



UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

CALCOLO DELLE PROBABILITÀ E STATISTICA

2021/2022

<i>Professore</i>	<i>Studente</i>
Aniello Buonocore	Vincenzo Tramo

Indice

Elementi di Probabilità	5
1 Elementi di Probabilità	5
1.1 Spazio degli esiti ed eventi	5
1.2 Assiomi della probabilità	6
1.3 Spazi di esiti equiprobabili	8
1.3.1 Calcolo combinatorio	11
1.4 Terna di probabilità o spazio di probabilità	14
1.5 Problema delle concordanze	22
1.5.1 Probabilità di osservare almeno una concordanza nelle n chiamate	22
1.5.2 Probabilità di osservare zero concordanze nelle n chiamate	22
1.5.3 Probabilità di osservare SOLO una concordanza nelle n chiamate	23
1.6 Probabilità condizionata	24
1.7 Formula delle alternative e formula di Bayes	26
1.7.1 Sistema completo di alternative e formula delle alternative generalizzata	28
1.7.2 Eventi indipendenti	32
Variabili aleatorie e valore atteso	35
2 Variabili aleatorie e valore atteso	35
2.1 Variabili aleatorie	35
2.2 Continuità e misurabilità	37
2.3 Definizione di valore atteso (o media) di un numero aleatorio discreto	40
2.4 La funzione generatrice dei momenti	41
2.4.1 Proprietà di funzione generatrice dei momenti	41
2.5 Funzione di distribuzione	42
2.5.1 Successione di eventi monotona crescente/decrescente	44
2.5.2 Proprietà della funzione di distribuzione	44
Modelli di variabili aleatorie discrete	47
3 Modelli di variabili aleatorie discrete	47
3.1 Variabili aleatorie di Bernoulli e binomiali	47
3.1.1 Distribuzione di probabilità binomiale con parametri n e p	49
3.1.2 Calcolo esplicito della distribuzione binomiale	51
3.2 Distribuzione geometrica - legge geometrica	52
3.3 Distribuzione di Pascal	53
3.4 Distribuzione di probabilità uniforme e discreta	53
3.5 Distribuzione di Poisson	54
3.5.1 Calcolo esplicito della distribuzione di Poisson	55
3.6 Numeri aleatori ipergeometrici	56
Modelli di variabili aleatorie continue	58
4 Modelli di variabili continue	58
4.1 Variabile aleatoria continua con legge uniforme	60
4.1.1 Definizione varianza - varianza legge uniforme	64
4.2 Variabile aleatoria esponenziale	65
4.2.1 Proprietà di assenza di usura	68
4.3 Variabile aleatoria normale o gaussiana	69
4.3.1 Funzione generatrice dei momenti di un numero aleatorio gaussiano standard	69
4.3.2 Media e varianza numero aleatorio gaussiano standard	70
4.3.3 Proprietà varianza (in generale)	70
4.3.4 Operazione di standardizzazione	71
4.3.5 Funzione generatrice dei momenti di un numero aleatorio normale	71
4.3.6 Calcolo punto massimo	72
4.3.7 Calcolo punti di flesso	72
Coppie e vettori di variabili aleatorie	76

5	Coppie e vettori di variabili aleatorie	76
5.1	Distribuzione congiunta per variabili aleatorie discrete	76
5.2	Distribuzione congiunta per variabili aleatorie continue	78
5.3	Distribuzione congiunta quando le variabili sono indipendenti	79
5.4	Distribuzioni condizionali	79
5.4.1	Distribuzioni condizionali caso discreto	79
5.4.2	Distribuzioni condizionali caso continuo	81
5.5	Valore atteso	81
5.5.1	Valore atteso di una funzione di variabile aleatoria	82
5.5.2	Proprietà valore atteso variabili aleatorie congiunte	82
5.5.3	La media è il miglior predittore di X rispetto all'errore quadratico medio	84
5.6	Covarianza	85
5.7	Disuguaglianza di Markov	89
5.7.1	Quand'è che un evento è quasi certamente 1 o quasi certamente impossibile?	89
5.8	Disuguaglianza di Chebyshev	89
5.9	La legge debole dei grandi numeri	90
5.9.1	Teorema di Bernoulli	91
	Statistica descrittiva	92
6	Statistica descrittiva	92
6.1	Definizione di statistica descrittiva	92
6.2	Tipo e modalità dei caratteri	92
6.3	Distribuzioni di frequenza	93
6.4	Rappresentare un carattere quantitativo continuo - Istogramma	93
6.5	Moda e quartili	94
6.6	Rappresentazioni grafiche	94
6.7	Indici di posizione	95
6.7.1	Medie analitiche	95
6.7.2	Centri	96
6.8	Indici di dispersione	98
6.9	Tabella indici di sintesi e indici di dispersione	99
6.10	Diagramma scatola con baffi	99
	Statistica	101
7	Statistica	101
7.1	Distribuzione delle statistiche campionarie	101
7.2	Il teorema del limite centrale	102
7.2.1	Distribuzione approssimata della media campionaria tramite il teorema del limite centrale	103
7.2.2	Quando un campione è abbastanza numeroso?	103
7.3	Funzione generatrice dei momenti della media campionaria	104
7.3.1	Caso in cui la genitrice è gaussiana	104
7.4	Varianza campionaria	104
7.5	Stimatori	105
7.5.1	Introduzione ai stimatori con un esempio	105
7.5.2	Correttezza dello stimatore S^2 per la varianza σ^2	105
7.5.3	Definizione di correttezza	106
7.5.4	Ordinamento tra gli stimatori corretti	106
7.5.5	Rischio quadratico medio	106
7.5.6	Ordinamento tra due stimatori generici (anche non corretti)	107
	Stima parametrica	108
8	Stima parametrica	108
8.1	Introduzione	108
8.2	Stimatori di massima verosimiglianza	108
8.2.1	Stimatori di massima verosimiglianza - caso continuo (variabile esponenziale)	108
8.2.2	Stimatori di massima verosimiglianza - caso discreto (variabile binomiale)	110
8.3	Stimatori di massima verosimiglianza - variabile gaussiana	112
8.3.1	Determinazione stimatori - caso μ incognita e σ^2 nota	112
8.3.2	Determinazione stimatori - caso μ noto e σ^2 incognita	112

8.3.3	Determinazione stimatori - caso μ incognita e σ^2 incognita	113
8.4	Stimatori di massima verosimiglianza - variabile di poisson	114
8.5	Stimatori di massima verosimiglianza - differenza caso discreto e caso continuo	114
8.6	Metodo dei momenti	114
8.7	Proprietà di consistenza degli stimatori	116

Elementi di Probabilità

1 Elementi di Probabilità

Il concetto di probabilità di un evento, quando si effettua un esperimento, è passibile di diverse interpretazioni (diversi modi di definire la probabilità di un evento):

- *Interpretazione di Laplace*: solo in caso di simmetria perfetta per i punti campione (che ne devono essere un numero finito). Abbiamo che:

$$A \in \mathcal{F}, \quad \mathbb{P}_e(A) = \frac{|A|}{|\Omega|} \geq 0$$

- *Interpretazione frequentistica (o statistica)*: la probabilità di un esito è considerata una proprietà dell'esito stesso. Si ripete l'esperimento sottostante un numero $n \in \mathbb{N}$ di volte e si considera il rapporto:

$$A \in \mathcal{F}, \quad \mathbb{P}_f(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \geq 0$$

con n_A che rappresenta il numero delle volte in cui A si è verificato. Lo svantaggio è che dobbiamo effettuare un certo numero di esperimenti per essere precisi. Inoltre, dobbiamo ripetere lo stesso identico esperimento e non sempre questo è possibile.

- *Interpretazione soggettivistica*: non si crede che la probabilità di un esito sia una proprietà oggettiva, ma piuttosto la precisazione del livello di fiducia che lo studioso ripone nel verificarsi dell'esito. La probabilità di un evento è il prezzo p che un individuo *equo e coerente* paga per ricevere 1 nel caso in cui A si verifichi.

In questo capitolo presentiamo le regole e gli assiomi della teoria della probabilità. Gli assiomi e le regole della teoria della probabilità valgono per tutte e tre le interpretazioni.

1.1 Spazio degli esiti ed eventi

Preliminarmente all'enunciare gli *assiomi di probabilità*, occorre introdurre il concetto di **spazio degli esiti**, e quello di **evento**.

Si consideri un esperimento il cui esito non sia prevedibile con certezza. Quello che normalmente si può fare comunque, è individuare la rosa degli esiti plausibili. L'insieme di tutti gli esiti possibili si dice *spazio campione* (oppure *spazio degli esiti*, in inglese *sample space* - indicato con S oppure con Ω).

Esempio 1.1.1

Se l'esito dell'esperimento consiste nella determinazione del sesso di un neonato, allora poniamo:

$$S = \{f, m\}$$

dove si intende che l'esito f rappresenta la nascita di una femmina, e l'esito m quella di un maschio.

Esempio 1.1.2

Se l'esperimento consiste in una gara tra sette cavalli denotati dai numeri 1, 2, 3, 4, 5, 6 e 7, allora:

$$S = \{\text{tutti gli ordinamenti di } (1, 2, 3, 4, 5, 6, 7)\}$$

In questo caso l'esito $(2, 3, 7, 6, 5, 4, 1)$ è quello in cui il cavallo 2 arriva primo, il 3 arriva secondo, il 7 terzo, e così via.

