Probabilità della somma logica di eventi	38
2.2	Probabilità condizionata	39
2.3	Probabilità del prodotto logico di eventi	41
2.4	Il teorema di Bayes	43
2.5	Variabili aleatorie	50
2.6	Indici di tendenza centrale per le variabili aleatorie	55
2.7	Variabili aleatorie multidimensionali	56
3	Distribuzioni notevoli e teoremi di convergenza	59
3.1	Distribuzione di Bernoulli	59

3.2	Distribuzione binomiale	60
3.3	Distribuzione di Poisson	61
3.3.1	Un'applicazione: ALOHA e CSMA	63
3.4	Distribuzione Uniforme	69
3.5	Distribuzione Esponenziale	70
3.6	Distribuzione di Weibull	73
3.6.1	Un'applicazione alla Teoria dell'Affidabilita'	73
3.7	Distribuzione Normale (di Gauss)	77
3.7.1	Errori di misura casuali e sistematici	80
3.8	Distribuzioni limite	81
3.9	Distribuzione Chi-quadro	81
3.10	Distribuzione t di Student	84
3.11	Legge dei grandi numeri	86
3.12	Teorema del limite centrale	87
4	Stime di parametri	88
4.1	Problema del campionamento	88
4.1.1	Strategie di campionamento	88
4.2	Principali distribuzioni campionarie	90
4.3	Stimatori puntuali	92
4.3.1	Altri metodi	93
4.4	Campionamento da una distribuzione normale	95
4.5	Stime intervallari	96
4.5.1	Intervallo di confidenza per la media	97
5	Verifica di Ipotesi	107
5.1	Caratteristiche generali di un test di ipotesi	108
5.2	Test parametrico sulla media di una popolazione normale	108
5.2.1	Z test bilatero con varianza nota	108

5.2.2	Z test unilatero con varianza nota	110
5.2.3	t test	111
5.2.4	t test per il confronto delle medie di due popolazioni normali	112
5.2.5	t test per dati accoppiati	113
5.3	Test parametrico sulla media di una popolazione non normalmente distribuita	114
5.3.1	Conclusione	114
5.4	Test non parametrici sulla bonta' dell'adattamento	116
5.4.1	Test di Kolmogorov - Smirnov	117
5.4.2	Test del Chi-quadro	122
6	Numeri Casuali	131
6.1	Generazione dei numeri casuali con densita' di probabilita' uniforme	132
6.1.1	RNG basato su ricorrenze lineari	132
6.1.2	RNG moltiplicativo congruenziale	133
6.2	Test statistici per i numeri casuali	134
6.3	Generazione di numeri casuali con assegnata densita' di probabilita'	137
6.3.1	Tecnica diretta	137
6.3.2	Tecnica di reiezione	138
6.3.3	Tecnica combinata	140
7	Il Metodo Monte Carlo	143
7.1	Richiami dei metodi per l'integrazione numerica	144
7.1.1	Integrale doppio	146
7.2	Il metodo MC "Hit or Miss"	148
7.2.1	Stima per l'intervallo di confidenza per I	150
7.3	Metodo Sample-Mean MC (MC della media)	152
7.4	Efficienza del metodo Monte Carlo	154
7.5	Tecniche di riduzione della varianza	155
7.5.1	Importance sampling (importanza del campionamento)	155

7.5.2	Control Variates	156
7.5.3	Stratified sampling (campionamento stratificato)	157
7.5.4	Antithetic Variates	158
8	Catene di Markov	162
8.1	Alcune definizioni	162
8.2	Calcolo di leggi congiunte	169
8.3	Classificazione degli stati	173
8.4	Problemi di assorbimento	175
8.5	Probabilita' invarianti	177
8.6	Stato stazionario	178
8.7	L'algoritmo di Metropolis	182
8.8	Un'applicazione alla teoria delle code	185
A	Il Processo di Poisson	192

Capitolo 1

Statistica descrittiva

Con il termine statistica descrittiva si intende una raccolta di metodi, strumenti matematici atti ad organizzare serie di dati per evidenziare in forma sintetica simmetrie, periodicità, leggi di altro genere ovvero di descrivere in maniera immediatamente comprensibile le informazioni dagli stessi dati. Sia $E = \{x_1, x_2, \dots, x_n\}$ un insieme di dati di numerosità n diamo alcune semplici definizioni.

i) modalità del carattere

I valori assumibili dai dati discreti sono di solito un numero limitato: per questo motivo conviene raggrupparli considerando l'insieme dei valori assumibili detti modalità del carattere. Sia N il numero dei valori assumibili allora denoto con $S = \{s_1, s_2, \dots, s_N\}$ l'insieme delle modalità.

ii) frequenza assoluta e relativa

$$f_j = n^\circ \text{ elementi di } E \text{ con valore } s_j \quad j = 1, 2, \dots, N$$
$$p_j = \frac{f_j}{n} \quad n = \text{numerosità di } E$$

iii) Frequenza cumulata assoluta e relativa

$$F_j = \sum_k f_k \quad \text{con } k \text{ tale che } s_k \leq s_j \quad j = 1, 2, \dots, N$$
$$P_j = \sum_k p_k \quad \text{con } k \text{ tale che } s_k \leq s_j \quad j = 1, 2, \dots, N$$

Tabella 1.1: Numero stanze in un campione di appartamenti.

3	4	2	6	5	2	4	4
2	5	4	4	5	7	5	4
5	7	8	4	3	6	2	3
5	2	7	2	4	8	4	2
6	5	4	4	6	5	3	3
8	5	2	5	6	5	5	4
2	6	4	5	5	7	3	4
3	3	3	4	4	3	4	6
4	3	7	4	4	6	4	2
4	4	6	3	2	3	5	4

Esempio 1.1 Nella tabella 1.1 sono elencati i dati relativi al numero di stanze possedute da 80 appartamenti scelti a caso tra tutti quelli che si trovano in un determinato quartiere di una data città.

Esercizio 1.1 Dalla tabella 1.1 creare una tabella come la seguente

n° stanze	freq. ass.	freq. rel	freq. cum. ass.	freq. cum. rel.
1				
2				
3				
4				
5				
6				
7				
8				

Utilizzare Excel (funzione conta.se) e fare gli istogrammi.

Tabella 1.2: Costo mq di un campione di appartamenti.

2,11	3,08	2,35	3,54	0,44	2,24	4,60	1,88
2,08	1,90	2,15	5,11	3,69	0,88	2,56	4,00
3,15	3,67	3,15	4,09	4,57	1,06	2,05	2,34
4,17	4,10	4,75	1,90	2,36	0,90	2,07	3,23
4,21	2,12	1,21	2,10	4,05	5,42	0,85	4,80
2,11	5,08	2,78	4,88	1,11	1,83	1,85	2,87
2,23	3,20	2,80	2,19	1,88	2,16	2,74	2,45
1,19	3,79	1,24	3,06	2,11	3,70	2,91	1,80
3,48	4,10	3,13	0,90	3,07	4,10	1,66	2,88
2,11	1,90	1,18	0,75	1,60	3,85	1,45	2,00

Può succedere che le modalità siano elevate, allora conviene considerare dei sottoinsiemi di S .
Chiamo classe C un sottoinsieme di S .

Chiamo partizione di S ogni famiglia di classi tra loro disgiunte la cui unione sia S ovvero

$$C_i \subseteq S \quad i = 1, \dots, k \in N$$

$$C_i \cap C_j = \phi \quad \forall i \neq j$$

$$\bigcup_{i=1}^k C_i = S$$

Esempio 1.2 Nella tabella 1.2 sono elencati i dati relativi al costo al metro quadro (in migliaia di Euro) di 80 appartamenti scelti a caso tra quelli che si trovano in un quartiere di una città italiana.

Volendo raggruppare in classi i dati riportati in questa tabella, la prima cosa che occorre fare è osservare quali sono i valori minimo e massimo in essa riportati. Essendo questi 0,44 e 5,42, possiamo arbitrariamente pensare all'insieme S dei valori assumibili come al sottoinsieme $[0.40, 5.50] \subset \mathbb{R}$. Se vogliamo suddividere S in 5 classi, potremmo ad esempio scegliere le

seguenti:

$$C_1 = (0.40, 1.50],$$

$$C_2 = (1.50, 2.30],$$

$$C_3 = (2.30, 3.00],$$

$$C_4 = (3.00, 4.00],$$

$$C_5 = (4.00, 5.50].$$

Esercizio 1.2 *Riferendosi alla tabella 1.2 creare una matrice con*

<i>costo mq</i>	<i>freq. ass.</i>	<i>freq. rel</i>	<i>freq. cum. ass.</i>	<i>freq. cum. rel.</i>
C_1				
C_2				
C_3				
C_4				
C_5				

Creare istogrammi con Excel.

1.1 Indici di tendenza centrale

Gli indici di tendenza centrale sono delle misure in grado di sintetizzare con un solo valore numerico, i valori assunti dai dati.

i) Media e Media pesata

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.1)$$

$$\bar{x}_p = \frac{1}{s} \sum_{i=1}^n s_i x_i \quad , s_i \text{ sono i pesi} \quad s = \sum_{i=1}^n s_i \quad (1.2)$$

Proposizione 1.1 *La media rende minima la funzione*

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

ovvero la media e' quel punto che dista di meno da tutti i punti della serie di dati $\{x_i\}$.

Dim.

$$f'(x) = 2 \sum_{i=1}^n (x - x_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^n x - \sum_{i=1}^n x_i = 0 \quad \Rightarrow \quad n x = \sum_{i=1}^n x_i$$

da cui si ricava

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

Bisogna adesso dimostrare che in \bar{x} $f(x)$ è minima.

$$f''(x) = 2 \sum_{i=1}^n 1 = 2n > 0$$

allora \bar{x} è l'unico punto di minimo di $f(x)$. ■

ii) Mediana (\hat{x})

È quel numero che “sta nel mezzo”. Ovvero, ordinati i dati $\{x_i\}$ in ordine crescente allora se n è dispari la mediana è l’elemento di posto

$$\frac{n+1}{2} \quad .$$

Se n è pari la mediana non è univocamente determinata. Una possibile approssimazione si ottiene prendendo la **media aritmetica dei due valori centrali**, ovvero quelli di posto

$$\frac{n}{2} \quad , \quad \frac{n}{2} + 1 \quad .$$

Un’altra approssimazione si ottiene utilizzando un’interpolazione lineare a partire dai due valori centrali (vedi esempio).

svantaggio: risistemare i dati in ordine crescente (non nella media)

vantaggio: non dipende dai valori estremi

iii) Moda

È quel valore che si ripete più volte nella serie di dati.

Dà un andamento qualitativo dei dati e non è garantito che sia un unico numero (distrib. bi o multi-modale).

iv) Quantili, percentili, Quartili

Supponiamo di avere un insieme di dati ordinati in modo crescente. Abbiamo già visto che la mediana è il *valore che sta nel mezzo*. In analogia possiamo definire

Definizione 1.1

Si chiama quantile di ordine $\alpha \in [0, 1]$, e lo si indica con q_α , un valore per cui alla sua sinistra compare almeno il $100 \alpha\%$ delle osservazioni e alla sua destra almeno il $100 (1 - \alpha)\%$.

Alle volte, si usa il termine percentile, al posto di quantile, in questo caso α è indicata come percentuale. Per esempio $q_{0.95}$ è il novantacinquesimo percentile.

Definizione 1.2

Si dicono primo quartile, secondo quartile e terzo quartile, e si indicano con Q_1, Q_2, Q_3 i quantili, rispettivamente, di ordine 0.25, 0.5, 0.75 e quindi

$$Q_1 = q_{0.25} \quad , \quad Q_2 = q_{0.50} \quad , \quad Q_3 = q_{0.75}$$

Il secondo quartile coincide con la mediana.

Calcolo dei quartili

Come già detto per la mediana, anche i quartili ed i percentili non sono univocamente determinati. Vediamo due modi per calcolarli:

1. *media aritmetica*

In analogia con quanto già visto per il calcolo della mediana

- fissiamo $\alpha = 0.25, 0.5, 0.75$ e calcoliamo $\alpha(n+1)$
- se $\alpha(n+1) = m \in \mathbb{N}$ allora $Q_\alpha = x_m$
- se $\alpha(n+1) \notin \mathbb{N}$ allora si prende la sua parte intera ¹, che è quel numero $m \in \mathbb{N}$ tale che $m < \alpha(n+1) < m+1$ e quindi, la media aritmetica

$$Q_\alpha = \frac{x_m + x_{m+1}}{2} \tag{1.3}$$

2. *interpolazione lineare*

Supponiamo di avere due coppie di dati $P_1(x_1, y_1)$, $P_2(x_2, y_2)$ e di voler stimare il valore y^* che corrisponde a $x^* \in]x_1, x_2[$. Con il metodo dell'interpolazione lineare si stima il valore y^* tramite la retta passante per i due punti P_1 e P_2 (detta retta interpolante), ovvero

$$y^* = y_1 + \frac{y_2 - y_1}{x_2 - x_1} (x^* - x_1) \quad . \tag{1.4}$$

Utilizziamo questo metodo per stimare la mediana:

¹la parte intera di un numero x si indica con $[x]$, mentre la sua parte frazionaria è $x - [x]$.

- fissiamo $\alpha = 0.25, 0.5, 0.75$ e calcoliamo $\alpha(n+1)$
- se $\alpha(n+1) = m \in \mathbb{N}$ allora $Q_\alpha = x_m$
- se $x^* = \alpha(n+1) \notin \mathbb{N}$ allora utilizziamo la retta interpolante (1.4).

Sia $m = [x^*]$ (parte intera) e $\beta = x^* - [x^*] = x^* - m$ (parte frazionaria), allora $P_1(m, x_m)$, $P_2(m+1, x_{m+1})$ e quindi avremo

$$y^* = Q_\alpha = x_m + \frac{x_{m+1} - x_m}{m+1 - m}(x^* - m) = x_m + (x_{m+1} - x_m)\beta \quad (1.5)$$

In EXCEL la funzione QUARTILE, viene calcolata usando un' opportuna interpolazione lineare.

Esempio 1.3 *Siano assegnati i seguenti dati, che rappresentano le età di un campione di persone*

$$E = \{x_i, i = 1 : 18\} = \{16, 18, 18, 19, 20, 20, 20, 20, 21, 21, 21, 22, 23, 25, 28, 30, 31, 37\}$$

La numerosità del campione è $n=18$. Si ottiene come media 18, mediana 21 e moda 20 (calcolare con EXCEL ed anche la distribuzione delle frequenze).

Per il calcolo dei quartili si ha:

1. Per $\alpha = 0.25$ (primo quartile) $0.25(18+1) = 4.75$ quindi $m = [4.75] = 4$ e la parte frazionaria $\beta = 0.75$. Usando la media aritmetica (1.3)

$$Q_{0.25} = \frac{x_4 + x_5}{2} = \frac{19 + 20}{2} = 19.5$$

mentre con l'interpolazione lineare (1.5)

$$Q_{0.25} = 19 + 0.75(20 - 19) = 19.75$$

2. Per $\alpha = 0.5$ (secondo quartile o mediana) $0.50(18+1) = 9.5$ quindi $m = 9$ e $\beta = 0.5$. Usando la (1.3)

$$Q_{0.5} = \frac{21 + 21}{2} = 21$$

che coincide con il valore ottenuto utilizzando la (1.5).

3. Per $\alpha = 0.75$ (terzo quartile) $0.75(18+1) = 14.25$ quindi $m = 14$ e $\beta = 0.25$. Usando la (1.3)

$$Q_{0.75} = \frac{25 + 28}{2} = 26.5$$

mentre con la (1.5) si ottiene

$$Q_{0.75} = 25 + 0.25(28 - 25) = 25.75$$

Da questa analisi possiamo concludere che il 25 % delle persone del campione hanno un'età minore o uguale a 19,5 anni, il 50 % minore o uguale a 21 e il 75 % minore o uguale a 26,5.

1.2 Indici di variabilità

Può accadere che 2 serie di dati abbiano stessa media e/o mediana, ma le 2 serie sono molto diverse.

Esempio 1.4

$$E_1 = \{0.5, 0.8, 2.0, 2.7, 4.0\} \quad \bar{x}_1 = \hat{x}_1 = 2$$

$$E_2 = \{1.4, 1.7, 2.0, 2.1, 2.8\} \quad \bar{x}_2 = \hat{x}_2 = 2$$

Però i dati di E_2 sono più omogenei (cioè vicini tra loro).

Occorre pertanto definire indici che misurino il grado di variabilità o dispersione.

a) Varianza (s^2)

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

ovviamente s^2 è tanto più grande, tanto più i dati sono “distanti” dalla media.

Esempio 1.5

$$s_{E_1}^2 = 1.64, \quad s_{E_2}^2 = 0.22 \quad (\text{omogeneo!})$$

Proposizione 1.2

$$s^2 = \overline{(x^2)} - (\bar{x})^2$$

Dim.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \tag{1.6}$$

$$= \frac{1}{n} \sum_{i=1}^n (x_i^2 + \bar{x}^2 - 2x_i\bar{x}) = \tag{1.7}$$

$$= \frac{1}{n} \left[\sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{x}^2 - 2 \sum_{i=1}^n x_i\bar{x} \right] = \tag{1.8}$$

$$= \frac{1}{n} \left[\sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2\bar{x} \sum_{i=1}^n x_i \right] = \tag{1.9}$$

$$= \overline{x^2} + (\bar{x}^2) - 2(\bar{x}^2) = \overline{x^2} - (\bar{x}^2) \quad \blacksquare \tag{1.10}$$

b) Scarto quadratico medio (Deviazione standard)

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

c) Scarto medio assoluto

$$s.a. = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

Cosa fare se 2 serie di dati hanno la stessa varianza?

Esempio 1.6

$$E_3 = \{3, 4, 5, 6, 7\} \quad \bar{x}_3 = 5 \quad s_3^2 = 2.5$$

$$E_4 = \{13, 14, 15, 16, 17\} \quad \bar{x}_4 = 15 \quad s_4^2 = 2.5$$

In entrambi i casi un dato rispetto al precedente varia di 1. Ma questa variazione è più importante nelle serie di dati E_3 (che sono numeri più piccoli rispetto ad E_4) che nella serie di dati E_4 .

È logico pensare che i dati di E_3 siano più dispersi di quelli di E_4 , anche perchè $\bar{x}_4 > \bar{x}_3$. Per questo motivo si definisce un **coefficiente di variazione** (c.v.)

$$c.v. = \frac{s}{\bar{x}}$$

Per E_3 $c.v. = \frac{\sqrt{2.5}}{5}$ per E_4 $c.v. = \frac{\sqrt{2.5}}{15}$. A valori maggiori del $c.v.$ corrisponde una maggiore variabilità dei dati.

Esempio 1.7 Nella seguente tabella vengono riportati il prezzo (in EURO/LITRO) di un particolare combustile e la frequenza con cui esso viene venduto giornalmente in un distributore.

<i>prezzi</i>	14.5	16.8	12.3	10.7	11.4	18.1	20.6	13.8
<i>frequenze</i>	7	5	8	12	10	6	4	11

Calcolare media, varianza, moda e mediana.

In questo caso:

- la media e' quella pesata (1.2) con x_i i prezzi ed i pesi s_i dati dalle frequenze.
- La varianza si calcola

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_p)^2 f_i}{\sum_{i=1}^k f_i}$$

dove $k = 8$.

- Per prima cosa si devono ordinare i dati in modo crescente nel prezzo. Sia N la somma delle frequenze. Essendo $N = 63$ (dispari) allora la mediana e' elemento di posto $64/2=32$ e quindi pari a 13.8

1.2.1 Dati raggruppati per classi

Supponiamo che gli n dati sono raggruppati in classi C_i ($i = 1, \dots, k$), come nell'esempio 1.2.

Detto m_i il valore centrale della classe C_i , allora la media e la varianza sono così definite

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k m_i f_i \quad , \quad s^2 = \frac{1}{n} \sum_{i=1}^k (m_i - \bar{x})^2 f_i \quad (1.11)$$

dove f_i è la frequenza assoluta della classe.

Esempio 1.8 *Calcolare la media e la varianza dell'esempio 1.2.*

In questo caso il numero di dati è $n = 80$, il numero di classi $k = 5$ e

$$C_1 = (0.40, 1.50] \quad , \quad m_1 = \frac{0.4 + 1.5}{2} = \dots$$

$$C_2 = (1.50, 2.30] \quad , \quad m_2 = \frac{1.5 + 2.3}{2} = \dots$$

$$C_3 = (2.30, 3.00] \quad , \quad m_3 = \frac{2.3 + 3.0}{2} = \dots$$

$$C_4 = (3.00, 4.00] \quad , \quad m_4 = \frac{3 + 4}{2} = \dots$$

$$C_5 = (4.00, 5.50] \quad , \quad m_5 = \frac{4 + 5.5}{2} = \dots$$

applicando la (1.11) si ottiene

Per il calcolo dei quartili si applica l'interpolazione lineare come nel seguente esempio.

Esempio 1.9 *La tabella che segue mostra la distribuzione di frequenza degli stipendi settimanali di 65 impiegati di una ditta. Calcolare la media, la varianza ed i quartili.*

Tabella

stipendi (classe)	Numero impiegati (frequenza)
50 - 59.99 €	8
60 - 69.99 €	10
70 - 79.99 €	16
80 - 89.99 €	14
90 - 99.99 €	10
100 - 109.99 €	5
110 - 119.99 €	2

Per il calcolo del valore medio e della varianza applichiamo la formula (1.11), dove, per esempio, il valore centrale della prima classe di stipendi e'

$$\frac{50 + 59.99}{2} = 54.995 \quad .$$

Nel nostro esempio, ho delle classi di stipendio con la relativa frequenza, che da' la posizione dei dati. Poiche' le classi sono intervalli numerici, utilizzeremo per il calcolo dei quartili l'interpolazione lineare. Il primo quartile Q_1 e' lo stipendio che occupa la posizione $n \times 0.25 = 65/4 = 16.25$. Per capire qual'e' lo stipendio che corrisponde a questa posizione, bisogna prendere la colonna delle frequenze. La prima classe corrisponde alla posizione 8 e quindi ne servono 8.25 della seconda classe (infatti $8 + 8.25 = 16.25$). Ora l'incremento di stipendio della seconda classe e' 10 € con frequenza di 10 e quindi se x e' l'incremento di stipendio con posizione 8.25 sara' valida la seguente proporzione

$$10 : 10 = x : 8.25 \quad \rightarrow \quad x = 8.25$$

pertanto

$$Q_1 = 59.99 + 8.25 = 68.24 \text{ €} \quad .$$

Analogamente Q_2 e' lo stipendio che occupa la posizione $n \times 0.5 = 65/2 = 32.5$. Dato che le prime due classi hanno frequenze totali pari a 18, dobbiamo prendere 14.5 della terza classe (infatti $18 + 14.5 = 32.5$) . L'incremento di stipendio della terza classe e' 10 con frequenza 16 e quindi se x e' l'incremento di stipendio con posizione 14.5 sara'

$$10 : 16 = x : 14.5 \quad \rightarrow \quad x = 9.06$$

pertanto

$$Q_2 = 69.99 + 9.06 = 79.05 \text{ €} \quad .$$

Analogamente Q_3 e' lo stipendio che occupa la posizione $65 \times 0.75 = 48.75$. Le prime 4 classi hanno frequenza totale 48, dobbiamo prendere 0.75 della quinta classe, cioe'

$$10 : 10 = x : 0.75 \quad \rightarrow \quad x = 0.75$$

pertanto

$$Q_3 = 89.99 + 0.75 = 90.74 \text{ €} \quad .$$

In conclusione, il 25% degli impiegati guadagna fino a 68.24 €, il 50% fino a 79.05 €, il 75% fino a 90.74 €

1.3 Indici di forma

Due indici statistici numerici che tengono conto della forma di una distribuzione di una serie di dati sono:

a) **Asimmetria:** E' una misura dello scostamento di una distribuzione dalla simmetria. Se la curva di frequenza di una distribuzione ha una coda piu' lunga a destra del massimo centrale, piuttosto che a sinistra, la distribuzione si dice *positivamente asimmetrica*. Se e' vero il contrario si dice *negativamente asimmetrica*.

Definizione 1.3

Date n osservazioni x_1, x_2, \dots, x_n e' detto indice di asimmetria (o skewness), la quantita'

$$sk = \frac{m_3}{\sqrt{m_2^3}} \quad , \quad m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad k = 2, 3, 4, \dots$$

dove m_k sono i momenti centrati di ordine k .

Questo indice indica se la distribuzione del campione e' simmetrica rispetto alla media ($sk=0$): se $sk > 0$ la distribuzione sara' piu' concentrata a sinistra, con una coda piu' lunga a destra, il contrario se $sk < 0$.

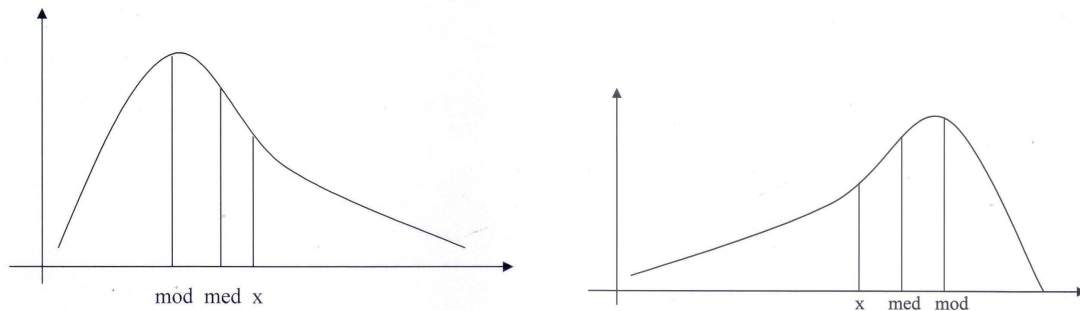


Figura 1.1: Figura sinistra: indice di asimmetria: $sk > 0$ (asimmetria a destra). Figura destra: indice di asimmetria: $sk < 0$ (asimmetria a sinistra)

a) **Curtosi:** E' una misura dell'appiattimento di una distribuzione di dati rispetto alla distribuzione normale (gaussiana).

Definizione 1.4

Date n osservazioni x_1, x_2, \dots, x_n e' detta Curtosi, la quantita'

$$\kappa = \frac{m_4}{m_2^2}$$

Si prova che, se $\kappa > 3$ allora la distribuzione (detta leptocurtica) e' piu' appuntita rispetto alla normale (con code piu' grandi), se $\kappa < 3$ (platicurtica) e' piu' appiattita (con code piu' piccole), infine se $\kappa = 3$ ha la stessa altezza di una normale.

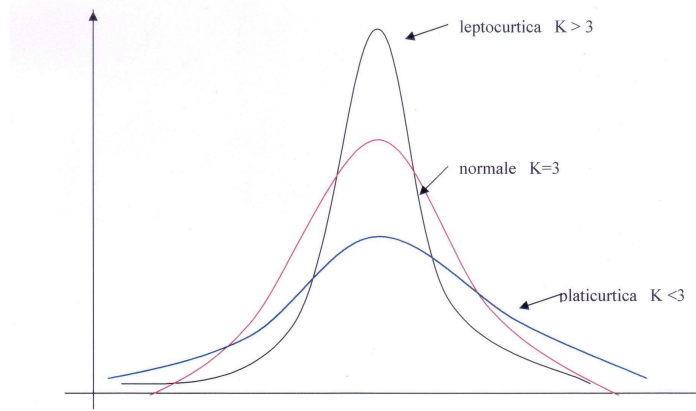


Figura 1.2: La curtosi

1.4 La correlazione tra due serie di dati

Supponiamo di avere due set di dati $\{x_i\}$, $\{y_i\}$ di numerosità n . Ci chiediamo se esiste una certa relazione tra questi dati ovvero se sono tra loro indipendenti. Per rispondere a questa domanda si può confrontare le variazioni delle coppie di dati rispetto ai rispettivi valori medi.

$$x_i - \bar{x} \quad y_i - \bar{y}.$$

È ovvio supporre che esista una relazione di dipendenza tra $\{x_i\}$ e $\{y_i\}$ se $x_i - \bar{x}$, $y_i - \bar{y}$ hanno lo stesso segno.

Quindi, tanto più i prodotti $(x_i - \bar{x})(y_i - \bar{y})$ hanno concordanza di segno, tanto più i dati considerati hanno forte dipendenza. Anche nel caso in cui a valori positivi di $(x_i - \bar{x})$ corrispondono valori negativi di $(y_i - \bar{y})$ o viceversa, denota una forte dipendenza tra i dati considerati.

Invece, se tutti i prodotti $(x_i - \bar{x})(y_i - \bar{y})$ hanno segni diversi, la loro somma risulta piccola in valore assoluto (ovvero tende a zero) e potrebbe esserci indipendenza tra le due serie di dati.

Definizione 1.5 Si definisce *covarianza*

$$c_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Teorema 1.1 *Si prova facilmente che $c_{xy} = \overline{xy} - \bar{x} \bar{y}$*

Definizione 1.6 Due serie di dati x_i, y_i sono *statisticamente incorrelati* se:

$$c_{xy} = 0 \quad .$$

Da quanto detto questo indice è tale che:

- i) $c_{xy} \in \mathbb{R}$
- ii) se $\{x_i\}, \{y_i\}$ sono fortemente correlati c_{xy} è grande in valore assoluto;
- iii) Se $c_{xy} > 0$ i due set di dati si dicono *correlati positivamente*, e questo significa che si muovono nella stessa direzione (all'aumentare dell'uno l'altro aumenta e viceversa). Viceversa se $c_{xy} < 0$ i due set di dati si dicono *correlati negativamente*.
- iv) se $\{x_i\}, \{y_i\}$ sono *statisticamente incorrelati*, dal teorema precedente, si ha che $\overline{xy} \simeq \bar{x} \bar{y}$ ovvero la media del prodotto delle due serie di dati (\overline{xy}) è uguale al prodotto delle medie delle singole serie di dati ($\bar{x} \bar{y}$). Ma questo non ci assicura che le due serie di dati siano tra loro indipendenti. Vedremo più avanti un esempio in cui le due serie di dati hanno covarianza piccola senza essere per questo indipendenti.

1.5 Analisi di regressione per una serie di dati

Assegnato un insieme E di coppie di dati ci domandiamo se esiste un legame funzionale del tipo

$$y = f(x)$$

che descriva bene la relazione tra i dati.

Un'analisi di questo tipo si chiama **analisi di regressione**.

A questo punto come si fa a determinare la f , che al suo interno contiene dei parametri in modo che questo legame sia buono?

1.5.1 Metodo dei minimi quadrati

Si cerca f tale che sia minima la funzione *residuo*

$$g(f) = \sum_{i=1}^n [f(x_i) - y_i]^2$$

1.5.2 Regressione lineare

In questo caso la funzione f è una retta

$$f(x) = mx + q$$

$$g(m, q) = \sum_{i=1}^n [mx_i + q - y_i]^2$$

(incognite m, q !)

Poiché $g(m, q)$ è una funzione di due variabili:

condizione necessaria e sufficiente affinché $P_ = (m_*, q_*)$ sia minimo relativo è che*

i)

$$\left. \frac{\partial g}{\partial m} \right|_{P_*} = 0 \quad , \quad \left. \frac{\partial g}{\partial q} \right|_{P_*} = 0$$

ii)

$$H(m_*, q_*) = g_{mm}g_{qq} - (g_{mq})^2 > 0$$

iii)

$$g_{qq}(m_*, q_*) > 0$$

dove

$$g_m = \frac{\partial g}{\partial m} = \lim_{h \rightarrow 0} \frac{g(m+h, q) - g(m, q)}{h}$$

$$g_{mm} = \frac{\partial^2 g}{\partial m^2} \quad , \quad g_{mq} = \frac{\partial^2 g}{\partial m \partial q}$$

$$\frac{\partial g}{\partial m} = 2 \sum_{i=1}^n (mx_i + q - y_i)x_i$$

$$\frac{\partial g}{\partial q} = 2 \sum_{i=1}^n (mx_i + q - y_i)$$

$$\frac{\partial^2 g}{\partial m^2} = 2 \sum_{i=1}^n x_i^2 > 0$$

$$\frac{\partial^2 g}{\partial q^2} = 2n > 0$$

$$\frac{\partial^2 g}{\partial m \partial q} = 2 \sum_{i=1}^n x_i > 0 \quad \text{Provare che } H(m, q) > 0!$$

Risolviamo ora il sistema:

$$\begin{cases} \sum_i (mx_i + q - y_i)x_i = 0 \\ \sum_i (mx_i + q - y_i) = 0 \end{cases}$$

$$\begin{cases} \sum_i mx_i^2 + \sum_i qx_i - \sum_i x_i y_i = 0 \\ \sum_i mx_i + nq - \sum_i y_i = 0 \end{cases}$$

dividendo per n si ottiene:

$$\begin{cases} \frac{m}{n} \sum_i x_i^2 + q\bar{x} - \frac{1}{n} \sum_i x_i y_i = 0 \\ m\bar{x} + q - \bar{y} = 0 \Rightarrow q = \bar{y} - m\bar{x} \end{cases}$$

$$m \frac{\sum_i x_i^2}{n} + \bar{x}\bar{y} - m(\bar{x})^2 - \frac{1}{n} \sum_i x_i y_i = 0 \quad , \quad m = \frac{\frac{1}{n} \sum_i x_i y_i - \bar{x}\bar{y}}{\frac{\sum_i x_i^2}{n} - (\bar{x})^2}$$

Da cui ricordando la covarianza

$$c_{xy} = \frac{1}{n} \sum_i x_i y_i - \bar{x}\bar{y}$$

e la varianza per la variabile x

$$s_x^2 = \frac{\sum_i x_i^2}{n} - (\bar{x})^2$$

da cui

$$m = \frac{c_{xy}}{s_x^2} \quad , \quad q = \bar{y} - \frac{c_{xy}}{s_x^2} \bar{x}$$

Il metodo applicato fornisce la retta che meglio approssima i dati, ma non il grado di approssimazione. Per questo motivo si introduce il **coefficiente di correlazione lineare (o di Pearson)**

$$r_{xy} = \frac{c_{xy}}{s_x s_y}$$

Proposizione 1.3

- i) $r_{xy} \in [-1, 1]$
- ii) se $r_{xy} = \pm 1$ i dati (x_i, y_i) sono perfettamente allineati con la retta di regressione
- ii) se $r_{xy} > 0$ la retta è ascendente

NB: nella pratica, se $|r_{xy}| < 0.9$, i dati si allontanano dall'andamento rettilineo.

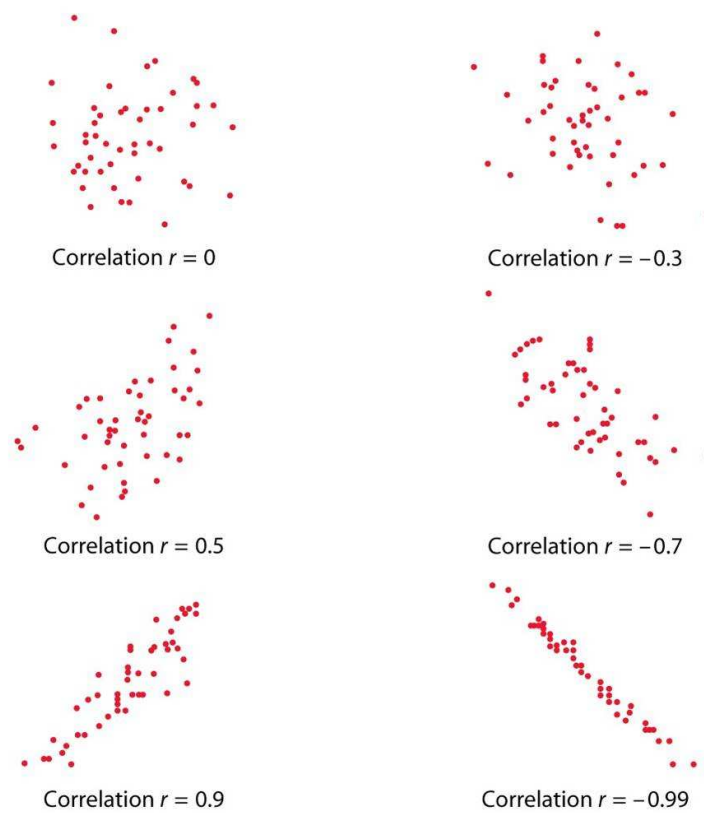


Figura 1.3: Indice di correlazione $r = r_{xy}$

Esempio 1.10 *Tabella carichi di rottura* Si deve controllare la resistenza di un campione di 15 travi di cemento, tutte ottenute dalla stessa gittata, misurando sia i carichi di prima lesione X_i che quelli di rottura finale Y_i (in Kg), come in tabella 1.3.

Con EXCEL calcolare: s_x , s_y , c_{xy} , r_{xy} , \bar{x} , \bar{y} , retta di regressione e grafico dei dati x,y con retta di regressione.

Esempio 1.11 *Assegnati i set di dati in tabella 1.4, calcolare la covarianza ed il coefficiente di Pearson. Si osservi che la covarianza tende a zero ma non il coefficiente di Pearson !*

Tabella 1.3: Carichi di rottura.

I^a lesione	rottura
2550	4650
2900	4650
3000	4700
3000	4750
3000	4775
3000	4775
3250	4800
3250	4950
3250	5050
3600	5100
4225	5100
4650	5150
4750	5175
5175	5250
5300	5300

Tabella 1.4: Esempio sulla covarianza ed indipendenza

x	0.185	0.22	0.233	0.247	0.255	0.2745
y	0.049	0.053	0.054	0.0565	0.058	0.0605

1.5.3 Parabola dei minimi quadrati

Regressione non lineare

In questo caso

$$f(x) = a + bx + cx^2$$

a questo punto devo rendere minimo

$$g(a, b, c) = \sum_{i=1}^n [a + bx_i + cx_i^2 - y_i]^2$$

allora

$$\frac{\partial g}{\partial a} = 0 \quad , \quad \frac{\partial g}{\partial b} = 0 \quad , \quad \frac{\partial g}{\partial c} = 0$$

$$\left\{ \begin{array}{l} \sum_i y_i = an + b \sum_i x_i + c \sum_i x_i^2 \\ \sum_i x_i y_i = a \sum_i x_i + b \sum_i x_i^2 + c \sum_i x_i^3 \\ \sum_i x_i^2 y_i = a \sum_i x_i^2 + b \sum_i x_i^3 + c \sum_i x_i^4 \end{array} \right.$$

Sistema di 3 equazioni in 3 incognite

Esempio 1.12 Trovare la soluzione del sistema lineare 3x3 con EXCEL

$$AX = B$$

Se $\det A \neq 0$ (teorema di Cramer) allora

$$\exists A^{-1} \quad t.c. \quad A^{-1}AX = A^{-1}B$$

quindi la soluzione e'

$$X = A^{-1}B$$

NB:

$$A \rightarrow n \times n, X \rightarrow n \times 1, B \rightarrow n \times 1$$

Esempio 1.13 Utilizzando la tabella dei carichi di rottura 1.3, calcolare la parabola dei minimi quadrati.