Esempio 1.1.3

Supponiamo di voler determinare il minimo dosaggio di un farmaco al quale un paziente reagisce positivamente. Una possibile scelta per lo spazio degli esiti di questo esperimento potrebbe essere l'insieme di tutti i numeri positivi, ovvero:

$$S = (0, \infty)$$

intendendo ovviamente che l'esito sarebbe x se il paziente reagisse a un dosaggio pari a x e a nessun dosaggio inferiore.

I sottoinsiemi dello spazio degli esiti si dicono *eventi*, quindi un evento E è un insieme i cui elementi sono esiti possibili. Se l'esito dell'esperimento è contenuto in E , diciamo che l'evento E si è verificato. L'*unione* $E \cup F$ di due eventi E e F dello stesso spazio degli esiti S , è definita come l'insieme formato dagli esiti che stanno o in E o in F . Quindi l'evento $E \cup F$ si verifica se *almeno uno* tra E e F si verifica.

In maniera simile è utile definire l'*intersezione* $E \cap F$ di due eventi E e F . Essa è l'insieme formato dagli esiti che sono presenti sia in E , sia in F . Come evento, rappresenta il verificarsi di *entrambi* gli eventi E e F . Quindi nell'esempio 1.1.3, se $E = (0, 5)$ è l'evento in cui il dosaggio cercato è minore di 5, e $F = (2, 10)$ è l'evento in cui esso è compreso tra 2 e 10, allora $E \cap F = (2, 5)$ è l'evento in cui esso è compreso tra 2 e 5. Nell'esempio 1.1.2, se $E = \{\text{tutti gli esiti che terminano con } 5\}$ è l'evento "il cavallo 5 arriva ultimo" e $F = \{\text{tutti gli esiti che cominciano con } 5\}$ è l'evento "il cavallo 5 arriva primo", allora chiaramente l'evento $E \cap F$ non contiene esiti possibili e non può avvenire mai. Per dare una denominazione ad un tale evento, ci riferiremo ad esso come l'*evento vuoto* e lo rappresenteremo con il simbolo \emptyset . Esso è quindi un evento che non contiene esiti possibili per l'esperimento. Se $E \cap F = \emptyset$, ovvero se E e F non possono verificarsi entrambi, li diremo eventi *mutuamente esclusivi* o *eventi disgiunti*.

Per ogni evento E , definiamo l'evento E^c , che diciamo *complementare* di E , come l'insieme formato dagli esiti di S che non stanno in E . Quindi E^c si verifica se e solo se non si verifica E . Si noti infine come valga la ovvia relazione $S^c = \emptyset$.

1.2 Assiomi della probabilità

Se si ripete molte volte un esperimento mettendosi sempre nelle stesse condizioni, si verifica empiricamente che la frazione di casi sul totale in cui si realizza un qualunque evento E tende - al crescere dei tentativi - ad un valore costante che dipende solo da E . Tutti sanno ad esempio, che se si lancia tante volte una moneta, il rapporto tra il numero di risultati *testa* e il numero di tentativi, man mano che aumentiamo il numero di lanci, tende ad un valore costante (cioè 0.5).

Quale che sia la definizione di probabilità che vogliamo abbracciare, vi è un comune accordo sulle regole che tali probabilità devono rispettare: da qui in poi il modo di procedere diviene allora esclusivamente astratto. Si associa ad ogni evento E sullo spazio degli esiti S , un numero che si denota $\mathbb{P}(E)$ e che si dice *probabilità* dell'evento E . Ciò non può essere fatto in maniera completamente libera: le probabilità dei vari eventi devono rispettare alcuni assiomi dal significato intuitivo.

$$0 \leq \mathbb{P}(E) \leq 1 \quad (\text{Assioma 1})$$

$$\mathbb{P}(S) = 1 \quad (\text{Assioma 2})$$

Inoltre per ogni successione di eventi mutuamente esclusivi E_1, E_2, \dots (cioè tali che $E_i \cap E_j = \emptyset$ quando $i \neq j$),

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbb{P}(E_i), \quad n = 1, 2, \dots, \infty \quad (\text{Assioma 3})$$

Il primo assioma afferma che ogni probabilità è un numero compreso tra 0 e 1. Il secondo stabilisce che l'evento S si verifica con probabilità 1, ovvero vi è assoluta certezza che si realizzi un esito contenuto in S , o ancora, S contiene necessariamente tutti gli esiti possibili del nostro esperimento. Il terzo assioma infine afferma che, preso un insieme finito o numerabile di eventi mutuamente esclusivi, la probabilità che se ne verifichi almeno uno è uguale alla somma delle loro probabilità.

Si può a questo punto notare che se si interpreta $\mathbb{P}(E)$ come la *frequenza relativa* dell'evento E quando l'esperimento è ripetuto un gran numero di volte, questa definizione soddisfa i predetti assiomi. Gli assiomi permettono di dedurre un gran numero di proprietà delle probabilità degli eventi. Ad esempio, possiamo notare che E e E^c sono eventi disgiunti, e quindi usando Assioma 2 e Assioma 3,

$$1 = \mathbb{P}(S) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c)$$

Proposizione 1.2.1

Per ogni evento $E \subset S$, vale la relazione

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E) \quad (1.4.1)$$

La probabilità che un evento qualsiasi non si verifichi è pari a uno meno la probabilità che si verifichi. Ad esempio, se sappiamo che la probabilità di ottenere *testa* lanciando una certa moneta è $3/8$, allora evidentemente la probabilità di ottenere *croce* dalla stessa moneta è $5/8$.

La prossima proposizione fornisce la probabilità dell'unione di due eventi in termini della loro probabilità singole e di quella dell'intersezione (si noti che questa rappresenta una estensione dell'Assioma 3 che funziona anche con eventi non mutuamente esclusivi).

Proposizione 1.2.2

Se E e F sono due eventi qualsiasi, allora:

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) \quad (1.2.2)$$

Dimostrazione. I diagrammi di Venn forniscono una dimostrazione molto intuitiva.

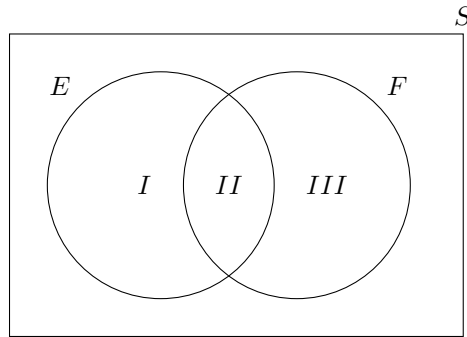


Figura 1: Illustrazione della Proposizione 1.2.2 con un diagramma di Venn.

Si osservi la Figura 1; poiché le regioni I , II e III sono disgiunte, si può applicare tre volte l'Assioma 3 per ottenere:

$$\mathbb{P}(E \cup F) = \mathbb{P}(I) + \mathbb{P}(II) + \mathbb{P}(III)$$

$$\mathbb{P}(E) = \mathbb{P}(I) + \mathbb{P}(II)$$

$$\mathbb{P}(F) = \mathbb{P}(II) + \mathbb{P}(III)$$

Confrontando le tre identità si vede che:

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(II)$$

e la dimostrazione è conclusa, poiché: $II = E \cap F$

□

1.3 Spazi di esiti equiprobabili

Per tutta una serie di esperimenti è naturale assumere che ogni esito di uno spazio S abbia la stessa probabilità di realizzarsi. Ciò può accadere solo se S è un insieme finito, e in questo caso, si può assumere senza perdita di generalità che sia $S = \{1, 2, \dots, N\}$; in queste ipotesi l'equiprobabilità degli esiti si scrive:

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \dots = \mathbb{P}(\{N\}) := p$$

Da Assioma 2 e Assioma 3 segue che:

$$1 = \mathbb{P}(S) = \mathbb{P}(1) + \mathbb{P}(2) + \dots + \mathbb{P}(N) = Np$$

da cui si deduce che $\mathbb{P}(\{i\}) = p = 1/N$, per tutti gli $i = 1, 2, \dots, N$. Da questo risultato e ancora dall'Assioma 3 si conclude che per ogni evento E ,

$$\mathbb{P}(E) = \frac{|E|}{N} \quad (\text{Proprietà di equiprobabilità})$$

In altre parole se si assume che ogni esito di S abbia la medesima probabilità, allora la probabilità di un qualunque evento E è pari al rapporto tra il numero di esiti contenuti in E e il numero totale di esiti di S .

Una conseguenza notevole di questo risultato è che occorre sapere contare efficacemente il numero degli esiti differenti appartenenti ad un evento. A questo scopo faremo uso della regola seguente.

Osservazione 1.3.1 (Principio di enumerazione o regola moltiplicativa)

Consideriamo la realizzazione di due diversi esperimenti (detti 1 e 2), che possono avere rispettivamente m e n esiti differenti. Allora complessivamente vi sono mn diversi risultati se si considerano entrambi gli esperimenti contemporaneamente.

Dimostrazione. L'enunciato si dimostra enumerando tutte le possibili coppie di risultati dei due esperimenti, che sono:

$$\begin{array}{cccc} (1, 1) & (1, 2) & \dots & (1, n) \\ (2, 1) & (2, 2) & \dots & (2, n) \\ \vdots & \vdots & & \vdots \\ (m, 1) & (m, 2) & \dots & (m, n) \end{array}$$

dove si intende che si ottiene il risultato (i, j) se nell'esperimento 1 si realizza l'esito i -esimo tra gli m possibili, e nell'esperimento 2 quello j -esimo tra gli n possibili. Siccome la tabella ottenuta ha m righe e n colonne, vi sono complessivamente mn esiti possibili. □

Generalizzazione del principio di enumerazione

Se si eseguono r esperimenti, ed è noto che il primo esperimento ammette n_1 esiti possibili, per ognuno dei quali il secondo esperimento ammette n_2 esiti diversi, inoltre se per ogni combinazione di esiti dei primi due esperimenti il terzo ammette n_3 esiti diversi, e così via, allora vi sono un totale di $n_1 \times n_2 \times \dots \times n_r$ combinazioni di esiti degli r esperimenti considerati tutti insieme.

Il principio di enumerazione ammette una utile generalizzazione, descritta nel riquadro presentato. Per illustrarne un'applicazione, proviamo a determinare il numero di modi diversi in cui si possono ordinare n oggetti. Per esempio, il numero di modi in cui si possono ordinare i tre simboli a , b e c sono sei, ovvero esplicitamente abc , acb , bac , bca , cab e cba . Ciascuno di questi ordinamenti prendono il nome di *permutazione* dei tre simboli considerati; le permutazioni di tre elementi sono perciò sei. Vediamo come questo risultato fosse deducibile dal principio di enumerazione generalizzato. Il primo simbolo della permutazione può essere scelto in tre modi diversi; per ogni scelta del primo simbolo, il secondo può essere preso tra i due restanti; il terzo e ultimo viene individuato per esclusione (una sola scelta). Quindi vi sono $3 \times 2 \times 1 = 6$ possibili permutazioni.

Supponiamo ora di avere n oggetti. Se ragioniamo in modo analogo, scopriamo che vi sono:

$$n(n-1)(n-2) \dots 3 \cdot 2 \cdot 1 =: n!$$

diverse permutazioni degli n oggetti.

Come formalizzare un esperimento aleatorio - Il gioco della zara con due dadi onesti

La **zara** è un gioco d'azzardo in uso nel Medioevo. Si gioca con tre dadi (in questo problema consideriamo due dadi per semplicità): a turno ogni giocatore chiama un numero da 3 a 18, quindi getta i dadi. Vince chi per primo ottiene il punteggio pari al numero chiamato. Vogliamo determinare la probabilità di uscita di ogni possibile numero ottenuto sommando la coppia di numeri (in questo caso da 2 a 12).

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)		2	3	4	5	6	7
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)		3	4	5	6	7	8
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)		4	5	6	7	8	9
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)		5	6	7	8	9	10
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)		6	7	8	9	10	11
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)		7	8	9	10	11	12

$\xrightarrow{\mathbf{Z}}$

Lo *spazio campione* (o *spazio degli esiti*) è $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$, quindi $|\Omega| = 36$. Notiamo che ogni esito dello spazio Ω ha la stessa probabilità di realizzarsi. Dunque, possiamo dire che per ogni elemento $x \in \Omega$, $\mathbb{P}(x) = \frac{1}{36}$, dunque Ω è uno *spazio equiprobabile*. Consideriamo adesso questo insieme $S_z = \{2, 3, 4, 5, \dots, 12\}$ detto *spettro della trasformazione* (in questo caso è una trasformazione che ad ogni coppia associa la sua somma $z \in \mathbb{R}$, chiamiamo questa trasformazione f). Analizzando la tabella di destra possiamo facilmente risalire alla probabilità di ogni evento.

$$\begin{aligned}
 f^{-1}(2) &= \{(1, 1)\} \\
 f^{-1}(3) &= \{(1, 2), (2, 1)\} \\
 f^{-1}(4) &= \{(1, 3), (2, 2), (3, 1)\} \\
 f^{-1}(5) &= \{(1, 4), (2, 3), (3, 2), (4, 1)\} \\
 f^{-1}(6) &= \{(5, 1), (4, 2), (3, 3), (2, 4), (1, 5)\} \\
 f^{-1}(7) &= \{(6, 1), (5, 2), (4, 3), (3, 4), (2, 5), (1, 6)\} \\
 f^{-1}(8) &= \{(6, 2), (5, 3), (4, 4), (3, 5), (2, 6)\} \\
 f^{-1}(9) &= \{(6, 3), (5, 4), (4, 5), (3, 6)\} \\
 f^{-1}(10) &= \{(6, 4), (5, 5), (4, 6)\} \\
 f^{-1}(11) &= \{(6, 5), (5, 6)\} \\
 f^{-1}(12) &= \{(6, 6)\}
 \end{aligned}$$

Dunque la probabilità di ogni numero (da 2 a 12) è:

$$\begin{aligned}
 \mathbb{P}_z(2) &= \frac{1}{36} = \mathbb{P}_z(12) \\
 \mathbb{P}_z(3) &= \frac{2}{36} = \mathbb{P}_z(11) \\
 \mathbb{P}_z(4) &= \frac{3}{36} = \mathbb{P}_z(10) \\
 \mathbb{P}_z(5) &= \frac{4}{36} = \mathbb{P}_z(9) \\
 \mathbb{P}_z(6) &= \frac{5}{36} = \mathbb{P}_z(8) \\
 \mathbb{P}_z(7) &= \frac{6}{36}
 \end{aligned}$$

Quindi abbiamo avuto un riscontro della Proprietà di equiprobabilità degli spazi equiprobabili. Abbiamo risolto il problema utilizzando anche il principio di enumerazione (1.3.1).

Esempio 1.3.1

Un corso di probabilità è frequentato da 10 studenti: 6 maschi e 4 femmine. Viene effettuato un esame, e i punteggi degli studenti sono tutti diversi. (a) Quante diverse classifiche sono possibili? (b) Se tutte le classifiche si pensano equiprobabili, qual è la probabilità che le quattro studentesse ottengano i punteggi migliori?

(a) Siccome ogni classifica è associata ad una precisa permutazione di dieci studenti, esse in tutto sono $10! = 3\,628\,800$. (b) Poiché vi sono $4!$ diverse classifiche delle studentesse tra di loro e $6!$ classifiche dei maschi, segue dal principio di enumerazione che vi sono $4! \times 6! = 24 \times 720 = 17\,280$ possibili classifiche in cui le studentesse occupano le prime 4 posizioni. Quindi la probabilità cercata è:

$$\frac{4!6!}{10!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{10 \cdot 9 \cdot 7 \cdot 6} = \frac{1}{210}$$

Esempio 1.3.2

Se in una stanza sono radunate n persone, qual'è la probabilità che non ve ne siano due che compiono gli anni lo stesso giorno dell'anno? Quanto grande deve essere n affinché tale probabilità sia minore di $1/2$?

Siccome ogni persona può celebrare il compleanno in uno qualsiasi dei 365 giorni, vi sono in tutto 365^n diversi esiti dell'esperimento consistente nel domandare a ciascun partecipante la data di nascita (stiamo ignorando la possibilità che qualcuno sia nato il 29 febbraio di un anno bisestile). Secondariamente, vi sono in tutto $365 \cdot 364 \cdot 363 \cdots (365 - n + 1)$ esiti che fanno sì che tutte le persone abbiano date di compleanno diverse. Infatti la prima persona può compiere gli anni in uno qualsiasi dei 365 giorni dell'anno; la seconda - non potendo usare la stessa data - può essere nata in uno dei 364 giorni rimanenti; la terza in uno dei 363 giorni diversi da quelli delle prime due, e così via fino all'ultima persona, che ha $365 - n + 1$ date libere in cui può compiere gli anni. Allora, assumendo che ciascun esito sia equiprobabile, la probabilità cercata è pari a:

$$\frac{365 \cdot 364 \cdot 363 \cdots (365 - n + 1)}{365^n}$$

1.3.1 Calcolo combinatorio

Sia $S = \{e_1, e_2, \dots, e_n\}$ un alfabeto e sia $k \in \mathbb{N}$ un numero intero positivo. Una k -selezione da S è una parola lunga k costruita a partire da S . Individuiamo due tipi di k -selezione:

- **Disposizioni:** sono k -selezioni in cui è importante l'ordine di presentazione degli elementi di S . A sua volta si dividono in:
 - *Disposizioni semplici (senza ripetizioni):* non è ammesso utilizzare un elemento più di una volta in un raggruppamento. Quando $k = n$ chiamiamo la disposizione *permutazione* (la permutazione è un caso particolare di disposizione semplice dove vengono considerati tutti gli elementi dell'insieme);
 - *Disposizioni con ripetizione:* è ammesso utilizzare un elemento più di una volta in un raggruppamento.
- **Combinazioni:** sono k -selezioni in cui NON è importante l'ordine di presentazione degli elementi di S . Si dividono in:
 - *Combinazioni semplici (senza ripetizioni);*
 - *Combinazioni con ripetizione.*

Disposizioni semplici

Si dice *disposizione semplice* di n elementi distinti su k ($n, k \in \mathbb{N}$, $0 < k \leq n$) una collezione di k degli n elementi che rispetti le seguenti proprietà:

1. ciascun raggruppamento contiene k elementi (è una k -selezione);
2. uno stesso elemento può figurare al più una volta in un raggruppamento;
3. due raggruppamenti sono da considerarsi distinti quando essi differiscono per almeno un elemento, o per l'ordine degli elementi.