Un modo grossolano per vedere la bontà dell'approssimazione è calcolare il *residuo*:

$$g(\bar{a}, \bar{b}, \bar{c}) = \sum_i [\bar{a} + \bar{b}x_i + \bar{c}x_i^2 - y_i]^2$$

- Confrontare il valore ottenuto con la retta di regressione
- Mostrare il grafico dei dati e la parabola di regressione

Esempio 1.14 *Popolazione USA: in tabella 1.5 sono assegnati per alcuni anni la popolazione degli Stati Uniti d'America.*

Si vogliono fare delle stime di crescita di questa popolazione. In particolare:

- Calcolare la retta di regressione e r_{xy} ;
- Calcolare la parabola di regressione;
- Approssimare i dati con la curva esponenziale

$$y = ae^{bx}$$

In questo caso ci possiamo ricondurre al caso della regressione lineare :

$$\ln y = \ln(ae^{bx})$$

$$\ln y = \ln a + \ln e^{bx} = \ln a + bx$$

con un cambiamento di variabili, si ha

$$\tilde{y} = \alpha + bx$$

- Approssimare i dati con una legge esponenziale $y = b m^x$ (utilizzare REGR.LOG)
- Grafico dei dati sperimentali con quelli di regressione.

Tabella 1.5: Popolazione USA

Anno	Popol(mln)
1840	17.1
1850	23.2
1860	31.4
1870	39.8
1880	50.2
1890	62.9
1900	76.0
1910	92.0
1920	105.7
1930	122.8
1940	131.7
1950	151.1
1960	179.3

Suggerimento: scegliere x in modo che:

.....

anno 1890 $\rightarrow x=-1$

anno 1900 $\rightarrow x=0$

anno 1910 $\rightarrow x=1$

anno 1920 $\rightarrow x=2$

.....

Capitolo 2

Elementi di Probabilità

La probabilità riguarda il nostro modo di conoscere. Es: qual'è la probabilità che domani piova? Qual'è la probabilità che faccia un ambo?

Ad ogni evento si associa un numero

$$P \in [0, 1]$$

detto *probabilità*. Come definirla?

1. Probabilità classica (*Laplace* ~ 1600)

$$P = \frac{n.casi\ favorevoli}{n.casi\ possibili}$$

supponendo tutti i casi egualmente possibili

- Esempio: lancio di una moneta, di un dado, etc.
- Difetti: Abbiamo un circolo vizioso; dire che gli eventi sono egualmente possibili è lo stesso di dire che sono ugualmente probabili ! Inoltre si deve conoscere il numero dei casi.

2. Probabilità frequentista (*VonMises* ~ 1900)

È il limite del rapporto tra gli esperimenti favorevoli ed il totale di quelli effettuati per n° di esperimenti $\rightarrow \infty$, e supponendo di ripetere gli esperimenti in condizioni identiche.

- Esempio: lanciando una monetina, calcolare la probabilità che esca testa

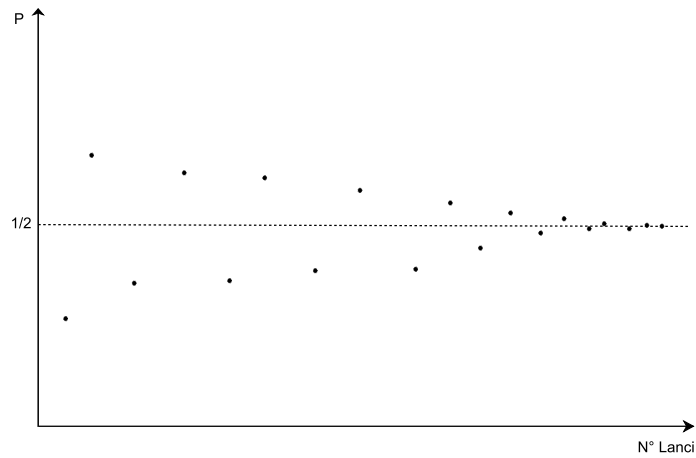


Figura 2.1: Lancio di una moneta

- Difetto: devo fare ∞ esperimenti nelle stesse condizioni!

3. Probabilità soggettivista (*DeFinetti* \sim 1930)

La probabilità dell'evento A è la misura del grado di fiducia che un individuo coerente attribuisce al verificarsi dell'evento

grado di fiducia: è il prezzo che l'individuo ritiene equo pagare nel fare una scommessa per vincere se l'evento si verifica.

coerente: l'individuo deve essere disposto ad essere sia lo scommettitore che il bookmaker una volta fissato il prezzo.

- Difetto: visione troppo soggettivista, variabile da persona a persona.

Possiamo dare la seguente definizione di **probabilità soggettiva dell'evento E**

$$P(E) = \frac{\text{prezzo da pagare}}{\text{somma ricevuta al verificarsi di } E}$$

4. Definizione assiomatica di probabilità (*Kolmogorov*1933)

Si abbandona una teoria sperimentale per la definizione di probabilità e si costruisce una teoria assiomatica logico-deduttiva.

In modo rigoroso occorre introdurre le algebre di Boole (σ -algebre)

• Esempio: Lancio del dado

l'esito del lancio sarà uno dei 6 numeri detti *eventi elementari*, mentre l'insieme degli eventi si chiama *spazio campione* Ω

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

a) Ω è lo spazio campione, formato dai possibili esiti.

b) $A \subseteq \Omega$ è l'evento elementare o composto

c) due eventi A e B sono *incompatibili* se sono 2 sottoinsiemi di Ω disgiunti (senza parti in comune ovvero $A \cap B = \emptyset$)

d) chiamo insiemi delle parti di Ω , $\mathcal{P}(\Omega)$ l'insieme dei suoi sottoinsiemi.

Questa identificazione tra eventi e sottoinsiemi di Ω permette di trasportare agli eventi le operazioni di \cap , \cup e il passaggio al complementare. Il significato intuitivo di queste operazioni riferite agli eventi è facile: se A e B sono sottoinsiemi di Ω corrispondenti a due eventi allora

i) $A \cap B$ corrisponderà all'evento (detto prodotto logico): “i due eventi associati ad A e B si verificano entrambi”;

ii) $A \cup B$ corrisponderà all'evento (detto somma logica): “uno almeno dei due eventi si verifica”;

iii) A^C corrisponderà all'evento: “l'evento associato ad A non si verifica”.

In questa identificazione Ω sarà l'evento certo, cioè quello che si verifica sicuramente, mentre l'insieme vuoto \emptyset corrisponde all'evento impossibile, quello che certamente non si verifica.

Definizione 2.1 *Probabilità secondo Kolmogorov*

Chiamo *misura di probabilità* ogni applicazione

$$P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}_+$$

(che associa ad ogni sottoinsieme di Ω un numero reale) con le proprietà

$$1) \quad \forall A \subseteq \Omega \quad \exists! P(A) \geq 0$$

$$2) \quad P(\Omega) = 1$$

3) Data la famiglia di eventi incompatibili $\{A_i, i = 1, \dots, N\}$ ¹ allora:

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

NB: La misura di probabilità $P(A)$ associa valori numerici non ai singoli elementi di Ω ma più in generale a sottoinsiemi di Ω . Esempio: probabilità che nel lancio del dado esca un numero dispari. Questa definizione tuttavia non specifica i valori associati ai singoli eventi.

Definizione 2.2 *Spazio di probabilità*

Chiameremo spazio di probabilità la terna (Ω, \mathcal{P}, P) .

Esempio 2.1 *Estrazione di una pallina da un'urna che contenga 3 colori*

$$\Omega = \{R, G, B\}$$

Posso definire la misura di probabilità in questo modo:

$$P(\emptyset) = 0$$

$$P(R) = P(B) = P(G) = \frac{1}{3}$$

$$P(R, G) = P(R, B) = P(B, G) = \frac{2}{3}$$

$$P(\Omega) = 1$$

¹Due eventi sono incompatibili se sono disgiunti, cioè $A_i \cap A_j = \emptyset$, $\forall i \neq j$.

dove $P(B, G)$ è la probabilità che in una estrazione si abbia o B o G . Questa definizione soddisfa le probabilità assiomatiche. Si possono dare anche altre misure di probabilità.

Esempio 2.2 *Probabilità di fare terno al lotto su una ruota.*

In questo caso lo spazio campione è

$$\Omega = \{1, 2, 3, \dots, 90\}$$

$A \subset \Omega$ dove A è formato da 3 numeri di Ω

allora definisco

$$P(A) = \frac{n.casi\ favorevoli}{n.casi\ possibili}$$

che è una misura di probabilità.

Il numero di casi possibili sono tutte le cinque estraibili da Ω , che sono

$$\binom{90}{5}$$

Il numero di casi favorevoli è dato dalle cinque che contengono i 3 numeri giocati, cioè che contengono 2 altri numeri fra i restanti 87:

$$\binom{87}{2}$$

$$\text{allora } P = \frac{\binom{87}{2}}{\binom{90}{5}} \simeq 0,0000851$$

Esempio 2.3 *A una corsa di cavalli una persona è disposta a pagare 90 Euro per ricevere 120 Euro in caso di vittoria di un determinato cavallo (evento E). Dalla definizione di probabilità soggettivista si ha:*

$$P(E) = \frac{90}{120} = \frac{3}{4}$$

Per la condizione di coerenza, deve essere disposto, scambiando i ruoli, a ricevere 90 Euro e pagare 120 Euro in caso di vittoria del cavallo. Si dice che la vittoria del cavallo è pagata 4 a 3. Quindi se si scommettono 30 Euro, in caso di vittoria del cavallo, si riceveranno $30 \times 4/3 = 40$ Euro.

2.1 Probabilità della somma logica di eventi

Abbiamo già visto che l'evento somma logica $A \cup B$ consiste nel fatto che almeno uno dei due eventi si verifica. Vale il seguente teorema:

Teorema 2.1 *Teorema della somma logica di due eventi*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

In particolare se i due eventi sono *incompatibili* allora

$$P(A \cup B) = P(A) + P(B) \quad . \quad (2.1)$$

Esempio 2.4 *Un'urna contiene 15 palline numerate da 1 a 15. Calcolare la probabilità che, estraendo una pallina,*

a) *esca un numero dispari o maggiore di 10;*

b) *esca un numero minore di 6 o maggiore di 10;.*

Definiamo gli eventi:

A = esce un numero dispari , B = esce un numero maggiore di 10

C = esce un numero minore di 6 .

Dalla definizione di probabilità classica

$$P(A) = \frac{8}{15} \quad , \quad P(B) = \frac{5}{15} \quad , \quad P(C) = \frac{5}{15} \quad , \quad P(A \cap B) = \frac{3}{15}$$

dal teorema segue

a)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{2}{3}$$

b) Essendo gli eventi B, C incompatibili

$$P(B \cup C) = P(B) + P(C) = \frac{2}{3}$$

2.2 Probabilità condizionata

Un concetto molto importante nella teoria della probabilità è quello di *Probabilità condizionata*. Supponiamo inoltre di sapere che un certo evento B si è verificato. In genere, questa informazione dovrebbe alterare le probabilità che assegniamo agli altri eventi. In particolare, se A è un altro evento, allora A si verifica se e solo se si verificano sia A che B ; di fatto, lo spazio campionario si è ridotto a B . Quindi, la probabilità di A , data la conoscenza del fatto che B si è verificato, dovrebbe essere proporzionale a $P(A \cap B)$. In ogni caso, la probabilità condizionata, dato il verificarsi di B dev'essere sempre una misura di probabilità, ovvero deve soddisfare gli assiomi di Kolmogorov. Ciò fa sì che la costante di proporzionalità debba essere $1 / P(B)$. Pertanto, si giunge alla seguente definizione

Definizione 2.3 - *Probabilità condizionata*

Sia Ω uno spazio campione e P una sua misura di probabilità.

Siano A e B due eventi con $P(B), P(A) > 0$ allora si chiama probabilità di A *condizionata dall'evento B* la quantità

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (2.2)$$

dove $P(A \cap B)$ è la probabilità che A e B si verifichino contemporaneamente.

Osservazioni:

1. la probabilità condizionata ha senso solo se l'evento B si può verificare (non è l'evento impossibile), altrimenti non ha senso.
2. la probabilità condizionata di due insiemi disgiunti è zero. Poiché per definizione, due insiemi si dicono disgiunti se non si intersecano mai, la probabilità dell'intersezione sarà zero (probabilità dell'evento impossibile). Dunque, se $A \cap B = \emptyset$, allora $P(A \cap B) = 0$ e $P(A | B) = 0$;

3. se $A \subset B$ allora sarà anche $A \cap B = A$ e quindi

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

Se B è l'insieme dei valori possibili (cioè coincide con Ω), ritroviamo la definizione di probabilità casi favorevoli/ casi possibili come un caso particolare della probabilità condizionata.

Teorema 2.2 *Sia Ω lo spazio campione e $A, B \subset \Omega$. Allora*

$$P(A|B) = \frac{\text{numero di elementi di } A \cap B}{\text{numero di elementi di } B}$$

Dim. Infatti se Ω è uno spazio finito e con $|A|$ indichiamo il numero degli elementi dell'evento A , si avrà

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

da cui

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

Esempio 2.5 Si lanci una coppia di dadi. Se la somma è 6 (evento B) si determini la probabilità che uno dei due dadi abbia come esito 2 (evento A).

In questo caso si avrà:

$$B = \{\text{somma dei dadi } 6\} = \{(1, 5)(2, 4)(3, 3)(4, 2)(5, 1)\} \quad \rightarrow \quad |B| = 5$$

$$A = \{\text{un } 2 \text{ si presenta su almeno un dado}\}$$

ne segue che

$$A \cap B = \{(2, 4)(4, 2)\} \quad \rightarrow \quad |A \cap B| = 2$$

quindi

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|} = \frac{2}{5}$$

2.3 Probabilità del prodotto logico di eventi

L'evento prodotto logico $A \cap B$ consiste nel fatto che i due eventi A e B si verificano entrambi.

Vale il seguente teorema:

Teorema 2.3 *Teorema della probabilità composta*

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

Il teorema è una conseguenza della definizione di Probabilità condizionata e dice che la probabilità che due eventi si verifichino contemporaneamente è pari alla probabilità di uno dei due eventi moltiplicato con la probabilità dell'altro evento condizionato al verificarsi del primo.

Esempio 2.6 *Una scatola contiene 3 palline bianche e 2 nere.*

Sia A l'evento "la prima pallina estratta è nera". Sia B l'evento "la seconda pallina estratta è nera."

Supponiamo che le palline una volta estratte non vengono reintrodotti nella scatola.

A e B sono eventi *dependenti*.

$$P(A) = \frac{2}{3+2} = \frac{2}{5} \longrightarrow \text{probabilità che la prima pallina sia nera}$$

$$P(B|A) = \frac{1}{3+1} = \frac{1}{4} \longrightarrow \text{probabilità condizionata che la 2ª pallina sia nera supposto che la 1ª sia nera.}$$

$P(A \cap B)$ è la probabilità che si verifichino A e B .

Quindi

$$P(A \cap B) = P(A)P(B|A) = \frac{1}{10}$$

Definizione 2.4 - Due eventi A e B sono *stocasticamente indipendenti* se

$$P(A|B) = P(A) \quad \text{ovvero} \quad P(B|A) = P(B)$$

\Downarrow

$$P(A \cap B) = P(A) \cdot P(B) \tag{2.3}$$

ovvero il presentarsi dell'evento B non influenza la probabilità dell'evento A (e viceversa).

Teorema 2.4 *Teorema della probabilità assoluta (o totale)*

Sia A_i una successione di insiemi che formano una partizione² dello spazio campionario Ω e B un qualsiasi evento (dipendente dagli eventi A_i), allora:

$$P(B) = \sum_{i=1}^N P(A_i \cap B) = \sum_{i=1}^N P(A_i)P(B|A_i) \quad (2.4)$$

Esempio 2.7 *Supponiamo di avere due urne A_1 e A_2 di cui la prima contiene 8 palline rosse e 2 nere, la seconda 8 palline rosse e 8 nere. Scegliamo un'urna a caso ed estraiamo una pallina. Qual'è la probabilità che la pallina estratta sia rossa?*

In questo caso lo spazio campionario Ω è l'unione delle due urne che formano una sua partizione. L'evento B = estraggo da una delle due urne una pallina rossa'.

$$P(A_1) = \frac{1}{2} \longrightarrow \text{probabilità di scegliere l'urna } A_1$$

$$P(A_2) = \frac{1}{2} \longrightarrow \text{probabilità di scegliere l'urna } A_2$$

$$P(B|A_1) = \frac{8}{8+2} = \frac{4}{5} \longrightarrow \text{probabilità condizionata di estrarre una pallina rossa supposto di scegliere l'urna } A_1.$$

$$P(B|A_2) = \frac{8}{8+8} = \frac{1}{2} \longrightarrow \text{probabilità condizionata di estrarre una pallina rossa supposto di scegliere l'urna } A_2.$$

Dalla (2.4) segue che

$$P(B) = \frac{1}{2} \frac{4}{5} + \frac{1}{2} \frac{1}{2} = \frac{13}{20}$$

²cioè $A_i \subset \Omega, A_i \cap A_j = \emptyset$, e $\forall i \neq j \bigcup_{i=1}^N A_i = \Omega$

2.4 Il teorema di Bayes

Spesso, l'osservazione di un dato sperimentale, che chiameremo evento B , è condizionata da una diversa situazione iniziale, che chiameremo ipotesi A_i . Per fare un esempio concreto, possiamo immaginare l'osservazione di un evento $B = \text{fulmine}$, che in qualche modo dipende dalla direzione del vento (ipotesi $A_i = \text{maestrale, libeccio, bora, ...}$). Supponiamo che una qualsiasi ipotesi escluda tutte le altre e che la somma di tutte le ipotesi riempia lo spazio totale delle ipotesi (la rosa dei venti che è lo spazio campionario Ω), cioè gli insiemi A_i formano una partizione di Ω . Possiamo calcolare la probabilità totale di osservare un fulmine $P(B)$ con il teorema della probabilità assoluta (2.4), supposto di conoscere :

- la probabilità $P(B|A_i)$ di vedere un fulmine in presenza di un certo tipo di vento ;
- le probabilità $P(A_i)$ che ci sia un certo vento.

Con queste informazioni possiamo calcolare $P(A_i|B)$ cioè *la probabilità che se c'è il vento A_i si veda un fulmine ?* La soluzione ci è data dal seguente teorema.

Teorema 2.5 *Teorema di Bayes (Londra 1702-1761)*

Sia A_i una successione di insiemi che formano una partizione dello spazio campionario Ω e B un qualsiasi evento (dipendente dagli eventi A_i), allora:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^N P(A_i)P(B|A_i)} \quad (2.5)$$

Dim. Dalla definizione di probabilità condizionata e dal teorema della probabilità composta si ha

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

infine dal teorema della probabilità assoluta, avremo anche

$$\frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^N P(A_i)P(B|A_i)} \quad .$$

Osservazione

Questo teorema può essere utilizzato in due modi diversi, con un approccio di tipo soggettivista quando non sono note a priori le probabilità $P(A_i)$, o con un approccio di tipo frequentista quando si assumono date le probabilità $P(A_i)$. In entrambi in casi, le probabilità condizionate provengono da una misura.

Questo teorema si applica in tantissimi casi pratici. In epidemiologia, si usa per ottenere la probabilità di malattia in un gruppo di persone con della caratteristica sulla base della percentuale complessiva di quella malattia, e delle probabilità di quella caratteristica in individui sani e malati. La domanda più familiare è in analisi di decisione clinica dove si usa per valutare la probabilità di una particolare diagnosi data l'aspetto di dei sintomi o risultato di prova.

Esempio 2.8 Una popolazione si compone del 40 % di fumatori (F) e per il 60 % di non fumatori (N). Si sa che il 25 % dei fumatori ed il 7 % dei non fumatori sono affetti da una malattia respiratoria cronica (M). Qual'è la probabilità che una persona affetta da malattia respiratoria sia un fumatore ?

In questo caso gli insiemi F e N formano una partizione A_i dello spazio campionario Ω (formato da tutta la popolazione). Inoltre avremo le seguenti probabilità:

$$P(F) = 0.4 \quad , P(N) = 0.6 \quad , P(M|F) = 0.25 \quad , P(M|N) = 0.07$$

$$P(M) = P(F)P(M|F) + P(N)P(M|N) = 0.142$$

Quindi applicando la formula di Bayes con $A_i = F$ e $B = M$:

$$P(F|M) = \frac{P(F)P(M|F)}{P(M)} = 0.704$$

Esempio 2.9 In un negozio di componenti hw per l'informatica, una fornitura di 100 moduli RAM proviene da tre costruttori differenti A, B, C . Sappiamo che il 50 % della fornitura viene dalla fabbrica A , l'altra metà è equamente distribuita tra le fabbriche B e C . I prodotti delle fabbriche hanno prezzi differenti, quelli della fabbrica A costano meno a discapito dei controlli di qualità, difatti il 5 % dei moduli prodotti risulta difettoso, per la fabbrica B tale percentuale scende al 2 % ed i difetti si riducono al 1 % per la fabbrica C . Calcolare la probabilità che, supposto il modulo risulti difettoso, sia stato prodotto da C .

In questo caso l'insieme Ω è formato dai 100 moduli ram e la sua partizione è formata dai moduli C_A, C_B, C_C che provengono dai tre costruttori, che hanno probabilità

$$P(C_A) = 0.5 \quad , P(C_B) = 0.25 \quad , P(C_C) = 0.25$$

Se inchiamo con E il modulo difettoso, allora avremo le seguenti probabilità condizionate

$$P(E|C_A) = 5/100 \quad , P(E|C_B) = 2/100 \quad , P(E|C_C) = 1/100$$

e dalla formula di Bayes otterremo (con $A_i = C_C$ e $B=E$)

$$P(C_C|E) = \frac{P(C_C)P(E|C_C)}{\sum_{i=A,B,C} P(C_i)P(E|C_i)} = 0.08$$

Esempio 2.10 *Il problema di Monty Hall (1975)*

In questo gioco, vengono mostrate a un giocatore tre porte chiuse; al di là di una c'è un'automobile e dietro ciascuna delle altre due si nasconde una capra. Al giocatore è permesso aprire una porta, e tenersi ciò che si trova di là da essa. Ad ogni modo, dopo che il giocatore ha selezionato una porta, ma non l'ha ancora aperta, il conduttore dello show (che conosce ciò che si trova dietro ogni porta) apre un'altra porta, rivelando una delle due capre, e offre al giocatore la possibilità di cambiare la propria scelta iniziale, passando all'unica porta restante. Convienne al concorrente cambiare porta per vincere l'auto ?

Il problema di Monty Hall è un noto paradosso della teoria della probabilità, legato al gioco a premi americano *Let's Make a Deal*. Il nome viene da quello del conduttore dello show, Maurice Halprin, noto con lo pseudonimo di Monty Hall. Questo è un tipico esempio di applicazione del teorema di Bayes. Si potrebbe pensare che con due porte chiuse si abbia una probabilità 50:50 per ognuna, e che quindi non ci sia motivo di cambiare porta. Ma non è così. Chiamiamo l'evento che la macchina si trovi dietro una certa porta rispettivamente A_1, A_2, A_3 .

All'inizio è ovviamente

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}.$$

Supponiamo che il giocatore scelga la porta numero 1 (senza aprirla). Allora il presentatore dovrà aprire la porta 2 o 3 con probabilità a priori del 50 %. Chiamo evento B "il presentatore apre la porta 3". Ora :

- se la macchina è dietro la porta 1, il presentatore sarà libero di scegliere la porta 2 o 3 casualmente. Quindi la probabilità che il presentatore, sapendo che la macchina è in 1, apra la porta 3 è $P(B|A_1) = \frac{1}{2}$;
- se la macchina è dietro la porta 2, il presentatore sarà obbligato ad aprire la porta 3 e quindi $P(B|A_2) = 1$;
- se la macchina è dietro la porta 3, il presentatore sarà obbligato ad aprire la porta 2 e quindi $P(B|A_3) = 0$.

Il teorema di Bayes permette di calcolare la probabilità che la macchina si trovi in A_i , aprendo il presentatore la porta 3

$$\begin{aligned}P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}; \\P(A_2|B) &= \frac{P(B|A_2)P(A_2)}{P(B)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}; \\P(A_3|B) &= \frac{P(B|A_3)P(A_3)}{P(B)} = \frac{0 \times \frac{1}{3}}{\frac{1}{2}} = 0\end{aligned}$$

da cui risulta evidente che si deve cambiare con la porta 2.

Esempio 2.11 *Filtri antispam*

Lo spam via email, detto anche *junk mail*, consiste in una serie di messaggi pubblicitari uguali inviati a migliaia di destinatari, in un lasso di tempo relativamente breve, utilizzando come mezzo di trasporto la posta elettronica. Esistono varie categorie di tecniche antispam utilizzate a livello del server. Una di questa e' rappresentata dall'uso di filtri statistici, alcuni dei quali basati sul Teorema di Bayes, denominati **filtri bayesiani**. Il filtro piu' semplice utilizza una sola parola (contenuta nel messaggio) per sapere se il messaggio e' di spam o no (in tal caso detto di *ham*). Per esempio la parola *sconto* ha una specifica probabilita' di occorrenza nelle email di spam, che il filtro non conosce a priori ma che deve calcolare in qualche modo in una fase di *apprendimento*. Solitamente e' l'utente che indica manualmente se un messaggio e' di spam. Ogni parola contenuta nella mail contribuisce alla probabilita' che quella mail sia spam (magari escludendo articoli, congiunzioni, etc..). Considerando molti messaggi sara' possibile avere una statistica di apprendimento e calcolare

$P(Word|Spam)$ = la probabilita' che la parola *sconto* compaia nel messaggio di spam

$P(Word|Ham)$ = la probabilita' che la parola *sconto* compaia nel messaggio non spam (ham)

inoltre, inizialmente, si supporra' che il messaggio da esaminare abbia una eguale probabilita' di essere si spam o no (*filtro imparziale*), ovvero

$P(Spam)$ = la probabilita' apriori che il messaggio sia di spam = 0.5

$P(Ham)$ = la probabilita' apriori che il messaggio non sia di spam = 0.5 .

Se adesso indichiamo con

$P(Spam|Word)$ = la probabilita' che il messaggio considerato sia di spam,

supposto che contenga la parola *sconto*

dalla formula di Bayes (2.5) avremo che

$$\begin{aligned}
P(\textit{Spam}|\textit{Word}) &= \frac{P(\textit{Word}|\textit{Spam})P(\textit{Spam})}{P(\textit{Word}|\textit{Spam})P(\textit{Spam}) + P(\textit{Word}|\textit{Ham})P(\textit{Ham})} \\
&= \frac{P(\textit{Word}|\textit{Spam})}{P(\textit{Word}|\textit{Spam}) + P(\textit{Word}|\textit{Ham})}
\end{aligned}$$

la precedente quantita' e' chiamata *spamicity* o *spaminess* della parola *sconto*. Ovviamente non e' esatto prendere in considerazione una sola parola per capire se un messaggio e' di spam. Dunque i filtri bayesiani piu' comuni combinano la *spaminess* di un insieme significativo di parole contenute nel messaggio in modo tale da calcolare la probabilita' globale che il messaggio sia spam o meno. Questa probabilita' globale cosi' calcolata e' comparata con una soglia: se e' piu' bassa il messaggio sara' di ham altrimenti sara' di spam.

2.5 Variabili aleatorie

Nei problemi di calcolo delle probabilità si considerano delle quantità che sono funzioni del risultato di un fenomeno casuale, come:

- *nel gioco della roulette, l'ammontare del nostro capitale dopo cinque partite;*
- *nel gioco del lotto, il valore della nostra vincita se punto un certo numero sulla ruota di PA;*
- *nel mercato azionario, il valore del mio capitale a fine anno;*
- *il valore del dollaro tra una settimana;*
- *la durata della vita di un fiore.*

Tutte queste sono variabili aleatorie (o casuali) ³, in quanto non conosciamo esattamente il loro valore. Però ad una variabile aleatoria, in una data prova, posso dire di conoscerla se si possono determinare tutti i valori che può assumere e ad ognuno di essi la relativa probabilità.

Definizione 2.5 - *Variabile aleatoria*

Assegnato uno spazio di probabilità (Ω, \mathcal{P}, P) , si chiama *variabile aleatoria* (o *casuale*) un'applicazione che ad ogni elemento ω di uno spazio campione Ω associa un numero reale

$$X(\omega) : \omega \in \Omega \rightarrow \mathbb{R}$$

tale che, per ogni $t \in \mathbb{R}$, l'insieme $\{\omega; X(\omega) \leq t\}$ sia in \mathcal{P} .

Quest'ultima affermazione equivale a dire che ha senso calcolare la probabilità che X assuma valori più piccoli di un assegnato t . In generale è fondamentale per le v.a. il calcolo delle probabilità del tipo $P\{\omega; X(\omega) \in B\}$ dove B è un sottoinsieme di \mathbb{R} . Per esempio, nel caso del gioco della roulette, se X è *l'ammontare del nostro capitale dopo cinque partite* e B è una somma maggiore del nostro capitale, allora $P\{\omega; X(\omega) \in B\}$ è la probabilità che X sia cresciuto dopo la scommessa.

³dal latino *alea* cioè il *gioco dei dadi*.

Per semplicità di notazione nel seguito scriveremo $\{X \leq t\}$ invece di $\{\omega; X(\omega) \leq t\}$ e $\{X \in B\}$ invece di $\{\omega; X(\omega) \in B\}$.

Definizione 2.6 - Una variabile aleatoria si dice *discreta* se l'insieme dei valori S da essa assumibili (detto *supporto* della variabile aleatoria) è un insieme finito.

$$S = \{s_1, s_2, s_3, \dots, s_n\}$$

Definizione 2.7 - *Distribuzione discreta di probabilità*

Sia S il supporto della variabile aleatoria discreta X . Allora definisco *distribuzione discreta di probabilità* la funzione

$$P_X(s) = \begin{cases} P(X = s) & \forall s \in S \\ 0 & \text{altrimenti} \end{cases}$$

La funzione P_X è una distribuzione di probabilità se e solo se valgono le proprietà:

$$P_X(s) \geq 0 \quad \forall s$$

$$\sum_{s \in S} P_X(s) = 1$$

Esempio 2.12 *Lancio del dado.*

X =punto che si può presentare nel lancio di un dado.

$$\begin{array}{ll} s_1 = 1 & P_1 = \frac{1}{6} \\ s_1 = 2 & \vdots \\ s_1 = 3 & \vdots \\ s_1 = 4 & \vdots \\ s_1 = 5 & \vdots \\ s_1 = 6 & P_6 = \frac{1}{6} \end{array}$$

Esempio 2.13 *Si lanciano 2 dadi. Studiare la variabile aleatoria*

X =somma dei 2 punti

$s_i =$	2	3	4	5	6	7	8	9	10	11	12
$P_i =$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$P_i = \frac{n.eventi\ favorevoli}{n.totale\ eventi}$$

La probabilità che la somma $s = 3$ deriva dagli eventi (1,2) (2,1)

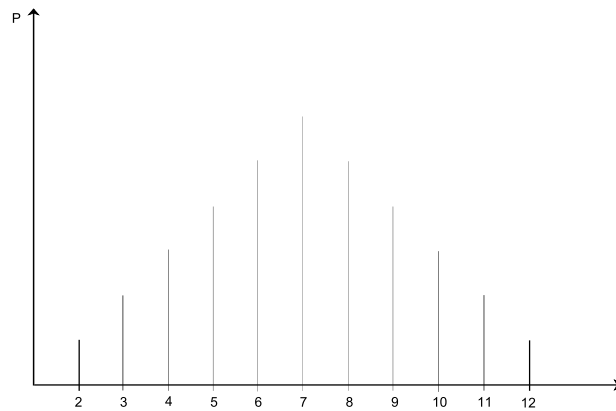


Figura 2.2: Lancio di 2 dadi

Poichè non si conosce a priori il valore assunto da una variabile aleatoria, si può fare una valutazione probabilistica sui valori che assumerà. Per ogni variabile aleatoria si introduce la **funzione di ripartizione**

Definizione 2.8 - La *funzione di ripartizione* $F_X : \mathbb{R} \rightarrow [0, 1]$ è definita:

$$\forall t \in \mathbb{R} \quad F_X(t) = P(X \leq t)$$

Quindi $F_X(t)$ rappresenta la probabilita' che la variabile aleatoria X assuma un valore minore o uguale a T . Nel caso discreto, ovviamente si avrà:

$$F_X(t) = \sum_{\substack{s \in \mathcal{S} \\ s \leq t}} P(s) \quad \forall t \in \mathbb{R}$$

Esempio 2.14 *Lancio di un dado*

s_i	P_i	$F(s_i)$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{6} + \frac{1}{6} = \frac{2}{6}$
3	$\frac{1}{6}$	$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$
6	$\frac{1}{6}$	$\frac{6}{6}$

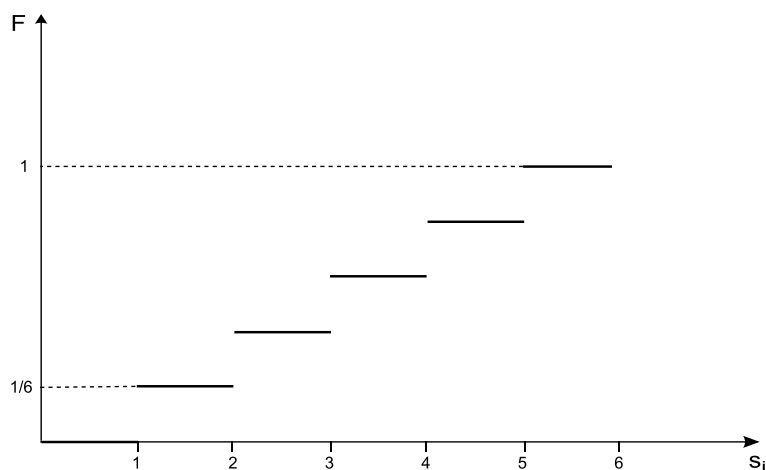


Figura 2.3: Lancio di 1 dado

Definizione 2.9 - Una variabile aleatoria X è **continua** se la corrispondente funzione di ripartizione $F_X(t)$ è continua.

Definizione 2.10 Una variabile aleatoria X è **assolutamente continua** se esiste una funzione $f_X(u)$

$$f_X(u) : \mathbb{R} \longrightarrow \mathbb{R}_+$$

tale che

$$F_X(t) = \int_{-\infty}^t f_X(u) du \quad \forall t \in \mathbb{R} \quad (2.6)$$

La funzione $f_X(u)$, quando esiste, è detta **densità di probabilità**.

osservazioni:

- se la densità di probabilità $f_X(u)$ esiste, allora la funzione di ripartizione $F_X(t)$ è una sua primitiva e rappresenta l'area di $f_X(u)$ nell'intervallo $]-\infty, t]$ (vedi figura 2.4);

•

$$\int_{-\infty}^{+\infty} f_X(u) du = 1$$

- $f_X(u)du$ è la probabilità che la variabile casuale X sia compresa tra $[u, u + du[$.

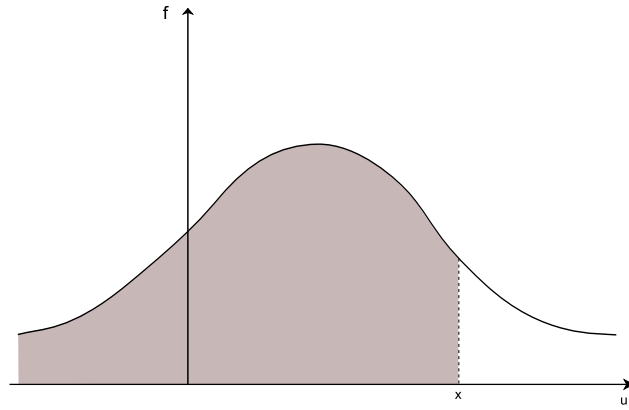


Figura 2.4: Densità di probabilità

2.6 Indici di tendenza centrale per le variabili aleatorie

Sintetizzano con un solo valore le caratteristiche della distribuzione

a) Valore atteso (\sim media)

$$\mu = E[X] = \sum_{s \in S} sp(s) \quad S = \text{supporto di } X$$

$$\mu = E[X] = \int_{-\infty}^{+\infty} uf(u)du$$

è una media dei valori di s pesata con la probabilità $p(s)$

b) Varianza

$$\sigma_X^2 = V[X] = \sum_{s \in S} (s - \mu)^2 p(s) \quad , \quad \sigma_X^2 = V[X] = \int_{-\infty}^{+\infty} [u - \mu]^2 f(u) du$$

c) Deviazione standard

$$\sigma_X = \sqrt{\sigma_X^2}$$

Elenchiamo alcune proprietà:

1.

$$\mu = E[aX + b] = aE[X] + b \quad \forall a, b \in \mathbb{R} \quad (2.7)$$

2.

$$V[aX + b] = a^2 V[X] \quad \forall a, b \in \mathbb{R} \quad (2.8)$$

3.

$$V[X] = E[X^2] - (E[X])^2 \quad \forall X \quad (2.9)$$

Dim. (2)

$$V[aX + b] = \int_{-\infty}^{+\infty} [au + b - \bar{\mu}]^2 f(u) du$$

sostituendo

$$\bar{\mu} = E[aX + b] = aE(X) + b = a\mu + b$$

si ha:

$$\begin{aligned} & \int_{-\infty}^{+\infty} [au + b - a\mu - b]^2 f(u) du = \\ & = a^2 \int_{-\infty}^{+\infty} [u - \mu]^2 f(u) du = a^2 V[X] \quad \blacksquare \end{aligned}$$

2.7 Variabili aleatorie multidimensionali

In certi casi l'esito dà luogo a una n-upla di valori.

◦ Esempio: Da un'urna contenente 6 palline numerate da 1 a 6, se ne estraggo 2 con rimpiazzo (se esce il 5 rimetto il 5). Il risultato di 2 estrazioni X, Y mi dà una distribuzione congiunta.

Definizione 2.11 - Una *variabile aleatoria bidimensionale* è una legge

$$(X, Y) \quad \Omega \rightarrow \mathbb{R}^2$$

dove Ω è lo spazio campione dove è definita la probabilità P .

Definizione 2.12 - *Funzione di ripartizione congiunta*

$$F_{X,Y} : \mathbb{R}^2 \rightarrow [0, 1]$$

$$F_{X,Y}(t, s) = P(\{X \leq t\} \cap \{Y \leq s\})$$

se F è assolutamente continua:

$$F_{X,Y}(t, s) = \int_{-\infty}^t \int_{-\infty}^s f_{X,Y}(u, v) du dv$$

Definizione 2.13 - *Funzione e densita' di ripartizione marginale di X*

Data una variabile aleatoria bidimensionale (X, Y) assolutamente continua, avente funzione di ripartizione congiunta $F_{X,Y}$ e densita' congiunta $f_{X,Y}$, si definiscono funzione di ripartizione marginale per X e densita' marginale per X , rispettivamente

$$F_X(t) = F_{X,Y}(t, +\infty) \quad , \quad f_X(t) = \int_{-\infty}^{+\infty} f_{X,Y}(t, s) ds$$

Definizione 2.14 Data una variabile aleatoria bidimensionale (X, Y) , si definisce Covarianza

$$Cov[X, Y] = E(X - E[X]) E(Y - E[Y]) = E[XY] - E[X] E[Y]$$

Proposizione 2.1

$$E[aX + bY] = aE[X] + bE[Y] \quad \forall a, b \in \mathbb{R} \quad (2.10)$$

Proposizione 2.2

$$V[X \pm Y] = V[X] + V[Y] \pm 2COV(X, Y) \quad (2.11)$$

Definizione 2.15 *Due variabili aleatorie*

$$\text{dove} \quad COV(X, Y) = E[(X - E[X])(Y - E[Y])]$$

COEFFICIENTE DI CORRELAZIONE LINEARE:

$$\rho_{X,Y} = \frac{COV(X, Y)}{\sqrt{V[X]V[Y]}} = \begin{cases} \text{se } \rho_{X,Y} = 0, & X \text{ e } Y \text{ sono incorrelate;} \\ \text{se } |\rho_{X,Y}| = 1, & X \text{ e } Y \text{ sono correlate da } Y=aX+b. \end{cases}$$

Definizione 2.16 - *Variabili aleatorie statisticamente indipendenti*

Data una variabile aleatoria bidimensionale (X, Y) , diremo che le due variabili X e Y , considerate singolarmente, sono *stocasticamente indipendenti* se e solo se

$$F_{X,Y}(t, s) = F_X(t)F_Y(s) \quad \forall (t, s) \in \mathbb{R}^2 \quad (2.12)$$

$$\Updownarrow$$

$$f_{X,Y}(t, s) = f_X(t)f_Y(s)$$

Proposizione 2.3 *Siano X, Y due variabili aleatorie **statisticamente indipendenti**, elenchi-amo alcune importanti proprieta'*

1.

$$E[XY] = E[X]E[Y] \quad \forall X, Y \quad (2.13)$$

2.

$$Cov[X, Y] = 0 \quad (2.14)$$

3.

$$V[X + Y] = V[X] + V[Y] \quad \forall X, Y \quad (2.15)$$

Dim.

$$\begin{aligned} V[X + Y] &= \int \int (u + v - E[X + Y])^2 f_{X,Y}(u, v) du, dv \\ E[X + Y] &= E[X] + E[Y] = \mu + \sigma \\ (u + v - \mu - \sigma)^2 &= [(u - \mu) + (v - \sigma)]^2 = \\ &= (u - \mu)^2 + (v - \sigma)^2 + 2(u - \mu)(v - \sigma) \\ V[X + Y] &= \int \int (u - \mu)^2 f + \int \int (v - \sigma)^2 f + 2 \int \int (u - \mu)(v - \sigma) f \\ V[X + Y] &= V[X] + V[Y] + 2COV(X, Y) \end{aligned}$$

Capitolo 3

Distribuzioni notevoli e teoremi di convergenza

3.1 Distribuzione di Bernoulli

Definizione 3.1 - Una variabile aleatoria X è *distribuita secondo una bernoulliana* di parametro $p \in [0, 1]$, se essa può assumere valori 1 e 0 rispettivamente con probabilità p e $1 - p$.

La distribuzione di probabilità è quindi:

$$P_X(s) = \begin{cases} p, & \text{se } s=1; \\ 1-p, & \text{se } s=0; \\ 0, & \text{altrimenti.} \end{cases}$$

La funzione di ripartizione $F_X(t) = \sum_{s \leq t} P_X(s)$ è:

$$F_X(t) = \begin{cases} 0, & \text{se } t < 0; \\ 1-p, & \text{se } 0 \leq t < 1; \\ 1, & t \geq 1. \end{cases}$$

Sono *variabili aleatorie di Bernoulli* tutte quelle che individuano il verificarsi di uno specifico evento e che valgono 1 se questo si verifica e 0 se non si verifica.

Es: lancio di una moneta

Si prova che

$$E[X] = \sum_s sP_X(s) = 0(1-p) + 1 \cdot p = p$$

$$V[X] = \sum_s [s - E[X]]^2 P_X(s) = (1-p)p$$

3.2 Distribuzione binomiale

Definizione 3.2 - Siano X_1, X_2, \dots, X_n n variabili bernoulliane di uguale parametro p e stocasticamente indipendenti tra loro. Sia

$$X = X_1 + X_2 + \dots + X_n$$

Questa variabile aleatoria è detta *distribuita secondo una binomiale* di parametri n e p :

$$X \simeq B(n, p)$$

Teorema 3.1 X può assumere $\forall k \in \mathbb{N}$ con $0 \leq k \leq n$ con probabilità

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

Dim.

La probabilità che k delle n variabili X_i assumano il valore 1 e le restanti $n - k$ assumano il valore 0 è

$$p^k (1-p)^{n-k}$$

Questa probabilità va moltiplicata per il n° di combinazioni possibili per cui k variabili valgano 1 e $(n - k)$ valgano 0

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

ne segue che

$$P_X(t) = \begin{cases} \binom{n}{t} p^t (1-p)^{n-t}, & \text{se } t=0,1,\dots,n; \\ 0, & \text{altrimenti.} \end{cases}$$

Poichè le n variabili X_i sono indipendenti, ne segue che

$$E[X] = E[X_1] + \dots + E[X_n] = np \quad (3.1)$$

$$V[X] = V[X_1] + \dots + V[X_n] = n(1-p)p \quad (3.2)$$

La distribuzione binomiale fu scoperta da Bernoulli nel 1700, per risolvere il problema:

Qual'è la probabilità che si presenti 2 volte testa in 6 lanci di una moneta?

$k = 2 \rightarrow$ n° di volte che si presenta testa

$n = 6 \rightarrow$ n° di lanci

$p = \frac{1}{2} \rightarrow$ probabilità

$$P(X = 2) = \binom{6}{2} \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{6-2} = \dots$$

3.3 Distribuzione di Poisson

Si può considerare come un caso particolare della distribuzione di Bernoulli e si ottiene quando:

- i) il numero di variabili X_i è $n \rightarrow \infty$
- ii) il parametro $p \rightarrow 0$ in modo che $\lambda = np = \text{costante}$.

Consideriamo quindi una v.a. $X \simeq B(n, \frac{\lambda}{n})$ e studiamo il suo comportamento per $n \rightarrow \infty$.

$$\begin{aligned} P(X = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

dove si è fatto uso dei limiti notevoli

$$\left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}, \quad \frac{n(n-1)\dots(n-k+1)}{n^k} \rightarrow 1.$$

Allora chiamiamo **distribuzione di Poisson** di parametro $\lambda > 0$

$$P_X(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda}, & k=0,1,2,3,\dots \\ 0, & \text{altrimenti.} \end{cases}$$

In pratica

- λ e' un qualsiasi valore positivo equivalente al numero di successi che ci si aspetta che si verifichino in un dato intervallo di tempo (la frequenza media di accadimento dell'evento osservato);
- k e' il numero delle occorrenze (successi) per cui si vuole prevedere la probabilita' (deve essere intero non negativo ($k=0,1,2,3,\dots$))
- $P_X(k)$ e' la probabilita' che si verifichino k successi, supposto di conoscere λ

Si prova che

$$F_X(t) = \sum_{\substack{k \in \mathbb{N} \\ k \leq t}} \frac{\lambda^k}{k!} e^{-\lambda} \quad \forall t \in \mathbb{R}$$

$$E[X] = np = \lambda \quad , \quad V[X] = \lambda - \frac{\lambda^2}{n} \rightarrow \lambda \quad \text{per } n \rightarrow \infty.$$

Questa distribuzione appare in maniera naturale come leggi di quantita' casuali X che rappresentano il numero di successi su un numero molto grande di prove ripetute indipendenti, in cui ciascuna delle quali la probabilita' di successo sia molto piccola. Per questo viene anche detta *statistica degli eventi rari*. Facciamo qualche semplice esempio di applicazione:

- ad una guardia medica arrivano in media 3.5 richieste ogni ora di interventi urgenti a domicilio. Calcolare la probabilita' che in una stessa ora arrivino 3, 4, oppure 5 chiamate urgenti. Il fenomeno puo' essere descritto utilizzando la statistica di Poisson, in quanto il n . di persone che potrebbe telefonare e' elevato ma che ciascuna di esse possa chiamare con una probabilita' piccola (ed indipendente dalle altre). Basta applicare la formula precedente con $\lambda = 3.5$ e $k = 3, 4, 5$, ottenendo

$$P(k=3) = 0.21579 \quad , \quad P(k=4) = 0.1888 \quad , \quad P(k=5) = 0.13217$$

- Un rivelatore di radioattivit  ambientale misura in numero di particelle ionizzanti al secondo. Supponiamo che siano stati rilevati 5348 conteggi in 30000 sec. ($\simeq 8.3$ h) cio 

$$r = \frac{5348}{30000} = 0.178 \quad \text{conteggi/sec.}$$

La probabilit  dei possibili numeri di conteggi che si osserveranno in un tempo T   descritta quindi da una poissoniana con parametro $\lambda = rT$, cio 

$$P(k) = \frac{(rT)^k}{k!} e^{-rT}$$

Riportiamo in tabella le probabilit  dei diversi numeri di conteggio per tempi di misura da 3 a 100 secondi.

conteggi	T = 3 s	T = 65 s	T = 12 s	T = 30 s	T = 100 s
0	58.63	34.37	11.81	0.48	
1	31.31	36.71	25.23	2.56	
2	8.36	19.60	26.94	6.84	
3	1.49	6.98	19.19	12.17	
4	0.20	1.86	10.25	16.25	0.01

3.3.1 Un'applicazione: ALOHA e CSMA

Negli anni 70 l'Universit  delle Hawaii sviluppo  una delle prime reti a commutazione di pacchetto, chiamata ALOHA (dal saluto hawaiano), per connettere le varie facolt  che si trovavano su varie isole. Il canale di comunicazione   una frequenza radio dove vengono trasmessi i pacchetti. La rete   formata da piu  nodi, cio  delle trasmittenti sintonizzate su quella frequenza radio. Ogni pacchetto contiene l'indirizzo di destinazione e viene trasmesso dal nodo quando   pronto, mentre tutti nodi ascoltano sempre il canale per vedere se c'  un messaggio per loro.

Questo protocollo ha pero  un difetto evidente, perche  se due o piu  pacchetti vengono trasmessi contemporaneamente, l'informazione contenuta si distrugge, ovvero si ha una collisione (vedi figura 3.1). Questo canale di comunicazione da  la possibilit  di verificare se il pacchetto   stato ricevuto correttamente o se c'  stata una collisione. Quindi, quando accade una collisione, tutti i nodi se ne accorgono (perche  ascoltano il canale). Per evitare che

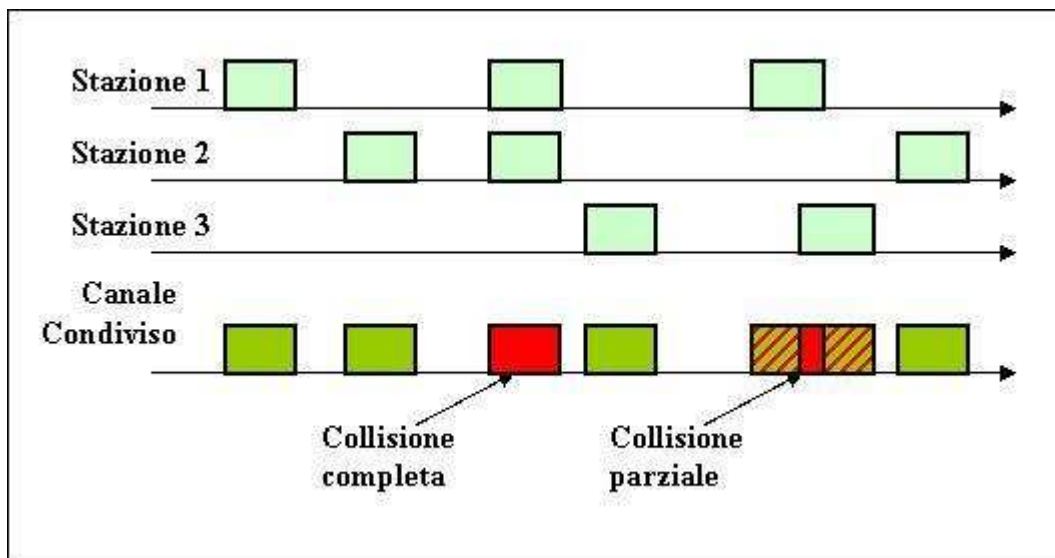


Figura 3.1: Esempio di collisione nell'ALOHA

la collisione si ripeta indefinitamente e' opportuno che le stazioni coinvolte tentino la loro ritrasmissione in tempi distinti, in modo da ridurre la probabilita' di nuove sovrapposizioni fra i due periodi di trasmissione. Poiche' le stazioni agiscono indipendentemente, il modo migliore per evitare la sovrapposizione delle ritrasmissioni e' che ogni stazione scelga casualmente, con opportuni vincoli, l'istante di tempo in cui provare a ritrasmettere. Cio' si attua utilizzando un meccanismo di back-off, secondo il quale la ritrasmissione viene effettuata dopo un ritardo selezionato casualmente compreso tra 0 e $(K-1)T$, dove T e' il tempo di trasmissione del messaggio e K può eventualmente dipendere dal numero di collisioni già avvenute.

Supponiamo che la trasmissione del pacchetto P inizi al tempo t_0 ed abbia tempo di trasmissione pari a T . Se una qualunque stazione inizia la propria trasmissione nell'intervallo compreso tra $[t_0 - T, t_0 + T[$ causa di sicuro una collisione. Quindi il periodo di tempo in cui il pacchetto P , risulta vulnerabile ad una collisione e' pari a $2T$ (vedi figura 3.2).

Vogliamo stimare il rendimento (o throughput) massimo del protocollo. Sia G il numero medio di pacchetti trasmessi nel tempo T , S il numero di pacchetti medio arrivati a destinazione nell'unita' di tempo e $\Lambda = G/T$ il tasso di arrivo. Allora il rendimento S e' dato dal numero medio di pacchetti generati nell'intervallo T per la probabilita' p che nel periodo di vulnerabilita'

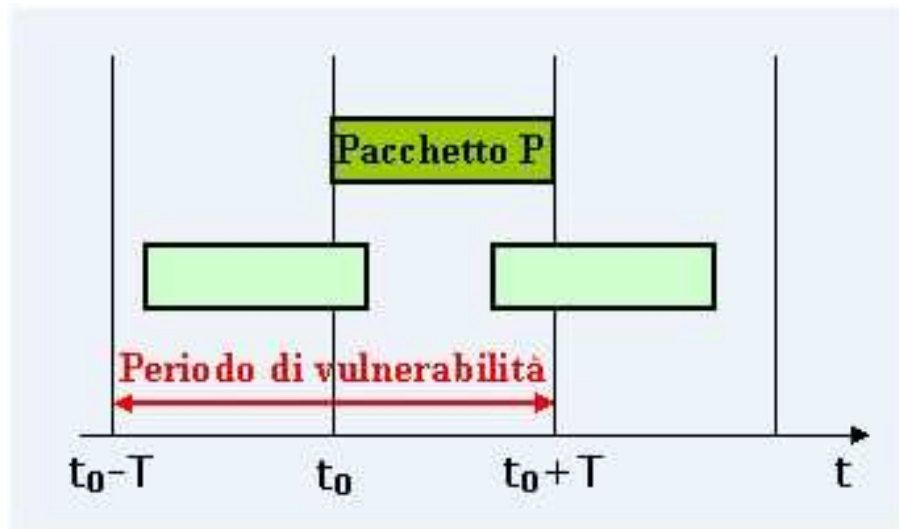


Figura 3.2: Periodo di vulnerabilità dell'ALOHA

(pari a $2T$) non si generino pacchetti, cioè

$$S = Gp(k=0) \quad .$$

Supponiamo che il n. di nodi che possono trasmettere un pacchetto è molto elevato, ma che la probabilità di trasmissione sia piccola (ed indipendente dalle altre). Questo significa che il processo è di Poisson di parametro $\lambda = \Lambda t$, cioè

$$p(k) = \frac{(\Lambda t)^k}{k!} e^{-\Lambda t}$$

Allora la probabilità che non ci siano altri arrivi ($k=0$) durante l'intervallo $]t_0 - T, t_0 + T[$ è uguale a

$$p(k=0) = e^{-2T\Lambda} = e^{-2TG/T} = e^{-2G}$$

da cui

$$S(G) = Ge^{-2G}$$

questa funzione assume il suo massimo per $G = \frac{1}{2}$ e quindi la resa massima è:

$$S_{max} = \frac{1}{2e} \simeq 0.184 \quad .$$

Un miglioramento delle prestazioni si ottiene con il protocollo Slotted Aloha (Roberts 1972).

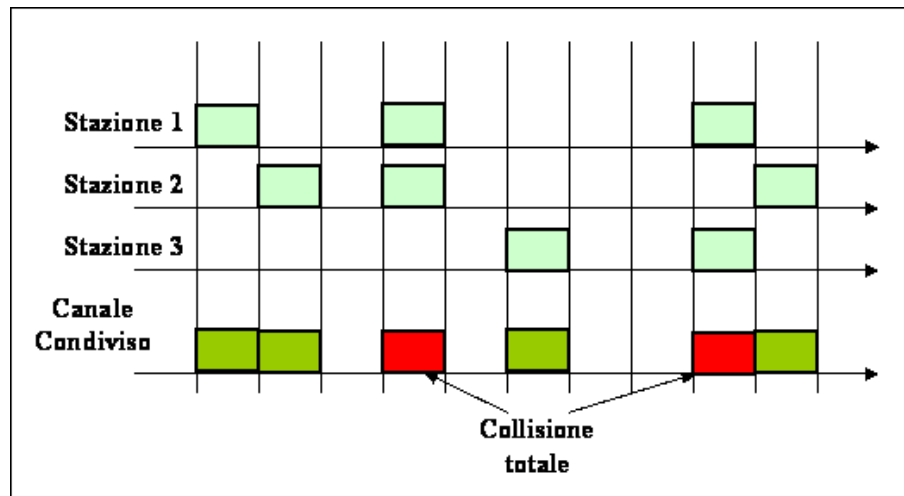


Figura 3.3: Esempio di collisione nello Slotted ALOHA

Rispetto al protocollo Aloha il tempo e' suddiviso in intervalli discreti chiamati slot. Ogni stazione e' vincolata a cominciare la propria trasmissione necessariamente all'inizio di uno slot temporale (vedi fig. 3.3). Se una stazione ad un certo istante e' pronta a trasmettere dovrà attendere necessariamente l'inizio del successivo slot. La conseguenza di tale caratteristica e' che due trasmissioni o collidono completamente all'interno dello stesso slot oppure non collidono affatto; il problema delle collisioni parziali osservato in Aloha risulta in questo modo eliminato. Come illustrato nella figura 3.4, il protocollo Slotted Aloha ha come conseguenza il

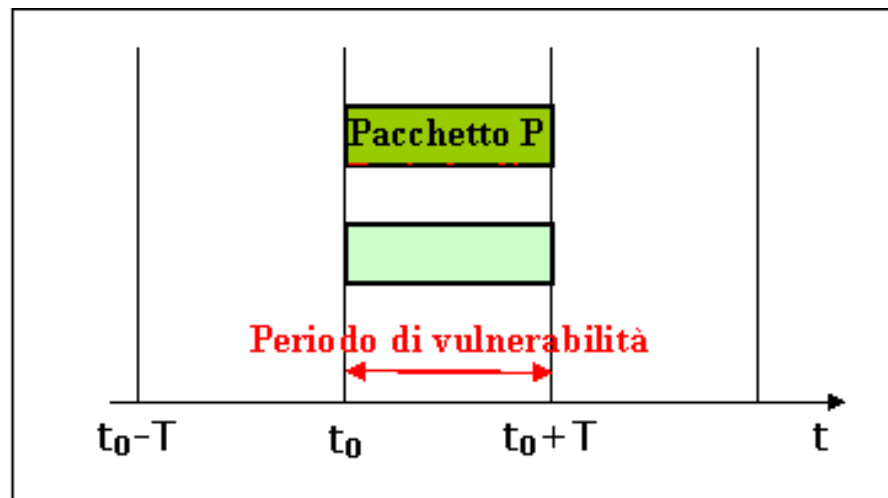


Figura 3.4: Periodo di vulnerabilita' nello Slotted ALOHA

dimezzamento del periodo di vulnerabilit , che in tal caso   pari a T, pertanto

$$S(G) = Ge^{-G}$$

che ha il suo massimo per $G = 1$ e quindi

$$S_{max} = \frac{1}{e} \simeq 0.37 \quad .$$

Un ulteriore miglioramanto si puo' ottenere dal fatto che tutti nodi sono capaci di monitorare continuamente il canale di trasmissione. Infatti la trasmissione puo' iniziare in un qualunque tempo, purch  il nodo trasmittente abbia sentito che il canale sia libero. Se il nodo che deve trasmettere sente che il canale   occupato, allora si tira indietro e prova ancora dopo un intervallo di tempo casuale (mentre nell'Aloha il tempo di trasmissione   casuale). Questo protocollo   noto come Carrier Sense Multiple Access (CSMA) di cui l'applicazione piu' famosa   la rete Ethernet.

Tuttavia si possono aver collisioni, perche' la velocita' di trasmissione del segnale   finita e un nodo che spedisce un pacchetto puo' pensare che il canale   libero, mentre un altro nodo ha cominciato la spedizione. Sia d il tempo di propagazione di ritardo nel canale, cio  il tempo che un segnale impiega per arrivare da un nodo ad un altro. Se il nodo A trasmette al tempo t ed un altro B trasmette nell'intervallo $[t - d, t + d]$, a causa del ritardo del segnale il nodo B non si accorge di A e si ha una collisione. Quindi la trasmissione di un pacchetto in un tempo in cui si pensa che il canale sia libero avr  successo se e solo se non ci saranno arrivi nell'intervallo di ampiezza $2d$. Quindi la probabilit  di avere $k=0$ pacchetti nell'intervallo $2d$  

$$p = e^{-2dG} \quad .$$

Il rendimento S della rete   il n. medio dei pacchetti arrivati con successo a destinazione. Questo numero   proporzionale

1. alla frazione di tempo in cui il canale   libero. Se il tempo di trasmissione del pacchetto vale 1 allora S   proporzionale alla frazione di tempo in cui il canale   occupato. Quindi la frazione di tempo in cui canale   libero   $(1-S)$;

2. al n. medio di pacchetti G generati quando il canale e' libero ;

3. la probabilita' di avere $k=0$ pacchetti nell'intervallo $2d$

Pertanto si ha:

$$S = (1 - S)Ge^{-2dG}$$

che risolta in S dara'

$$S(G) = \frac{Ge^{-2dG}}{1 + Ge^{-2dG}} \quad .$$

Questa funzione raggiunge il suo massimo per $G = 1/(2d)$ e quindi

$$S_{max} = \frac{1}{1 + 2de}$$

e poiche' d e' piccolo, il rendimento e' migliore dei precedenti protocolli. Se pero' G supera di molto $1/(2d)$, allora il rendimento tendera' a zero.

Per ultimo accenniamo che queste analisi di rendimento valgono del caso stazionario, mentre e' piu' realistico avere una analisi dipendente dal tempo al fine di studiare la stabilita' della rete di comunicazione.

3.4 Distribuzione Uniforme

E' la distribuzione che assume un valore costante in un intervallo $[a, b]$.

Definizione 3.3 Una variabile aleatoria X ha *distribuzione uniforme* in $[a, b] \subset \mathbb{R}$, ovvero $X \simeq U[a, b]$, se e' assolutamente continua con densita' di probabilita'

$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{se } t \in [a, b]; \\ 0 & \text{altrimenti} \end{cases} \quad (3.3)$$

da cui la funzione di ripartizione (2.6)

$$F_X(t) = \begin{cases} 0, & \text{se } t < a; \\ \frac{t-a}{b-a}, & \text{se } t \in [a, b]; \\ 1 & \text{se } t > b. \end{cases}$$

Si puo' verificare con semplici integrazioni che

$$E[X] = \int_{-\infty}^{+\infty} t f_X(t) dt = \frac{a+b}{2} \quad , \quad V[X] = \frac{(b-a)^2}{12} \quad .$$

3.5 Distribuzione Esponenziale

Questa distribuzione e' importante nello studio di quelle variabili che descrivono i tempi di attesa al verificarsi di un evento (p.es. il tempo di attesa in una coda)

Definizione 3.4 Una variabile aleatoria X è detta *distribuita secondo un' Esponenziale di parametro* $\lambda \in \mathbb{R}_+$, ovvero $X \simeq \text{Exp}(\lambda)$, se e' assolutamente continua con densita' di probabilità

$$f_X(t) = \begin{cases} 0, & \text{se } t < 0; \\ \lambda \exp(-\lambda t), & \text{se } t \geq 0; \end{cases} \quad (3.4)$$

da cui la funzione di ripartizione (2.6)

$$F_X(t) = \begin{cases} 0, & \text{se } t < 0; \\ 1 - \exp(-\lambda t), & \text{se } t \geq 0; \end{cases}$$

Si puo' verificare con semplici integrazioni che

$$E[X] = \frac{1}{\lambda} \quad , \quad V[X] = \frac{1}{\lambda^2} \quad .$$

La regione dell'importanza di questa distribuzione in vari campi applicativi sta nel fatto che questa e' l'unica distribuzione continua che soddisfa la proprieta' di **non-memoria** (memory-less), cioe' di soddisfare la seguente eguaglianza

$$P(X > s + t \mid X > s) = P(X > t) \quad , \forall t, s \in \mathbb{R}_+ \quad (3.5)$$

Esempio 3.1 Consideriamo il gioco del lotto ed indico con X la variabile aleatoria {numero minimo di estrazioni su una data ruota affinché esca il numero 34}. In questo caso s, t saranno numeri naturali.

- $P(X > s)$ e' la probabilita' che nelle prime s prove non esca il numero 34;
- $P(X > s + t)$ e' la probabilita' che nelle prime $s + t$ prove non esca il numero 34;
- $P(X > s + t | X > s)$ e' la probabilita' che, supposte che nelle prime s prove non sia uscito il numero 34, nelle successive t non esca in numero 34.

La proprieta' di non-memoria afferma che la probabilita' che non si verifichi alcun successo (ovvero estrazione del numero 34) fino alla prova $s + t$ (supposto che non si sia verificato nelle prime s prove) non dipende da s ossia da quanto si e' atteso (ovvero il ritardo), ma solo dal numero t di prove ancora da effettuarsi. E' come se, ad ogni estrazione, il numero 34 non avesse memoria di quello che e' accaduto nel passato: infatti ogni numero ha sempre la stessa probabilita' di essere estratto (a meno che non ci sia frode, con bussolotti riconoscibili al tatto o altri imbrogli). Già ai primi dell'ottocento, Laplace scrisse: ' Quando un numero non esce da molto tempo, i giocatori corrono a coprirlo di danaro, essi ritengono che quel numero reticente debba uscire al primo colpo, a preferenza di altri, ma il passato non può avere alcuna influenza sull'avvenire '.

Esempio 3.2

Supponiamo che X sia il tempo di vita di una macchina soggetta a guasti e supponiamo che essa non sia guastata fino al tempo s . La proprieta' di non-memoria afferma che, supposto che la macchina si arrivata al tempo s senza guasti, la probabilita' che la macchina non si guasti per un ulteriore intervallo di tempo pari a t (calcolato a partire da s), non dipende dal tempo s ovvero da quello che e' successo nel passato. In altre parole quello che e' successo in precedenza non incide in alcun modo su quello che possiamo aspettarci per il futuro.

Esempio 3.3

Sia X il tempo di attesa per essere visitati dal medico. Supponimo di sapere che la durata media della visita sia 30 minuti e che il paziente precedente e' entrato da 20 minuti. Allora, se il processo (di attesa) e' senza-memoria, in media dovremo aspettare 30 minuti, proprio come se il paziente precedente fosse appena entrato. Che la visita precedente sia in corso da 1 minuto o da 100, la nostra attesa prevista e' la stessa.

Teorema 3.2 *Se una variabile aleatoria X e' distribuita secondo un' Esponenziale di parametro $\lambda \in \mathbb{R}_+$ allora soddisfa la proprieta' di non-memoria (3.5)*

Dim. Infatti dalla definizione di probabilita' condizionata (2.2)

$$P(X > t + s | X > s) = \frac{P(X > t + s \cap X > s)}{P(X > s)} = \frac{P(X > t + s)}{P(X > s)}$$

ma poiche'

$$P(X > t + s) = 1 - P(X \leq t + s) = 1 - \int_{-\infty}^{t+s} f_X(t) dt = 1 - F_X(t + s) = \exp(-\lambda(t + s))$$

$$P(X > s) = 1 - P(X \leq s) = 1 - \int_{-\infty}^s f_X(t) dt = 1 - F_X(s) = \exp(-\lambda s)$$

allora

$$P(X > t + s | X > s) = \exp(-\lambda t) = 1 - F_X(t) = 1 - \int_{-\infty}^t f_X(t) dt = P(X > t) \quad \square$$

3.6 Distribuzione di Weibull

Definizione 3.5 Una variabile aleatoria X è detta *distribuita secondo una Weibull di parametri* $\alpha, \beta \in \mathbb{R}_+$ se e' assolutamente continua con densita' di probabilità

$$f_X(t) = \begin{cases} 0, & \text{se } t < 0; \\ \alpha\beta t^{\beta-1} \exp(-\alpha t^\beta), & \text{se } t \geq 0; \end{cases}$$

da cui la funzione di ripartizione (2.6)

$$F_X(t) = \begin{cases} 0, & \text{se } t < 0; \\ 1 - \exp(-\alpha t^\beta), & \text{se } t \geq 0; \end{cases}$$

La determinazione della media e varianza e' abbastanza complessa. Si puo' provare che

$$E[X] = \alpha^{-\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right) \quad , \quad V[X] = \alpha^{-\frac{2}{\beta}} \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right) \right]^2 \right\}$$

dove $\Gamma(z)$ e' la funzione Gamma di Eulero. La distribuzione di Weibull si riconduce a quella esponenziale quando il parametro $\beta = 1$.

Come la distribuzione esponenziale descrive la *durata di vita* di un fenomeno privo di memoria, così la distribuzione di Weibull può descrivere la durata di vita per un fenomeno la cui *probabilità di morire* può variare nel tempo, in funzione di β .

La distribuzione di Weibull è stata comunemente ritenuta adeguata per la rappresentazione della statistica delle velocità medie del vento campionato in un sito eolico.

3.6.1 Un'applicazione alla Teoria dell'Affidabilità

Questa distribuzione trova molte applicazioni nella Teoria dell'Affidabilità. La rottura di un sistema o di un suo particolare componente e' un fenomeno di natura completamente casuale. Sia X 'la lunghezza dell'intervallo di tempo dall'attivazione di un componente (calcolato a partire dal tempo $t = 0$) fino al suo guasto', essa e' modellizzata da una variabile aleatoria continua che assume valori non negativi, avente funzione di ripartizione $F_X(t)$ ¹.

¹ $F_X(t) = P(X \leq t)$ rappresenta la probabilità che X assuma valori minori o uguali a t

Definizione 3.6 Chiamiamo *funzione di affidabilit * (o *reliability function*) di un componente la probabilit  che quest'ultimo sia ancora funzionante al tempo t , vale a dire:

$$R_X(t) = P(X > t) = \frac{N_s(t)}{N}$$

dove $N_s(t)$ 'n. componenti sopravvissuti al tempo t ', e N 'n. iniziale componenti'.

Da questa definizione segue che

$$R_X(t) = P(X > t) = 1 - P(X \leq t) = 1 - F_X(t) \quad . \quad (3.6)$$

Supponiamo che la variabile aleatoria X sia assolutamente continua (vedi (2.6)), ovvero che esista la corrispondente densit  di probabilit  $f_X(t)$. Quindi $f_X(t) dt$ rappresenta la probabilit  che $X \in [t, t + dt[$ ovvero che il macchinario si guasti nell'intervallo $[t, t + dt[$.

Definizione 3.7 Chiamiamo *funzione di intensit  di rottura* (o *tasso di guasto*)

$$\lambda(t) = \frac{f_X(t)}{R_X(t)} = \frac{f_X(t)}{1 - F_X(t)} \quad . \quad (3.7)$$

Propriet 

- Si dimostra che il tasso di guasto $\lambda(t)$ rappresenta la densit  di probabilit  che se il sistema   ancora attivo all'istante t , esso si guasti nell'immediato futuro, cio  nell'intervallo $[t, t + dt[$.
- Poich  $f_X(t) = F'_X(t) = -R'_X(t)$ dalla eq.(3.7) segue che

$$\lambda(t) = -\frac{1}{R_X(t)} \frac{d}{dt} R_X(t)$$

e quindi possiamo approssimare la derivata

$$\lambda(t) \simeq \frac{N_s(t) - N_s(t + \Delta t)}{N_s \Delta t} \quad (3.8)$$

che   un modo operativo di calcolare il tasso di guasto.

Teorema 3.3 *La funzione $\lambda(t)$ determina univocamente la $F_X(t)$.*

Dim: Infatti essendo $F'_X(t) = f_X(t)$, dalla eq.(3.7) si ha

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{F'(t)}{1 - F(t)} = -\frac{d}{dt} \log[1 - F(t)]$$

da cui integrando tra 0 e t si ha

$$\int_0^t \lambda(t) dt = -\log[1 - F(t)] + \log[1 - F(0)] = -\log[1 - F(t)] \quad \text{perche' } F(0) = 0$$

e quindi

$$F(t) = 1 - \exp \left\{ - \int_0^t \lambda(s) ds \right\} \quad \square \quad (3.9)$$

L'importanza pratica del teorema e' che, misurando il tasso di guasto con la (3.8), si puo' risalire alla $F_X(t)$ e quindi alla sua densita' di probabilita'

Esempio 3.4 *Supponiamo che il tasso di guasto di un sistema meccanico sia una funzione lineare del tempo t cioe' $\lambda(t) = a + bt$. Dalla eq.(3.9) si trova facilmente che*

$$F(t) = 1 - \exp(-at - bt^2/2) \quad , \quad f(t) = F'(t) = (a + bt) \exp(-at - bt^2/2)$$

Nota la $f(t)$ posso calcolare la probabilita' di guasto in qualunque intervallo di tempo.

Nel caso in cui $f_X(t)$ sia la distribuzione di Weibull si prova facilmente che :

$$\lambda(t) = \alpha\beta t^{\beta-1}$$

da cui

- per $\beta < 1$ il tasso di guasto diminuisce nel tempo (alta mortalita' infantile/rodaggio)
- per $\beta = 1$ il tasso di guasto e' invariante nel tempo (memoryless o danneggiamento casuale)
- per $\beta > 1$ il tasso di guasto aumenta con il tempo (invecchiamento)

Solitamente il tasso di guasto di un componente e' la 'somma' di tre funzioni di Weibull che danno complessivamente la forma di una vasca da bagno (vedi fig. 3.6): superato il periodo di rodaggio ($\beta < 1$), il componente puo' essere danneggiato casualmente ($\beta = 1$) e da un certo tempo in poi invecchia ($\beta > 1$).

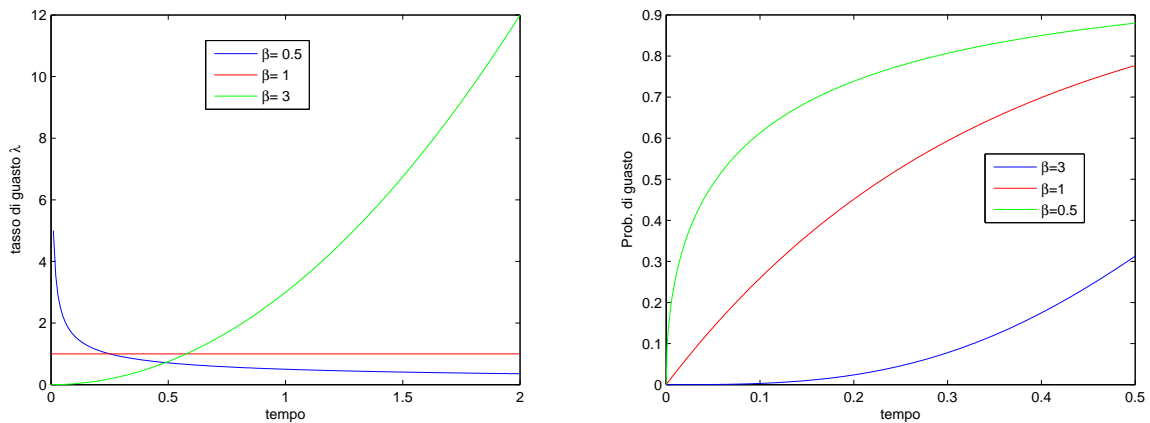


Figura 3.5: **Sinistra:** Tasso di guasto λ con distribuzione di Weibull. **Destra :** Funzione di ripartizione di Weibull.

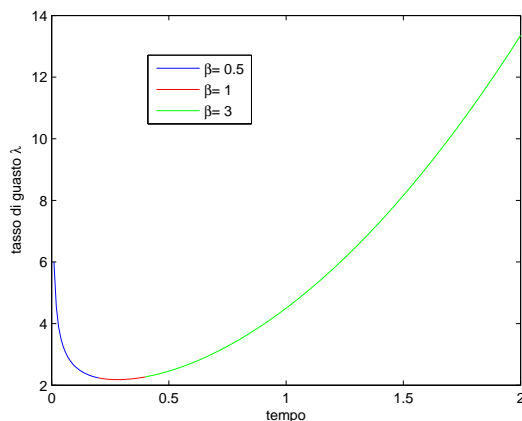


Figura 3.6: Tasso di guasto λ per un componente reale.

3.7 Distribuzione Normale (di Gauss)

È la distribuzione di variabili aleatorie continue più conosciuta per le applicazioni.