Le disposizioni semplici di n elementi presi k per volta sono in totale:

$$D_{n,k} = n \cdot (n-1) \cdot \dots \cdot (n-k+1) \cdot \frac{(n-k)}{(n-k)} \cdot \frac{(n-k+1)}{(n-k+1)} \cdot \dots \cdot \frac{2}{2} \cdot \frac{1}{1} = \frac{n!}{(n-k)!}$$

Esempio 1.3.3

La Corsa Tris detta anche, più semplicemente, Tris è una gara ippica (al trotto o al galoppo) legata ad un concorso nazionale che premia gli scommettitori che indovino i primi tre piazzati della corsa nell'esatto ordine di arrivo. L'alfabeto sono numeri che rappresentano i cavalli $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Vogliamo contare quanti podii possibili ci sono. Siccome l'ordine di rappresentazione degli elementi dell'alfabeto all'interno della k -selezione (in questo caso 3-selezione) è importante, dobbiamo contare quante disposizioni semplici ci sono con $k = 3$ e $n = 9$:

$$D_{3,9} = \frac{9!}{6!} = 9 \cdot 8 \cdot 7 = 504$$

Disposizioni con ripetizione

Si dice *disposizione con ripetizione* (o *reimmissione*) di n elementi distinti su $k > 0$ (intero positivo) posizioni una collezione di k degli n elementi che rispetti le seguenti proprietà:

1. ciascun raggruppamento contiene k elementi (è una k -selezione);
2. due qualsiasi raggruppamenti sono da considerarsi distinti quando essi differiscono per almeno un elemento, o per l'ordine degli elementi.

Rispetto ad una disposizione semplice, quindi, in una disposizione con ripetizione ogni elemento può essere ripetuto. Possiamo affermare quindi che ogni sotto procedura lavora sullo stesso insieme. Le disposizioni con ripetizione di n su k saranno:

$$D_{n,k}^{(r)} = n^k$$

Esempio 1.3.4

Il Totocalcio (acronimo di Totalizzatore calcistico) è un concorso a premi istituito nel 1946 e gestito dall'Amministrazione Autonoma dei Monopoli di Stato (AAMS), il cui obiettivo è la previsione degli esiti di 13 partite di calcio (14 fino al 2021). L'alfabeto è uguale a $\{1, x, 2\}$. Siccome è possibile ripetere lo stesso elemento in un raggruppamento (colonna con 14 esiti, una 14-selezione), ogni schedina compilata è una disposizione con ripetizione. Vogliamo contare tutte le possibili colonne.

$$D_{3,14}^{(r)} = 3^{14}$$

Permutazione semplice

Si chiamano *permutazioni semplici* di n elementi distinti, tutti i raggruppamenti diversi che si possono formare con gli elementi dati, rispettando le seguenti proprietà:

1. ciascun raggruppamento contiene n elementi;
2. uno stesso elemento può figurare al più una volta in un raggruppamento;
3. due raggruppamenti sono da considerarsi distinti quando essi differiscono per almeno un elemento, o per l'ordine degli elementi.

Una permutazione semplice è un caso particolare di disposizione semplice (dove essenzialmente $k = n$). Le permutazioni semplici di n elementi ne sono:

$$P_n = D_{n,n} = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1 = n!$$

Esempio 1.3.5

Vogliamo contare quanti sono gli anagrammi della parola "case":

$$P_4 = 4! = 24$$

Permutazione con ripetizioni

Per capire al meglio cos'è una *permutazione con ripetizioni* partiamo con un esempio:

Esempio 1.3.6

Vogliamo contare tutti gli anagrammi della parola "statistica". Notiamo che ci sono tre t , due s , due a , due i e una c . Le lettere che si ripetono producono anche permutazioni uguali (ad esempio se scambiamo le due lettere "s", la parola non cambia, generiamo una permutazione che è uguale a quella di partenza). Dunque la soluzione è dividere il numero totale di permutazioni semplici per il prodotto dei fattoriali delle "cardinalità" delle lettere (ad esempio la lettera t compare tre volte, dunque esistono $3!$ permutazioni che danno lo stesso risultato):

$$P_{10}^{(r)} = \frac{10!}{3! 2! 2! 2! 1!}$$

Dunque in generale (il coefficiente in questo caso è chiamato *coefficiente multinomiale*. Quando abbiamo al denominatore soltanto due lettere che si ripetono, allora necessariamente il coefficiente diventa *binomiale* ♦):

$${}_{n_1, n_2, \dots, n_s} P_n = \frac{n!}{n_1! n_2! \dots n_s!} = \binom{n}{n_1 \ n_2 \ \dots \ n_s} \quad \text{dove } n_1 + n_2 + \dots + n_s = n$$

Combinazione semplice

Si dice *combinazione semplice* di n elementi distinti su k posizioni ($n, k \in \mathbb{N}$, $0 < k \leq n$) una collezione di k degli n elementi che rispetti le seguenti proprietà:

1. ciascun raggruppamento contiene k elementi;
2. uno stesso elemento può figurare al più una volta in un raggruppamento;
3. due raggruppamenti sono da considerarsi diversi soltanto quando differiscono tra loro almeno per un elemento.

L'ordine degli elementi non ha importanza in una combinazione. Le combinazioni semplici di n elementi su k posti sono:

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Siccome l'ordine non conta nelle combinazioni, non dobbiamo considerare le permutazioni delle k -selezioni, per questo dividiamo il totale delle disposizioni semplici per il numero totale delle possibili permutazioni di una k -selezione poiché ogni gruppo di lettere fissato viene contato $k!$ volte (una per ogni sua permutazione). Alla fine troviamo il *coefficiente binomiale* di n su k : notiamo che $k + n - k = n$ ♦. Il coefficiente binomiale ha due significati: conta il numero delle combinazioni semplici ma conta anche il numero di permutazioni con ripetizione dove si ripetono solo due lettere. Esiste quindi una corrispondenza biunivoca tra le combinazioni semplici e le permutazioni con ripetizione dove si ripetono solo due lettere. Ogni permutazione con ripetizioni con due sole lettere, si può trasformare come combinazione semplice e viceversa.

Esempio 1.3.7

Consideriamo l'alfabeto $\{a, b, c, d, e, f\}$. Vogliamo contare tutte le combinazioni semplici di lunghezza 4 sull'alfabeto dato.

$$C_{6,4} = \binom{6}{4} = \binom{6}{4,2} = {}_{4,2}P_6^{(r)}$$

Il problema lo si può risolvere in maniera del tutto equivalente calcolando le permutazioni con ripetizioni dove si ripetono solo due lettere. Prendiamo un nuovo alfabeto $\{P, A\}$ dove P sta per *presenza* e A sta per *assenza*. Esiste una corrispondenza biunivoca tra combinazioni semplici e permutazioni con ripetizioni con due sole lettere secondo la codifica rappresentata nella tabella sottostante. Nota bene le stringhe che si trovano a sinistra sono combinazioni semplici e dunque l'ordine non ci interessa (la stringa "abdf" è uguale alla stringa "adbf" e così via... ogni stringa identifica una classe, cioè un insieme di stringhe che sono considerate identiche). Contare le stringhe che stanno nella tabella equivale a contare le stringhe che stanno fuori.

		Alfabeto					
		a	b	c	d	e	f
a b d f	↔	P	P	A	P	A	P
a b c d	↔	P	P	P	P	A	A

Table 1: Corrispondenza biunivoca tra combinazioni semplici e permutazioni con ripetizioni con due sole lettere

Combinazioni con ripetizione

Si dice *combinazione con ripetizione* di n elementi distinti su $k > 0$ (intero positivo) posizioni, una collezione di k degli n elementi che rispetti le seguenti proprietà:

1. ciascun raggruppamento contiene k elementi;
2. due raggruppamenti sono da considerarsi diversi soltanto quando differiscono tra loro almeno per un elemento.

Uno stesso elemento può quindi comparire più di una volta e l'ordine non conta. Le combinazioni con ripetizione di n elementi su k posti sono:

$$C_{n,k}^{(r)} = {}_{n-1,k}P_{n+k-1}^{(r)} = \binom{n+k-1}{k}$$

Esempio 1.3.8

Vogliamo contare tutte le combinazioni con ripetizione a 4 posti sull'alfabeto $\{a, b, c, d, e, f\}$. La corrispondenza biunivoca tra combinazioni con ripetizione e permutazioni con ripetizioni di due lettere può essere applicata anche qui, anzi qui è fondamentale per capire come contare le combinazioni con ripetizione. A differenza della tabella precedente, poiché le lettere possono essere ripetute in uno stesso raggruppamento nelle combinazioni con ripetizione, abbiamo bisogno di ulteriori informazioni che ci permettano di capire *quante volte una lettera si ripete* in una k -selezione (e questa è rappresentata dalla colonna delle *Ripetizioni*) e *quando fermarci* (ci permette di capire quant'è lunga la nostra k -selezione e questa informazione è rappresentata dalla colonna *Block*).

In questo esempio, $n = 6$ e $k = 4$ (tabella di sotto). Le stringhe di destra solo lunghe 9 ($n + k - 1$ dato che l'ultima informazione, cioè *Block*, è sempre fissa e non viene considerata, ecco il motivo del -1) e sono

Alfabeto						Ripetizioni			Block
a	b	c	d	e	f				
a a a a \longleftrightarrow	P	A	A	A	A	P	P	P	A
b b e e \longleftrightarrow	A	P	A	A	P	P	A	P	A
a b c d \longleftrightarrow	P	P	P	P	A	A	A	A	A

Table 2: Corrispondenza biunivoca tra combinazioni con ripetizione e permutazioni con ripetizioni di due lettere

permutazioni con ripetizione di due lettere. Le lettere che si ripetono sono A e P . La lettera A si ripete 5 volte (in generale $n - 1$) mentre la lettera P si ripete 4 volte (in generale k dato che ogni P rappresenta la presenza di una lettera dell'alfabeto nativo nella nostra k -selezione). Notiamo infine che $5 + 4 = 9$ che è proprio la lunghezza della nostra stringa nella tabella (senza considerare l'ultima colonna *Block*) e quindi viene rispettato il vincolo $n - 1 + k = n + k - 1$.

1.4 Terna di probabilità o spazio di probabilità

Sia ξ un esperimento non deterministico (non prevedibile, aleatorio). In particolare poniamo ξ uguale al lancio di un dado onesto. Lo spazio degli esiti è quindi $\Omega = \{1, 2, 3, 4, 5, 6\}$. Ci troviamo nella situazione in cui ogni esito ha la stessa probabilità di verificarsi. Ci mettiamo nell'ambito delle scommesse. Supponiamo che sia possibile scommettere su questi due eventi A e B :

- $A = \{2, 4, 6\}$, cioè esce un punteggio *pari*
- $B = \{5, 6\}$, cioè esce un punteggio *alto*

Quali sono tutti i possibili eventi di questo gioco? Sono tutti gli eventi ottenibili a partire dagli eventi A e B . E' possibile scommettere su $A \cup B$ ad esempio, oppure su $A \cap B$:

$$\begin{aligned}\mathbb{P}(A) &= \frac{3}{6} = \frac{1}{2} \\ \mathbb{P}(B) &= \frac{2}{6} = \frac{1}{3} \\ \mathbb{P}(A \cap B) &= \frac{1}{6} \\ \mathbb{P}(A \cup B) &= \frac{4}{6} = \frac{2}{3}\end{aligned}$$

Ma sono eventi anche A^c oppure B^c (punteggio *medio o basso*). Come possiamo ottenere la *famiglia degli atomi*, cioè tutti quegli eventi non ulteriormente scomponibili di questo gioco? Possiamo ricorrere alla tecnica molto semplice del **metodo degli atomi**. Ad esempio, la famiglia degli atomi di questo gioco è ottenibile in questo modo (applicando il metodo degli atomi):

$$\begin{aligned}\boxed{A \cap B} &= \{6\} \\ \boxed{A \cap B^c} &= \{2, 4\} \\ \boxed{A^c \cap B} &= \{5\} \\ \boxed{A^c \cap B^c} &= \{1, 3\}\end{aligned}$$

Gli eventi che contengono un solo elemento vengono chiamati *eventi singoletti*.

Ed ecco quindi che gli eventi A e B vengono chiamati **eventi generatori** e la loro aggregazione definisce la **famiglia degli eventi generatori**, che in questo caso è:

$$\mathcal{G} := \{A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c\}$$

Definiamo adesso l'**operazione sigma** (denotata con σ) sulla famiglia degli eventi generatori che è la **famiglia degli eventi generata dalla famiglia degli eventi generatori** applicando \cup , \cap e il complemento c in tutti i modi possibili. In questo caso otteniamo:

$$\begin{aligned}\sigma(\mathcal{G}) &= \{\{5\}, \{6\}, \{1, 3\}, \{2, 4\}, \{5, 6\}, \{1, 3, 5\}, \{2, 4, 5\}, \\ &\{1, 3, 6\}, \{2, 4, 6\}, \{1, 2, 3, 4\}, \{1, 3, 5, 6\}, \{2, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5\}, \Omega, \emptyset\}\end{aligned}$$

Un altro esempio di famiglia di eventi generatori dove l'applicazione dell'operazione sigma su di essa corrisponde esattamente alle parti di Ω , cioè $P(\Omega)$:

$$\mathcal{G} := \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$$

$$\sigma(\mathcal{G}) = P(\Omega)$$

Notiamo che comunque prendiamo due elementi dalla famiglia degli eventi generata dalla famiglia degli eventi generatori ed eseguiamo una qualsiasi operazione insiemistica su questi due elementi, il risultato continua ad appartenere alla famiglia.

Definizione 1.4.1 (Algebra)

Sia \mathcal{A} una famiglia di sottoinsiemi di Ω . Si dice che \mathcal{A} è un'algebra se e solo se:

$$\boxed{a_1.} \quad \Omega \in \mathcal{A}$$

$$\boxed{a_2.} \quad A \in \mathcal{A}, \quad A^c \in \mathcal{A}$$

$$\boxed{a_3.} \quad A_1, A_2 \in \mathcal{A}, \quad A_1 \cup A_2 \in \mathcal{A}$$

Proposizione 1.4.1 (L'intersezione su un'algebra è un'operazione chiusa)

Sia \mathcal{A} un'algebra. Siano $A_1, A_2 \in \mathcal{A}$. Allora $A_1 \cap A_2 \in \mathcal{A}$.

Dimostrazione.

Possiamo portare $A_1 \cap A_2$ in questa forma grazie alle leggi di De Morgan:

$$A_1 \cap A_2 = (A_1^c \cup A_2^c)^c$$

$A_1^c \in \mathcal{A}$ per $\boxed{a_2}$. Stessa cosa per A_2^c . Dunque $A_1^c \cup A_2^c \in \mathcal{A}$ per $\boxed{a_3}$. Quindi $(A_1^c \cup A_2^c)^c \in \mathcal{A}$ per $\boxed{a_2}$. □

Dalla Proposizione 1.4.1 possiamo generalizzare e possiamo certamente dire: sia \mathcal{A} un'algebra, sia $n \in \mathbb{N}$: $n \geq 2$ e siano $A_1, A_2, \dots, A_n \in \mathcal{A}$, allora:

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c \in \mathcal{A}$$

Definizione 1.4.2 (Sigma algebra)

Sia \mathcal{F} una famiglia di sottoinsiemi di Ω . Si dice che \mathcal{F} è una **sigma algebra** (o σ algebra) se e solo se:

$$\boxed{\Sigma_1.} \quad \Omega \in \mathcal{F}$$

$$\boxed{\Sigma_2.} \quad A \in \mathcal{F}, \quad A^c \in \mathcal{F}$$

$$\boxed{\Sigma_3.} \quad (A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}, \quad \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F} \quad (\text{proprietà di stabilità rispetto all'unione numerabile})$$

Proposizione 1.4.2 (L'intersezione su una sigma algebra è un'operazione chiusa)

Sia \mathcal{F} una sigma algebra e $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F}$. Allora:

$$\bigcap_{n \in \mathbb{N}} A_n = \left(\bigcup_{n \in \mathbb{N}} A_n^c \right)^c \in \mathcal{F}$$

Esempio 1.4.1

Esperimento aleatorio ξ : si lancia una moneta un numero indefinito di volte. Lo spazio degli esiti è $\Omega = \{T, C\}^\infty$ cioè il prodotto cartesiano $\{T, C\} \times \{T, C\} \times \dots$ infinite volte. L'elemento generico di Ω è una successione ordinata di esiti testa/croce, cioè $\underline{\omega} = (\omega_1, \omega_2, \dots) = (\omega_n)_{n \in \mathbb{N}} \in \Omega$. Il punto campione è dunque una successione.

Consideriamo l'evento A : "due T consecutivi prima di due C consecutive". Esiti che fanno avverare questo evento possono essere:

$$\begin{aligned} A_1 &: T_1 T_2 \\ A_2 &: C_1 T_2 T_3 \\ A_3 &: T_1 C_2 T_3 T_4 \\ A_4 &: C_1 T_2 C_3 T_4 T_5 \\ &\vdots \end{aligned}$$

Dunque $A = A_1 \cup A_2 \cup A_3 \cup A_4 \cup \dots$. Possiamo identificare la terna $(\Omega, \mathcal{F} = \sigma(\mathcal{G}), \mathbb{P})$ dove la famiglia degli eventi generatori è $\mathcal{G} = (T_n)_{n \in \mathbb{N}}$. Ci basta questo perché con il complemento otteniamo tutti gli altri eventi (che sono sempre successioni). Notiamo che l'unione di tutti questi eventi è infinita. Questo è un esempio che mette in luce l'importanza della *proprietà di stabilità rispetto all'unione numerabile* di una sigma algebra (cioè \sum_3).

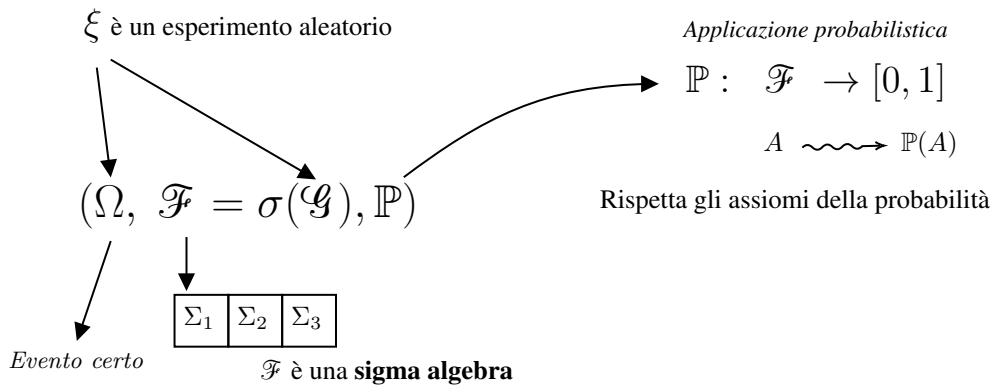


Figura 2: Schema generale terna probabilistica

Assiomi della probabilità

- $M_1.$ $A \in \mathcal{F}, \mathbb{P}(A) \geq 0$
- $M_2.$ $\mathbb{P}(\Omega) = 1$
- $M_3.$ $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{F} : A_i \cap A_j \text{ } i \neq j, \mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$

Teorema 1 (L'insieme vuoto è un evento (appartiene ad \mathcal{F}))

$$\emptyset \in \mathcal{F}$$

Dimostrazione.