Definizione 3.8 Una variabile aleatoria X è detta *distribuita secondo una normale* di parametri $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ ovvero $X \simeq N(\mu, \sigma)$ se ha densità di probabilità

$$f_X(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \frac{(u - \mu)^2}{\sigma^2} \right]$$

Proposizione 3.1 - *Proprietà analitiche*

i) f è simmetrica rispetto alla retta $u = \mu$

ii) ha un max in $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}} \right)$ ed è crescente per $x < \mu$

iii) ha asintoto orizzontale $y = 0$

iv) ha due flessi in

$$\left(\mu - \sigma, \frac{1}{\sigma\sqrt{2\pi e}} \right), \left(\mu + \sigma, \frac{1}{\sigma\sqrt{2\pi e}} \right)$$

v)

$$\int_{-\infty}^{+\infty} f_X(u) du = 1$$

- all'aumentare di σ il max diminuisce in altezza

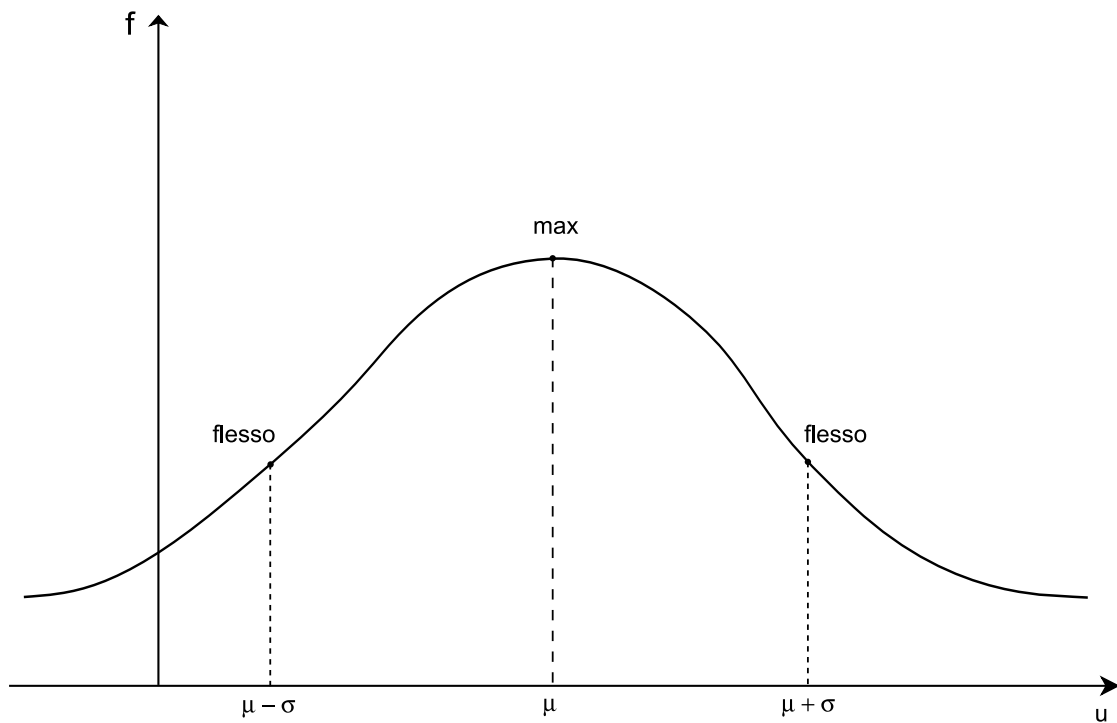


Figura 3.7: Distribuzione Normale - funzione $f_X(u)$

- se $\mu = 0$ è simmetrica rispetto all'asse y
- $E[X] = \mu$, $V[X] = \sigma^2$

La funzione di ripartizione F_X associata a N è

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[-\frac{1}{2} \frac{(u - \mu)^2}{\sigma^2} \right] du$$

questo integrale non si risolve analiticamente ma tramite approssimazioni

Definizione 3.9 - *Distribuzione normale standardizzata*

$$X \simeq N(0, 1) \quad \mu = 0, \sigma^2 = 1$$

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

Si può passare da $N(\mu, \sigma)$ a $N(0, 1)$ con un cambio di variabile

$$z = \frac{x - \mu}{\sigma}$$

$$\underbrace{P(a \leq X \leq b)}_{N(\mu, \sigma)} = \underbrace{P\left(\frac{a - \mu}{\sigma} \leq z \leq \frac{b - \mu}{\sigma}\right)}_{N(0, 1)}$$

FUNZIONI EXCEL 3.1

- DISTRIB.NORM → calcola fissato t, μ, σ :

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{u - \mu}{\sigma}\right)^2\right] du$$

- DISTRIB.NORM.ST → calcola fissato t :

$$F_X(t) = P(Z \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] du \quad (3.10)$$

Esercizio 3.1

Calcolare $P(9.2 \leq X \leq 11.35)$ con $N(\mu, \sigma)$, $\mu = 6.5$, $\sigma = 1$

$$\int_{9.2}^{11.35} \dots = ?$$

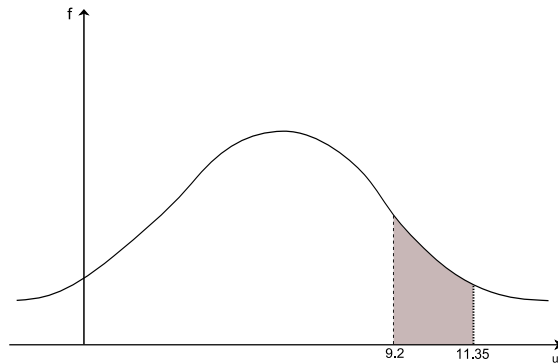


Figura 3.8: Distribuzione Normale - area

Si prova che

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.683$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997$$

Quindi ogni variabile casuale $X \simeq N(\mu, \sigma)$ assumerà valori compresi tra $[\mu - 3\sigma, \mu + 3\sigma]$ con probabilità $\simeq 1$.

Perchè è così importante la distribuzione normale?

3.7.1 Errori di misura casuali e sistematici

Quando si fa una misura con uno strumento si commette un errore più o meno grande. Questi errori sono classificati in

- *errori sistematici*: sono dovuti allo strumento e al metodo di misura. “Spingono” il risultato sempre nella stessa direzione. Si eliminano cambiando lo strumento.
- *errori casuali*: sono di natura stocastica. Si eliminano rifacendo la misura N volte.

Esempio 3.5 *Faccio una misura e creo una tabella*

<u>Intervallo</u>	22-23	23-24	24-25	25-26	26-27
<u>N° misure</u>	1	3	1	4	1

N° totale di misure = 10

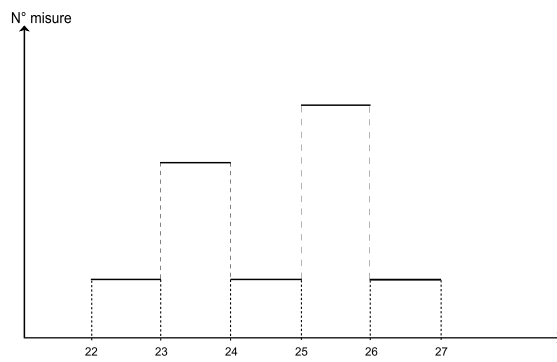


Figura 3.9: Errori nelle misure

per $n \sim 1000$

L'istogramma tende ad una curva continua, che nel caso delle misure e' una gaussiana.

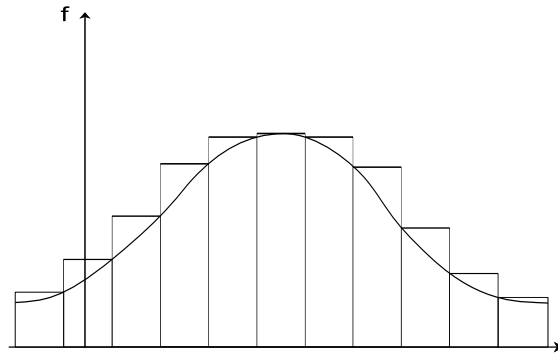


Figura 3.10: Errori nelle misure - limite

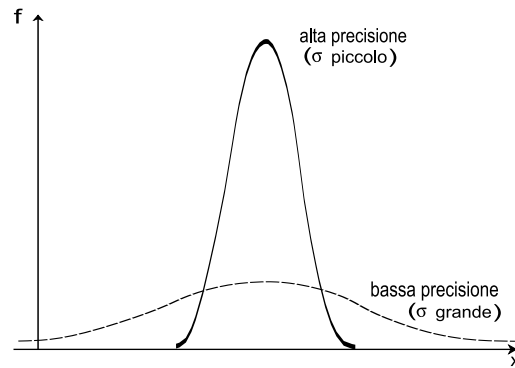


Figura 3.11: Errori nelle misure - precisione

3.8 Distribuzioni limite

E' possibile provare analiticamente le seguenti proprieta'

$$B(n, p) \xrightarrow{n \rightarrow \infty} N(\mu, \sigma) \xleftarrow{\lambda \rightarrow \infty} Pois(\lambda)$$

3.9 Distribuzione Chi-quadro

Siano X_i $i = 1, \dots, n$ variabili aleatorie con distribuzione normale $N(0, 1)$ stocasticamente indipendenti tra loro. Sia X una variabile aleatoria:

$$X = X_1^2 + X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2$$

X è detta *distribuita secondo una χ^2* con gradi di libertà $X \simeq \chi_n^2$.

Si prova che la densità di X è

$$f_X(t) = \begin{cases} 0, & \text{se } t < 0; \\ \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n}{2}} t^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}t\right), & \text{se } t \geq 0 \end{cases}$$

con $\Gamma(\alpha) = \int_0^{+\infty} e^{-x} x^{\alpha-1} dx$, che è la funzione Gamma di Eulero.

Si prova che

$$E[X] = n, \quad V[X] = 2n$$

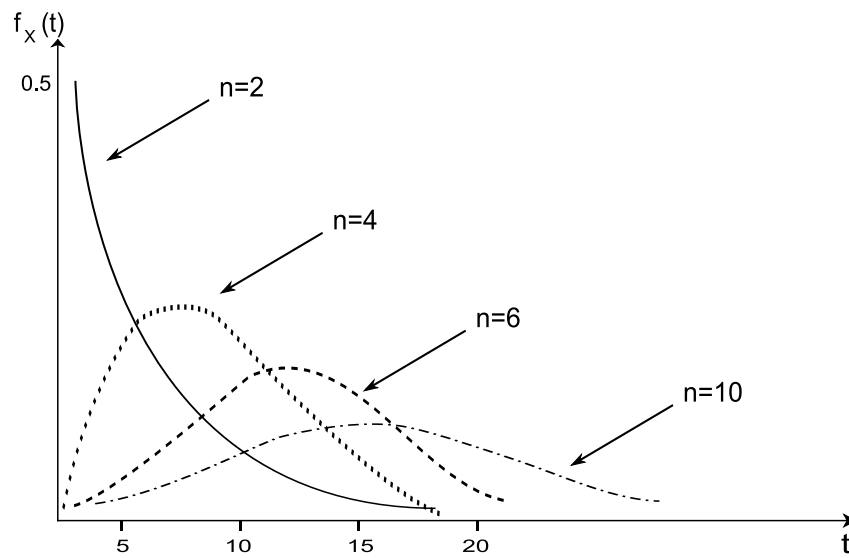


Figura 3.12: Distribuzione Chi-Quadro

FUNZIONI EXCEL 3.2

- **DISTRIB.CHI**(x_0 ; n)

dove x_0 valore positivo assegnato, n grado di libertà

DISTRIB.CHI = $P(X > x_0)$ che è la probabilità che la variabile aleatoria assuma valori maggiori di x_0 (*distribuzione a una coda*)

- `INV.CHI(α ; n)` fissata la probabilità α ed n , mi restituisce quel punto x_0 (che è il quantile) tale che alla sua destra l'area della curva è α . In pratica è la funzione inversa di `DISTRIB.CHI`

3.10 Distribuzione t di Student

Siano $Z \simeq N(0, 1)$ e $Y \simeq \chi_n^2$ due variabili indipendenti. Definisco:

$$X := \frac{Z}{\sqrt{\frac{Y}{n}}}$$

che è una variabile aleatoria *distribuita secondo una t di Student* con n **gradi di libertà**

$$X \simeq t_n$$

Si prova che X ha densità di probabilità:

$$f_X(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}$$

Si prova che

$$E[X] = 0 \quad \text{se } n > 1$$

$$V[X] = \frac{n}{n-2} \quad \text{se } n > 2$$

cioè l'esistenza di E e V dipende dal parametro n .

FUNZIONI EXCEL 3.3

- DISTRIB.T(x_0 ; n ; coda)

dove x_0 valore positivo assegnato, n grado di libertà, coda = (1,2) .

Se coda = 1 , DISTRIB.T = P ($X > x_0$) che è la probabilità che la variabile aleatoria assuma valori maggiori di x_0 (*distribuzione a una coda*)

Se coda = 2 , DISTRIB.T = P ($|X| > x_0$) che è la probabilità che la variabile aleatoria assuma valori $X < -x_0 \cup X > x_0$ (*distribuzione a due code*)

- INV.T(α, n) da' il quantile ovvero, fissato α ed il grado di liberta' n , questa funzione EXCEL restituisce un numero t_E che e' quel punto tale che

$$P(|t| > t_E) = \alpha$$

ovvero t_E e' punto della distribuzione tale che l'area delle due code (a destra e sinistra di t_E) abbiano area α . Essendo la distribuzione simmetrica rispetto all'asse delle y, avremo che

$$P(|t| > t_E) = \alpha \quad \rightarrow \quad P(-t_E < t < t_E) = 1 - \alpha \quad \rightarrow \quad P(t < t_E) = 1 - \frac{\alpha}{2} \quad (3.11)$$

cioe' t_E e' quel punto della distribuzione alla cui sinistra cade un'area pari a $1 - \frac{\alpha}{2}$. Pertanto

$$t_{1-\frac{\alpha}{2}}(n) = \text{INV.T}(\alpha, n) \quad (3.12)$$

- Funzione di ripartizione

La funzione di ripartizione rappresenta l'area nell'intervallo $]-\infty, u]$, cioe'

$$F_n(u) = \int_{-\infty}^u f_X(t) dt \quad . \quad (3.13)$$

Poiche' la funzione di DISTRIB.T necessita di $x_0 = u$ positivo, avremo

$$F_n(u) = \begin{cases} u \geq 0 & , \quad 1 - \text{DISTRIB.T}(u, n, 1) \\ u < 0 & , \text{DISTRIB.T}(|u|, n, 1) \end{cases} \quad (3.14)$$

Per $n \geq 30$ la t Student approssima una $N(0, 1)$.

Questa distribuzione si deve a Gosset, che era uno statistico che lavorava presso la fabbrica di birra Guinness (~ 1900). Egli dovette usare lo pseudonimo di Student, perché la birreria presso la quale era impiegato vietava ai propri dipendenti di pubblicare articoli affinché questi non divulgassero segreti di produzione.

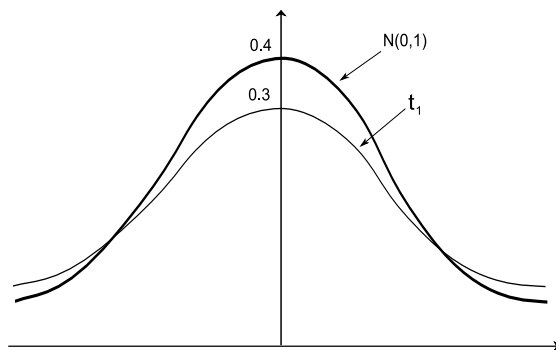


Figura 3.13: Distribuzione T di Student

3.11 Legge dei grandi numeri

Hp:

Supponiamo di avere una successione $\{X_i\}$ di variabili aleatorie statisticamente indipendenti con identica funzione di ripartizione. Definiamo la nuova variabile aleatoria:

$$\overline{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (\text{media campionaria})$$

supponiamo inoltre che

$$E[X_i] = \mu \quad , \quad V[X_i] = \sigma^2 \quad i = 1, \dots, n \quad (3.15)$$

Ts:

$$\overline{X}_n \xrightarrow{d} M$$

dove

a) M è una variabile aleatoria che assume il valore μ con probabilità 1

b) \xrightarrow{d} è un particolare tipo di convergenza

Sia $\begin{cases} X_n & \text{variabile aleatoria con funzione di ripartizione } F_n \\ X & \text{variabile aleatoria con funzione di ripartizione } F \end{cases}$

allora $X_n \xrightarrow{d} X \Leftrightarrow \lim_{n \rightarrow \infty} F_n = F$

La legge dice che medie molto grandi di variabili aleatorie tendono alla media vera, ovvero

$$E[\overline{X}_n] \rightarrow \mu$$

3.12 Teorema del limite centrale

La legge dei grandi numeri ci dice che $\bar{X}_n \rightarrow M$, ma non ci dice con quale rapidità (cioè a partire da quale valore di n).

Sotto le stesse ipotesi della legge dei grandi numeri, allora:

$$\bar{X}_n \xrightarrow{d} X \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (3.16)$$

cioè per n grande \bar{X}_n si distribuisce come una normale $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, i cui parametri dipendono da quelli delle variabili X_i (cioè da μ e σ).

Dim. Dalla definizione di variabili aleatorie statisticamente indipendenti (2.11), dalle (2.8) e (3.15) si ha

$$V[\bar{X}_n] = V\left[\sum_{i=1}^N \frac{X_i}{n}\right] = \frac{1}{n^2} V\left[\sum_{i=1}^N X_i\right] = \frac{1}{n^2} \sum_{i=1}^N V[X_i] = \frac{1}{n^2} \sum_{i=1}^N \sigma^2 = \frac{\sigma^2}{n}$$

e quindi la deviazione standard della media campionaria è pari a σ/\sqrt{n} . Per dimostrare che la media campionaria si distribuisce come una normale di parametri μ e σ/\sqrt{n} , bisogna sfruttare la disequazione di Chebychev, che esula dai nostri scopi.

Nella pratica il teorema del Limite Centrale ci dice che per n *sufficientemente grande*, la variabile aleatoria media campionaria ha come funzione di distribuzione una normale di parametri μ e σ/\sqrt{n} , indipendentemente dalla distribuzione della popolazione. Ciò accade per $n \geq 30$.

Capitolo 4

Stime di parametri

La statistica *inferenziale* consente di dedurre caratteristiche particolari di una popolazione analizzando un numero finito e piccolo (preferibilmente) di suoi individui detto *campione*.

Quando le caratteristiche che si vogliono individuare sono esprimibili numericamente, allora prendono il nome di *parametri*.

4.1 Problema del campionamento

Devo determinare le caratteristiche di una popolazione con un numero limitato di individui. Infatti, se la popolazione è vasta, per risparmiare tempo e denaro conviene analizzare un piccolo campione.

4.1.1 Strategie di campionamento

i) **Casuale:** Associa ad ogni individuo un numero e con un generatore di numeri casuali ne estraggo un certo numero.

Attenzione: se la statistica va fatta sugli abitanti di una città, non si deve fare il campionamento ad esempio tra abbonati al telefono o quelli che si incontrano per strada. Non si prenderebbero in considerazione chi non ha telefono e chi esce poco.

- ii) **Stratificato:** La popolazione è suddivisa in gruppi con stesse caratteristiche (es. età, sesso, etc..)
- ii) **A grappoli:** Si suddivide la popolazione in gruppi *eterogenei*, in modo che ogni singolo gruppo rappresenti l'intera popolazione.

Sia X il carattere della popolazione su cui si è interessati a fare delle inferenze (es. peso, altezza, etc... su una popolazione di persone). Il valore assunto da questo carattere varia a seconda dell'individuo considerato e viene indicato con x . Quindi X è una variabile aleatoria con *distribuzione sconosciuta*, che corrisponde a quella che si otterrebbe facendo ricorso alle tecniche della statistica descrittiva e potendo quindi utilizzare l'intera popolazione.

Definizione 4.1 - *Campione casuale di numerosità n*

È una n -upla (X_1, X_2, \dots, X_n) di variabili aleatorie indipendenti (estratte da una popolazione) aventi ognuna la stessa distribuzione del carattere X della popolazione. I valori assunti da questa n -upla

$$(x_1, x_2, \dots, x_n)$$

sono le misure fatte e sono dette *realizzazioni* di (X_1, X_2, \dots, X_n) .

Definizione 4.2 - *Parametro e stima*

Un *parametro* è un valore numerico che descrive una caratteristica di una popolazione, ed è una grandezza associata ad una sua distribuzione (quale il valore atteso e la varianza).

Una *stima* del parametro è una misura fatta sul campione.

Esempio 4.1

X ="costo al mq degli appartamenti della città"

Sia $(X_1, X_2, \dots, X_{80})$ un campione di numerosità 80. Si considera il parametro valore atteso:

μ ="costo medio al mq degli appartamenti della città"

Ovviamente non si conosce μ , perchè non si hanno i dati relativi a tutti gli appartamenti della città.

Una stima del parametro μ può essere fatta con la media sugli 80 valori

$$\bar{X}_n = \frac{1}{80} \sum_i^n x_i$$

presumibilmente il valore vero di μ sarà diverso da \bar{X}_n .

Allora le stime sono anch'esse delle variabili aleatorie definite in funzione del campione

$$H_n = f(X_1, \dots, X_n)$$

Esse prendono il nome di *statistiche campionarie* e le loro distribuzioni sono chiamate *distribuzioni campionarie*.

4.2 Principali distribuzioni campionarie

Sia X il carattere di una popolazione con distribuzione cumulativa F , valore atteso μ , varianza σ sconosciuti. Vediamo come stimare questi due parametri incogniti.

Definizione 4.3 - *Media campionaria n -esima di un campione casuale*

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad (4.1)$$

Proposizione 4.1

Se le variabili aleatorie X_i del campione sono tutte indipendenti e

$$E[X_i] = \mu, \quad V[X_i] = \sigma^2 \quad i = 1, n$$

allora

$$E[\bar{X}_n] = \mu, \quad V[\bar{X}_n] = \frac{\sigma^2}{n} \quad (\text{dimostrare}) \quad (4.2)$$

In tal caso il valore atteso \bar{X}_n non dipende dalla numerosità del campione, mentre la sua varianza ne dipende. Questo significa che la media campionaria sarà più vicina al valore incognito μ quanto più è grande la numerosità del campione.

Osservazione:

Per la legge dei grandi numeri ed il teorema del limite centrale, per $n \geq 30$, \bar{X}_n si può approssimare con una variabile aleatoria avente distribuzione normale di parametri μ e $\frac{\sigma}{\sqrt{n}}$:

$$\bar{X}_n \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Definizione 4.4 - *Varianza campionaria n-esima*

Sia (X_1, \dots, X_n) un campione estratto da una popolazione avente distribuzione F , media μ e deviazione standard σ . Si definisce varianza campionaria n-esima:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad . \quad (4.3)$$

La distribuzione di S_n^2 si chiama *distribuzione della varianza campionaria n-esima*.

Si prova che

$$E[S_n^2] = \frac{n-1}{n} \sigma^2 \quad (4.4)$$

$$V[S_n^2] = \frac{1}{n} \left(E[X^4] - \frac{n-3}{n-1} \sigma^4 \right)$$

Ancora per il teorema del limite centrale si prova che per $n \geq 30$, la sua distribuzione si può approssimare con

$$N(\bar{\mu}, \bar{\sigma})$$

$$\bar{\mu} = \frac{n-1}{n} \sigma^2$$

$$\bar{\sigma} = V[S_n^2] \quad .$$

A causa del fattore $\frac{n-1}{n}$ in (4.4), si preferisce considerare una nuova statistica

Definizione 4.5 - *Varianza campionaria n-esima corretta*

$$\hat{S}_n^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X}_n)^2 \quad (4.5)$$

da cui ovviamente

$$E[\hat{S}_n^2] = \sigma^2 \quad . \quad (4.6)$$

Abbiamo visto che una stima del valore atteso μ di una popolazione e' data dalla media campionaria \bar{X}_n , mentre una stima di σ e' data dalla varianza campionaria S_n^2 o quella corretta \hat{S}_n^2 .

FUNZIONI EXCEL 4.1

- DEV.ST = $\hat{S}_n = \sqrt{\frac{n}{n-1}} S_n$ da usare per un campione
- DEV.ST.POP = $S_n = \sqrt{\frac{1}{n} \sum_i (X_i - \bar{X}_n)^2}$ da usare su tutta la popolazione

4.3 Stimatori puntuali

Sia θ un parametro incognito di una popolazione X . Estratto dalla popolazione un campione di numerosita' n , chiamo **estimatore puntuale** $\hat{\theta}$ un numero costruito a partire dalle realizzazioni (x_1, x_2, \dots, x_n) del campione casuale. Per esempio la media campionaria e' un estimatore puntuale del parametro valore di aspettazione, come anche la varianza campionaria.

Definizione 4.6 - *Estimatore corretto o non-distorto (unbiased)*

*Un estimatore θ di una variabile aleatoria si dice **corretto o non-distorto (unbiased)**, se il suo valore di aspettazione coincide con il valore vero.*

Dalle proprieta' (4.2)₁, (4.6) ne segue che la media campionaria e la varianza campionaria corretta sono estimatori non-distorti, mentre la varianza campionaria e' un estimatore distorto.

Osserviamo che per stimare un parametro θ possiamo definire diversi estimatori corretti. Un criterio per stabilire quale sia preferibile e' il seguente

Definizione 4.7 *Siano $H_{1,n}, H_{2,n}$ due estimatori corretti del parametro θ di una popolazione. Allora diremo che H_1 e' piu' efficiente di H_2 se vale*

$$V[H_{1,n}] \leq V[H_{2,n}]$$

Esempio 4.2 Si considerino i seguenti estimatori della media μ di una popolazione

$$H_{1,n} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad , \quad H_{2,n} = \frac{X_1}{n} + \frac{X_2 + X_3 + \dots + X_n}{2(n-1)}$$

proviamo che il primo e' piu' efficiente del secondo.

Si osservi che il secondo estimatore da' piu' importanza alla prima componente del campione, a cui e' assegnato un peso di valore $\frac{1}{2}$ invece di $\frac{1}{2(n-1)}$. Supposto che tutte le variabili aleatorie del campione siano indipendenti (si veda (4.2)), e' facile vedere che i due estimatori sono corretti, cioe'

$$E[H_{1,n}] = E[H_{2,n}] = \mu$$

mentre

$$V[H_{1,n}] = \frac{\sigma^2}{n} \leq V[H_{2,n}] = \left[\frac{1}{4} + \frac{1}{4(n-1)} \right] \sigma^2 \quad \square$$

4.3.1 Altri metodi

Per effettuare delle stime puntali di parametri esistono altri metodi. Quelli principali sono il *metodo dei momenti* e di *massima verosimiglianza*. Nel seguito considereremo solo il secondo metodo con il seguente esempio

Esempio 4.3 Supponiamo di estrarre da una popolazione X distribuita secondo un'Esponenziale (3.4) di parametro incognita λ , il campione

$$(3.0, 4.1, 2.8, 5.5, 1.5, 2.2, 6.01, 2, 3.2, 0.9) \quad .$$

Vogliamo stimare λ con il metodo di massima verosimiglianza.

Siccome i campioni sono indipendenti, allora

$$\mathcal{L}(x_1, \dots, x_{10}, \lambda) = \prod_{i=1}^{10} \lambda e^{-\lambda x_i} = \lambda^{10} e^{-30.4\lambda}$$

La stima del parametro si ottiene determinando quel valore di λ per cui risulta massima la funzione \mathcal{L} , ovvero

$$\mathcal{L}' = \lambda^9 e^{-30.4\lambda} (10 - 30.4\lambda) = 0 \quad \rightarrow \quad \lambda = 0.329$$

Osservazione

Nel nostro caso la media campionaria $\bar{x} = 3.04$. Per una variabile aleatoria con distribuzione esponenziale sappiamo che

$$E[X] = \frac{1}{\lambda}$$

e se poniamo $E[x] \simeq \bar{x}$ otteniamo ancora $\lambda = 0.329$. Questa osservazione però non può essere generalizzata a tutte le stime ottenute con il metodo di massima verosimiglianza.

Esempio 4.4 *Supponiamo di effettuare N misure della stessa grandezza fisica x_i , che siano tra loro statisticamente indipendenti ed inoltre affette da errori casuali distribuiti secondo la legge di Gauss. La densità di probabilità corrispondente all'evento casuale costituito dall'osservazione degli N valori (applicando il teorema della probabilità composta)*

$$\prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{x^* - x_i}{2\sigma_i^2}\right)$$

dove x^* è il valore vero (incognito) e σ_i gli errori quadratici medi (noti). Detta funzione di verosimiglianza

$$\mathcal{L}(x_1, \dots, x_N, x) = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{x - x_i}{2\sigma_i^2}\right)$$

la stima più verosimile di x^* è quella che rende massima L , rispetto alla variabile incognita x

Si prova che questo massimo esiste ed è unico e vale

$$\bar{x} = \frac{1}{K} \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \quad , \quad K = \sum_{i=1}^N \frac{1}{\sigma_i^2} \quad (4.7)$$

4.4 Campionamento da una distribuzione normale

Consideriamo una popolazione normalmente distribuita $N(\mu, \sigma)$ ed estraiamo da essa un campione di numerosita' n . Valgono i seguenti teoremi:

Teorema 4.1 *Sia (X_1, \dots, X_n) un campione estratto da $N(\mu, \sigma)$, allora la media campionaria \bar{X}_n e' ancora distribuita secondo una normale ma con parametri μ e σ/\sqrt{n} , ovvero*

$$\bar{X}_n \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad . \quad (4.8)$$

Osservazione

Notiamo che il risultato (4.8) vale per qualsiasi n . Un simile risultato e' ottenuto con il teorema del limite centrale (3.16), dove pero' l'unica ipotesi e' che $n \geq 30$.

Teorema 4.2 *Sia (X_1, \dots, X_n) un campione estratto da $N(\mu, \sigma)$ di cui supponiamo di non conoscere μ . Allora la varianza campionaria S_n^2 e' tale che*

$$S_n^2 \simeq \frac{\sigma^2}{n-1} \chi_{n-1}^2 \quad . \quad (4.9)$$

Teorema 4.3 *Sia (X_1, \dots, X_n) un campione estratto da $N(\mu, \sigma)$. Allora la variabile aleatoria*

$$T_n = \frac{\bar{X}_n - \mu}{\frac{\hat{S}_n}{\sqrt{n}}} \simeq t_{n-1} \quad (4.10)$$

Osservazione

Per calcolare T_n occorre conoscere μ (media su tutta la popolazione) ed anche la media e la varianza campionaria, ma non la σ . Inoltre T_n e' distribuita secondo una t di Student che ha come solo parametro n (cioè il grado di libertà), e che quindi non dipende più dai parametri μ e σ .

4.5 Stime intervallari

La stima di un parametro di una popolazione data da un solo numero è detta *stima puntuale* (es. valore medio), mentre se è data da 2 numeri si dice *stima intervallare*.

Ad esempio: “La misura di una distanza è 5,28 mt (stima puntuale)”; “La misura di una distanza è compresa tra gli estremi $5,28 \pm 0,03$ mt (stima intervallare)”.

Sia θ un parametro puntuale di una popolazione (sconosciuta) e $\hat{\theta}$ una sua realizzazione. Un intervallo del tipo

$$I = [\hat{\theta} - e_1, \hat{\theta} + e_2] \subseteq \mathbb{R}$$

conterrà il valore θ con maggiore o minore probabilità a seconda dell'ampiezza.

se e_1, e_2 sono grandi, $P(\theta \in I) \simeq 1$

se $e_1 \simeq e_2 \simeq 0$, $P(\theta \in I) \simeq 0$

Definizione 4.8 - *Intervallo di confidenza per il parametro θ*

Fissato $\alpha \in [0, 1]$, si chiama *intervallo di confidenza con livello di fiducia α* , quell'intervallo:

$$[\hat{\theta} - e_1, \hat{\theta} + e_2]$$

tale che

$$P(\theta \in [\hat{\theta} - e_1, \hat{\theta} + e_2]) = 1 - \alpha$$

solitamente

$$\alpha = 0,1 \text{ (90\%)} \quad \alpha = 0,05 \text{ (95\%)} \quad \alpha = 0,01 \text{ (99\%)}$$

A questo punto si costruiscono intervalli di confidenza per i parametri *media* e *varianza* di una popolazione

4.5.1 Intervallo di confidenza per la media

a) *Popolazione non normalmente distribuita e varianza σ^2 nota*

Dal teorema del limite centrale sappiamo che la media campionaria \bar{X}_n è approssimabile (per $n \geq 30$) ad una variabile aleatoria con distribuzione normale con media μ (incognita) e deviazione standard $\frac{\sigma}{\sqrt{n}}$ (nota)

$$\bar{X}_n \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

che si può normalizzare definendo

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \simeq N(0, 1)$$

I valori assunti da Z dipendono dal campione (X_1, \dots, X_n) .

Si fissa un livello di fiducia α e si cerca quell'intervallo

$$[-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}]$$

tale che

$$P(Z \in [-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}]) = 1 - \alpha$$

e per la gaussiana normalizzata avrò:

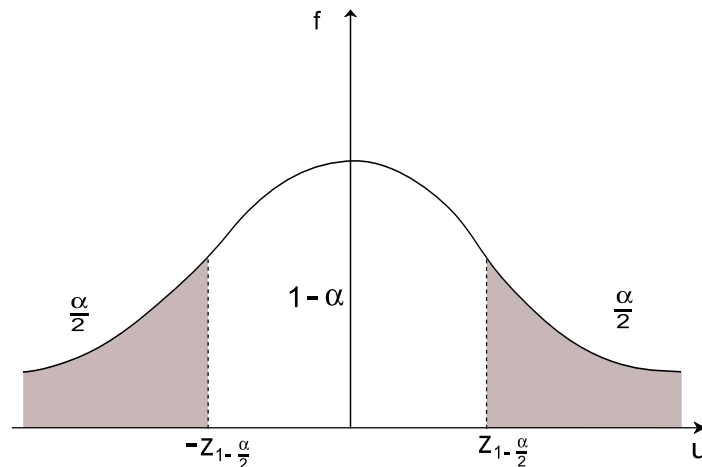


Figura 4.1: Intervallo di confidenza

$(1 - \alpha)$ è l'area staccata dal segmento $[-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}]$ sulla curva e quindi al di fuori di esso ho due regioni simmetriche aventi area $\frac{\alpha}{2}$, infatti $1 - \alpha + \frac{\alpha}{2} + \frac{\alpha}{2} = 1$.

$Z_{1-\frac{\alpha}{2}}$ è detto *quantile* della distribuzione normale standardizzata ed è tale che alla sua destra lascia un'area pari a $\frac{\alpha}{2}$ e alla sua sinistra un'area pari a $1 - \alpha + \frac{\alpha}{2} = 1 - \frac{\alpha}{2}$.

Essendo

$$P(X \leq z) = \int_{-\infty}^z f(u)du$$

allora avremo

$$P(X \leq Z_{1-\frac{\alpha}{2}}) = \int_{-\infty}^{Z_{1-\frac{\alpha}{2}}} f(u)du = 1 - \frac{\alpha}{2} \quad .$$

Una volta conosciuto il quantile

$$P(Z \in [-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}]) = 1 - \alpha$$

$$Z = \frac{\frac{\bar{X}_n - \mu}{\sigma}}{\frac{1}{\sqrt{n}}}$$

\Downarrow

$$P\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \in [-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}]\right) = 1 - \alpha$$

che risolta rispetto al parametro μ incognito:

$$-Z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{1-\frac{\alpha}{2}}$$

$$-Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\bar{X}_n - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$P\left(\bar{X}_n - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Quindi assegnata la media campionaria \bar{X}_n allora il valore medio (vero) μ della popolazione, con probabilità $1 - \alpha$, sta nell'intervallo

$$\left[\bar{X}_n - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] . \quad (4.11)$$

FUNZIONI EXCEL 4.2

Con EXCEL 2010 è possibile calcolare facilmente l'intervallo di confidenza in questo caso. Basta chiamare la funzione CONFIDENZA.NORM che richiede come argomenti α, σ e la dimensione del campione n , il cui risultato è $Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

FUNZIONI EXCEL 4.3

Se vogliamo calcolare il solo quantile $Z_{1-\frac{\alpha}{2}}$, allora si possono utilizzare altre funzioni

- DISTRIB.NORM.ST $= P(X \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$
Fissato z questa funzione mi restituisce l'area della Normale nell'intervallo $]-\infty, z]$, ovvero

$$z \rightarrow P(X \leq z)$$

- INV.NORM.ST Fissata l'area della Normale, questa funzione mi restituisce la z che corrisponde a quest'area, ovvero

$$P(X \leq z) \rightarrow z$$

quindi INV.NORM.ST è l'inversa di DISTRIB.NORM.ST .

In definitiva, per calcolare il quantile $Z_{1-\frac{\alpha}{2}}$, poichè per definizione esso lascia alla sua sinistra un'area pari a $1 - \frac{\alpha}{2}$:

$$\text{Fisso } \alpha \longrightarrow \text{calcolo } 1 - \frac{\alpha}{2} \longrightarrow \text{INV.NORM.ST}\left(1 - \frac{\alpha}{2}\right) \longrightarrow Z_{1-\frac{\alpha}{2}}$$

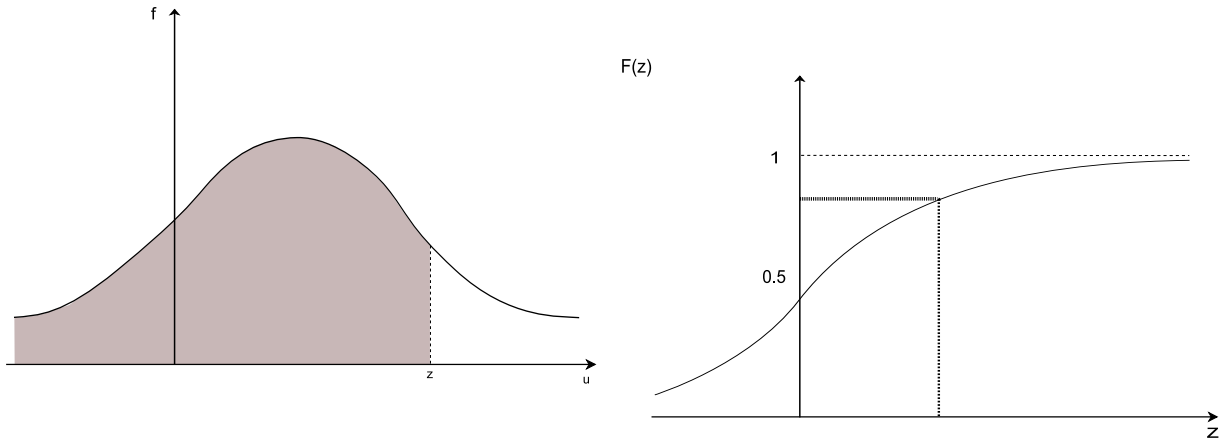


Figura 4.2: DISTRIB.NORM.ST e INV.NORM.ST

b) *Popolazione non normalmente distribuita e varianza σ^2 sconosciuta*

Si ragiona come nel caso a) sostituendo a σ una sua stima:

$$\sigma \sim \hat{S}_n = \sqrt{\frac{n}{n-1}} S_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Si prova che per n grande anche la variabile aleatoria:

$$\hat{Z} = \frac{\bar{X}_n - \mu}{\frac{\hat{S}_n}{\sqrt{n}}}$$

è distribuita come una normale standardizzata ed il valore medio μ della popolazione sta in

$$\left[\bar{X}_n - Z_{1-\frac{\alpha}{2}} \frac{\hat{S}_n}{\sqrt{n}}, \bar{X}_n + Z_{1-\frac{\alpha}{2}} \frac{\hat{S}_n}{\sqrt{n}} \right] \quad (4.12)$$

con probabilità $1 - \alpha$

c) *Popolazione normalmente distribuita e varianza σ^2 nota*

In questo caso prova che la media campionaria \bar{X}_n è una variabile aleatoria con $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Si ragiona come nel caso a) con la differenza che n può essere qualunque (nel caso a) invece era $n \geq 30$).

d) *Popolazione normalmente distribuita e varianza σ^2 sconosciuta*

La variabile aleatoria che si usa in questo caso e'

$$T_n = \frac{\overline{X}_n - \mu}{\frac{\widehat{S}_n}{\sqrt{n}}} \simeq t_{n-1} \quad (4.13)$$

che per la proprieta' (4.10) e' distribuita secondo una t di Student con $n - 1$ **gradi di liberta'**. Fissato $\alpha \in [0, 1]$ determino quel valore $t_{1-\frac{\alpha}{2}}$:

$$P\left(-t_{1-\frac{\alpha}{2}} \leq T_n \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Ovvero l'area della densità di probabilità di Student $f_X(t)$ compresa in $\left[-t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}\right]$ è $1 - \alpha$

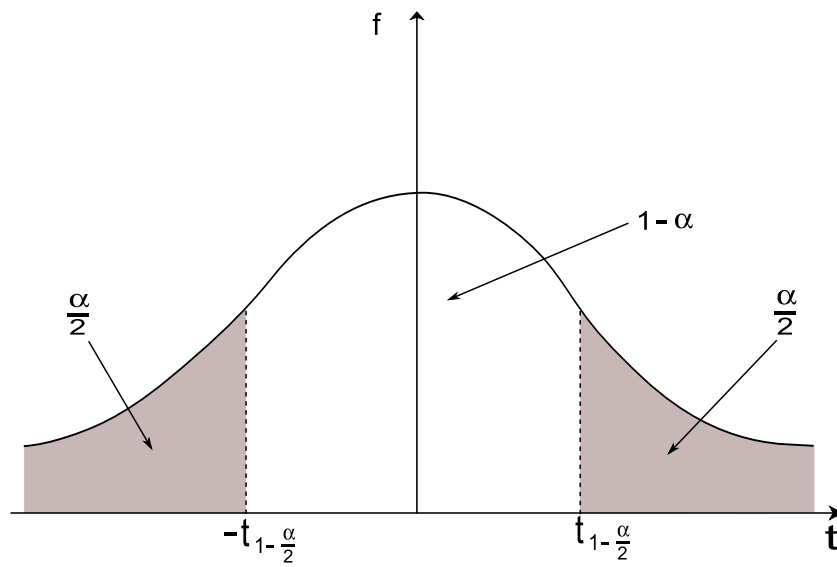


Figura 4.3: Densità di probabilità di Student

$t_{1-\frac{\alpha}{2}}$ è il *quantile* della t di Student ed è tale che lascia alla sua sinistra un'area di $1 - \frac{\alpha}{2}$ (per il calcolo con EXCEL del quantile si veda l'eq.(3.12).