\mathcal{F} è una sigma algebra e dunque vale l'assioma $\boxed{\sum_1}$, cioè $\Omega \in \mathcal{F}$. Questo implica, per l'assioma $\boxed{\sum_2}$, che $\Omega^c \in \mathcal{F}$. Ma $\Omega^c = \emptyset \in \mathcal{F}$. Dunque:

$$\text{Per } \boxed{\sum_1} \Omega \in \mathcal{F} \xRightarrow{\boxed{\sum_2}} \Omega^c = \emptyset \in \mathcal{F}$$

□

Teorema 2 (\mathcal{F} è anche un algebra)

Sia $n \in \mathbb{N} : n > 1$ e $A_1, A_2, \dots, A_n \in \mathcal{F}$. Allora, $\bigcup_{i=1}^n A_i \in \mathcal{F}$.

Dimostrazione.

Sia $(B_n)_{n \in \mathbb{N}}$ tale che per ogni $i = 1, \dots, n$ $B_i := A_i \in \mathcal{F}$ e per ogni $i > n$ poniamo $B_i = \emptyset$. Possiamo dire quindi che $(B_n)_{n \in \mathbb{N}}$ è numerabile. Dunque, applicando $\boxed{\sum_3}$ a $(B_n)_{n \in \mathbb{N}}$ possiamo dire che:

$$\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{F} \iff \bigcup_{i=1}^n A_n \in \mathcal{F}$$

□

Teorema 3 (L'intersezione finita di eventi appartiene a \mathcal{F})

Sia $n \in \mathbb{N} : n > 1$ e siano $A_1, A_2, \dots, A_n \in \mathcal{F}$, allora $\bigcap_{i=1}^n A_i \in \mathcal{F}$.

Dimostrazione.

Possiamo scrivere per la legge di De Morgan che:

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c \in \mathcal{F}$$

Per il Teorema 2 l'unione finita di eventi è un evento e per l'assioma $\boxed{\sum_2}$ il complemento di un evento è un evento. □

Teorema 4 (Un evento complementato due volte non cambia)

$$A \in \mathcal{F}, (A^c)^c = A$$

Teorema 5 (Trasformazione di un evento in unione disgiunta di eventi)

Siano $A, B \in \mathcal{F}$, allora :

$$(i) A \cup B = A \dot{\cup} (B \cap A^c)$$

$$(ii) A = (A \cap B) \dot{\cup} (A \cap B^c)$$

Dimostrazione (ii).

$$A = A \cap \Omega = A \cap (B \dot{\cup} B^c) = (A \cap B) \dot{\cup} (A \cap B^c)$$

□

Dimostrazione (i).

$$A \cup B = A \cup (B \cap A \dot{\cup} B \cap A^c) = A \cup (B \cap A^c)$$

Nell'ultimo passaggio non consideriamo $B \cap A$ perché questi elementi già li abbiamo in A e quindi non contribuisce all'unione finale.

□

Teorema 6 (La probabilità dell'insieme vuoto è uguale a 0)

$$\mathbb{P}(\emptyset) = 0$$

Dimostrazione.

Sia $(A_n)_{n \in \mathbb{N}} = \emptyset$. Possiamo dire che $A_i \cap A_j = \emptyset$ con $i \neq j$. Ma allora per l'assioma di probabilità $\boxed{M_3}$ possiamo dire che:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) &\stackrel{\boxed{M_3}}{=} \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) \\ \iff \mathbb{P}(\emptyset) &= \sum_{n \in \mathbb{N}} \mathbb{P}(\emptyset) \\ \iff \mathbb{P}(\emptyset) &= \mathbb{P}(\emptyset) \sum_{n \in \mathbb{N}} 1 \\ \iff \mathbb{P}(\emptyset) &= 0 \end{aligned}$$

□

Teorema 7 (La probabilità dell'unione finita disgiunta di eventi è uguale alla somma delle loro probabilità)

Sia $n \in \mathbb{N}$ e siano $A_1, A_2, \dots, A_n \in \mathcal{F}$ con $A_i \cap A_j = \emptyset$ con $i \neq j$

$$\text{allora } \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i)$$

Dimostrazione.

Sia $(B_n)_{n \in \mathbb{N}}$ tale con per ogni $i = 1, \dots, n$ $B_i := A_i$ e per $i > n$, $B_i := \emptyset$ allora dall'assiamo $\boxed{\sum_3}$ possiamo dire che:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_i\right) &= \sum_{n \in \mathbb{N}} \mathbb{P}(B_i) \\ &= \sum_{i=1}^n \mathbb{P}(B_i) \\ &= \sum_{i=1}^n \mathbb{P}(A_i) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \end{aligned}$$

□

Corollario 1 (La probabilità dell'unione disgiunta di due eventi è uguale alla somma delle loro probabilità)

$$A_1, A_2 \in \mathcal{F} : A_1 \cap A_2 = \emptyset, \mathbb{P}(A_1 \dot{\cup} A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$$

Dimostrazione.

E' il Teorema 7 per $n = 2$.

□

Corollario 2 (La probabilità di un evento complementato è il suo complemento ad uno)

$$A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

Dimostrazione.

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \dot{\cup} A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) \text{ dunque } \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

□

Corollario 3 (La probabilità di un evento è minore o uguale di uno)

$$A \in \mathcal{F}, \mathbb{P}(A) \leq 1$$

Dimostrazione.

$$0 \leq \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

□

Teorema 8 (Un evento A contenuto in un evento B ha probabilità minore o uguale dell'evento B)

$$A, B \in \mathcal{F} : A \subseteq B \text{ allora } \mathbb{P}(A) \leq \mathbb{P}(B)$$

Dimostrazione.

$$B = A \cup B \cap A^c \text{ e quindi } \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \implies \mathbb{P}(B) \geq \mathbb{P}(A)$$

□

Teorema 9 (Principio di inclusione-esclusione per n=2)

$$A, B \in \mathcal{F}, \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Dimostrazione.

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$$

D'altra parte:

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) \implies \mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Concludiamo quindi la dimostrazione sostituendo:

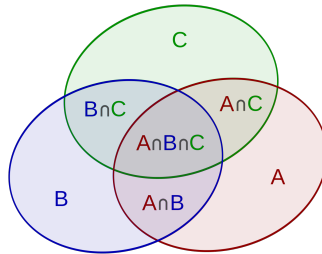
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

□

Teorema 10 (Principio di inclusione-esclusione per n=3)

$A, B, C \in \mathcal{F}$ allora :

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



Dimostrazione.

Per la proprietà associativa dell'intersezione possiamo scrivere: $A \cup B \cup C = (A \cup B) \cup C$

$$\begin{aligned} \mathbb{P}[(A \cup B) \cup C] &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}[(A \cup B) \cap C] = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}[(A \cup B) \cap C] = \\ &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C \cup B \cap C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - [\mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C)] \\ &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - [\mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C)] + \mathbb{P}(A \cap B \cap C). \end{aligned}$$

□

Teorema 11 (Principio di inclusione-esclusione generalizzato)

Consideriamo una famiglia finita di insiemi finiti : A_1, A_2, \dots, A_n .

Per la probabilita' dell'unione di tale famiglia si ha :

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots (-1)^{n-1} \mathbb{P}(A_1 \cap \dots \cap A_n) = \\ &= \sum_{i=1}^n (-1)^{i+1} \sum_{1 \leq j_1 < \dots < j_i \leq n} \mathbb{P}\left(\bigcap_{k=1}^i A_{j_k}\right)\end{aligned}$$

1.5 Problema delle concordanze

Sia n un numero intero e si abbia a disposizione un mazzetto di n cartoncini. Essi vengono numerati da 1 a n sul fronte inizialmente bianco; sul dorso i cartoncini presentano la stessa immagine. L'esperimento ξ consiste nel (i) mischiare i cartoncini, (ii) appoggiare su un tavolo il mazzetto mostrandone il dorso, (iii) mostrare il fronte del cartoncino più in alto "chiamando" il numero "uno" e riporre accanto al mazzetto il cartoncino mostrandone il fronte, (iv) procedere allo stesso modo fino all'ultimo cartoncino. Per $i = 1, 2, \dots, n$, si designa con C_i (concordanza) l'evento che si verifica quando si presenta il cartoncino con numero i alla i -ma chiamata.

1.5.1 Probabilità di osservare almeno una concordanza nelle n chiamate

L'evento "almeno una concordanza nelle n chiamate" risulta essere l'unione di tutte le concordanze: $\bigcup_{i=1}^n C_i$. Ovviamente, nulla vieta che nella stessa effettuazione dell'esperimento si possano verificare due o più concordanze; pertanto, è necessario applicare la formula di inclusione-esclusione:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) &= \sum_{i=1}^n \mathbb{P}(C_i) - \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{P}(C_i \cap C_j) + \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \mathbb{P}(C_i \cap C_j \cap C_k) \\ &\quad + \dots + (-1)^{n-1} \mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_n) \end{aligned} \quad (1)$$

Si osservi che ci sono $n!$ possibili mischiature; risulta:

$$i = 1, 2, \dots, n, \quad \mathbb{P}(C_i) = \frac{(n-1)!}{n!}.$$

Infatti, bloccando solo il cartoncino numerato con i all' i -mo posto dall'alto si verifica C_i ma tutti gli altri possono permutare in tutti i modi possibili. Con lo stesso ragionamento, si ha:

$$i, j = 1, 2, \dots, n : i < j, \quad \mathbb{P}(C_i \cap C_j) = \frac{(n-2)!}{n!}.$$

e:

$$i, j, k = 1, 2, \dots, n : i < j < k, \quad \mathbb{P}(C_i \cap C_j \cap C_k) = \frac{(n-3)!}{n!}.$$

Allo stesso modo per i successivi addendi nella (1). In particolare, dal momento che $(n-n)! = 1$, si ha:

$$\mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_n) = \frac{1}{n!}.$$

La formula (1) diventa:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) &= \frac{(n-1)!}{n!} \sum_{i=1}^n 1 - \frac{(n-2)!}{n!} \sum_{i=1}^n \sum_{j=i+1}^n 1 + \frac{(n-3)!}{n!} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n 1 + \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{(n-1)!}{n!} \binom{n}{1} - \frac{(n-2)!}{n!} \binom{n}{2} + \frac{(n-3)!}{n!} \binom{n}{3} + \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-1} \frac{1}{n!}. \end{aligned} \quad (2)$$

Il passaggio intermedio nella (2) segue da questo ragionamento (valido per il secondo addendo ma facilmente generalizzabile agli altri): il numero degli addendi 1 è uguale al numero delle combinazioni semplici di lunghezza 2 ottenute dall'insieme $S_n = 1, 2, \dots, n$ in quanto il valore dell'indice j è strettamente maggiore del valore dell'indice i .

In conclusione, si ha:

$$\mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = \sum_{i=1}^n (-1)^{i-1} \frac{1}{i!}. \quad (3)$$

1.5.2 Probabilità di osservare zero concordanze nelle n chiamate

Si indichi con $E_{0,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento ξ non si osserva alcuna concordanza (ovvero, 0 concordanze); ovviamente, risulta che:

$$E_{0,n} = \left(\bigcup_{i=1}^n C_i\right)^c \implies \mathbb{P}(E_{0,n}) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n C_i\right). \quad (4)$$

In definitiva, dalla (3) e dalla (4), si ottiene:

$$\mathbb{P}(E_{0,n}) = \sum_{i=0}^n (-1)^i \frac{1}{i!}. \quad (5)$$

Prima di terminare il paragrafo si osservi che, indicato con $N_{0,n}$ il numero delle mischiate che non fanno verificare alcuna concordanza, dalla (5) si ricava:

$$\mathbb{P}(E_{0,n}) = \frac{N_{0,n}}{n!} \implies N_{0,n} = n! \cdot \mathbb{P}(E_{0,n}). \quad (6)$$

1.5.3 Probabilità di ossevare SOLO una concordanza nelle n chiamate

Si indichi con $F_{1,1,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento ξ si osserva esattamente una concordanza e questa avviene alla prima chiamata. Ovviamente, indicato con $M_{1,1,n}$ il numero delle mischiate che fanno verificare solo la concordanza nella prima chiamata e zero concordanze nelle successive, risulta che $M_{1,1,n} = N_{0,n-1}$: infatti, dopo la concordanza alla prima chiamata, dalla seconda alla n -chiamata non si verificano altre concordanze. Allora, in virtù della (6), si ha:

$$\mathbb{P}(F_{1,1,n}) = \frac{M_{1,1,n}}{n!} = \frac{N_{0,n-1}}{n!} = \frac{(n-1)! \cdot \mathbb{P}(E_{0,n-1})}{n!} = \frac{1}{n} \mathbb{P}(E_{0,n-1}). \quad (7)$$

Si osservi che il risultato ottenuto nella (7) non dipende dal fatto che è stato richiesto di osservare la concordanza alla prima chiamata in quanto lo stesso ragionamento è valido per la richiesta di osservare un'unica concordanza nella generica chiamata. In altri termini, indicato con $F_{1,i,n}$ l'evento che si presenta quando in una esecuzione dell'esperimento ξ si osserva esattamente una concordanza e questa avviene alla i -ma chiamata, allora si ha:

$$i = 1, 2, \dots, n, \quad \mathbb{P}(F_{1,i,n}) = \mathbb{P}(F_{1,1,n}) = \frac{1}{n} \mathbb{P}(E_{0,n-1}). \quad (8)$$

Dopo di ciò, se $E_{1,n}$ è l'evento che si presenta quando in una esecuzione dell'esperimento ξ si osserva una sola concordanza, risulta:

$$\mathbb{P}(E_{1,n}) = \sum_{i=1}^n \mathbb{P}(F_{1,i,n}) = \frac{1}{n} \mathbb{P}(E_{0,n-1}) \sum_{i=1}^n 1 = \mathbb{P}(E_{0,n-1}) = \sum_{i=0}^{n-1} (-1)^i \frac{1}{i!}. \quad (9)$$

1.6 Probabilità condizionata

In questa sezione presentiamo e sviluppiamo uno dei concetti fondamentali della teoria della probabilità - quello di probabilità condizionata. L'importanza che ha è duplice. In primo luogo, accade spesso di volere calcolare delle probabilità quando si è in possesso di informazioni parziali sull'esito dell'esperimento, o di volerle ricalcolare una volta ottenute nuove informazioni. Quelle di questo tipo sono probabilità condizionate. Secondariamente vi è una sorta di bonus nel fatto che a volte il modo più semplice di determinare la probabilità di un evento complesso, consiste nel condizionarlo al realizzarsi o meno di un evento accessorio.

Per illustrare questo concetto, immaginiamo di tirare due dadi. Lo spazio degli esiti di questo esperimento può essere descritto da:

$$\Omega = \{(i, j) \mid i = 1, 2, \dots, 6, j = 1, 2, \dots, 6\}$$

dove si intende che si ottiene l'esito (i, j) se il risultato del primo dado è i e quello del secondo è j . Supponiamo che ciascuno dei 36 esiti di Ω abbia la stessa probabilità, ovvero $\frac{1}{36}$ (in queste ipotesi si dice che i due dadi sono onesti). Supponiamo infine che il primo dado sia risultato in un 3. Allora, possedendo questa informazione, qual è la probabilità che la somma dei due dadi valga 8? Dato che il primo dado ha totalizzato un 3, vi sono solo 6 risultati possibili per l'esperimento, che sono $(3, 1), (3, 2), (3, 3), (3, 4), (3, 5)$ e $(3, 6)$. Inoltre, siccome in origine ciascuno di questi esiti aveva la stessa probabilità di realizzarsi, essi dovrebbero essere ancora equiprobabili. Ciò significa che, se il primo dado ha dato un 3, allora la probabilità (condizionata) di ciascuno degli esiti possibili $(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)$ è $\frac{1}{6}$, mentre la probabilità (condizionata) degli altri 30 elementi di Ω è 0. Se ne conclude che la probabilità cercata è $\frac{1}{6}$.

Se denotiamo con E e F rispettivamente l'evento che la somma dei due dadi valga 8 e l'evento che il primo dado risulti in un 3, allora la probabilità che abbiamo appena calcolato si dice **probabilità condizionata di E dato F** , e si denota con:

$$\mathbb{P}(E|F)$$

Con un ragionamento analogo a quello dell'esempio è possibile trovare una formula generale per $\mathbb{P}(E|F)$ valida per qualunque coppia di eventi $E, F \in \mathcal{F}$. Infatti, se si è verificato l'evento F , affinché si verifichi anche E , deve verificarsi un esito che si trova sia in E che in F , cioè che appartiene all'intersezione $E \cap F$. In secondo luogo essendosi verificato F , questo evento diviene il nuovo (ridotto) spazio degli esiti e per questo la probabilità condizionata dell'evento $E \cap F$ sarà pari al rapporto tra la sua probabilità e quella di F . In formula:

$$\mathbb{P}(E|F) := \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Si noti che questa equazione ha senso solo se $\mathbb{P}(F) > 0$ e infatti in caso contrario $\mathbb{P}(E|F)$ non si definisce. La definizione di probabilità condizionata che compare nell'equazione è compatibile con l'interpretazione frequentistica della probabilità degli eventi. Supponiamo di realizzare un numero molto elevato n di ripetizioni di un esperimento. Poiché $\mathbb{P}(F)$ è il limite della frazione di prove in cui si verifica F , su un numero elevato n di tentativi, saranno circa $n\mathbb{P}(F)$ quelli in cui si realizza F . Analogamente saranno approssimativamente $n\mathbb{P}(E \cap F)$ quelli in cui si realizzano sia E che F . Perciò limitatamente agli esperimenti che hanno visto la realizzazione di F , la frazione di quelli per i quali ha avuto luogo anche l'evento E è circa uguale a:

$$\frac{n\mathbb{P}(E \cap F)}{n\mathbb{P}(F)} = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Le approssimazioni fatte divengono esatte quando n tende all'infinito, e quindi l'equazione che definisce la probabilità condizionata è la corretta definizione di probabilità di E qualora vi sia verificato F .

Esempio 1.6.1

Una confezione contiene 5 transistor guasti (non funzionano per niente), 10 difettosi (funzionano correttamente per qualche ora e poi si guastano) e 25 accettabili. Si sceglie un transistor a caso. Qual è la probabilità che sia accettabile se inizialmente funziona?

Sappiamo che non si tratta di uno dei 5 guasti, perché per il momento sta funzionando. Consentendoci un rilassamento nella notazione¹, la quantità cercata si può esprimere come

$$\begin{aligned} \mathbb{P}(\text{accettabile} | \text{non guasto}) &= \frac{\mathbb{P}(\text{accettabile}, \text{non guasto})}{\mathbb{P}(\text{non guasto})} \\ &= \frac{\mathbb{P}(\text{accettabile})}{\mathbb{P}(\text{non guasto})} \end{aligned}$$

¹Sarebbe infatti più corretto scrivere $\mathbb{P}(\{\text{accettabile}\} | \{\text{non guasto}\})$, ma alla lunga esagerare con le parentesi distrae l'attenzione. Si noti anche che la virgola nell'argomento di $\mathbb{P}(\cdot)$ denota l'intersezione degli eventi descritti ai suoi lati. Questo tipo di notazione è assai comune e sarà usata ancora.

dove la seconda uguaglianza segue perché i transistor contemporaneamente accettabili e non guasti sono esattamente quelli accettabili. Assumendo allora che i 40 transistor possano essere scelti con uguale probabilità, si ottiene:

$$\mathbb{P}(\text{accettabile}|\text{non guasto}) = \frac{25/40}{35/40} = \frac{5}{7}$$

E' utile notare che si sarebbe arrivati al medesimo risultato operando direttamente sullo spazio degli esiti ridotto. Infatti, sapendo che il pezzo scelto non è guasto, il problema si riduce a calcolare con che probabilità un transistor scelto da una confezione con 25 pezzi accettabili e 10 difettosi, risulti accettabile. Questa probabilità è ovviamente $25/35$.