Dalla formula precedente, poichè

$$T_n = \frac{\overline{X}_n - \mu}{\frac{\widehat{S}_n}{\sqrt{n}}}$$

si avrà

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

$$\Downarrow$$

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

ovvero, conoscendo \overline{X}_n e \widehat{S}_n , il parametro μ è compreso nell'intervallo

$$\left[\overline{X}_n - t_{1-\frac{\alpha}{2}} \frac{\widehat{S}_n}{\sqrt{n}}, \overline{X}_n + t_{1-\frac{\alpha}{2}} \frac{\widehat{S}_n}{\sqrt{n}}\right] \quad (4.14)$$

con probabilità $1 - \alpha$.

Poichè la variabile aleatoria T_n è distribuita secondo una t di Student per qualsunque valore di n , si parlerà di **statistica per piccoli campioni**. Se n è grande

$$t_n \rightarrow N(0, 1).$$

FUNZIONI EXCEL 4.4

Con EXCEL 2010 è possibile calcolare facilmente l'intervallo di confidenza in questo caso. Basta chiamare la funzione CONFIDENZA.T che richiede come argomenti α , \widehat{S}_n e la dimensione del campione n , il cui risultato è $t_{1-\frac{\alpha}{2}} \frac{\widehat{S}_n}{\sqrt{n}}$.

Una nota storica

William Gosset (1876-1937, il cui pseudonimo era Student) era un chimico inglese assunto dalla famosa birreria Guinness di Dublino ed eseguiva analisi statistiche su campioni dei prodotti per la mansione che oggi verrebbe chiamata controllo di qualità'. In generale, rilevare un campione costa sempre tempo e denaro. Per questo motivo, spesso Gosset era costretto ad usare per le sue indagini statistiche un numero ridotto di campioni. Gosset si accorse che, avendo una popolazione distribuita secondo una normale di parametri μ e σ incogniti, se prendo un piccolo campione esso non è ancora distribuito secondo una normale. Gosset scoprì che la variabile casuale T_n (4.13) nella sola incognita μ , è distribuita secondo una t_{n-1} .

osservazione:

Per intervalli di fiducia che hanno più di un parametro, $\exists t_1 \neq t_2$ tali che:

$$P(X \in [-t_1, t_2]) = 1 - \alpha$$

Nei casi precedenti si sono considerati intervalli simmetrici. La natura del problema determina la scelta dell'intervallo:

- i) se si vuole commettere il minimo errore nella scelta di μ si sceglierà l'intervallo simmetrico
- ii) se si vuole controllare che μ non raggiunga valori troppo grandi si preferirà un intervallo unilatero:

$$X \in] - \infty, t]$$

Esercizio 4.1 Consideriamo la tabella (1.3) dei carichi di rottura delle travi , relativa ad un campione di $n=15$.

Supponiamo che detto campione sia estratto da una popolazione con distribuzione normale di parametri incogniti μ e σ . Abbiamo visto in (4.10), che T_n e' distribuita come una t di Student con **$n-1$** gradi di liberta'.

a) Calcolare l'intervallo di confidenza per μ con livello di fiducia $\alpha = 0.01$

Questo intervallo sarà:

$$\left[\bar{X}_{15} - t_{1-\frac{\alpha}{2}} \frac{\hat{S}_{15}}{\sqrt{15}} \quad , \quad \bar{X}_{15} + t_{1-\frac{\alpha}{2}} \frac{\hat{S}_{15}}{\sqrt{15}} \right]$$

con

$$\bar{X}_{15} = \dots$$

$$\hat{S}_{15} = \dots$$

$$\alpha = 0.01$$

$$\mathbf{n} = 15$$

Per il calcolo del quantile $t_{1-\frac{\alpha}{2}}(n-1)$, come detto in (3.11), si utilizzi INV.T(α , $n-1$), oppure si utilizzi la funzione CONFIDENZA.T che dà come risultato $t_{1-\frac{\alpha}{2}} \frac{\hat{S}_{15}}{\sqrt{15}}$. Si ottiene [4770 , 5120]

b) Supponiamo di essere interessati ad una determinazione di μ che non superi un certo valore con $\alpha = 0.01$

Allora devo determinare $t_{1-\alpha}(n-1)$ tale che

$$P(T_{15} \in] - \infty, t_{1-\alpha}]) = 1 - \alpha$$

\Downarrow

$$P \left(-\infty \leq \mu \leq \bar{X}_{15} + t_{1-\alpha} \frac{\hat{S}_{15}}{\sqrt{15}} \right) = 1 - \alpha$$

Si trova $] - \infty, 5099]$, cioè con probabilità 0.99 il parametro μ non supera 5099.

Esercizio 4.2 Riprendere la tabella (1.2) costo al mq di 80 appartamenti e calcolare:

1. la media campionaria \overline{X}_{80}
2. la varianza campionaria \widehat{S}_{80}
3. per $\alpha = 0.05$ calcolare $Z_{1-\frac{\alpha}{2}}$ per la normale standardizzata
4. calcolare l'intervallo di fiducia per la media

$$\left[\overline{X}_{80} - Z_{1-\frac{\alpha}{2}} \frac{\widehat{S}_{80}}{\sqrt{80}} \quad , \quad \overline{X}_{80} + Z_{1-\frac{\alpha}{2}} \frac{\widehat{S}_{80}}{\sqrt{80}} \right]$$

5. con livello di fiducia $\alpha = 0.05$ stimare il limite superiore per la media μ cioè

$$P\left(\widehat{Z} \in]-\infty, Z_{1-\alpha}]\right) = 1 - \alpha$$

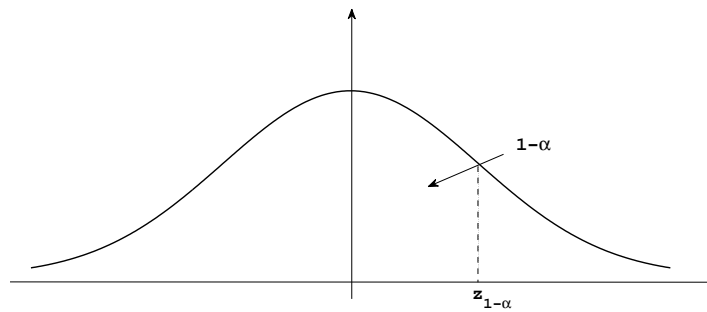


Figura 4.4: Limite superiore per la media

6. É evidente dalla formula (4.12) che l'intervallo di confidenza (a parità di α) dipende da n . Determinare n in modo che (con $\alpha = 0.05$) l'intervallo di confidenza simmetrico abbia ampiezza non superiore a 0.03

$$\overline{X}_n - t_{1-\frac{\alpha}{2}} \frac{\widehat{S}_n}{\sqrt{n}} \quad \overbrace{\hspace{1.5cm}}^{0.03} \quad \overline{X}_n + t_{1-\frac{\alpha}{2}} \frac{\widehat{S}_n}{\sqrt{n}}$$

Si possono determinare anche intervalli di confidenza per la varianza σ^2 della popolazione, ma è più complicato. Si hanno risultati significativi soltanto quando la popolazione è *normalmente distribuita*.

Capitolo 5

Verifica di Ipotesi

Un'affermazione relativa ad una caratteristica di una popolazione e' detta **ipotesi statistica** quando viene formulata in base all'esperienza o con considerazioni teoriche. *L'altezza media degli uomini italiani è 165 cm.* Questa è un'ipotesi da controllare con un'analisi statistica detta **verifica di ipotesi**. Per effettuare tale verifica si utilizzano i **test di ipotesi** che si suddividono in

1. Test parametrici: si riferiscono ad ipotesi relative ai parametri della distribuzione della popolazione. Tipici test parametrici sono
 - sulla media di una popolazione
 - sulla varianza di una popolazione
 - sulla differenza delle medie di due popolazioni
 - sulla differenza delle varianze di due popolazioni
 - test di incorrelazione
2. Test non parametrici: riguardano il tipo di distribuzione ipotizzabile o altre caratteristiche non esprimibili come parametri. In pratica sono dei test che non richiedono assunzioni sulla distribuzione dei dati, quali
 - test sulla bontà dell'adattamento
 - test per il confronto delle distribuzioni di due popolazioni
 - test per l'indipendenza

5.1 Caratteristiche generali di un test di ipotesi

Gli ingredienti che occorrono per un test di ipotesi sono:

- una popolazione statistica X su cui fare il test.
- un'ipotesi nulla H_0 , che è quella da convalidare sulla base del campione (X_1, X_2, \dots, X_n) di numerosità n estratto da X .
- un'ipotesi alternativa H_1 , che è l'ipotesi da considerare valida se si rifiuta H_0 .
- una distribuzione campionaria $T = T(X_1, \dots, X_n)$ di cui è nota la distribuzione quando H_0 è vera.
- una regione di accettazione \bar{C} , che è l'insieme dei valori assumibili dalla statistica T che portano all'accettazione di H_0 .
- una regione critica C , che è l'insieme dei valori assumibili dalla statistica T che portano al rifiuto di H_0 (accettazione di H_1).
- un livello di significatività (o ampiezza del test) denotato con α , che permette di individuare la regione di accettazione (o critica). Esso è tale che quando H_0 è vera, allora la statistica T assume valori nella regione critica con probabilità α .

Infine ricordiamo che l'obiettivo di una verifica di ipotesi non è quello di dire se questa ipotesi sia vera o falsa, ma piuttosto di dire se l'ipotesi fatta sia compatibile con i dati raccolti.

5.2 Test parametrico sulla media di una popolazione normale

5.2.1 Z test bilatero con varianza nota

Supponiamo di estrarre un campione aleatorio (X_1, \dots, X_n) proveniente da una popolazione normale con **media μ incognita** e varianza σ^2 nota. Fissata una costante μ_0 vogliamo verificare

l'ipotesi nulla:

$$H_0 \quad : \quad \mu = \mu_0$$

contro l'ipotesi alternativa

$$H_1 \quad : \quad \mu \neq \mu_0 \quad .$$

Un estimatore naturale di μ e' la media campionaria

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad .$$

Siccome la popolazione e' normale, per la proprieta' di campionamento (4.8), sappiamo che

$$\bar{X}_n \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \rightarrow \quad \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \simeq N(0, 1) \quad .$$

Se fosse vera la H_0 , scegliendo per μ il valore μ_0 la variabile aleatoria

$$U = \frac{\bar{X}_n - \mu_0}{\sqrt{\sigma^2/n}} \simeq N(0, 1) \tag{5.1}$$

ovvero U dovrebbe essere distribuita secondo una normale standardizzata. Fissato un livello di significativita' α

$$\left\{ \begin{array}{l} \text{accettiamo } H_0 \quad \text{se} \quad -Z_{1-\frac{\alpha}{2}} < U < Z_{1-\frac{\alpha}{2}} \\ \text{rifiutiamo } H_0 \quad \text{se} \quad U < -Z_{1-\frac{\alpha}{2}} \quad \cup \quad U > Z_{1-\frac{\alpha}{2}} \end{array} \right. \tag{5.2}$$

cioe' accettiamo H_0 se U cade all'interno della normale standardizzata delimitata da $[-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}]$ avente area $1 - \alpha$, mentre rifiutiamo H_0 se U cade in una delle due code di destra e sinistra aventi area complessiva α . Ovviamente al variare di α cambia il risultato del test. Per evitare di rifare il test con diversi valori di α , si introduce il $p - value$

Definizione 5.1 - $p - value$

E' il più basso livello di significatività a cui l'ipotesi nulla può essere rifiutata.

Nel nostro caso fissato U il corrispondente $p - value$ e' l'area delle due code $]-\infty, -U] \cup [U, +\infty[$, ovvero se $F_X(t)$ e' la funzione di ripartizione della normale standardizzata (3.10), allora

$$p - value = 2(1 - F_X(|U|)) \quad . \tag{5.3}$$

Calcolato il $p - value$, ovvero l'area delle due code, questa va confrontata con l'area della distribuzione normale (α) :

$$\begin{cases} \text{se } p - value \geq \alpha & \text{l'ipotesi nulla va accettata al livello } \alpha \\ \text{se } p - value < \alpha & \text{l'ipotesi nulla va rigettata al livello } \alpha \end{cases} \quad (5.4)$$

In pratica se il $p - value$ e' un numero molto piccolo allora si scartera' l'ipotesi nulla.

Esempio 5.1 *Esaminando un campione di 10 batterie di una ditta, e' stata riscontrata una durata media di 20.7 ore. Sapendo che le durate delle batterie sono variabili aleatorie con distribuzione normale e deviazione standard $\sigma = 3.5$ ore, possiamo affermare che la durata delle batterie e' di 22 ore ?*

In questo caso $H_0 : \mu = \mu_0 = 22$. Calcoliamo la variabile U (5.1), otterremo

$$U = -1.1746$$

Per calcolare $F_X(|U|) = F_X(1.1746)$ utilizzo $\text{DISTRIB.NORM.ST}(Z=1.1746) = 0.88$ da cui il $p\text{-value} = 2*(1-F_X(|U|)) = 0.24$. Quindi per ogni $\alpha \leq 0.24$ possiamo accettare l'ipotesi nulla. Poichè sarebbe stato assurdo eseguire un test con un livello di significativita' maggiore di 0.24, ne concludiamo che e' senz'altro opportuno accettare H_0 .

5.2.2 Z test unilatero con varianza nota

In questo caso avremo:

$$H_0 \quad : \quad \mu = \mu_0 \quad (\text{oppure } \mu \leq \mu_0)$$

contro l'ipotesi alternativa

$$H_1 \quad : \quad \mu > \mu_0.$$

In tal caso si ragiona come al caso precedente, solo che la regione critica sara' ad una sola coda.

Riassumendo:

Test	Ipotesi H_0	Ipotesi H_1	Regione rifiuto	$p - value$
due code	$\mu = \mu_0$	$\mu \neq \mu_0$	$ U > Z_{1-\frac{\alpha}{2}}$	$2 (1 - F_X(U))$
una coda	$\mu = \mu_0$ (oppure $\mu \leq \mu_0$)	$\mu > \mu_0$	$U > Z_{1-\alpha}$	$1 - F_X(U)$
una coda	$\mu = \mu_0$ (oppure $\mu \geq \mu_0$)	$\mu < \mu_0$	$U < -Z_{1-\alpha}$	$F_X(U)$

Tabella 5.1: Z test

Esempio 5.2 *Le solite batterie di una ditta hanno una durata dichiarata di 22 ore. Esaminando il solito campione di 20 batterie, e' stata riscontrata una durata media di 20.7 ore. Sapendo che le durate delle batterie sono variabili aleatorie con distribuzione normale e deviazione standard $\sigma = 3.5$ ore, possiamo affermare con livello di significativita' del 5 % che la durata delle batterie sia inferiore a 22 ore ?*

In questo caso $H_0 : \mu = \mu_0 = 22$ e $H_1 : \mu < \mu_0$. Calcoliamo la variabile U (5.1), otterremo

$$U = -1.661 \quad .$$

Per calcolare $F_X(U) = F_X(-1.661)$ utilizzo DISTRIB.NORM.ST(-1.661) = 0.048 da cui il p-value = $F_X(U) = 0.048$. Poiche' $\alpha = 0.05 > 0.048$ rifiuteremo l'ipotesi nulla e accetteremo quella alternativa: la durata media delle batterie e' inferiore a quanto dichiarato.

5.2.3 t test

In questo caso non conosciamo ne' la media μ che la varianza σ^2 della popolazione normale. Assegnato il solito campione posso calcolare la media campionaria (4.1) e la varianza campionaria corretta (4.5). Poiche' dal teorema (4.10) sappiamo che

$$T_n = \frac{\bar{X}_n - \mu}{\frac{\hat{S}_n}{\sqrt{n}}} \simeq t_{n-1}$$

possiamo definire la variabile aleatoria

$$U_t = \frac{\bar{X}_n - \mu_0}{\frac{\hat{S}_n}{\sqrt{n}}} \simeq t_{n-1} \quad . \quad (5.5)$$

Poichè la t di Student ha forma simile alla normale standardizzata (in particolare e' simmetrica rispetto l'origine), valgono i ragionamenti fatti per lo Z test. Avremo ancora la tabella 5.1, dove al posto di Z avremo la t_{n-1} e la F_{n-1} che e' la funzione di ripartizione della t di Student $n - 1$ gradi di liberta', che si calcola con la (3.14).

Esempio 5.3 *Le solite batterie di una ditta hanno una durata dichiarata di 22 ore. Esaminando il solito campione di 20 batterie, e' stata riscontrata una durata media di 20.7 ore. Sapendo che le durate delle batterie sono variabili aleatorie con distribuzione normale e varianza campionaria $\hat{S}_n = 3.5$ ore, possiamo affermare con livello di significativita' del 5 % che la durata delle batterie sia inferiore a 22 ore ?*

In questo caso $H_0 : \mu = \mu_0 = 22$ e $H_1 : \mu < \mu_0$. Calcoliamo la variabile U_t (5.5), otterremo

$$U_t = -1.661 \quad , \quad \text{dalla (3.14)} \quad p\text{-value} = F_{19}(|-1.661|) = 0.05656 \quad .$$

Poiche' $\alpha = 0.05 < 0.05656$ non possiamo rifiutare l'ipotesi nulla.

5.2.4 t test per il confronto delle medie di due popolazioni normali

Supponiamo che X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_m siano dei campioni indipendenti provenienti da due popolazioni normali di medie incognite rispettivamente μ_X, μ_Y e varianza σ_X^2, σ_Y^2 . Se calcoliamo le rispettive medie campionarie \bar{X} e \bar{Y} ovviamente questi due numeri non coincideranno. Allora ci chiediamo: la differenza fra queste medie e' significativa, oppure sono uguali con un certo livello di significativita' ? Cio' dipende solo da fluttuazioni casuali dovute al fatto che usiamo dei campioni di dimensioni ridotte? Anche in questo caso posso rifare il t test con l'ipotesi nulla:

$$H_0 \quad : \quad \mu_X = \mu_Y$$

contro l'ipotesi alternativa

$$H_1 \quad : \quad \mu_X \neq (>)\mu_Y \quad .$$

FUNZIONI EXCEL 5.1

Possiamo utilizzare la funzione TEST.T di EXCEL, nel seguente modo:

- Matrice 1 e Matrice 2 sono i dati dei due campioni;
- coda =(1,2): se coda =2 test bilatero ($H_1 : \mu_A \neq \mu_B$), se coda = 1 test unilatero;
- tipo =(1,2,3) : se tipo =3 i campioni hanno varianza diversa, se tipo =2 i campioni hanno varianza uguale ¹.

Il risultato di questa funzione e' il $p - value$ e quindi la decisione va presa secondo la solita regola (5.4).

Esercizio 5.1 *In un esperimento di laboratorio supponiamo di prendere due campioni,*

$$A = \{7, 8, 9, 8, 8, 7, 6, 7, 8, 7, 7, 8, 8, 7, 8, 6, 8, 7, 6, 7\}$$

$$B = \{7, 8, 7, 8, 9, 8, 8, 9, 7, 6, 8, 6, 9, 8, 9, 8, 8, 7\}$$

vogliamo sapere se le due medie provengono dalla stessa popolazione.

Si trova che $\bar{X}_A = 7.35$, $\hat{S}_A^2 = 0.66$, ed anche $\bar{X}_B = 7.77$, $\hat{S}_B^2 = 0.88$. Quindi le due medie sono diverse. Se applico la funzione TEST.T (coda=2, tipo=2) ottengo $p = 0.14$ che e' un valore piu' grande di 0.05. Questo significa che non posso rifiutare l'ipotesi nulla, cioe' la differenza tra le medie dipende da fattori casuali dovute al fatto che usiamo campioni di dimensioni ridotte.

5.2.5 t test per dati accoppiati

Siamo interessati a sapere se un l'installazione di un dispositivo contro l'inquinamento possa influire i consumi di alcune auto (ipotesi nulla). Siano X_i, Y_i i consumi dell'auto i -esima prima e dopo l'installazione del dispositivo. Questi due campioni non possono essere ovviamente considerati indipendenti: infatti se il consumo dell'auto X_1 e' molto elevato, ci aspettiamo che lo sia anche quando installiamo il dispositivo. Un possibile approccio per questa verifica di

¹In effetti in questo caso si dovrebbe preliminarmente effettuare un test di ipotesi per sapere se le varianze sono uguali.

ipotesi e' di prendere come variabile aleatoria la differenza dei consumi per ogni auto prima e dopo l'installazione del dispositivo $W_i = X_i - Y_i$. Se non vi fosse nessuna influenza del dispositivo, le W_i avrebbero media nulla

In questo caso avremo

$$H_0 : \mu_W = 0 \quad , \quad H_1 : \mu_W \neq 0 \quad .$$

Per eseguire questo test si utilizza ancora la funzione TEST.T con tipo = 1.

5.3 Test parametrico sulla media di una popolazione non normalmente distribuita

In questo caso per poter definire un test occorre avere un campione di numerosita' $n \geq 30$. Infatti, dal teorema del limite centrale 3.16 sappiamo che la media campionaria e' distribuita secondo una normale di parametri μ e σ/\sqrt{n} , quando σ e' nota, o \hat{S}_n/\sqrt{n} . Ci si puo' ricondurre al caso con popolazione normalmente distribuita e varianza nota.

5.3.1 Conclusione

Sinora abbiamo visto test di ipotesi che riguardavano la media di una popolazione, ma e' possibile effettuare altri test sulla varianza di una popolazione, che rimandiamo a test specifici.

Nella tabella 5.2 riassumiamo i vari procedimenti da eseguire per effettuare un test di ipotesi sulla media di una popolazione. Osserviamo che il procedimento d) vale per $n < 30$ e per questo la statistica di Student prende in nome di statistica per piccoli campioni. Questo test va fatto se la popolazione da cui estraiamo il campione è rigorosamente normale. Nel mondo reale, non saremo mai sicuri che ciò accada. Quindi applicare il t test a piccoli campioni presi da una popolazione di cui non sappiamo la distribuzione può dare risposte totalmente errate.

Se $n \geq 30$ il teorema del limite centrale ci permette di applicare i procedimenti a) e b) a qualsiasi popolazione. Si osservi che, sempre per $n \geq 30$, il procedimento d) coincide con il

procedimento	ipotesi	Statistica test	distribuzione	test
a)	$n \geq 30, \sigma \text{ nota}$	$Z = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	distribuzione normale	Z test
b)	$n \geq 30, \sigma \text{ incognita}$	$Z = \frac{\bar{X}_n - \mu_0}{\frac{S_n}{\sqrt{n}}}$	distribuzione normale	Z test
c)	$n < 30, \text{pop. norm. } \sigma \text{ nota}$	$Z = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	distribuzione normale	Z test
d)	$n < 30, \text{pop. norm. } \sigma \text{ incognita}$	$T = \frac{\bar{X}_n - \mu_0}{\frac{S_n}{\sqrt{n}}}$	t_{n-1} Student	t test

Tabella 5.2: Tabella riassuntiva test di ipotesi sulla media

b), perché sappiamo dal teorema del limite centrale che la media campionaria è distribuita secondo una normale ed inoltre la distribuzione t di Student coincide con quella normale.

5.4 Test non parametrici sulla bontà dell'adattamento

Questi test permettono di rispondere alla domanda:

Possiamo affermare che la popolazione X esaminata è distribuita secondo una specifica funzione di ripartizione F ?

Il test è detto *di bontà dell'adattamento* perchè si chiede se la distribuzione specificata F è adatta a descrivere la popolazione X

Definizione 5.2 - *Ipotesi nulla*

$$H_0 : F_X(t) = F(t) \quad \forall t \in \mathbb{R}$$

dove $F_X(t)$ è la reale distribuzione della popolazione X , mentre F è la distribuzione da noi specificata.

Definizione 5.3 - *Ipotesi alternativa*

$$H_1 : F_X(t) \neq F(t) \quad \text{per almeno un } t \in \mathbb{R}$$

5.4.1 Test di Kolmogorov - Smirnov

Si deve presupporre che $F(t)$ sia **continua**.

Sia X una popolazione e (X_1, X_2, \dots, X_n) un suo campione di numerosità n estratto dalla popolazione. Si chiama *funzione di ripartizione empirica* della popolazione X la funzione:

$$\hat{F}_{X_n}(t) = \frac{1}{n} \sum_{i=1}^n U_{[0,t]}(X_i) \quad \forall t \in \mathbb{R}$$

dove

$$U_{[0,t]}(X_i) = \begin{cases} 1 & \text{se } X_i \in [0, t]; \\ 0 & \text{altrimenti.} \end{cases}$$

che è una funzione crescente a gradino.

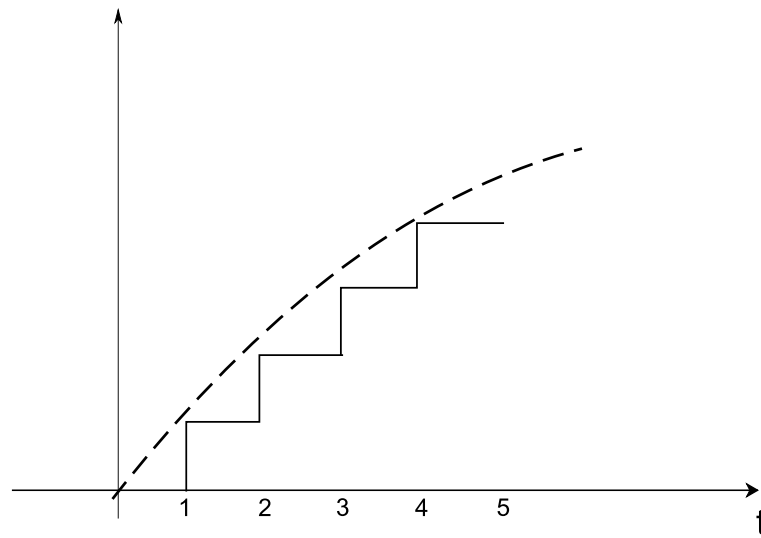


Figura 5.1: Funzione di ripartizione empirica

Sia X il carattere della popolazione su cui siamo interessati a fare deduzioni statistiche (ad es. il peso e l'altezza delle persone).

Ovviamente il valore di X varia al variare dell'elemento considerato e si può pensare come una variabile aleatoria la cui distribuzione è sconosciuta.

Un campione casuale di numerosità n è una n -upla (X_1, X_2, \dots, X_n) di variabili aleatorie distribuite con il carattere della popolazione. I valori (x_1, x_2, \dots, x_n) assunti da questa n -upla (che sono le misure effettivamente fatte) sono una *realizzazione* di (X_1, X_2, \dots, X_n)

Esempio 5.4 Sia la popolazione una colonia di batteri e

$$X = \text{“tempo di riproduzione della colonia”}$$

Dalla popolazione estraggo 10 campioni

$$(X_1, X_2, \dots, X_{10})$$

la cui realizzazione (ovvero le misure fatte sul tempo di riproduzione) dà i valori (in opportune unità di tempo):

$$(3.0, 4.1, 2.8, 5.5, 1.5, 2.2, 6.0, 1.2, 3.2, 0.9)$$

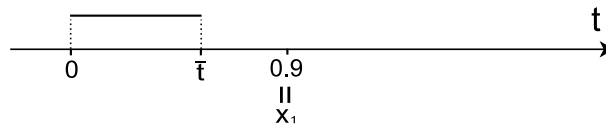
Si deve costruire la funzione di ripartizione empirica. Si riordinano quindi in modo *crescente* i dati:

$$\begin{aligned} x_1 = 0.9, \quad x_2 = 1.2, \quad x_3 = 1.5, \quad x_4 = 2.2, \quad x_5 = 2.8 \\ x_6 = 3.0, \quad x_7 = 3.2, \quad x_8 = 4.1, \quad x_9 = 5.1, \quad x_{10} = 6.0 \end{aligned}$$

$$\widehat{F}_{X_n}(t) = \frac{1}{10} \sum_{i=1}^{10} U_{[0,t]}(X_i)$$

$$U_{[0,t]}(X_i) = \begin{cases} 1 & \text{se } X_i \in [0, t]; \\ 0 & \text{altrimenti.} \end{cases}$$

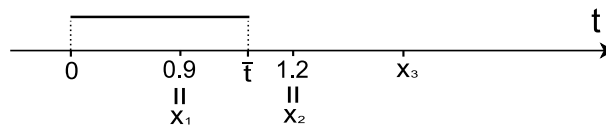
1) se $\bar{t} \in [0, 0.9[$



$$\left\{ \begin{array}{l} x_1 = 0.9 \notin [0, \bar{t}] \\ x_2 = 1.2 \notin [0, \bar{t}] \\ \vdots \\ x_{10} = 6.0 \notin [0, \bar{t}] \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} U(x_1) = 0 \\ U(x_2) = 0 \\ \vdots \\ U(x_{10}) = 0 \end{array} \right.$$

$$\widehat{F}_{X_{10}}(\bar{t}) = 0 \quad \text{per } \bar{t} \in [0, 0.9[$$

2) se $\bar{t} \in [0.9, 1.2[$

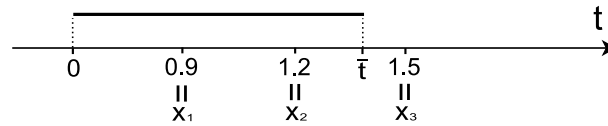


$$x_1 = 0.9 \in [0, \bar{t}], \quad x_2 = 1.2 \notin [0, \bar{t}], \quad \dots \quad x_{10} = 6.0 \notin [0, \bar{t}]$$

$$\begin{array}{ccc} \Downarrow & \Downarrow & \Downarrow \\ U(x_1) = 1 & U(x_2) = 0 & \dots \quad U(x_{10}) = 0 \end{array}$$

$$\widehat{F}_{X_{10}}(\bar{t}) = \frac{1}{10} \quad \text{per } \bar{t} \in [0.9, 1.2[$$

3) se $\bar{t} \in [1.2, 1.5[$



$$x_1 = 0.9 \in [0, \bar{t}], \quad x_2 = 1.2 \in [0, \bar{t}], \quad x_3 = 1.5 \notin [0, \bar{t}], \quad \dots \quad x_{10} = 6.0 \notin [0, \bar{t}]$$

\Downarrow

$$U(x_1) = 1$$

\Downarrow

$$U(x_2) = 1$$

\Downarrow

$$U(x_3) = 0$$

\dots

\Downarrow

$$U(x_{10}) = 0$$

$$\hat{F}_{X_{10}}(\bar{t}) = \frac{1}{10}(1 + 1) = \frac{2}{10} \quad \text{per } \bar{t} \in [1.2, 1.5[$$

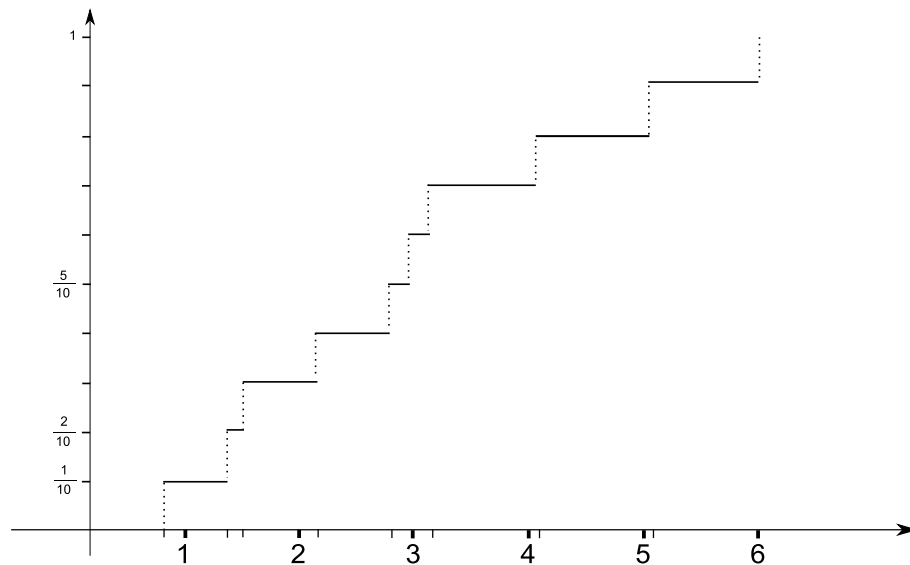


Figura 5.2: Funzione di ripartizione empirica per l'esempio

La $\hat{F}_{X_{10}}(t)$ è tale che in corrispondenza di ciascuno di questi valori essa compie un “salto” di $\frac{1}{10}$, partendo da 0 e arrivando a 1 in corrispondenza dell'ultimo valore ($t = 6$).

Il test di Kolmogorov si basa sulla statistica

$$D_n = \sup_{t \in \mathbb{R}} |F(t) - \hat{F}_{X_n}(t)|$$

che è il valore più grande delle distanze tra la funzione di distribuzione empirica e quella teorica.

La distribuzione statistica di D_n è stata studiata dagli autori del test al variare di n , fornendo

apposite tabelle per determinare i quantili.

È intuitivo aspettarsi che D_n assume valori piccoli se H_0 è vera, mentre assume valori grandi se H_0 è falsa.

Fissato il livello di significatività α , si prova che la regione critica (in cui si rifiuta H_0) è

$$C =]d_{1-\alpha}, 1]$$

dove $d_{1-\alpha}$ è il quantile della distribuzione, cioè quel valore per cui risulta

$$P(D_n \leq d_{1-\alpha}) = 1 - \alpha$$

dove $d_{1-\alpha}$ si calcola con apposite tavole.

In pratica:

- i) si calcola D_n e dalle tavole $d_{1-\alpha}$ (fissato α).
- ii) se $d_{1-\alpha} < D_n \leq 1$, si è nella regione critica quindi H_0 viene rifiutata.
- iii) se $D_n \leq d_{1-\alpha}$, H_0 viene accettata con livello di significatività α .

Esercizio 5.2

- Sovrapporre il grafico della distribuzione empirica dell'esempio precedente $\hat{F}_{X_{10}}(t)$, con il grafico della funzione di ripartizione

$$F(t) = \begin{cases} 0 & \text{se } t < 0; \\ 1 - \exp(-0.35t) & \text{se } t \geq 0. \end{cases}$$

- Provare che $D_n = \sup |F(t) - \hat{F}_{X_{10}}(t)| = ?$ e trovare per quale valore di t si ha la massima distanza.
- Fissato $\alpha = 0.05$ ($n = 10$), dalla tabella trovare $d_{1-\alpha}$ e vedere se H_0 può essere accettata o no.

Per la tabella del Test di Kolmogorov Smirnov si veda il seguente URL

<http://www.eridlc.com/onlinetextbook/index.cfm?fuseaction=textbook.appendix&FileName=Table7>

5.4.2 Test del Chi-quadro

Questo test **non necessita** dell'ipotesi che $F(t)$ sia una funzione continua.

Sia X una popolazione da cui si estrae un campione (X_1, X_2, \dots, X_n) di numerosità n . Sia $F(t)$ la funz. di ripartizione che si vuole testare quale possibile distribuzione della popolazione X .

Si chiama *supporto* di $F(t)$:

$$\text{supp } F = \{t \in [a, b] \mid F(t) \neq 0\}$$

suddividiamo il supporto di F in K intervalli o classi:

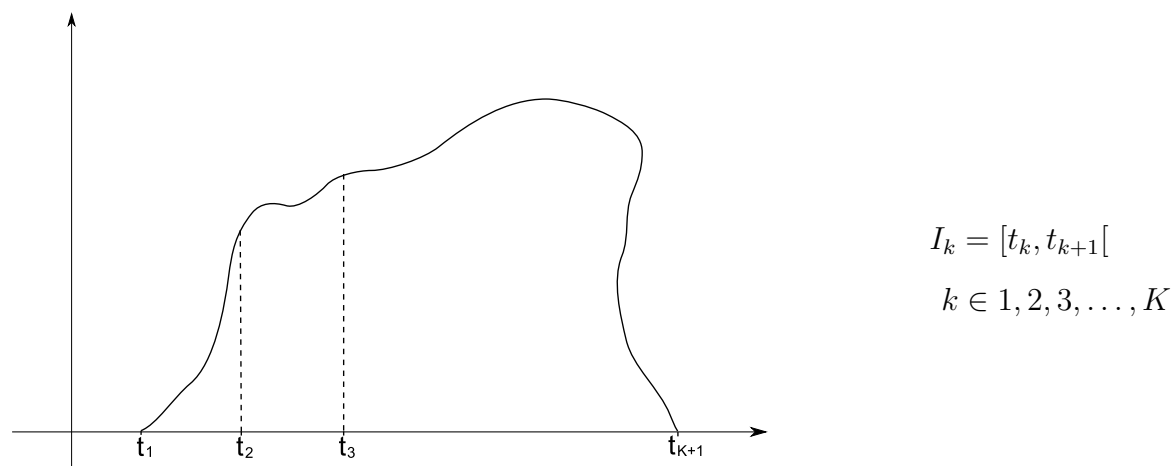


Figura 5.3: Supporto di $F(t)$

Definiamo:

n = numerosità del campione

O_k = n° elementi del campione che cadono in I_k (frequenze osservate)

p_k = probabilità teorica che il singolo campione X_i cada in I_k se H_0 è vera (cioè $F_X(t) = F(t) \forall t$)

$A_k = p_k n$ (frequenze attese).

Possiamo costruire la seguente tabella, detta ad un'entrata, perché frequenze osservate occupano una sola riga

Tabella del Chi-quadro ad un'entrata

Campione	X_1	X_2	X_3	X_K
freq. osservate	O_1	O_2	O_3	...	O_K
freq. attese	A_1	A_2	A_3	...	A_K

osservazione:

O_k sono le frequenze osservate negli intervalli I_k mentre A_k quelle attese. Se H_0 (vedi Definizione (5.2)) è vera le differenze $|A_k - O_k|$ sono piccole. Allora considero la statistica

$$W = \sum_{k=1}^K \frac{(O_k - A_k)^2}{A_k} \quad (5.6)$$

Teorema 5.1 *Se H_0 è vera e $O_k \geq 5$, allora W è distribuita come una distribuzione χ^2 , con:*

$K - 1$ gradi di libertà se la funzione di ripartizione F è stata decisa arbitrariamente senza fare uso di dati campionari.

$K - r - 1$ gradi di libertà se nella funzione di ripartizione compaiono r parametri stimati con dati campionari.

Si dimostra che, fissato il livello di significatività α , la regione critica della statistica è

$$C =]\chi_{1-\alpha}^2, +\infty[$$

dove $\chi_{1-\alpha}^2$ è il *quantile* della distribuzione χ^2 con $K - 1$ o $K - r - 1$ gradi di libertà, e si può determinare con l'utilizzo di una tabella.

Ricordiamo che il quantile $\chi_{1-\alpha}^2$ lascia alla sua destra un'area pari ad α (che è l'area della regione critica C , vedi figura 5.4), cioè:

$$P(W > \chi_{1-\alpha}^2) = \alpha$$

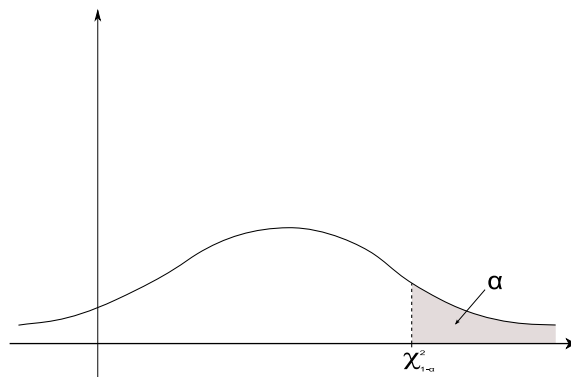


Figura 5.4: Grafico χ^2

In pratica, se:

$W < \chi^2_{1-\alpha}$, il test su H_0 viene accettato con livello α (cade fuori dalla regione critica). L'ipotesi nulla e' vera con una prob. di errore di α ovvero l'ipotesi nulla e' vera con una probabilita' di $1 - \alpha$.

$W > \chi^2_{1-\alpha}$, il test su H_0 viene rifiutato con livello α (cade nella regione critica). L'ipotesi nulla e' falsa una prob. di errore di α ovvero l'ipotesi nulla e' falsa con una probabilita' di $1 - \alpha$.

Per il calcolo del quantile $\chi^2_{1-\alpha}$ esistono delle apposite tabelle.

osservazione:

Poiché basta che $O_k \geq 5$, la statistica si adatta per i piccoli campioni.

Tabelle di contingenza

Abbiamo finora supposto che le frequenze osservate O_k occupino solo un vettore riga. Può succedere che queste frequenze occupino una tabella $H \times K$, che prende il nome di *tabella di contingenza a doppia entrata*. In questo caso avremo la matrice delle frequenze osservate O_{hk} e quella delle frequenze attese A_{hk} . In analogia al caso precedente si definirà una statistica W , dove i gradi di libertà saranno $(H - 1)(K - 1)$ se la funzione di ripartizione F è stata decisa arbitrariamente senza fare uso di dati campionari.

La matrice delle frequenze attese vengono stabilite secondo una particolare ipotesi nulla.

Test del chi-quadro per l'indipendenza

Supponiamo di avere una popolazione e che ogni membro venga classificato secondo due caratteristiche X e Y che assumano rispettivamente H e K valori possibili. Preso un elemento a caso della popolazione, indico con P_{hk} la probabilità che la sua caratteristica X assuma il valore h e quella Y il valore k , ovvero

$$P_{hk} = P(X = h \cap Y = k) \quad h = 1, \dots, H, \quad k = 1, \dots, K.$$

Per ogni elemento della popolazione posso anche definire le distribuzioni marginali

$$p_h = P(X = h) = \sum_k P_{hk}, \quad q_k = P(Y = k) = \sum_h P_{hk}.$$

In generale le due caratteristiche X e Y di ogni elemento della popolazione sono dipendenti. Vogliamo verificare invece che le due caratteristiche siano **una indipendente dall'altra** (vedi (2.3)), ovvero

$$\text{Ipotesi nulla } H_0 : \quad P_{hk} = P(X = h) P(Y = k) = p_h q_k \quad \forall h, k$$

$$\text{Ipotesi alternativa } H_1 : \quad P_{hk} \neq p_h q_k \quad \text{per qualche } h, k.$$

Per calcolare queste probabilità, consideriamo un campione della popolazione di numerosità n , e definiamo :

- N_{hk} numero di elementi del campione per cui $(X = h \cap Y = k)$
- $N_h = \sum_k N_{hk}$, $M_k = \sum_h N_{hk}$
- $p_h = \frac{N_h}{n}$, $q_k = \frac{M_k}{n}$

da cui le frequenze attese, in caso di indipendenza, sono

$$A_{hk} = n p_h q_k = \frac{N_h M_k}{n} \quad (5.7)$$

su cui applicheremo il test.

Esempio 5.5 *Per verificare la qualita' della produzione in una fabbrica, un ingegnere controlla il numero di pezzi difettosi prodotti da tre macchine diverse ottenendo la seguente tabella di dati*

	macchina 1	macchina 2	macchina 3
buoni	150	140	200
difettosi	25	40	20

Si puo' ritenere che la qualita' di pezzi difettosi non dipenda dalla macchina che si utilizza ?

In questo caso la popolazione dei pezzi prodotti viene classificata tramite le due caratteristiche

$X = \text{qualita' del prodotto} = \{\text{buono}, \text{difettoso}\}$, $h = 1, 2$

$Y = \text{tipo di macchina} = \{\text{macchina1}, \text{macchina2}, \text{macchina3}\}$, $k = 1, 2, 3$.

Partendo da questa tabella si costruisce la tabella delle frequenze attese A_{hk} , ovvero delle frequenze che si avrebbero nell'ipotesi di indipendenza, la (5.7) diventa

$$A_{hk} = \frac{(\text{totale riga } h) \times (\text{totale colonna } k)}{\text{totale generale}}$$

dove 'totale generale' e' la somma di tutti i pezzi prodotti. Quindi in analogia a (5.6) si costruisce la statistica W , che adesso sara'

$$W = \sum_{h=1}^H \sum_{k=1}^K \frac{(O_{hk} - A_{hk})^2}{A_{hk}^2} \quad (5.8)$$

e si esegue il test come nel caso precedente.

Esempio 5.6 Si lancia una moneta 200 volte e si osservano 115 teste e 85 croci. Verificare l'ipotesi che la moneta è buona con vari livelli di significatività.

$\hookrightarrow H_0$: la moneta è buona , $\hookrightarrow H_1$: la moneta non è buona

Poichè ho 2 risultati, ho $K = 2$ che suddivido in

$$\text{Testa} \quad \left\{ \begin{array}{l} k = 1 \\ O_1 = 115 > 5 \end{array} \right. \quad \text{Croce} \quad \left\{ \begin{array}{l} k = 2 \\ O_2 = 85 > 5 \end{array} \right.$$

inoltre il n° di lanci è $n=200$.

La probabilità teorica (se la moneta è buona) è:

$$p_1 = p_2 = \frac{1}{2}$$

allora

$$W = \frac{(O_1 - A_1)^2}{A_1} + \frac{(O_2 - A_2)^2}{A_2} = 4.5$$

Il grado di libertà è $K - 1 = 2 - 1 = 1$

i) Fisso $\alpha = 0.05$

$$\chi^2_{1-\alpha} = 3.84$$

$W > \chi^2_{1-\alpha}$, l'ipotesi che la moneta sia buona con livello $\alpha = 0.05$ viene rifiutata

l'ipotesi nulla e' falsa con prob. di errore del 5%

l'ipotesi nulla e' falsa con prob. del 95%

ii) Fisso $\alpha = 0.01$

$$\chi^2_{1-\alpha} = 6.63$$

$W < \chi^2_{1-\alpha}$, l'ipotesi che la moneta sia buona con livello $\alpha = 0.01$ non può essere rifiutata

l'ipotesi nulla e' vera con prob. di errore del 1%

l'ipotesi nulla e' vera con prob. del 99%

Conclusione: Risultati discordi \implies probabilmente la moneta non è buona, quindi fare altri test.

FUNZIONI EXCEL 5.2

- TEST.CHI : questa funzione effettua solamente il test del chi-quadro per l'indipendenza

$Int_effettivo$ = intervallo dati con frequenza osservata (115, 85)

$Int_previsto$ = intervallo dati con frequenze teoriche $np_k \equiv (100, 100)$

W e χ^2 **non vengono visualizzati!**

Risultato formule:

$$R = P(W > \chi^2)$$

ovvero la probabilità che W assuma valore maggiore della statistica test (che non viene calcolata). In pratica R è il $p - value$ (5.4) !

Scelto un livello α

1. $R < \alpha$ rifiuto l'ipotesi nulla con livello α (ovvero l'ipotesi nulla è falsa con prob. di errore di α)
2. $R > \alpha$ accetto l'ipotesi nulla con livello α (ovvero l'ipotesi nulla è vera con prob. di errore di α)

Nel nostro esempio $\rightarrow R = 0.03389 \begin{cases} < 0.05, & \text{rifiuto;} \\ > 0.01, & \text{accetto.} \end{cases}$

Esercizio 5.3 *Si lancia una dado 2000 volte e si osservano i seguenti risultati:*

1	→	388	4	→	316
2	→	322	5	→	344
3	→	314	6	→	316

Si può pensare che il dado sia equilibrato? (ad occhio si vede che il risultato 1 è apparso un numero di volte superiore agli altri casi).

Dati:

$n = 2000$ (numerosità del campione)

$K = 6$ (6 risultati)

$I_1 = [1, 2[$, $I_2 = [2, 3[$, $I_3 = [3, 4[$, $I_4 = [4, 5[$, $I_5 = [5, 6[$, $I_6 = [6, 7[$

La probabilità teorica è

$$p_k = \frac{1}{6} \quad k = 1, \dots, 6$$

Esercizio 5.4 *Un gruppo di pazienti si è lamentato di non dormire bene. Ad alcuni sono state date delle pillole di sonnifero ad altri zucchero (dicendo però che si trattava di sonnifero). In una tabella sono messi il risultati della prova:*

	Hanno dormito	Non hanno dormito
sonnifero SI	44	10
sonnifero NO	81	35

Supposto che tutti abbiano detto la verità, provare l'ipotesi che non ci sia differenza tra le pillole di zucchero e quelle di sonnifero (e determinare il livello di significatività).

Esercizio 5.5 Una tabella di 250 numeri casuali semplici ha mostrato la seguente distribuzione:

n°	0	1	2	3	4	5	6	7	8	9
frequenza	17	31	29	18	14	20	35	30	20	36

Questa tabella differisce significativamente dalla distribuzione uniforme attesa?

Esercizio 5.6 Da un'inchiesta fatta si è determinata la seguente tabella “età guidatore - n° incidenti”:

		<i>ETA'</i>				
		(21-30)	(31-40)	(41-50)	(51-60)	(61-70)
numero	0	748	821	786	720	672
incidenti	1	74	60	51	66	50
	2	31	25	22	16	15
	> 2	9	10	6	5	7

Con livello di significatività 0.05 e 0.01, provare l'ipotesi che il n° di incidenti non dipende dall'età del guidatore.

Capitolo 6

Numeri Casuali

I generatori di numeri casuali (Random Number Generator) sono fondamentali per varie applicazioni:

- i) Esperimenti statistici, analisi di algoritmi
- ii) Simulazione di sistemi stocastici
- iii) Analisi numerica basata su metodi Monte Carlo
- iv) Algoritmi probabilistici
- v) Computer Games (es. Virtual Casinò)
- vi Crittografia
- vii) Protocolli di comunicazione sicura
- viii) Gambling machines

Librerie software che contengono RNG:

Excel, Matlab, Fortran, VisualBasic, Java, ...

6.1 Generazione dei numeri casuali con densita' di probabilita' uniforme

I numeri casuali possono derivare da:

- a) fenomeni naturali (ad es. tempo di decadimento delle particelle α nell'uranio)
- b) algoritmi matematici che simulano i fenomeni naturali

Questi algoritmi hanno carattere deterministico a parità di seme iniziale, la sequenza di output è sempre la stessa. Inoltre la sequenza di numeri generata si ripete con un periodo T , per questo si parla di *numeri pseudocasuali*.

Un buon RNG deve avere:

- i) periodo T molto grande
- ii) generazione efficiente dei numeri: veloce e uso di poche risorse
- iii) sequenza riproducibile
- iv) portabilità del codice

6.1.1 RNG basato su ricorrenze lineari

$$X_{n+1} = (a_0x_n + a_1x_{n-1} + \dots + a_jx_{n-j}) \mod P$$

l'operatore $p \mod q$ è il resto della divisione $\frac{p}{q}$

Questo generatore ha bisogno di $j + 1$ numeri iniziali (x_0, x_1, \dots, x_j) ed è caratterizzato da un periodo T , che nel migliore dei casi

$$T \leq P^{j+1}$$

quindi il valore di T e le caratteristiche dei numeri dipendono dai valori di a_j e P .

6.1.2 RNG moltiplicativo congruenziale

È un caso particolare del precedente con:

$$a_j = 0 \quad j \geq 1$$

$$X_{n+1} = \lambda x_n \mod P$$

dove $P = 2^\beta$ o 10^β (a seconda che il computer è binario o decimale) e β è la lunghezza della parola in bit.

Una volta ottenuta la successione x_i , una ulteriore funzione di OUT mi dà il numero random, compreso nell'intervallo $[0,1]$:

$$u_i = g(x_i)$$

Visual Basic

$$x_{i+1} = (1140671485x_i + 12820163) \mod 2^{24}$$

$$u_i = \frac{x_i}{2^{24}}$$

Java

$$x_{i+1} = (25214903917x_i + 11) \mod 2^{48}$$

$$u_i = \frac{\left\{ 2^{27} \left[\frac{x_{2i}}{2^{22}} \right] + \left[\frac{x_{2i+1}}{2^{21}} \right] \right\}}{2^{53}}$$

6.2 Test statistici per i numeri casuali

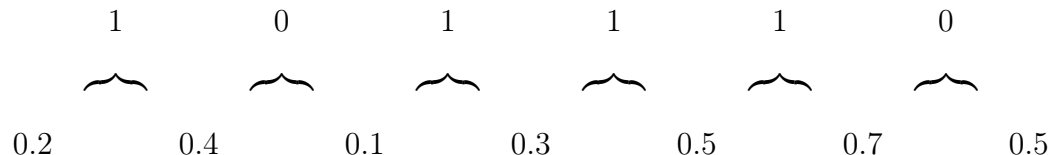
a) Uniformità

Il test del χ^2 può essere effettuato per testare l'uniformità del generatore. Divido l'intervallo $[0, 1]$ in k intervalli disgiunti. L'ipotesi da verificare è che in ogni intervallino cadano lo stesso n° di numeri casuali (esercizio per casa).

b) Test per il grado di casualità

b₁) Up and Down test

Assegnata una sequenza di N numeri casuali, si assegna 1 se si ha un incremento del numero, altrimenti si assegna 0:



La sequenza di 1,0 che ho è

1 0 1 1 1 0

Posso anche osservare per quanto tempo la sequenza è crescente o no (lunghezza del numero). Nel nostro caso

1 0 1 1 1 0

Run Up di lunghezza 1

Run Down di lunghezza 1

Run Up di lunghezza 3

Run Down di lunghezza 1

In una buona sequenza di numeri casuali non dovremmo avere un numero elevato di *run* di grande lunghezza.

È stata messa a punto una statistica per dei buoni numeri casuali, sulla base del numero di *run* e della loro lunghezza

n° <i>run</i> lunghezza 1	$\frac{N+1}{12}$
n° <i>run</i> lunghezza 2	$\frac{11N-14}{12}$
.....	
n° <i>run</i> lunghezza k con $3 \leq k < N-1$	$\frac{2[(k^2+3k+1)N - (k^3+3k^2-k-4)]}{(k+3)!}$
n° <i>run</i> lunghezza $N-1$	$\frac{2}{N!}$

Ci sono delle tavole che danno le regioni critiche al variare del n° di *run* e della loro lunghezza.

*b*₂) **Gap test**

Assegnata una sequenza di numeri casuali

$$\{X_i\} \quad i = 1, \dots, N$$

diremo che ogni sottosuccessione

$$x_j, \dots, x_{j+r}$$

di $r+1$ numeri casuali, rappresenta un *gap di lunghezza r* se

$$\begin{cases} x_j \text{ e } x_{j+r} & \text{stanno tra } \alpha \text{ e } \beta \text{ (} 0 \leq \alpha < \beta \leq 1 \text{)} \\ x_{j+i} \quad i = 1, \dots, r-1 & \text{non stanno in } [\alpha, \beta] \end{cases}$$

Il primo e l'ultimo stanno in $[\alpha, \beta]$, gli altri no.

Per avere una buona sequenza di numeri casuali, si è calcolato che la probabilità di ottenere un gap di lunghezza r è

$$P(r) = (0.9)^r (0.1)$$

Si può applicare il χ^2 per comparare il valore teorico del n° di gap di lunghezza r e quello dato dal generatore di numeri casuali.

Esercizio 6.1 *Generare con EXCEL numeri casuali con l'algoritmo congruenziale:*

$$x_{n+1} = (ax_n + c) \bmod m$$

Se tutti i parametri sono interi, il resto della divisione di un intero per m sarà un numero compreso tra 0 e m-1. Segue che la successione genererà al più m numeri prima di ripetersi. Scegliere come parametri iniziali c=0 e

a	m
32	17
1111	11111111
33331	12345678
...	...

Generare i numeri alternatamente su due colonne, per poterli prendere a coppie e rappresentarli su un piano, cioè

x_0	x_1
x_2	x_3
x_4	x_5
x_6	x_7
...	...

In questo modo ci accorgeremo di particolari regolarità che sono indice di non causalità'.

6.3 Generazione di numeri casuali con assegnata densità di probabilità

6.3.1 Tecnica diretta

Sia $f(x)$ una funzione con le seguenti proprietà:

i) $f(x)$ è continua in $[a, b]$, $f(x) \geq 0$

ii) $\int_a^b f(x)dx = 1$

iii) $F(x) = \int_a^x f(t)dt$ funzione integrale

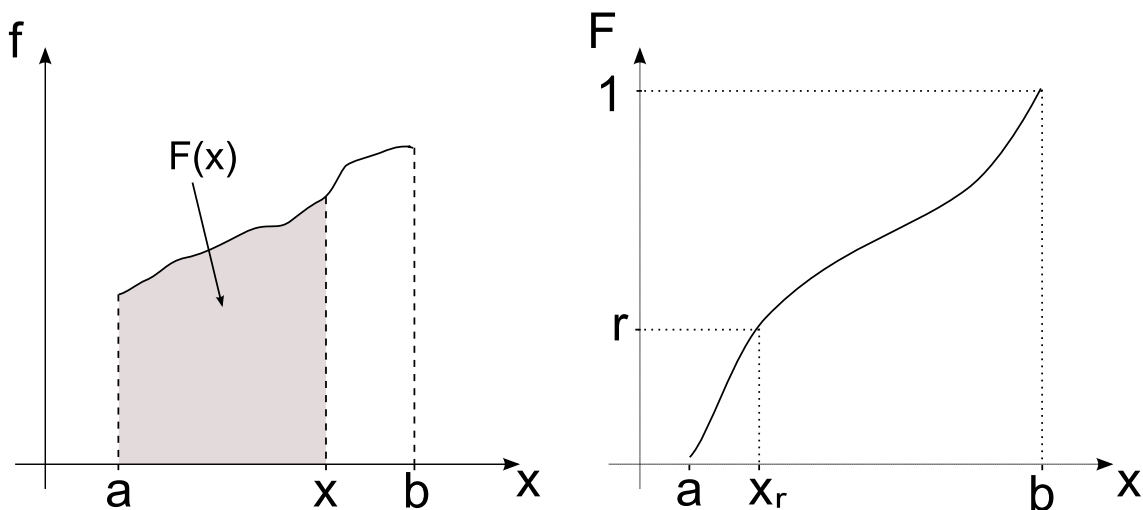


Figura 6.1: Funzione integrale $F(x)$ e funzione continua $f(x)$

$F(a) = 0$, $F(b) = 1$ per la (iii), $F(x)$ è crescente perchè $f(x) \geq 0$

Si sceglie un numero casuale uniforme $r \in [0, 1]$ ed in corrispondenza cerco un numero $x_r \in [a, b]$ tale che

$$r = F(x_r) = \int_a^{x_r} f(t)dt$$

e se è possibile farlo

$$x_r = F^{-1}(r)$$

Si vuole provare che x_r è un numero che ha densità di probabilità $f(x)$. Per semplicità si omette il pedice r .

Sia $P(x)$ la densità di probabilità associata a x , cioè $P(x)dx$ è la probabilità che $x \in [x, x+dx]$.

Poichè x è stato ottenuto da r , che è un numero casuale con distribuzione uniforme:

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

\Downarrow

$$P(x) = f(x)$$

Se $f(x)$ non è normalizzata, basta prendere

$$r = \frac{\int_a^{x_r} f(t)dt}{\int_a^b f(t)dt}$$

osservazione 1:

Questa procedura ha due problemi pratici:

i) si deve poter calcolare $\int f dx$

ii) si deve poter calcolare l'inversa di $F(x)$

osservazione 2:

e $f = \text{costante}$ $r = \frac{x_r - a}{b - a}$ $x_r = a + (b - a)r$

6.3.2 Tecnica di reiniezione

Sia $f(x)$ limitata in $[a, b]$ e $C \in \mathbb{R}_+$ tale che

$$f(x) \leq C \quad \forall x \in [a, b]$$

Si prendono 2 numeri casuali con distribuzione uniforme in $[0, 1]$

$$r_1, r'_1 \tag{6.1}$$

e quindi si calcola

$$x_1 = a + (b - a)r_1, \quad f_1 = r'_1 C$$

che sono ancora numeri casuali con distribuzione uniforme, rispettivamente in $[a, b]$ e $[0, C]$.

Se accade che

$$f_1 \leq f(x_1) \quad (6.2)$$

allora si prende x_1 come nuovo numero casuale con distribuzione $f(x)$, altrimenti si ritorna a (6.1) fino a quando si verifica la (6.2)

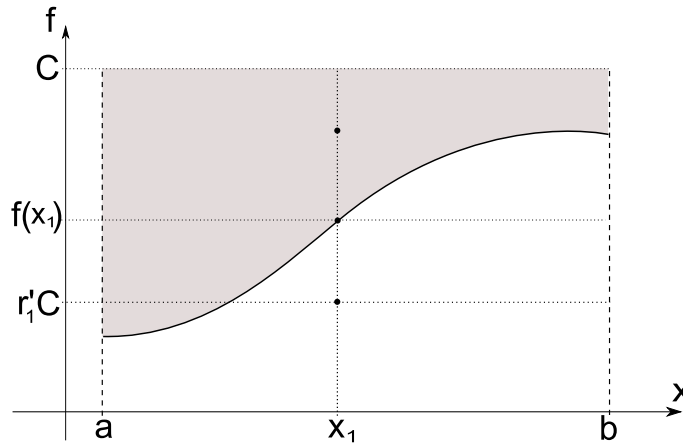


Figura 6.2: Tecnica di reiezione - Regione di accettazione della x

In pratica si scarta x_1 se cade nella parte che sta fuori della $f(x)$, altrimenti si accetta.

Proviamo che la successione così generata ha densità di probabilità $f(x)$.

La probabilità che $x_i \in [x, x + dx]$ sia accettato, è

$$\begin{array}{ccc} \text{Probabilità che } x_i \text{ stia in } [x, x + dx] & \times & \text{Probabilità che } x_i \text{ sia accettato} \\ \downarrow & & \downarrow \\ \sim dx \text{ (distribuzione uniforme)} & & \sim f(x_i) \end{array}$$

Quindi la probabilità è $f(x_i)dx$, ovvero x_i , ha densità di probabilità $f(x)$.

osservazione:

se $f(x)$ è molto piccata, occorre fare molte prove per accettare una x . Si ha l'ipotesi che $[a, b]$ sia limitato.

6.3.3 Tecnica combinata

Sia $g(x)$ una seconda densità di probabilità abbastanza semplice per permettere la generazione di numeri casuali distribuita secondo essa e tale che

$$f(x) \leq kg(x) \quad x \in [a, b]$$

- i) generiamo x_1 numero casuale con densità di probabilità $g(x)$ e r_1 con distribuzione uniforme $\in [0, 1]$
- ii) se $r_1 kg(x_1) < f(x_1)$ si accetta x_1 , altrimenti si ritorna a i)

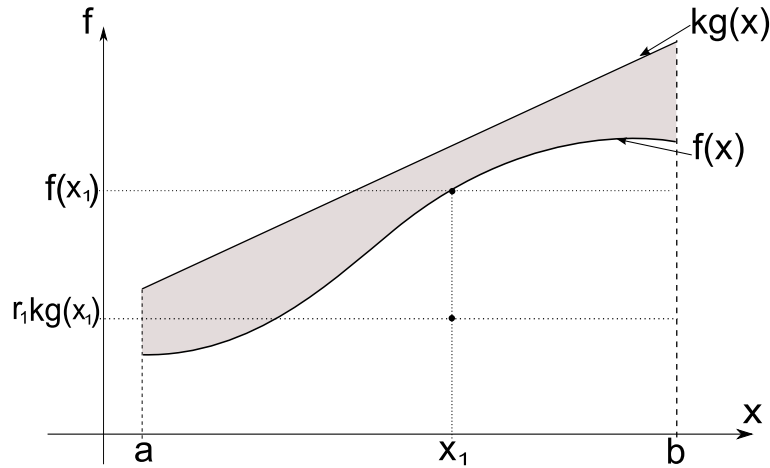


Figura 6.3: Tecnica combinata - Regione di accettazione della x

Al solito si scarta x_1 se cade nella parte evidenziata, ma adesso la zona di scarto è più piccola. NB: se $g(x) = 1$ si ha il caso b), quindi tanto più $kg(x)$ è vicino a $f(x)$, meno tentativi si faranno.

Proviamo che la successione delle x_i così generate ha funzione di distribuzione proprio $f(x)$.

Infatti la probabilità che $x_i \in [x, x + dx]$ sia accettato è

$$\text{Probabilità che } x_i \text{ stia in } [x, x + dx] \quad \times \quad \text{Probabilità che } x_i \text{ sia accettato}$$

$$\begin{array}{ccc} \downarrow & & \downarrow \\ \left(\begin{array}{c} x_i \text{ e' stato generato} \\ \text{con distribuzione} \\ \text{uniforme} \end{array} \right) \simeq g(x)dx & \times & \simeq \frac{f(x)}{kg(x)} \end{array}$$

↓

$$\frac{f(x)}{kg(x)} \times g(x)dx = \frac{f(x)dx}{k} \quad (6.3)$$

cioè è proporzionale a $f(x)dx$.

Se si fa l'integrale della (6.3):

$$\int \frac{f(x)dx}{k} = \frac{1}{k} \int f(x)dx = \frac{Area}{k}$$

si ha la probabilità totale di accettazione. Quindi lo step i) verrà eseguito in media k volte. Un buon generatore deve avere $k \simeq 1$

Esercizio 6.2 Sia $f(x) = 3x^2 \quad x \in [0, 1]$

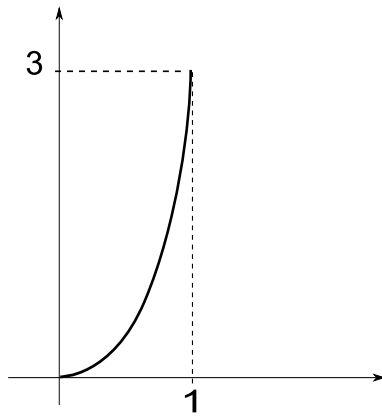


Figura 6.4: Funzione esempio

1. usare la tecnica diretta per generare un numero casuale con distribuzione $f(x) = 3x^2$
2. usare la tecnica di reiezione. Si osservi che

$$f(x) = 3x^2 \leq 3 \quad \forall x \in [0, 1]$$

calcolare il n° medio di tentativi per generare un numero casuale.

3. usare la tecnica combinata. Si osservi che

$$f(x) = 3x^2 \leq 3x \quad \forall x \in [0, 1]$$

quindi

$$k = 3 \quad g(x) = x$$

calcolare il n° medio di tentativi per generare un numero casuale e confrontarlo con il caso 2.

Capitolo 7

Il Metodo Monte Carlo

Con il termine di **Metodo Monte Carlo** oppure *metodo MC*, vengono in generale denominate tutte quelle tecniche che fanno uso di variabile aleatorie artificiali (cioe' generate al calcolatore) per la risoluzione di problemi matematici. Sicuramente questo non e' il modo piu' efficiente per trovare la soluzione di un problema, in quanto la procedure del campionamento simulato porta ad un risultato che e' sempre affetto dall'errore statistico. Nella pratica pero' ci si trova spesso davanti a situazioni in cui e' troppo difficile utilizzare i tradizionali strumenti numerici o analitici ed in tutti questi casi il metodo MC diventa l'unica alternativa possibile.

L'applicazione di questo metodo non e' ristretta solamente ai problemi di natura statistica, come forse si potrebbe pensare dato l'utilizzo di distribuzioni di probabilita', ma include tutti quie casi in cui si riesce a trovare un collegamento tra il problema in esame ed il comportamento di un certo sistema aleatorio. Per esempio il valore di un integrale definito, che certamente non e' una grandezza casuale, puo' essere calcolato anche usando dei numeri casuali.

Le basi teoriche del metodo MC risalgono al 1700: il matematico francese G.L. Leclerc conte di Buffon lo utilizzo' per il calcolo di π (Problema dell'ago di Buffon). Bisogna poi arrivare al 1940, quando questo metodo fu utilizzato da Enrico Fermi per studiare i processi di di assorbimento e diffusione dei materiali fissili, che porto' alla costruzione della prima bomba atomica.

Un impulso decisivo all'impiego del metodo si e' avuto con avvento di calcolatori sempre piu' potenti ed a buon mercato, grazie ai quali e' possibile studiare un sistema prima di realizzarlo in laboratorio. Attualmente si hanno applicazioni nei piu' svariati campi di ricerca, dalla fisica alla chimica all'economia.

7.1 Richiami dei metodi per l'integrazione numerica

$$I = \int_a^b f(x)dx = F(b) - F(a)$$

i) Approssimazione rettangolare

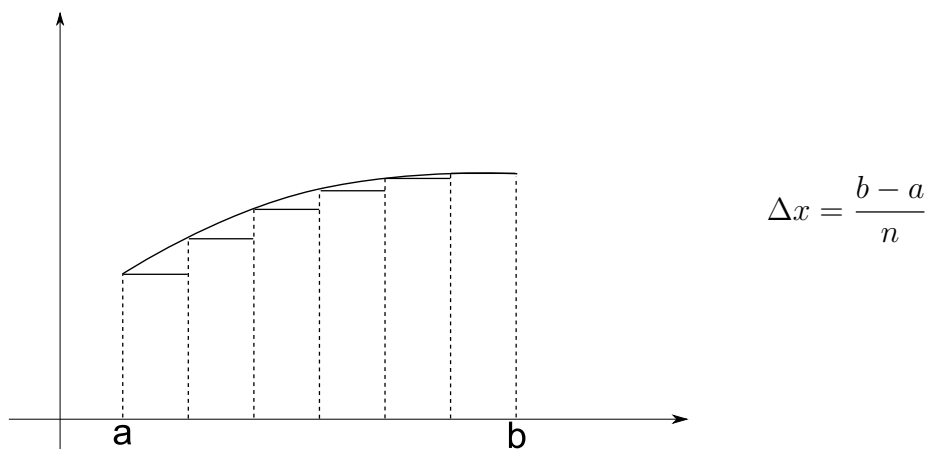


Figura 7.1: Approssimazione rettangolare dell'integrale

$$R_n = \sum_{i=0}^n f(x_i) \Delta x \quad , \quad x_i = a + i \Delta x \quad , \quad i = 0, 1, 2, \dots, n$$

Errore $\rightarrow E = \int_a^b [f(x) - p(x)]dx$ dove $p(x)$ è il polinomio che interpola la funzione (in questo caso la retta $y = y_0$)

$$E = \frac{b-a}{2} \Delta x f'(\xi) \quad \quad E \propto \Delta x \Rightarrow \propto \frac{1}{n}$$

dove ξ e' un opportuno punto dell'intervallo $[a, b]$.

ii) **Regola del trapezio** Si considera il trapezio inscritto sotto la curva $y = f(x)$

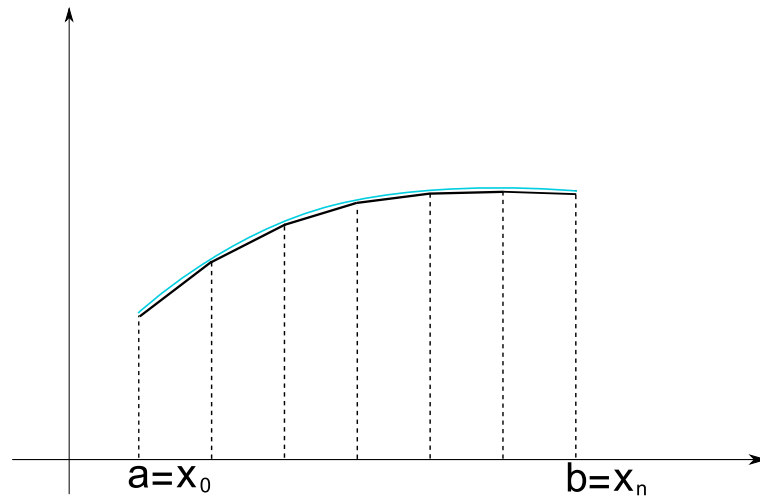


Figura 7.2: Approssimazione integrale trapezio

$$T_n = \left[\frac{1}{2}f(x_0) + \sum_{i=1}^{n-1} f(x_i) + \frac{1}{2}f(x_n) \right] \Delta x$$

$$E = -\frac{(b-a)}{12} (\Delta x)^2 f''(\xi) \quad E \propto (\Delta x)^2 \Rightarrow E \propto \frac{1}{n^2}$$

iii) **Regola di Simpson**

Considero un polinomio interpolante (formula di Newton forward) di ordine 4:

$$S_{2n} = \frac{\Delta x}{3} (f(x_0) + 4f(x_1) + \dots + 4f(x_{2n-1}) + f(x_{2n}))$$

$$E = -\frac{(b-a)}{180} (\Delta x)^4 f^{IV}(\xi) \quad E \propto (\Delta x)^4 \Rightarrow E \propto \frac{1}{n^4}$$

7.1.1 Integrale doppio

$$f(x, y) : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$$

$z = f(x, y)$ è una superficie in \mathbb{R}^3

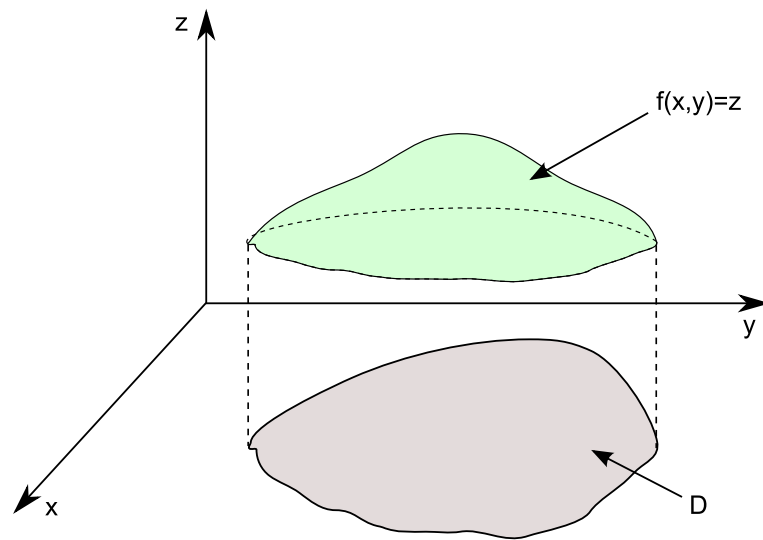


Figura 7.3: Integrale doppio

$\iint f(x) dx dy$ è il volume del cilindroide con base D ed altezza $f(x, y) = z$

Per un integrale $1D \rightarrow E \simeq \frac{1}{n^\alpha}$

Per un integrale $d - D \rightarrow E \simeq \frac{1}{n^{\frac{\alpha}{d}}}$

1. Regola di Simpson

$$d = 1 \quad E \propto \frac{1}{n^4}$$

$$d = 2 \quad E \propto \frac{1}{n^2}$$

$$d = 3 \quad E \propto \frac{1}{n^{\frac{4}{3}}}$$

2. Integrazione Monte Carlo

$$E \propto \frac{1}{\sqrt{n}} \quad \forall d$$

Quindi nel caso $d = 1$ tutti gli algoritmi numerici lavorano meglio del MC. Mentre per d sufficientemente grande il MC ne batte qualcuno. Infatti, supponiamo di applicare Simpson per $d = 6$, quindi

$$E \propto \frac{1}{n^{\frac{4}{6}}} = \frac{1}{n^{\frac{2}{3}}}$$

$$E_S < E_{MC} \quad .$$

Per utilizzare Simpson occorre fare una suddivisione per ogni asse (d=6 quindi ne ho 6) e nel caso ad esempio in cui faccio 50 suddivisioni, devo valutare

$$f(x_1, x_2, x_3, x_4, x_5, x_6)$$

almeno 50^6 volte, il che consuma molta CPU. Questo numero non può essere grande a piacere ma circa $10^{10} \div 10^{15}$, che limita di fatto l'utilizzo di solutori deterministici.

Un solutore MC invece, valutando un numero ragionevole di $f(\dots)$, può ottenere una migliore approssimazione.

Inoltre se il dominio è irregolare, i metodi deterministici funzionano male, mentre il MC non dipende dalla forma del dominio.

7.2 Il metodo MC “Hit or Miss”

Sia $f(x)$ definita in $[a, b]$ e limitata, cioè $\exists c \in \mathbb{R}$ t.c.

$$0 \leq f(x) \leq c$$

Consideriamo il rettangolo

$$\Omega = \{(x, y) \mid a \leq x \leq b, 0 \leq y \leq c\}$$

e quindi il vettore (x, y) di numeri casuali uniformemente distribuiti in Ω . La probabilità p che il vettore (x, y) cada sotto la curva $f(x)$ e'

$$p = \frac{\text{area } S}{\text{area } \Omega} = \frac{\int_a^b f(x) dx}{c(b-a)} = \frac{I}{c(b-a)} \quad (7.1)$$

Se adesso genero N vettori casuali

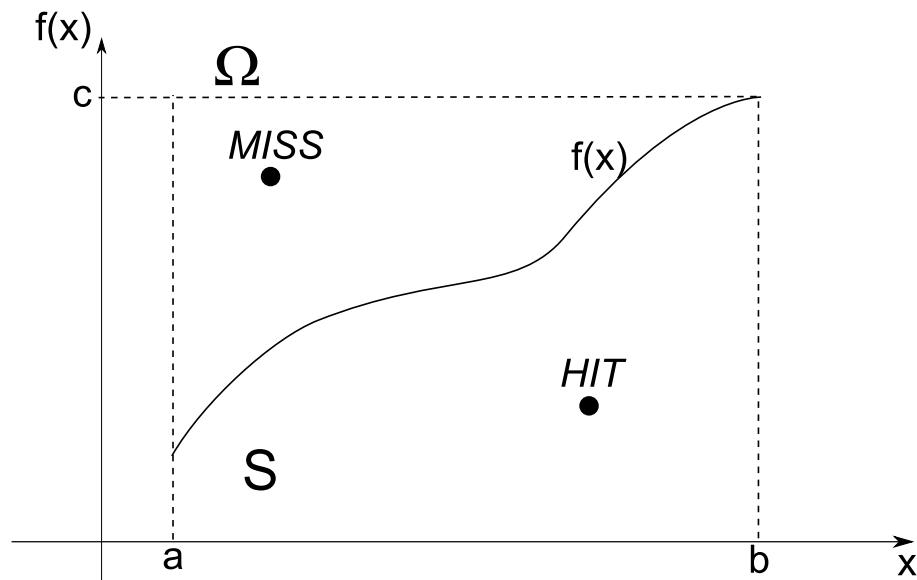


Figura 7.4: Metodo Hit or Miss

$$(x_i, y_i) \quad i = 1, 2, \dots, N$$

il parametro p si può stimare con

$$\hat{p} = \frac{N_H}{N}$$

N_H = numero *hits*, cioè il n° di volte in cui il punto (x_i, y_i) cade sotto la curva.

L'integrale si può stimare

$$I = \int_a^b f(x)dx \sim \theta_1 = c(b-a) \frac{N_H}{N}$$

Poichè la prova (hit or miss) assume i valori 0,1 con probabilità p , ho una variabile aleatoria con *distribuzione binomiale* con parametro p . Per questa distribuzione si ha $E[X] = Np$, $V[X] = N(1-p)p$ (vedi (3.1),(3.2)) e sfruttando l'equazione (7.1), si avra'

$$E[\theta_1] = c(b-a)E\left[\frac{N_H}{N}\right] = \frac{c(b-a)}{N}E[N_H] = c(b-a)p = I \quad (7.2)$$

$$\begin{aligned} V[\theta_1] &= V\left[c(b-a)\frac{N_H}{N}\right] = [c(b-a)]^2 V\left[\frac{N_H}{N}\right] = \\ &= c^2(b-a)^2 \frac{1}{N^2} V[N_H] = c^2(b-a)^2 \frac{1}{N} p(1-p) \end{aligned} \quad (7.3)$$

poichè $p = \frac{I}{c(b-a)}$

$$V[\theta_1] = \frac{I[c(b-a) - I]}{N}$$

da cui la deviazione standard:

$$\begin{aligned} \sigma_{\theta_1} = \sqrt{V[\theta_1]} &= \frac{\sqrt{I[c(b-a) - I]}}{\sqrt{N}} \rightarrow \begin{array}{l} \text{Dipende da } I \\ \text{che e' l'incognita} \end{array} \\ \sigma_{\theta_1} &\sim \frac{1}{\sqrt{N}} \end{aligned}$$

Osservazioni

- i. L'estimatore θ_1 si dice **unbiased** (non distorto) perche' vale l'eq.(7.2), cioe' $E[\theta_1] = I$;
- ii. la deviazione standard va a zero lentamente ($\propto 1/\sqrt{N}$).

Teorema 7.1 *Quante prove devo effettuare con il metodo "hit or miss" affinchè:*

$$P[|\theta_1 - I| < \epsilon] \geq \alpha ?$$

Si prova che

$$N \geq \frac{(1-p)p [c(b-a)]^2}{(1-\alpha)\epsilon^2}$$

7.2.1 Stima per l'intervallo di confidenza per I

Per N abbastanza grande, dal teorema del limite centrale sappiamo che:

$$\hat{\theta}_1 = \frac{\theta_1 - I}{\sigma_{\theta_1}}$$

è distribuita secondo la normale standardizzata $N(0, 1)$ e quindi l'intervallo di confidenza per I con livello di fiducia α , secondo la (4.12), è :

$$\left[\theta_1 - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} (b-a)c, \quad \theta_1 + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} (b-a)c \right] \quad (7.4)$$

Algoritmo “Hit or Miss”:

1. Generare una successione di $2N$ numeri casuali $\{U_i\} \quad i = 1, 2, \dots, N$
2. Disporre i numeri casuali in N coppie

$$(U_1, U'_1), (U_2, U'_2), \dots, (U_N, U'_N)$$

in modo che ogni numero casuale $\{U_i\}$ è usato una sola volta.

3. Calcolare $x_i = a + U_i(b-a)$ e $f(x_i) \quad i = 1, 2, \dots, N$
4. Contare il n° di volte N_H in cui $f(x_i) \geq cU'_i$
5. Stimare l'integrale I con

$$\theta_1 = c(b-a) \frac{N_H}{N}$$

6. Fissato il livello di fiducia α , calcolare l'intervallo di confidenza con l'eq.(7.4)

Esercizio 7.1 Calcolo di π

Per calcolare π occorre calcolare l'area del primo quadrante AS del cerchio di raggio unitario

$$x^2 + y^2 = 1 \quad , y = \sqrt{1 - x^2}$$

e moltiplicarla per 4. Applichiamo quindi l'algoritmo Hit or Miss.

1. Poiche' $x, y \in [0, 1]$, bastera' generare N coppie di numeri casuali uniformi (U, U') con la funzione di EXCEL *casuale()*
2. Per ogni coppia di numeri casuali (U, U') la funzione EXCEL $SE(\sqrt{1 - U^2} \geq U'; 1; 0)$ restituisce il valore 1 se Hit, 0 se Miss
3. Contare il numero N_H degli Hits con la funzione EXCEL *CONTA.SE(A:B;1)*
4. Un estimatore di AS e'

$$AS = \frac{N_H}{N}$$

ed una stima della sua deviazione standard sara' data dalla eq.(7.3) con $p \simeq AS$

$$\sigma = \sqrt{\frac{AS(1 - AS)}{N}}$$

L'intervallo di confidenza per AS sara' dato dall'eq.(7.4). Se fissiamo un livello di fiducia $\alpha = 0.003$ (pari al 99.7 %), questo intervallo sara' $[AS - 3\sigma, AS + 3\sigma]$.

5. Infine $\pi = 4AS$

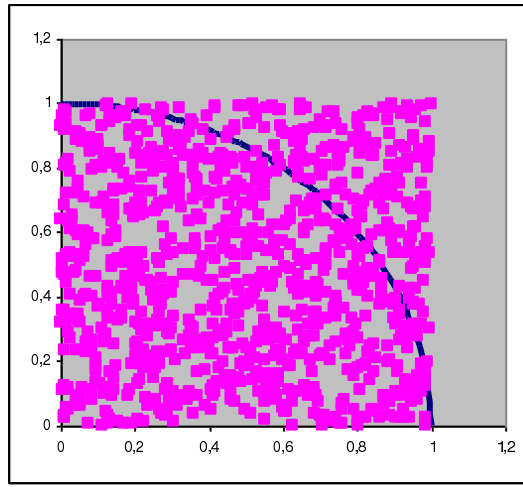


Figura 7.5: Calcolo di π con l'algoritmo Hit or Miss

7.3 Metodo Sample-Mean MC (MC della media)

Per calcolare $I = \int_a^b f(x)dx$, lo si può riscrivere come:

$$I = \int_a^b \frac{f(x)}{g(x)} g(x) dx$$

con $g(x) > 0$ quando $f(x) \neq 0$. Segue che:

$$I = E \left[\frac{f(x)}{g(x)} \right]$$

dove x è una *variabile casuale* distribuita con funzione di distribuzione $g(x)$.

Per semplicità assumiamo:

$$g(x) = \begin{cases} \frac{1}{b-a}, & a < x < b; \\ 0, & \text{altrimenti.} \end{cases}$$

Ne segue che:

i) la variabile casuale x con distribuzione $g(x)$, essendo questa uniforme, sarà

$$x_i = a + U_i(b - a)$$

con U_i numero casuale uniforme $\in [0, 1]$

ii)

$$I = E \left[\frac{f(x)}{\frac{1}{b-a}} \right] = (b-a)E[f(x)] \quad (7.5)$$

Uno stimatore di I è

$$\theta_2 = (b-a) \frac{1}{N} \sum_{i=1}^N f(x_i)$$

Esso si dice *non distorto* (o *unbiased*) perchè

$$E[\theta_2] = E \left[\frac{b-a}{N} \sum_i f(x_i) \right] = \frac{b-a}{N} E \left[\sum_i f(x_i) \right] =$$

per la 7.5

$$= \frac{b-a}{N} \sum_i \frac{I}{b-a} = \frac{b-a}{N} \frac{N}{b-a} I = I$$

iii)

$$V[\theta_2] = V \left[\frac{1}{N} (b-a) \sum_i f(x_i) \right] = \frac{(b-a)^2}{N^2} V \left[\sum_i f(x_i) \right] =$$

poichè $f(x_i)$ è incorrelata:

$$= \frac{(b-a)^2}{N^2} \sum_i V[f(x_i)]$$

ricordando che

$$V[Y] = \int (Y - \bar{Y})^2 h(x) dx = \int Y^2 h(x) dx - \left(\int Y h(x) dx \right)^2 \quad \text{con } h(x) = \frac{1}{b-a}$$

si ha:

$$\begin{aligned} &= \frac{(b-a)^2}{N^2} \sum_i \left\{ \int f^2 \frac{1}{b-a} dx - \left(\int f \frac{1}{b-a} dx \right)^2 \right\} = \\ &= \frac{1}{N} \left\{ (b-a) \int f^2(x) dx - I^2 \right\} \end{aligned}$$

Algoritmo Sample - Mean MC

1. Generare una sequenza $\{U_i\}$ di N numeri casuali uniformi $\in [0, 1]$
2. $x_i = a + U_i(b - a) \quad i = 1, \dots, N$
3. calcolare $f(x_i)$ e quindi

$$\theta_2 = (b - a) \frac{1}{N} \sum_{i=1}^N f(x_i)$$

7.4 Efficienza del metodo Monte Carlo

Si è visto che il metodo “Hit or Miss” produce una stima di I θ_1 , mentre quello “Sample-Mean” un'altra stima θ_2 tale che:

$$E[\theta_1] = E[\theta_2] = I$$

$$V[\theta_1] \neq V[\theta_2]$$

Siano t_1 e t_2 i tempi di calcolo per θ_1 e θ_2 . Diremo che il primo metodo è *più efficiente* del secondo se :

$$\epsilon = \frac{t_1 V[\theta_1]}{t_2 V[\theta_2]} < 1$$

Teorema 7.2

$$V[\theta_2] \leq V[\theta_1]$$

Dim.

Infatti, poichè $V[\theta_1] = \frac{I}{N}[c(b - a) - I]$ e $V[\theta_2] = \frac{1}{N} \left[(b - a) \int_a^b f^2(x) dx - I^2 \right]$

$$V[\theta_1] - V[\theta_2] = \frac{1}{N}(b - a) \left[cI - \int_a^b f^2(x) dx \right]$$

poichè $f(x) \leq c \longrightarrow f(x) \cdot f(x) \leq f(x) \cdot c$ e integrando $\int_a^b f^2(x) dx \leq cI$

$$V[\theta_2] \leq V[\theta_1] \quad \blacksquare$$

Se $t_1 \simeq t_2$, allora il metodo di sampling è *più efficiente* del metodo hit or miss.

7.5 Tecniche di riduzione della varianza

Si può ridurre la varianza solo se si hanno informazioni sul problema da trattare. Per esempio si possono avere informazioni per mezzo di una simulazione grossolana del problema.

7.5.1 Importance sampling (importanza del campionamento)

Si voglia calcolare un integrale multiplo

$$I = \int f(x)dx$$

L'idea principale di questa tecnica consiste nel concentrare la distribuzione dei punti del campione nella parte di D che è più importante invece di distribuirli uniformemente in D . L'integrale si può quindi scrivere:

$$I = \int \frac{f(x)}{g(x)}g(x)dx = E \left[\frac{f(x)}{g(x)} \right]$$

dove x è una variabile aleatoria con distribuzione $g(x)$ tale che $g > 0 \quad \forall x \in D$.

Quindi

$$\xi = \frac{f(x)}{g(x)}, \quad E[\xi] = I$$

è una stima di I con varianza

$$V[\xi] = E[\xi^2] - [E(\xi)]^2 = \int \left[\frac{f(x)}{g(x)} \right]^2 g(x)dx - I^2 = \int \frac{[f(x)]^2}{g(x)}dx - I^2$$

Per stimare I prendo dei numeri casuali $\{x_i\}$ con distribuzione $g(x)$ e calcolo

$$\theta_3 = \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{g(x_i)}$$

Problema: come scegliere $g(x)$ per rendere minima la varianza di θ_3 ?

Teorema 7.3 *Il minimo di $V[\theta_3]$ è uguale a*

$$V[\xi_0] = \left(\int |f(x)|dx \right)^2 - I^2$$

e si ottiene quando la variabile aleatoria x è distribuita con densità di probabilità

$$g(x) = \frac{|f(x)|}{\int |f(x)|dx}$$

Osservazione:

Se $f(x) > 0$ allora la distribuzione ottimale è

$$f_x(x) = \frac{f(x)}{I}, \quad \text{con } V[\xi_0] = 0$$

Questa funzione però non si può utilizzare perchè contiene I che è l'incognita del problema.

In pratica si vede che la varianza può essere ridotta se $g(x)$ è scelta in modo da avere una forma simile a $|f(x)|$

7.5.2 Control Variates

In questo caso si cerca di migliorare la varianza confrontando con qualche modello analitico noto.

Siano Y e C due variabili random tali che

$$E[C] = \mu_C$$

Allora $\forall \beta \in \mathbb{R}$ definisco

$$Y(\beta) = Y - \beta(C - \mu_C) = Y + \beta(\mu_C - C)$$

che ovviamente è tale che

$$E[Y(\beta)] = E[Y]$$

Poiché dalle proprietà delle variabili aleatorie (vd Capitolo 2)

$$\begin{cases} V[X + Y] = V[X] + V[Y] + 2Cov(X, Y) \\ V[aX] = a^2V[X] \end{cases}$$

$$V[Y(\beta)] = V[Y] + V[\beta(\mu_C - C)] + 2Cov(Y, \beta(\mu_C - C))$$

$$\cdot V[\beta(\mu_C - C)] = \beta^2 V[\mu_C - C] = \beta^2 \sum_i [\mu_C - C_i - (\overline{\mu_C - C})]^2 = \beta^2 V[C]$$

$$\cdot Cov(Y, \beta(\mu_C - C)) = E \{ (Y - \bar{Y})(\beta(\mu_C - C)) \} = -\beta E \{ (Y - \bar{Y})(C - \mu_C) \}$$

In totale

$$V[Y(\beta)] = V[Y] + \beta^2 V[C] - 2\beta Cov(Y, C)$$

Quindi possiamo considerare

$$V[y(\beta)] = f(\beta)$$

Si trova quindi il minimo

$$f'(\beta) = 2\beta V[C] - 2Cov(Y, C) = 0$$

$$f''(\beta) = 2\beta > 0$$

Il valore che minimizza la $f(\beta)$ e quindi $V[Y(\beta)]$ è:

$$\beta^* = \frac{Cov(Y, C)}{V[C]}$$

e il valore minimo è

$$V[Y(\beta^*)] = (1 - \rho_{YC}^2)V[Y]$$

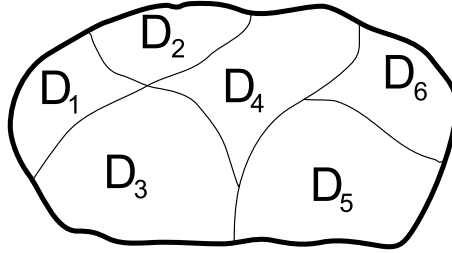
più C e Y sono correlate, più si riduce la varianza. ($\rho^2 \rightarrow 1$)

7.5.3 Stratified sampling (campionamento stratificato)

Si suddivide la regione di integrazione D in m sotto regioni disgiunte D_i , cioè

$$D_i \cap D_j = 0 \quad i \neq j$$

$$D = \bigcup_{i=1}^m D_i$$



e quindi

$$I = \int_D \phi(x)g(x)dx = \sum_{i=1}^m \int_{D_i} \phi(x)g(x)dx$$

ogni integrale della sommatoria si può stimare con un semplice MC (ad esempio *sampling*).

L'idea è di fare più campionamento in quelle sotto regioni di D che sono più importanti. In questo modo si ottiene una riduzione della varianza, invece di scegliere la $g(x)$ ottimale.

Esistono teoremi che ci danno una stima sulla varianza ottenuta con questo metodo.

7.5.4 Antithetic Variates

Con questa tecnica si cercano due stimatori unbiased dell'integrale I , tale che abbiano una forte correlazione negativa.

Siano Y' e Y'' due stimatori unbiased di I cioè:

$$E[Y'] = E[Y''] = I = \int f(x) dx$$

allora

$$Y = \frac{1}{2}(Y' + Y'')$$

è tale che

$$E[Y] = E\left[\frac{1}{2}(Y' + Y'')\right] = \frac{1}{2}(E[Y'] + E[Y'']) = I$$

quindi anche Y è unbiased. Calcoliamo la varianza ricordando le proprietà viste in (2.10),(2.11):

$$\begin{cases} V[X + Y] = V[X] + V[Y] + 2Cov(X, Y) \\ V[aX] = a^2V[X] \end{cases}$$

Quindi

$$V\left[\frac{1}{2}(Y' + Y'')\right] = \frac{1}{4}V[Y' + Y''] = \frac{1}{4}V[Y'] + \frac{1}{4}V[Y''] + \frac{1}{2}Cov(Y', Y'')$$

Se $Cov(Y', Y'') < 0$ allora la varianza diminuisce.

Teorema 7.4

$$I = \int_0^1 f(x)dx = \frac{1}{2} \int_0^1 [f(x) + f(1-x)]dx$$

Dim.

Infatti se sostituisco $1 - x = t$ e $-dx = -dt$,

$$\int_0^1 f(1-x)dx = - \int_1^0 f(t)dt = \int_0^1 f(t)dt = \int_0^1 f(x)dx$$

allora vale la proprietà. ■

In base al teorema precedente, scegliamo come estimatori dell'integrale I

$$Y' = f(x) , \quad Y'' = f(1-x)$$

$$Y = \frac{1}{2}[f(x) + f(1-x)].$$

Un estimatore di Y è (Sample-Mean):

$$\theta_A = \frac{1}{2N} \sum_{i=1}^N [f(U_i) + f(1-U_i)]$$

dove $\{U_i\}$ sono N numeri casuali con distribuzione uniforme in $[0, 1]$.

Il tempo di calcolo di θ_A è doppio rispetto a quello del metodo di sampling a causa della doppia valutazione della $f(x)$.

Affinchè questo stimatore sia *più efficiente* di quello del metodo di sampling θ_2 , occorre che:

$$\epsilon = \frac{t_A V[\theta_A]}{t_2 V[\theta_2]} < 1$$

e poichè $t_A = 2t_2 \Rightarrow V[\theta_A] \leq \frac{1}{2} V[\theta_2]$.

Teorema 7.5 *Se $f(x)$ è continua e monotona non crescente con derivata prima continua, allora*

$$V[\theta_A] \leq \frac{1}{2} V[\theta_2]$$

Infine osserviamo che $Cov(Y', Y'') < 0$ perchè nello stimatore θ_A per calcolare la $f(x)$ utilizzo i numeri casuali

$$U_i \quad , \quad 1 - U_i \quad \quad i = 1, \dots, N$$

che ovviamente sono correlati.

Esercizio 7.2 Riduzione della varianza

Assegnato l'integrale

$$I = \int_0^1 \cos\left(\frac{\pi}{2}x\right) dx = \frac{2}{\pi}$$

i) Calcolare I con l'algoritmo *MC Hit or Miss*. In questo caso

$$\theta_{HM} = \xi = \frac{N_H}{N}$$

e l'intervallo di confidenza e' dato dall'eq.(7.4). Iterare il calcolo per n. di punti $N=1000$
: 1000 : 50000;

ii) Calcolare I con l'algoritmo *MC Sample-Mean*. Si ha

$$\theta_{SM} = \langle \xi \rangle = \frac{1}{N} \sum_{i=1}^N \xi_i, \quad \xi_i = \cos\left(\frac{\pi}{2}U_i\right)$$

n. di punti $N = 1000 : 1000 : 50000$; U_i n. casuale con distribuzione uniforme in $[0, 1]$.

iii) Calcolare I con l'algoritmo *Importance Sampling*. Per trovare la funzione $g(x)$ 'vicina' a $f(x)$, scrivo la corrispondente formula di Taylor

$$f(x) = \cos\left(\frac{\pi}{2}x\right) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + O(3) = 1 - \frac{\pi^2}{8}x^2 + O(3)$$

e scelgo

$$g(x) = \frac{3}{2}(1 - x^2) \quad \text{tale che} \quad \int_0^1 \frac{3}{2}(1 - x^2) dx = 1 \quad .$$

Quindi :

a) for $N = 1000$: 1000 : 50000;

b) genero N numeri casuali $\{y_i\}$ con distribuzione $g(x)$ usando la tecnica di reiezione;

c) calcolo

$$\theta_{IS} = \langle \xi \rangle = \frac{1}{N} \sum_{i=1}^N \xi_i, \quad \xi_i = \frac{\cos\left(\frac{\pi}{2}y_i\right)}{\frac{3}{2}(1 - y_i^2)}$$

d) goto a)

iv) Calcolare I con la tecnica *Anthitetic variates*.

$$\theta_A = \langle \xi \rangle = \frac{1}{N} \sum_{i=1}^N \xi_i, \quad \xi_i = \frac{1}{2} \left[\cos \left(\frac{\pi}{2} U_i \right) + \cos \left(\frac{\pi}{2} (1 - U_i) \right) \right]$$

dove $N = 1000:1000:50000$; U_i n. casuale con distribuzione uniforme in $[0, 1]$.

L'intervallo di confidenza per gli estimatori $\theta_{SM}, \theta_{IS}, \theta_A$ si puo' calcolare grazie all'eq. (4.12).

In particolare se scegliamo il livello di fiducia $\alpha = 0.003$ (pari al 99.7%), otterremo l'intervallo

$$\frac{1}{N} \sum_{i=1}^N \xi_i \pm \Delta, \quad \Delta = 3 \sqrt{\frac{1}{N} \left(\frac{1}{N} \sum_{i=1}^N \xi_i^2 - \left[\frac{1}{N} \sum_{i=1}^N \xi_i \right]^2 \right)}.$$

Graficare il valore dell'integrale e di Δ nei casi i), ii), iii), iv) al variare di N .

Capitolo 8

Catene di Markov

Un processo stocastico e' formalmente definito da un insieme di variabili casuali $\{X_t\}$, dove il parametro t e' il tempo. In generale il comportamento futuro del sistema dipende dal suo passato. Cioe' gli stati X_{t_1} e X_{t_2} , a due istanti differenti, sono variabili casuali dipendenti. Una classe molto interessante di processi stocastici e' quella di Markov, in cui lo stato futuro dipende solo da quello presente e non dalla sua storia.

8.1 Alcune definizioni

Definizione 8.1 Proprieta' di Markov. Dato lo stato corrente X_t , la distribuzione di ogni stato futuro X_y (con $y > t$), non dipende dalla sua storia passata, $\{X_u : u < t\}$, ma solo da quello presente.

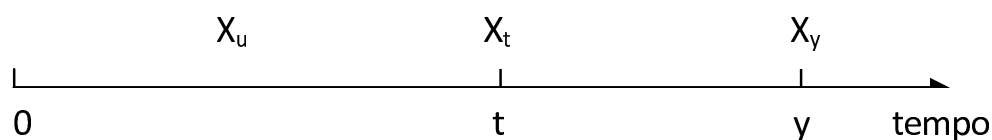


Figura 8.1: Passato, presente e futuro

Definizione 8.2 L'insieme dei valori presi dalle variabili casuali X_t si chiama **Spazio degli stati**.

Noi saremo interessati a studiare processi stocastici con spazio degli stati discreti $(0,1,2,\dots)$ ed in cui anche il parametro $t \in \mathbb{N}$ sia discreto. In tal caso il processo sara' detto **Catena di Markov a tempo discreto** (DTCM).

Quindi se il sistema si trova nello stato n , il suo stato al tempo $n+1$ non dipende dagli stati ai tempi $0,1,2,\dots,n-1$

$$P(X_{n+1} = j | X_0, X_1, \dots, X_n) = P(X_{n+1} = j | X_n). \quad (8.1)$$

L'evoluzione del sistema e' completamente descritta dalla **probabilita' di transizione ad un passo** che e' una matrice Q , i cui elementi $q_{i,j}(n)$ ci dicono che il sistema si muovera' nello stato j al tempo successivo $n+1$, supposto che al tempo n si trovi nello stato i

$$q_{i,j}(n) = P(X_{n+1} = j | X_n = i) \quad , i, j, n = 0, 1, 2, \dots$$

Definizione 8.3 La catena si dice **temporalmente omogenea** se questa probabilita' di transizione non dipende dall'istante di tempo, cioe'

$$q_{i,j}(n) = q_{i,j} \quad , i, j, n = 0, 1, \dots$$

Nel seguito considereremo solamente Catene di Markov a tempo discreto e temporalmente omogenee.

Poiche' il sistema deve essere in qualche stato ad ogni tempo di osservazione, ne segue che la somma degli elementi di ogni riga della matrice sia uguale a uno, cioe'

$$\sum_{j=0}^{\infty} q_{i,j} = 1 \quad , i = 0, 1, 2, \dots \quad (8.2)$$

Un modo di rappresentare una catena di Markov e' con un grafo orientato (o diretto) in cui i vertici corrispondono agli stati della catena e gli archi mostrano le probabilita' di transizione non nulle.

Esempio 8.1 *Un macchinario puo' essere in due stati rotto oppure operativo (0 e 1 rispettivamente). Se e' rotto al tempo n , sara' rotto od operativo al tempo $n+1$ con probabilita' 0.3 e 0.7 rispettivamente. Se e' operativo al tempo n , sara' rotto od operativo al tempo $n+1$ con probabilita' 0.1, 0.9 rispettivamente (per $n = 0, 1, 2, \dots$).*

Questo comportamento soddisfa la proprieta' di Markov. Gli elementi della matrice di transizione sono

$$q_{0,0} = \text{rotto} \rightarrow \text{rotto} = 0.3 \quad , \quad q_{0,1} = \text{rotto} \rightarrow \text{operativo} = 0.7$$

$$q_{1,0} = \text{operativo} \rightarrow \text{rotto} = 0.1 \quad , \quad q_{1,1} = \text{operativo} \rightarrow \text{operativo} = 0.9$$

cioe'

$$Q = \begin{pmatrix} 0.3 & 0.7 \\ 0.1 & 0.9 \end{pmatrix}$$

ed il grafo associato e'

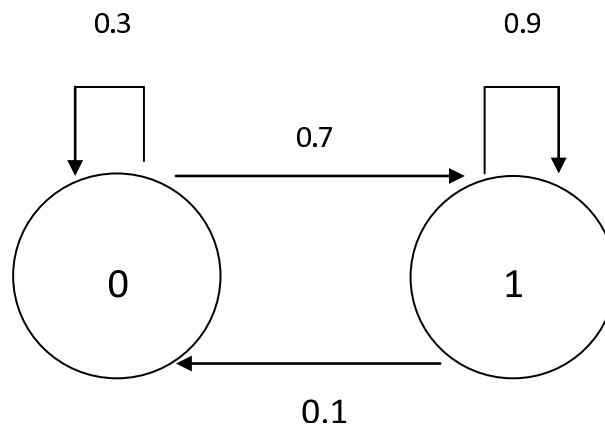


Figura 8.2: Grafo del macchinario

Esempio 8.2 (*Gambler's ruin*) *Un giocatore d'azzardo scommette 1 Euro alla volta nel lancio della moneta. Vince 1 Euro se esce testa e lo perde se esce croce. Il giocatore finisce di scommettere se il suo capitale diventa 0 o 100 Euro. Se $X_0 < 100$ e' il capitale iniziale del giocatore e X_n e' il capitale dopo n scommesse, allora questo valore dipende dal capitale precedente X_{n-1} ma non da tutti i valori precedenti. La successione $\{X_n : n = 0, 1, 2, \dots\}$ e' una catena di Markov.*

Vediamo come e' fatta questa catena. Quanto vale

$$q_{0,0} = P(X_{n+1} = 0 | X_n = 0)$$

cioe' qual'e' la probabilita' che se il capitale e' 0 Euro al tempo n rimanga 0 Euro al tempo $n+1$? Ovviamente 1, perche' in tal caso il gioco si arresta. Analogamente $q_{100,100} = 1$. Quindi

$$q_{1,0} = P(X_{n+1} = 0 | X_n = 1)$$

e' la probabilita' che se il capitale e' 1 Euro al tempo n diventa 0 Euro al tempo $n+1$: questa probabilita' e' 0.5 (probabilita' del lancio della moneta). Analogamente

$$q_{1,1} = P(X_{n+1} = 1 | X_n = 1) = 0 \quad , \quad q_{1,2} = P(X_{n+1} = 2 | X_n = 1) = 0.5 \quad \dots\dots$$

In definitiva la matrice di transizione e'

$$Q = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & \dots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \quad (8.3)$$

Il grafo associato e'

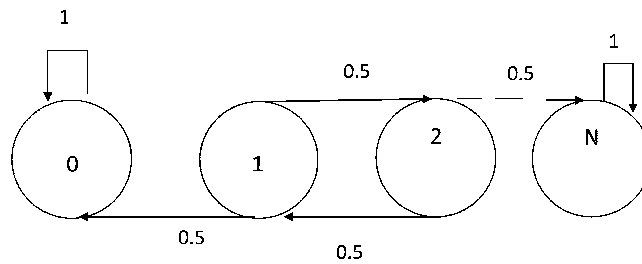


Figura 8.3: Grafo *Gambler's ruin*

Esempio 8.3 *Un canale di comunicazione a tempo discreto, ha un buffer finito di dimensione N . I messaggi che arrivano vengono immagazzinati nei nodi del buffer fino a che c'è spazio. Durante l' n -esimo tempo, o un nuovo messaggio arriva (ammesso che ci sia spazio per esso), con probabilit  α , o uno dei messaggi nel buffer (supposto che ce ne siano)   trasmesso, con probabilit  β , oppure non succede niente.*

Questo sistema pu  essere modellato con una catena di Markov che   nello stato i quando ci sono i messaggi nel buffer ($i = 0, 1, 2, \dots, N$). Dallo stato i , al tempo successivo, sono possibili le seguenti transizioni

- da $i \rightarrow i - 1$ (eccetto $i=0$), con probabilit  β ;
- da $i \rightarrow i + 1$ (eccetto quando $i = N$), con probabilit  α ;
- da $i \rightarrow i$, con probabilit  $1 - \alpha - \beta$

La matrice di transizione Q   una matrice tridiagonale cos  fatta:

$$q_{0,0} = 1 - \alpha \quad , \quad q_{0,1} = \alpha \tag{8.4}$$

$$q_{i,i} = 1 - \alpha - \beta \quad , \quad q_{i,i+1} = \alpha \quad , \quad q_{i,i-1} = \beta \quad , \quad i = 1, \dots, N - 1 \tag{8.5}$$

$$q_{N,N} = 1 - \beta \quad , \quad q_{N,N-1} = \beta \tag{8.6}$$

Il grafo associato  

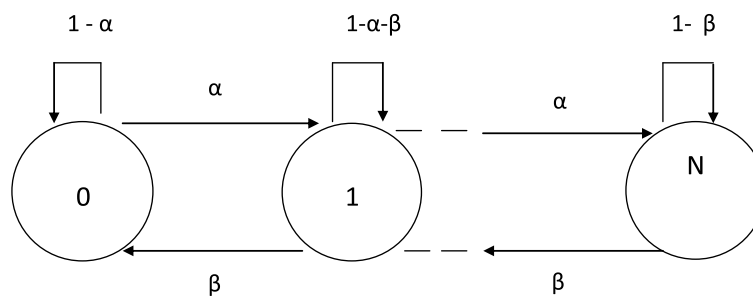


Figura 8.4: Grafo del buffer

Il buffer considerato fa parte della categoria piu' generale dei PROCESSI DI NASCITA E MORTE (birth-death). Essi sono catene di Markov in cui due stati successivi differiscono solo di una unità. Cioè questi processi sono caratterizzati dalla proprietà che se ci si trova in uno stato, le uniche possibilità sono quelle o di rimanere nello stesso stato o di spostarsi in uno stato che differisce di una unità da quello di partenza. Tali processi sono ideali per caratterizzare l'evoluzione di una coda. In essa infatti gli utenti arrivano uno alla volta e si accodano per ricevere il servizio.

8.2 Calcolo di leggi congiunte

Consideriamo adesso la transizione fatta dalla catena di Markov $X = \{X_n : n = 0, 1, 2, \dots\}$ in **due passi**. Poiche' la catena e' temporalmente omogenea, le corrispondenti probabilita' non dipendono dall'indice temporale. Definiamo

$$q_{i,j}^{(2)} = P(X_{n+2} = j | X_n = i), \quad i, j, n = 0, 1, \dots$$

Se ci vogliamo muovere da uno stato i ad uno j in due passi, la catena deve transitare da uno stato intermedio k , dopo il primo passo. Quindi possiamo scrivere

$$\begin{aligned} q_{i,j}^{(2)} &= \sum_{k=0}^{\infty} P(X_{n+2} = j | X_n = i, X_{n+1} = k) P(X_{n+1} = k | X_n = i) = \\ &= \sum_{k=0}^{\infty} P(X_{n+2} = j | X_n = i, X_{n+1} = k) q_{i,k} \end{aligned} \quad (8.7)$$

ma per la proprieta' di Markov eq.(8.1)

$$P(X_{n+2} = j | X_n = i, X_{n+1} = k) = P(X_{n+2} = j | X_{n+1} = k) = q_{k,j}$$

e quindi la (8.7) diventa

$$q_{i,j}^{(2)} = \sum_{k=0}^{\infty} q_{i,k} q_{k,j}$$

Si noti che il secondo membro dell'equazione precedente rappresenta l'elemento (i,j) della matrice Q^2 . Piu' in generale vale il seguente teorema:

Teorema 8.1 Sia $q_{i,j}^{(s)}$ la probabilita' di transizione da uno stato i ad uno j ad s **passi**,

$$q_{i,j}^{(s)} = P(X_{n+s} = j | X_n = i), \quad s = 1, 2, \dots$$

Allora la matrice di transizione $Q^{(s)}$ i cui elementi sono $q_{i,j}^{(s)}$ e' data da ¹

$$Q^{(s)} = Q^s$$

¹dove Q^s e' la potenza di una matrice quadrata, cioe' $s-1$ prodotti (riga-colonna) successivi della matrice per se stessa

Esercizio 8.1 *Se riprendiamo l'esempio 8.2, grazie a questo teorema possiamo calcolare la probabilita' che il giocatore perda tutto il suo capitale in non piu' di 50 scommesse, supposto che il suo capitale iniziale sia 10 Euro.*

Infatti questa e' data da

$$q_{10,0}^{(50)} = Q^{50}|_{10,0}$$

Esercizio 8.2 *Ancora con l'esempio 8.2, ed al teorema precedente possiamo calcolare il valore medio del suo capitale dopo 50 scommesse, supposto che il suo capitale iniziale sia 10 Euro.*

Infatti avremo

$$E(X_{50}|X_0 = 10) = \sum_{j=0}^{100} j q_{10,j}^{(50)}$$

Con la matrice di transizione (8.3), si possono fare i calcoli facilmente. Tuttavia, un conto simile per altri modelli potrebbe essere computazionalmente molto costoso.

Sia E l'insieme degli stati di una catena di Markov e supponiamo che $E = (1, 2, 3, \dots)$ oppure $E = (1, 2, 3, \dots, m)$ a seconda che E sia infinito oppure finito di cardinalità m . Poiché le variabili aleatorie X_n al tempo n , assumono valori in E , le probabilità che uno stato di E sia occupato, sono caratterizzate da un vettore riga di dimensione pari alla cardinalità di E $v = (v_1, v_2, \dots)$, con

$$v_k = P(X_n = k) \quad .$$

Quindi al tempo $n = 0$ la variabile aleatoria X_0 può assumere i valori $(1, 2, \dots, m)$ con probabilità

$$v_1 = P(X_0 = 1) \quad , \quad v_2 = P(X_0 = 2) \quad , \dots, \quad v_m = P(X_0 = m)$$

Affinché questa sia una legge di probabilità dovranno essere soddisfatte le condizioni

- $v_k \geq 0 \quad \forall k = 1, 2, \dots$
- $\sum_{k \in E} v_k = 1$

Supponiamo che al tempo $n = 0$ X_0 abbia legge data dal vettore v : è possibile calcolare al tempo n la legge w per X_n . Infatti

$$w_k = P(X_n = k) = \sum_{h \in E} P(X_n = k | X_0 = h) P(X_0 = h) = \sum_{h \in E} q_{hk}^{(n)} v_h$$

cioè i due vettori v e w sono legati dalla relazione

$$w = vQ^n \quad . \tag{8.8}$$

Conoscendo le probabilità di occupazione degli stati v al tempo $n = 0$ e la matrice di transizione ad un passo Q , l'equazione (8.8) ci permette di trovare le probabilità di occupazione degli stati al tempo n .

Più in generale se $0 < n_1 < n_2 < n_3 < \dots < n_k$, possiamo calcolare la seguente legge congiunta

$$P(X_{n_1} = i_1, X_{n_2} = i_2, \dots, X_{n_k} = i_k)$$

sfruttando la ricorrenza, ottenendo

$$P(X_{n_1} = i_1, X_{n_2} = i_2, \dots, X_{n_k} = i_k) = \sum_k v_k q_{ki_1}^{(n_1)} q_{i_1 i_2}^{(n_2 - n_1)} \dots q_{i_{k-1} i_k}^{(n_k - n_{k-1})} \tag{8.9}$$

la legge congiunta (8.9) e' determinata univocamente dalla funzione di transizione Q e dalla legge iniziale v .

Esercizio 8.3 *Consideriamo la catena di Markov con spazio degli stati $E=\{1, 2, 3, 4\}$ e matrice di transizione*

$$Q = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad (8.10)$$

Supposto di partire dallo stato 2, dopo 2 passi, quali sono le probabilità di essere in uno stato di E ?

Le probabilità di transizione dopo $n = 2$ passi si ottengono facendo il prodotto delle matrici $Q^2 = Q Q$, ovvero

$$Q^2_{ij} = \begin{pmatrix} \frac{7}{16} & 0 & \frac{9}{16} & 0 \\ \frac{5}{16} & \frac{1}{4} & \frac{7}{16} & 0 \\ \frac{3}{8} & 0 & \frac{5}{8} & 0 \\ \frac{1}{4} & \frac{5}{18} & \frac{13}{36} & \frac{1}{9} \end{pmatrix} \quad (8.11)$$

Al tempo iniziale avremo $v = (0, 1, 0, 0)$ e quindi dopo due passi

$$w = vQ^2 = \left(\frac{5}{16}, \frac{1}{4}, \frac{7}{16}, 0 \right)$$

8.3 Classificazione degli stati

Sia E l'insieme degli stati di una catena di Markov e $C \subset E$.

Definizione 8.4 Se $i, j \in E$ diciamo che **i comunica con j** (ovvero $i \rightarrow j$) se esiste $n > 0$ tale che $q_{i,j}^{(n)} > 0$.

Questo equivale al fatto che, nel grafo che rappresenta il diagramma di transizione della catena, esista un percorso che porta dallo stato i allo stato j in n passi. Ovviamente questa proprietà non è simmetrica.

Definizione 8.5 Un sottoinsieme di stati C si dice **chiuso** se gli stati di C non comunicano con gli stati che stanno fuori di C (ovvero del complementare di C).

Ovviamente l'insieme di tutti gli stati è chiuso.

Definizione 8.6 Un sottoinsieme di stati C chiuso si dice **irriducibile** se tutti i suoi stati comunicano tra loro.

Definizione 8.7 Una catena di Markov si dice **irriducibile** se tutti i suoi stati comunicano tra loro, ovvero se E è l'unica classe irriducibile

Definizione 8.8 Uno stato di una catena di Markov si dice **assorbente** se costituisce da solo una classe irriducibile.

Definizione 8.9 Un stato i di una catena di Markov si dice **ricorrente** se il sistema essendo stato una volta in i , la catena ritornerà in i con probabilità uno. Se indichiamo con ρ_{ii} questa probabilità, avremo che $\rho_{ii} = 1$.

Definizione 8.10 Un stato i di una catena di Markov che non è ricorrente si dice **transitorio**, e in questo caso $\rho_{ii} < 1$

Definizione 8.11 Un stato j di una catena di Markov si dice **periodico** con periodo $m > 1$, se i ritorni consecutivi allo stato j avvengono solamente con multipli di m passi, cioè

$$P(X_{n+s} = j | X_n = j) = 0 \quad , \text{ se } s \neq mk \quad \text{per qualche } k \geq 1$$

Se non esiste qualche $m > 1$ che soddisfi la relazione precedente, la catena si dirà **aperiodica**.

Teorema 8.2 *Se $i \rightarrow j$ e $j \rightarrow h$, allora $i \rightarrow h$*

Dim. Infatti per ipotesi esistono $n, m > 0$ tali che

$$q_{ij}^{(n)} > 0 \quad , \quad q_{jh}^{(m)} > 0$$

poiche' dal teorema 8.1 la matrice di transizione associata a $q_{ih}^{(n+m)}$ e' Q^{n+m} , ed essendo

$$Q^{n+m} = Q^n Q^m$$

dove il prodotto tra le due matrici e' quello riga per colonna, segue che

$$q_{ih}^{(n+m)} = \sum_k q_{ik}^{(n)} q_{kh}^{(m)} \geq q_{ij}^{(n)} q_{jh}^{(m)} > 0$$

che implica $i \rightarrow j$. ■

Esempio 8.4 *Classifichiamo gli stati dell'esempio 8.2 (gambler's ruin)*

Supponiamo per semplicita' che la matrice abbia dimensione sei ²

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (8.12)$$

Da questa matrice e' chiaro che :

- a) lo stato 1 comunica con 0 e 2
- b) lo stato 2 comunica con 1 e 3
- c) lo stato 3 comunica con 2 e 4

²Gli elementi della matrice Q sono $q_{ij}, i, j = 0, 1, 2, 3, 4, 5$

d) lo stato 4 comunica con 3

e dal teorema 8.2 segue che

$$1 \rightarrow 2, \quad 2 \rightarrow 3, \quad 3 \rightarrow 4 \quad \implies \quad 1 \rightarrow 4 \quad .$$

Così ragionando si prova che se $0 < i < 5$, allora lo stato i comunica con tutti gli altri stati.

Gli stati $i = 0$ e $i = 5$ costituiscono invece delle classi irriducibili, perché

$$q_{00}^{(n)} = 1 \quad , \quad q_{55}^{(n)} = 1 \quad \forall n$$

e quindi comunicano solo con se stessi. Ne segue che questi stati sono assorbenti. Infine tutti gli stati i sono transitori esclusi gli stati 0 e 5: infatti si prova dalle considerazioni precedenti, che partendo da i si ritorna nello stesso stato con una probabilità minore di 1. Per quanto riguarda gli stati 0 e 5, comunicando solo con se stessi, saranno ricorrenti.

8.4 Problemi di assorbimento

Consideriamo una catena di Markov con un numero finito di stati E e sia C un suo sottinsieme chiuso. Essendo C chiuso, se la catena raggiungerà C allora resterà per sempre in C , ovvero se $X_k \in C$ ne segue che $X_n \in C \quad \forall n \geq k$. Vogliamo quindi calcolare la *probabilità di assorbimento* λ_i che, partendo da uno stato $X_i \in E$, si arrivi in C cioè

$$\lambda_i = P \{X_n \in C, \text{ per qualche } n > 0 | X_i \in E\} \quad .$$

Ci sono due casi banali ovvero se $i \in C$ (in tal caso $\lambda_i=1$) o se i appartiene ad un'altra classe disgiunta da C ($\lambda_i=0$). Il caso più interessante è quando i sia uno stato transitorio che non si trova in C . Indichiamo con D l'insieme degli stati transitori che non fanno parte di C . Allora si prova il seguente teorema

Teorema 8.3 *Le probabilità λ_i , $i \in D$ sono soluzioni del seguente sistema lineare*

$$\lambda_i = \sum_{h \in C} q_{ih} + \sum_{j \in D} q_{ij} \lambda_j \quad (8.13)$$

Si osservi che la prima sommatoria rappresenta la probabilita' di fare una transizione al primo passo da i in C , mentre la seconda e' la probabilita' di passare al primo passo in un altro stato $j \in D$ e di essere poi assorbiti in C partendo da j .

Esempio 8.5 *Un commesso viaggiatore parte da una certa localita' A e si muove negli altri posti secondo lo schema riportato in figura. Le frecce indicano dove puo' andare. Assumendo che le scelte indicate dalle frecce siano equiprobabili, si descriva il viaggio del commesso tramite una catena di Markov, classificandone gli stati. E' piu' probabile che il commesso finisca il suo viaggio in E_1 o in E_2 ?*

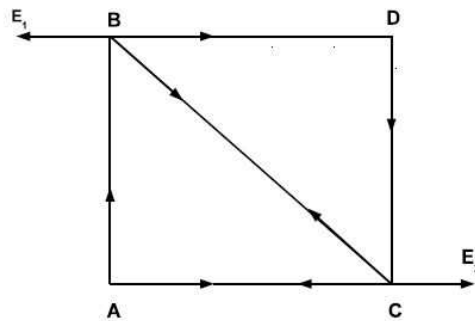


Figura 8.5: Problema del commesso viaggiatore

In questo caso la catena di Markov e' formata da 6 stati $\{A, B, C, D, E_1, E_2\}$ e la matrice di transizione associata e'

$$Q = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (8.14)$$

L'insieme degli stati transitori e' $T = \{A, B, C, D\}$, mentre quello degli assorbenti e' $S = \{E_1, E_2\}$. Calcoliamo quindi la probabilita' che se il sistema parte da uno stato transitorio

arrivi in uno assorbente. Poiche' gli stati assorbenti sono due, bisogna applicare la (8.13) due volte. La probabilita' che partendo da uno stato T si arrivi in E_1 si ottiene risolvendo il sistema

$$\lambda_i = q_{i,5} + \sum_{j=1,4} q_{ij} \lambda_j \quad , \quad i = 1, 2, 3, 4$$

ottenendo $\lambda_1 = 0.44, \lambda_2 = 0.55, \lambda_3 = 0.33$. La probabilita' che partendo da uno stato T si arrivi in E_2 si ottiene risolvendo il sistema

$$\bar{\lambda}_i = q_{i,6} + \sum_{j=1,4} q_{ij} \bar{\lambda}_j \quad , \quad i = 1, 2, 3, 4$$

ottenendo $\bar{\lambda}_1 = 0.55, \bar{\lambda}_2 = 0.44, \bar{\lambda}_3 = 0.66$. Siccome la probabilita' massima e' $\bar{\lambda}_3 = 0.66$, e' piu' probabile che il commesso finisca il suo giro in E_2 .

8.5 Probabilita' invarianti

Definizione 8.12 *Sia v una probabilita' su E : allora essa si dice **invariante o stazionaria** se accade*

$$v = vQ \tag{8.15}$$

Supponiamo che la legge di probabilita' v di X_0 sia invariante. Allora la legge di probabilita' w di X_n dalla (8.8) coincidera' con quella iniziale

$$w = vQ^n = vQQ^{n-1} = vQ^{n-1} = vQQ^{n-2} = \dots = v \quad .$$

Vale il seguente teorema:

Teorema 8.4 *(Markov-Kakutani) Una matrice di transizione Q su un insieme di stati E finito ammette sempre almeno una probabilita' invariante.*

Esempio 8.6 *I vettori $v_1 = (1, 0, 0, 0, 0, 0), v_2 = (0, 0, 0, 0, 0, 1)$ sono probabilita' invarianti per l'esempio 8.4.*

Definizione 8.13 *Una matrice di transizione Q si dice **regolare** se esiste un intero positivo m tale che*

$$q_{ij}^{(m)} > 0 \quad , \quad \forall i, j \in E$$

Definizione 8.14 *Una catena di Markov si dice regolare se e' tale la sua matrice di transizione.*

Se la catena e' regolare, tutti gli stati comunicano tra di loro e quindi la catena e' irriducibile e gli stati ricorrenti (non vale il viceversa).

Una condizione sufficiente che ci aiuta a capire se una catena di Markov e' regolare e' data dal seguente teorema

Teorema 8.5 (*Criterio di regolarita'*) *Se tutti gli stati comunicano tra loro e se inoltre esiste $h \in E$ tale che $q_{hh} > 0$ allora la catena e' regolare.*

8.6 Stato stazionario

Vogliamo studiare una catena di Markov per tempi molto grandi. Quando l'istante di osservazione e' molto lontano dal punto iniziale, la probabilita' di trovare la catena in uno stato j , p_j , non dipende dallo stato iniziale

$$\lim_{n \rightarrow \infty} P(X_n = j | X_0 = i) = \lim_{n \rightarrow \infty} q_{i,j}^{(n)} = p_j \quad , j = 0, 1..$$

Quando le probabilita' limite p_j esistono (e la loro somma fa uno), esse si chiamano **distribuzione di equilibrio** o **distribuzioni dello stato stazionario** di una catena di Markov. Il problema principale della teoria delle catene di Markov e' di sapere se esiste una distribuzione di equilibrio ed in tal caso determinarla. Un teorema molto utile e' il seguente

Teorema 8.6 (*Markov*) *Data una catena di Markov il cui insieme degli stati sia finito e pari a N , se inoltre la matrice di transizione Q e' regolare, allora esiste ed e' unica la probabilita' invariante $p = (p_1, p_2, ..., p_N)$ di Q tale che*

$$\lim_{n \rightarrow \infty} q_{ij}^{(n)} = p_j \quad .$$

Vediamo qual'è la conseguenza di questo teorema. Supponiamo che la nostra catena abbia distribuzione iniziale v . Allora la legge di probabilità associata a X_n sarà data dalla eq.(8.8). Allora, per n che tende all'infinito

$$P(X_n = j) = (vQ^n)_j = \sum_{i \in E} v_i q_{ij}^{(n)} \rightarrow \left(\sum_{i \in E} v_i \right) p_j = p_j$$

e quindi qualunque sia la distribuzione iniziale v la X_n converge in legge alla distribuzione invariante p . Dal punto di vista computazionale questo è molto importante: infatti per calcolare la distribuzione limite ci sono due possibilità

- i. usare l'equazione (8.8) e quindi moltiplicare Q per se stessa n volte;
- ii. il teorema precedente ci garantisce che esiste la probabilità limite ed inoltre ci dice che questa è anche invariante per Q (vedi eq.(8.15)), allora basta risolvere il seguente sistema lineare vincolato

$$p_j = \sum_{i \in E} p_i q_{i,j} \quad , \text{ (equazione di bilancio) } \quad (8.16)$$

$$\sum_{i \in E} p_j = 1 \quad \text{ (equazione di normalizzazione) } \quad (8.17)$$

Commento

L' eq.(8.16) si può riscrivere

$$p_j = \sum_{i=0, i \neq j}^N p_i q_{i,j} + p_j q_{j,j}$$

da cui

$$p_j(1 - q_{j,j}) = \sum_{i=0, i \neq j}^N p_i q_{i,j} \quad (8.18)$$

Dall'eq.(8.2) si ha

$$\sum_{j=0, j \neq i}^N q_{i,j} + q_{i,i} = 1$$

da cui scambiando j con i

$$\sum_{i=0, i \neq j}^N q_{j,i} + q_{j,j} = 1$$

e sostituendo nella eq.(8.18) si ha :

$$p_j \sum_{i=0, i \neq j}^N q_{j,i} = \sum_{i=0, i \neq j}^N p_i q_{i,j} . \quad (8.19)$$

Analizziamo il primo membro di questa equazione: p_j e' la probabilita' che il sistema sia nello stato j , mentre $q_{j,i}$ rappresenta la probabilita' di transizione che il sistema esca dallo stato j . Allora il primo membro rappresenta la frazione media di passi con cui la catena fa una transizione fuori dallo stato j . Analogamente il secondo membro rappresenta la frazione media di passi con cui la catena fa una transizione nello stato j .

Quindi nello stato stazionario

$$\text{numero medio transizioni } j \rightarrow i \simeq \text{numero medio transizioni } i \rightarrow j$$

e per questo motivo la (8.16) si chiama equazione di bilancio.

Esercizio 8.4 *Riprendiamo l'esempio (8.1) del macchinario e calcoliamo le probabilita' dello stato stazionario.*

La catena di Markov associata e' regolare perche' tutti gli stati comunicano tra loro ed inoltre esiste un h tale che $q_{hh} > 0$ (criterio di regolarita').

Calcoliamo l'eq.(8.19) per $j = 0$, otterremo:

$$p_0 q_{0,1} = p_1 q_{1,0} \quad \Rightarrow \quad 0.7 p_0 = 0.1 p_1 , \quad (8.20)$$

inoltre dall'equazione di normalizzazione (8.17)

$$p_0 + p_1 = 1 \quad . \quad (8.21)$$

Da cui risolvendo il sistema (8.20),(8.21) si otterra'

$$p_0 = \frac{1}{8} \quad , p_1 = \frac{7}{8} \quad .$$

Partendo la nostra macchina da una certa condizione iniziale, a regime dopo molti cicli (ovvero nello stato stazionario) la probabilita' che sia nello stato rotto e' $p_0 = \frac{1}{8}$, mentre quella che sia nello stato operativo e' $p_1 = \frac{7}{8}$.

Esercizio 8.5 *Riprendiamo l'esempio (8.3) del buffer e calcoliamo le probabilita' dello stato stazionario.*

Anche in questo caso la catena di Markov e' regolare, perche' vale il criterio di regolarita' 8.5. Sotto l'ipotesi che

$$\gamma = \frac{\alpha}{\beta} < 1 \quad (8.22)$$

il sistema (8.19) si puo' risolvere per ricorrenza, ottenendo

$$p_j = \frac{\gamma^j - \gamma^{j+1}}{1 - \gamma^{N+1}} \quad j = 1, 2, 3, \dots, N \quad (8.23)$$

dove ricordiamo che p_j rappresenta la probabilita' che la catena si trovi nello stato (nodo) j . A partire da queste probabilita' dello stato stazionario, possiamo calcolare il numero medio di elementi nel buffer, che e' dato da

$$L = \sum_{j=1}^N j p_j \quad .$$

8.7 L'algoritmo di Metropolis

Definizione 8.15 (*Bilancio dettagliato*) Una probabilita' p su E si dice reversibile se accade che

$$p_i q_{ij} = p_j q_{ji} \quad , \forall i, j \in E$$

Commento

Il significato di questa equazione e' analogo a quello dell'eq.(8.16).

Teorema 8.7 *Se p e' reversibile allora e' anche invariante.*

Dim. Infatti avremo che

$$\sum_{i \in E} p_i q_{ij} = \sum_{i \in E} p_j q_{ji} = p_j \sum_{i \in E} q_{ji} = p_j \cdot \blacksquare$$

Sia Q una matrice di transizione su E che sia irriducibile (ovvero tutti gli stati comunicano) e simmetrica³ e π una probabilita' su E tale che $\pi_i > 0$ per ogni $i \in E$. Definiamo a partire da Q e π una nuova matrice P i cui elementi sono

$$p_{ij} = \begin{cases} q_{ij} & \text{se } \pi_j \geq \pi_i \\ q_{ij} \frac{\pi_j}{\pi_i} & \text{se } \pi_j < \pi_i \\ 1 - \sum_{j \neq i} p_{ij} & \text{se } j = i \end{cases} \quad (8.24)$$

Si puo' dimostrare che P e' una matrice di transizione, e che π e' reversibile per P . Vale il seguente teorema

Teorema 8.8 (*Metropolis, Rosenbluth, Teller*) *Se π non e' la distribuzione uniforme e X_n e' una catena di Markov con matrice di transizione P (8.24), allora X_n converge a π per $n \rightarrow \infty$.*

Commento

Assegnata una qualunque matrice di transizione Q irriducibile e simmetrica, l'eq.(8.24) permette di costruire una nuova matrice di transizione P , la cui catena di Markov associata ha una prefissata distribuzione limite (π).

³Cioe' $q_{ij} = q_{ji}$.

Le applicazioni dell'algoritmo sono molteplici. Citiamo le piu' importanti:

- **Calcolo di integrali.**

Abbiamo visto che l'algoritmo permette di definire una successione di variabili aleatorie che asintoticamente convergono ad un'assegnata densita' di probabilita' $f(X)$: in altre parole, stiamo generando dei numeri casuali aventi come distribuzione asintotica $f(X)$.

Questi punti X_i possono essere utilizzati per valutare integrali definiti ponendo

$$\int g(X)f(X) dx \simeq \frac{1}{N} \sum_{i=1}^N g(X_i) \quad .$$

Poiche' $f(X)$ e' campionata solo asintoticamente, la stima dell'integrale e' *distorta* (biased) da una quantita' che puo' essere fatta piccola, eliminando alcune X_i all'inizio del run. Quanti bisogna scaricarne dipende dal problema particolare. Il vantaggio principale di questo metodo e' che permette di campionare densita' di probabilita' a piu' dimensioni molto complicate, in modo semplice ma dispendioso dal punto di vista del calcolo. Per es. in meccanica quantistica (nei superfluidi) occorre valutare integrali a piu' dimensioni molto complicati e questo algoritmo diventa uno strumento molto importante.

- **Fisica statistica.**

La Fisica Statistica studia i sistemi con un gran numero di particelle ($N \simeq 10^{23}$). Un sistema di questo tipo ha energia totale pari alla somma di quella cinetica e potenziale:

$$E = \sum_{i=1}^N \frac{1}{2} m_i \vec{v}_i^2 + \sum_{i < j \leq N} \Phi(|\vec{r}_i - \vec{r}_j|)$$

dove $m_i, \vec{r}_i, \vec{v}_i$ sono rispettivamente massa, posizione e velocita' della i-esima particella e Φ e' l'energia potenziale (che specifica la forza agente tra due particelle). Per un gas di particelle, in condizioni vicini all'equilibrio, si prova che la densita' di probabilita' e' una funzione di Boltzmann

$$f(\vec{R}) = \frac{\exp \left[-\frac{\sum \Phi(r_{ij})}{k_B T} \right]}{\int \exp \left[-\frac{\sum \Phi(r_{ij})}{k_B T} \right] d\vec{R}}$$

dove T e' la temperatura, k_B la costante di Boltzmann, $\vec{R} = \{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N\}$ e $r_{ij} = |\vec{r}_i - \vec{r}_j|$. Da questa densita' si puo' calcolare l'energia media del sistema, ma non analiticamente.

Una possibile soluzione e' quella di utilizzare l'algoritmo di Metropolis campionando questa densita' e calcolando le opportune medie.

- **Analisi di immagini.**

Questo algoritmo puo' essere usato per la generazione di tessiture, nonche' per la conservazione delle proporzioni delle immagini.

- **Ricerca del minimo assoluto.**

In questo caso l'algoritmo prende il nome di *Simulating annealing* (ricottura simulata), in cui la legge che definisce la probabilita' π e' proporzionale ad una funzione di distribuzione di Boltzmann.

8.8 Un'applicazione alla teoria delle code

Il processo assunto alla base della maggior parte dei sistemi a coda si puo' schematizzare da dei clienti (clients) richiedenti un dato servizio da un server. Questi clienti entrando nel queueing system formano una coda. A certi istanti, un membro della coda viene scelto come prossimo cliente da servire, secondo una certa politica nota come disciplina della coda (per esempio la disciplina potrebbe essere FIFO, LIFO, ecc.). Quando il servizio richiesto dal cliente viene svolto dal server, il customer puo' uscire dal sistema a coda.

I sistemi a coda (o **queueing systems**) sono identificati da 4 simboli:

- la *prima lettera* indica la natura del **Processo degli arrivi**. I valori tipici sono:
 - M**: MEMORYLESS (proprietà fondamentale dei processi markoviani), indica che il processo degli arrivi e' un processo di Poisson (vedi Appendice).
 - G**: GENERAL, indica che il processo degli arrivi e' caratterizzato da una distribuzione di probabilità generale. In questo caso non si conosce l'andamento della funzione di distribuzione di probabilità degli arrivi, ma si conoscono solamente il valor medio e il valore quadratico medio.
 - D**: DETERMINISTIC, indica che il processo degli arrivi e' caratterizzato da una distribuzione di probabilità deterministica.
- la *seconda lettera* indica la natura della distribuzione di probabilità dei tempi di servizio. I valori possibili, anche in questo caso, sono M, G, D e il significato e uguale a quello spiegato precedentemente, con la differenza che tali simboli si riferiscono alla distribuzione di probabilità del **Processo delle Partenze**.
- Il *terzo simbolo* indica il numero dei servers del sistema a coda.
- Il *quarto simbolo* indica il **Numero massimo di clients nel sistema**. In genere questo simbolo non e' indicato in quanto per default e' infinito.

Nel seguito, per brevit , tratteremo un sistema a coda di tipo M/M/1 (vedi figura 8.6). Questo   dunque un sistema caratterizzato da un solo server (terzo simbolo), dove la probabilit  che arrivino n clients in un intervallo di tempo t   data da un Processo di Poisson con parametro λ (tasso di arrivo clients/sec, vedi appendice)

$$p_n(t) = \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \quad . \quad (8.25)$$

ed anche la probabilit  di servizio   un Processo di Poisson con parametro μ (tasso di servizio clients/sec.).

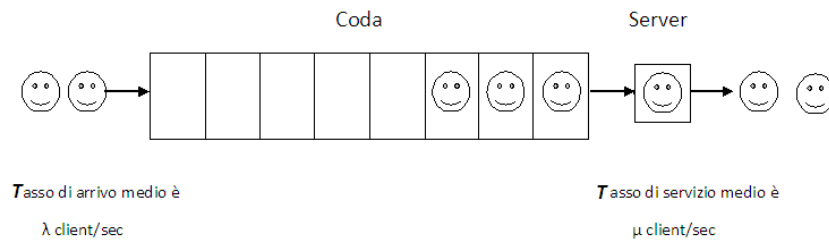


Figura 8.6: Sistema a coda di tipo M/M/1

Se pensiamo il tempo come una variabile discreta allora

$$t = k\delta \quad , k \in [0, 1, 2, \dots[$$

con δ un numero piccolo positivo.

Sia X_k la variabile aleatoria discreta **numero di clients che sono presenti nel server** (coda + server) al tempo (discreto) k . I valori assunti da X_k (ovvero gli stati del sistema) sono $E = \{0, 1, 2, \dots, \infty\}$, ovvero

- se $X_k = 0$ avro' 0 client al tempo k
- se $X_k = 1$ avro' 1 client al tempo k
- se $X_k = 2$ avro' 2 client al tempo k
-

Al successivo tempo $k + 1$

- a) se $X_k = 0$, al tempo $k + 1$ o non arrivano clients ($X_{k+1}=0$) oppure arriva 1 client ($X_{k+1}=1$)
- b) se $X_k = 1$, al tempo $k + 1$ o non arrivano clients ($X_{k+1}=1$) oppure arriva 1 client ($X_{k+1}=2$) oppure un client viene servito ($X_{k+1}=0$)
- c)

Definiamo

- $P(X_k = n)$ la probabilita' che ci siano n clients nel sistema (coda + server) al tempo k .
- $q_{ij}(k)$ e' la probabilita' di transizione (al tempo k) da uno stato $i \in E$ a uno $j \in E$, ovvero

$$q_{ij}(k) = P(X_{k+1} = j | X_k = i) \quad i, j \in E \quad .$$

Siccome le probabilita' di arrivo e di servizio sono di Poisson, vale la Proprieta' di Markov e quindi il sistema puo' essere descritto da una catena di Markov il cui grafo e' quello di figura (8.4), dove la matrice di transione e' data dalle equazioni (8.4)-(8.6).

Poiche' la probabilita' di arrivo di un client e' caratterizzata dalla distribuzione di Poisson con parametro λ , mentre quella di partenza (o servizio) dal parametro μ , allora (vedi appendice)

$$\alpha = \lambda \delta \quad , \quad \beta = \mu \delta \quad .$$

Abbiamo gia' visto nell'esercizio (8.5) che le probabilita' limite della catena sono date dalla eq.(8.23), cioe'

$$p_j = \frac{\gamma^j - \gamma^{j+1}}{1 - \gamma^{N+1}}$$

dove γ e' il **fattore di utilizzazione**

$$\gamma = \frac{\lambda}{\mu} \simeq \frac{\text{tasso di arrivo}}{\text{tasso di servizio}} \quad (8.26)$$

che ovviamente deve essere minore di 1.

Consideriamo il caso in cui $N \rightarrow \infty$. Essendo $\gamma < 1$ allora

$$p_0 = \lim_{N \rightarrow \infty} \frac{1 - \gamma}{1 - \gamma^{N+1}} = 1 - \gamma$$

poiche' p_0 rappresenta la probabilita' di avere 0 clients nel sistema, allora γ rappresenta la probabilita' che il server sia occupato.

Calcoliamo il n. medio di clients nel sistema

$$\begin{aligned}
 C &= \sum_{j=0}^{\infty} j p_j = \sum_{j=0}^{\infty} j \gamma^j (1 - \gamma) = \gamma(1 - \gamma) \sum_{j=0}^{\infty} j \gamma^{j-1} = \gamma(1 - \gamma) \sum_{j=0}^{\infty} \frac{d}{d\gamma} (\gamma^j) = \\
 &= \gamma(1 - \gamma) \frac{d}{d\gamma} \left[\sum_{j=0}^{\infty} (\gamma^j) \right] = \gamma(1 - \gamma) \frac{d}{d\gamma} \left[\frac{1}{1 - \gamma} \right] = \gamma(1 - \gamma) \frac{1}{(1 - \gamma)^2} = \frac{\gamma}{1 - \gamma} \quad (8.27)
 \end{aligned}$$

Questa equazione e' graficata nella figura 8.7: si nota che per $\gamma \rightarrow 1$ il sistema diventa instabile.

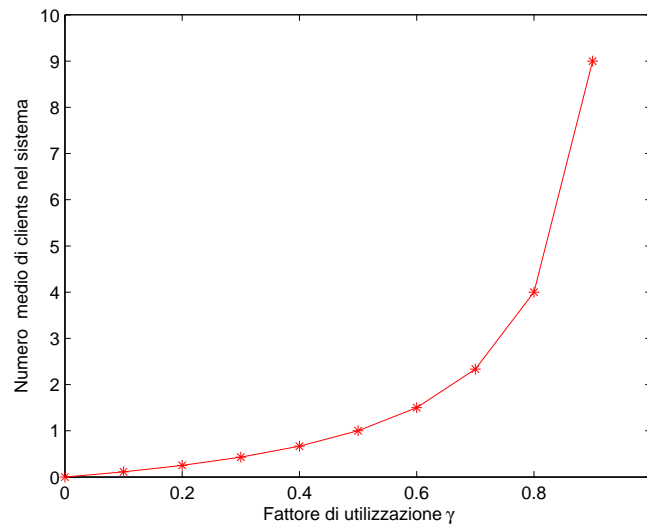


Figura 8.7: Fattore di utilizzazione γ

Conoscendo adesso il numero medio di clients C , possiamo dare alcune quantita' che caratterizzano il sistema:

- **tempo di attesa totale** T cioe' l'intervallo di tempo medio tra l'arrivo e la partenza di un client, incluso il tempo di servizio. Vale il seguente **risultato di Little**

$$T = \frac{C}{\lambda} = \frac{\gamma/\lambda}{1-\gamma} = \frac{1}{\mu-\lambda} \quad .$$

- **tempo medio di attesa in coda** W per un client: e' dato dal tempo medio di attesa T meno il tempo di servizio, cioe'

$$W = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\gamma}{\mu-\lambda}$$

- **numero medio di clients in coda** N_q

$$N_q = \lambda W = \frac{\gamma^2}{1-\gamma}$$

Esercizio 8.6 *Vogliamo confrontare le prestazioni di due centralini che hanno stessa probabilita' di servizio μ : essi pero' differiscono per il numero di chiamate che ciascuno puo' tenere in attesa (ovvero la dimensione del buffer) che saranno rispettivamente m e k . Quando arriva una nuova chiamata il centralino, se occupato, la mette in attesa a meno che il suo buffer sia pieno; in tal caso la chiamata viene respinta ed il centralino diventa saturo.*

Sfruttiamo i risultati precedentemente ottenuti per il sistema a coda di tipo M/M/1. Se il buffer del primo centralino ha dimensione massima m , allora p_m e' la probabilita' che esso sia pieno. Questa e' data dall'eq. (8.23) con $j = m$, ovvero

$$p_m = \frac{\gamma^m - \gamma^{m+1}}{1 - \gamma^{m+1}}$$

analogamente per il centralino di dimensione k avremo che

$$p_k = \frac{\gamma^k - \gamma^{k+1}}{1 - \gamma^{k+1}} \quad .$$

Supponiamo che il tasso di servizio $\mu = 0.75$ e che $k = 4$, $m = 8$. Calcoliamo p_4 e p_8 per due diversi valori del tasso di arrivo λ , si ottiene la seguente tabella

λ	p_4	p_8
0.5	0.0082	0.00016
0.66	0.0758	0.0133

Nel caso in cui $\lambda=0.5$ avremo che $p_4 \simeq p_8 \simeq 0$ %, ovvero la probabilita' che i due centralini si saturino (non accettino chiamate) e' circa zero. Se $\lambda=0.66$ (ovvero c'e' piu' traffico) $p_4 = 7.5\%$ e $p_8 = 1.3\%$, ovvero il primo centralino ha piu' probabilita' di essere saturato.

Esercizio 8.7 *Consideriamo un sistema di servizio costituito da un casello autostradale con una singola postazione. Se e' previsto il transito di 300 automobili in un' ora e per effettuare il pagamento servono in media 30 secondi, possiamo immediatamente vedere che il sistema non e' stabile. La frequenza media di terminazione di servizio infatti sar  di $3600/30 = 120 [h^{-1}]$ cioe' la capacita' del casello e' di 120 automobili l'ora. Quindi $\lambda = 300[h^{-1}]$, $\mu = 120[h^{-1}]$ e la condizione $\lambda < \mu$ non e' rispettata.*

In un sistema non stabile come questo appena visto la occupazione della fila di attesa tendera' a crescere indefinitamente. Questo significa che le probabilita' di stato non si potranno stabilizzare: osservando il sistema dopo un certo tempo aumentera' la probabilita' di trovare il sistema negli stati con piu' utenti nella fila di attesa e diminuira' la probabilita' di trovare il sistema negli stati con meno utenti nella fila di attesa (non vale l'equazione di bilancio dettagliato). Dal punto di vista matematico questo corrisponde a non poter trovare delle probabilita' di stato finite che mantengono in equilibrio il sistema (la sommatoria diverge). Osserviamo infine che in un sistema non stabile non si conserva il flusso degli utenti: nell' esempio visto prima arrivano al casello 300 macchine l'ora, mentre ne escono 120. In queste condizioni NON si applica la legge di Little, che ha bisogno della condizione di equilibrio statistico.

Appendice A

Il Processo di Poisson

Il processo di Poisson e' un processo di conteggio per il numero di eventi che sono accaduti fino ad un certo tempo. Vediamo come si definisce.

Definizione A.1 *Un processo stocastico $\{N(t), t \geq 0\}$ e' un processo di conteggio, se rappresenta un numero di eventi che sono accaduti nell'intervallo di tempo $[0, t]$.*

Per esempio $N(t)$ e' il numero di clienti che sono entrati alla Posta nell'intervallo $[0, t]$ per avere un servizio.

Un processo di conteggio deve soddisfare alle seguenti proprieta':

1. $N(t) \geq 0$
2. $N(t) \in \mathbb{N}_0$
3. $\forall s < t$ allora $N(s) \leq N(t)$ (non decrescente)
4. se $s < t$ allora $N(t) - N(s)$ e' il numero di eventi che sono accaduti nell'intervallo $]s, t]$.

Definizione A.2 *Un processo di conteggio $N(t)$ ha incrementi indipendenti se il numero di eventi che sono accaduti nell'intervallo di tempo $]s, t]$ pari a $N(t) - N(s)$, e' indipendente dal numero di eventi accaduti prima del tempo s .*

Questa e' la proprieta' di non-memoria. Inoltre, il numero di eventi che accade in intervalli di tempi disgiunti (non intersecanti) sono indipendenti.

Definizione A.3 Un processo di conteggio $N(t)$ ha incrementi stazionari se la distribuzione del numero di eventi che accade in ogni intervallo di tempo dipende solamente dall'ampiezza dell'intervallo.

Quindi $N(t)$ ha incrementi stazionari se accade che $N(t_2 + s) - N(t_1 + s) = N(t_2) - N(t_1)$ $\forall t_1 < t_2$ e $\forall s$.

Definizione A.4 Un processo di conteggio $N(t)$ e' un Processo di Poisson con tasso $\lambda > 0$ se

i) $N(0) = 0$;

ii) ha incrementi indipendenti;

iii) il numero di eventi in ogni intervallo di tempo avente ampiezza t ha una distribuzione di Poisson con media λt . Ovvero

$$P(N(t+s) - N(s) = n) = (\lambda t)^n \frac{e^{-\lambda t}}{n!} \quad \forall s, t, n = 0, 1, 2, \dots \quad (\text{A.1})$$

Osserviamo che il Processo di Poisson ha incrementi stazionari: infatti dalla eq.(A.1) si vede che il numero di conteggi nell'intervallo $]s, t+s]$ dipende dall'ampiezza dell'intervallo solamente che e' $t+s-s=t$.

In particolare, la probabilita' che nell'intervallo $[0, \tau]$ ci siano n arrivi viene indicata con

$$p(n) = (\lambda \tau)^n \frac{e^{-\lambda \tau}}{n!}, n = 0, 1, 2, \dots \quad (\text{A.2})$$

Definizione A.5 Sia T_1 l'ampiezza dell'intervallo di tempo di arrivo del primo evento, T_2 l'ampiezza dell'intervallo di tempo di arrivo dell'evento successivo al primo etc. In questo modo si definisce la successione dei tempi di interarrivo $\{T_n\}$.

Proposizione A.1 La successione dei tempi di interarrivo ha una distribuzione esponenziale con media $1/\lambda$.

Dim. Infatti il primo evento arrivera' nell'intervallo $[0, t_1]$ (di ampiezza $T_1 = t_1$), quindi nell'intervallo $[0, t]$ con $t < T_1$ non ci saranno arrivi, cioe'

$$P(T_1 > t) = P(N(t) - N(0) = 0) = e^{-\lambda t} \Rightarrow T_1 \simeq \text{Exp}(\lambda) \quad .$$

Il secondo evento arrivera' nell'intervallo $[t_1, t_2]$ di ampiezza $T_2 = t_2 - t_1$ e quindi

$$P(T_2 > t | T_1 = s) = P(N(t + s) - N(s) = 0 | T_1 = s) =$$

ma poiche' il primo evento e' certamente arrivato a T_1

$$= P(N(t + s) - N(s) = 0 | T_1 = s) = P(N(t + s) - N(s) = 0) = e^{-\lambda t} \Rightarrow T_2 \simeq \text{Exp}(\lambda)$$

e cosi' via. Poiche' questi intervalli T_n sono disgiunti allora questa successione avra' incrementi indipendenti e sara' distribuita secondo un esponenziale.

Definizione A.6 Il tempo di arrivo, detto anche tempo di attesa, dell' n -esimo evento e' la somma dei primi n tempi di interarrivo

$$S_n = T_1 + T_2 + \dots + T_n$$

Si puo' provare che S_n ha una distribuzione di tipo Gamma(n, λ).

Proposizione A.2 Dalla eq.(A.1) se $t = \delta \ll 1$, sviluppando in serie l'esponenziale e trascurando i termini quadratici in δ , otterremo ¹

- $P(N(\delta + s) - N(s) = 0) = 1 - \lambda\delta + o(\delta)$
- $P(N(\delta + s) - N(s) = 1) = \lambda\delta + o(\delta)$
- $P(N(\delta + s) - N(s) \geq 2) = o(\delta)$

Dim. Ricordiamo che per una funzione derivabile $m + 1$ volte vale la formula di Taylor:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \dots + \frac{1}{m!}f^{(m)}(x_0)(x - x_0)^m + R_m(x)$$

dove $R_m(x)$ e' la funzione resto tale che

$$\lim_{x \rightarrow x_0} \frac{R_m(x)}{(x - x_0)^m} = 0 \quad \Leftrightarrow \quad R_m(x) = o[(x - x_0)^m] \quad .$$

L'eq.(A.1) per $t = \delta$ si scrive

$$f_n(\delta) = P(N(\delta + s) - N(s) = n) = (\lambda\delta)^n \frac{e^{-\lambda\delta}}{n!}$$

e per $n = 0$ e $\delta_0 = 0$ applicando la formula di Taylor (con $m = 1$) avremo

$$f_0(\delta) = e^{-\lambda\delta} = e^{-\lambda\delta}|_0 + \left(\frac{d}{d\delta} e^{-\lambda\delta} \right)_0 \delta + R_1(\delta) = 1 - \lambda\delta + o(\delta)$$

in modo analogo si dimostrano le altre due relazioni.

¹Dove la funzione $o(\delta)$ (notazione di Landau) e' tale che $\lim_{\delta \rightarrow 0} \frac{o(\delta)}{\delta} = 0$, quindi $\delta^2 \in o(\delta)$ e $\delta \notin o(\delta)$.