Esempio 1.6.2

L'organizzazione per cui lavora il signor Jones organizza una cena tra uomini per i dipendenti e i loro figli. Sono invitati i dipendenti padri di figli maschi, assieme al minore fra i loro figli maschi. Jones ha due figli, ed è invitato alla cena. Qual è la probabilità condizionata che entrambi i suoi figli siano maschi?

Lo spazio degli esiti è $\Omega = \{(m, m), (m, f), (f, m), (f, f)\}$, dove con (m, f) si intende che il figlio maggiore è maschio e la minore è femmina; prima di condizionare, tutti gli esiti sono equiprobabili. L'informazione che Jones è invitato alla cena equivale a sapere che almeno uno dei suoi figli è maschio, quindi che non si è verificato l'evento (f, f) . Denotando con A e B gli eventi "almeno un figlio è maschio" e "entrambi i figli sono maschi", la quantità cercata è $P(B|A)$, ovvero (si noti come, volendo essere precisi, ciascuna parentesi sia necessaria):

$$\begin{aligned}\mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(\{m, m\})}{\mathbb{P}(\{m, m\}, \{m, f\}, \{f, m\})} \\ &= \frac{1/4}{3/4} = \frac{1}{3}\end{aligned}$$

Molte persone pensano erroneamente che la probabilità che entrambi i figli siano maschi sia $1/2$, anziché $1/3$, essendo convinte che il figlio di Jones che non partecipa alla cena abbia la stessa probabilità di essere maschio o femmina. Si rammenti tuttavia che inizialmente i quattro esiti erano equiprobabili, e il sapere che almeno un figlio è maschio equivale ad escludere l'esito (f, f) . Questo ci lascia con tre esiti equiprobabili, mostrando che vi sono il doppio delle possibilità che l'altro figlio di Jones sia femmina piuttosto che maschio. La risposta sarebbe stata $1/2$ ad esempio se avessimo avuto l'informazione che il minore dei figli di Jones è maschio (ci si convinca di questa affermazione, quindi si affronti il Problema 32 - risposta: dobbiamo eliminare anche l'esito (m, f)). Facendo poi i calcoli otteniamo $\frac{1}{2}$.

Se si moltiplicano entrambi i membri dell'equazione della probabilità condizionata per $\mathbb{P}(F)$, si trova quella che viene chiamata **legge delle probabilità composte** o **teorema della probabilità composta**:

$$\mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F)$$

Parafrasandola, questa equazione dice che la probabilità che E e F si verifichino entrambi è pari quella che si verifichi F per la probabilità condizionata di E dato che si è verificato F . Questa formula mostra la sua utilità quando si vuole calcolare la probabilità di una intersezione, come illustra l'esempio seguente.

Esempio 1.6.3

Il signor Perez è convinto che vi sia il 30% di probabilità che la sua azienda apra un nuovo ufficio a Phoenix. Nel caso ciò si verifichi, egli stima di avere un 60% di probabilità di assumere il ruolo dirigenziale nella nuova filiale. Che probabilità vi è che egli divenga il manager del nuovo ufficio di Phoenix?

Se denotiamo con U l'evento "viene aperto un nuovo ufficio a Phoenix" e con M l'evento "Perez viene promosso manager a Phoenix", allora la probabilità cercata è $\mathbb{P}(U \cap M)$, ovvero:

$$\begin{aligned}\mathbb{P}(U \cap M) &= \mathbb{P}(M|U)\mathbb{P}(U) \\ &= 0.6 \times 0.3 = 0.18\end{aligned}$$

Quindi vi è una probabilità del 18% che Perez divenga il manager a Phoenix.

La probabilità condizionata può essere descritta anche con la terna delle probabilità. Dalla terna $(\Omega, \mathcal{F}, \mathbb{P})$ passiamo a $(B, B \cap \mathcal{F}, \mathbb{P}_B)$ e quindi descriviamo la probabilità condizionata di un evento A dato B in questo modo:

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

1.7 Formula delle alternative e formula di Bayes

Siano E e F due eventi qualsiasi. E' possibile esprimere E come:

$$E = (E \cap F) \cup (E \cap F^c)$$

Infatti ogni punto che appartiene all'evento E , o sta sia in E sia in F , oppure sta in E ma non in F . Inoltre, visto che $E \cup F$ e $E \cup F^c$ sono eventi disgiunti, si ha per il Teorema 7 e per il Teorema della probabilità composta:

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) \\ &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)\mathbb{P}(F^c) \\ &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)[1 - \mathbb{P}(F)]\end{aligned}$$

Esempio 1.7.1

Una società di assicurazioni ritiene che la popolazione possa essere divisa in due categorie: quella delle persone inclini a provocare incidenti e quella delle persone non inclini. I rilevamenti statistici effettuati mostrano che una persona incline agli incidenti ha un incidente in un anno con probabilità 0.4, mentre questa probabilità si riduce a 0.2 per l'altra categoria. Assumendo che il 30% della popolazione sia incline agli incidenti, quanto vale la probabilità che un nuovo assicurato abbia un incidente entro un anno dalla stipula del contratto assicurativo?

Otteniamo la probabilità richiesta condizionando alla categoria di appartenenza del nuovo assicurato. Se denotiamo con A_1 l'evento "avrà un incidente entro un anno" e con H l'evento "è incline ad avere incidenti", otteniamo per $\mathbb{P}(A_1)$:

$$\begin{aligned}\mathbb{P}(A_1) &= \mathbb{P}(A_1|H)\mathbb{P}(H) + \mathbb{P}(A_1|H^c)\mathbb{P}(H^c) \\ &= 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26\end{aligned}$$

□

Negli esempi che seguono mostriamo come rivalutare la probabilità dell'evento condizionante (F nella notazione dell'equazione di inizio paragrafo), alla luce di informazioni addizionali (come il verificarsi dell'evento E).

Esempio 1.7.2 (Formula di Thomas Bayes)

Riconsideriamo l'Esempio 1.7.1 e supponiamo che il nuovo assicurato abbia un incidente entro un anno dalla stipula del contratto. Qual è la probabilità che appartenga alla categoria delle persone inclini agli incidenti?

Nell'esempio iniziale assumevamo per un nuovo assicurato una probabilità del 30% che fosse incline ad avere incidenti, quindi, $\mathbb{P}(H) = 0.3$. Tuttavia con la nuova informazione che A_1 si è verificato, possiamo stimare più correttamente questa probabilità, nel modo seguente:

$$\begin{aligned}\mathbb{P}(H|A_1) &= \frac{\mathbb{P}(H \cap A_1)}{\mathbb{P}(A_1)} \\ &= \frac{\mathbb{P}(A_1|H)\mathbb{P}(H)}{\mathbb{P}(A_1)} \quad \text{(Formula di Bayes)} \\ &= \frac{\mathbb{P}(A_1|H)\mathbb{P}(H)}{\mathbb{P}(A_1 \cap H) + \mathbb{P}(A_1 \cap H^c)} \\ &= \frac{\mathbb{P}(A_1|H)\mathbb{P}(H)}{\mathbb{P}(H)\mathbb{P}(A_1|H) + \mathbb{P}(H^c)\mathbb{P}(A_1|H^c)} \\ &= \frac{0.3 \times 0.4}{0.26} = \frac{0.12}{0.26} \approx 0.4615\end{aligned}$$

Esempio 1.7.3

In una prova a risposte multiple, nel rispondere ad una domanda uno studente può conoscere la risposta, oppure provare a indovinarla. Sia p la probabilità che conosca la risposta e $1 - p$ la probabilità che tiri a indovinare. Si

assuma che, se prova ad indovinare, risponda correttamente con probabilità $\frac{1}{m}$, dove m è il numero di alternative nelle scelte multiple. Qual è la probabilità condizionata che egli conoscesse la risposta a una domanda alla quale ha risposto correttamente?

Siano C e K rispettivamente gli eventi "sceglie la risposta giusta" e "conosce la risposta giusta". Per calcolare:

$$\mathbb{P}(K|C) = \frac{\mathbb{P}(K \cap C)}{\mathbb{P}(C)}$$

Notiamo subito che:

$$\mathbb{P}(K \cap C) = \mathbb{P}(C|K)\mathbb{P}(K) = 1 \times p = p$$

Per trovare $\mathbb{P}(C)$, condizioniamo al fatto che sapesse la risposta o meno.

$$\begin{aligned}\mathbb{P}(C) &= \mathbb{P}(C|K)\mathbb{P}(K) + \mathbb{P}(C|K^c)\mathbb{P}(K^c) \\ &= p + (1/m)(1-p)\end{aligned}$$

Quindi la quantità richiesta è:

$$\mathbb{P}(K|C) = \frac{p}{p + (1/m)(1-p)} = \frac{mp}{1 + (m-1)p}$$

Così ad esempio, se $p=1/2$ e $m=5$, la probabilità che lo studente conoscesse la risposta, considerato il fatto che ha risposto correttamente è $5/6$.

□

Esempio 1.7.4

Una particolare analisi del sangue è efficace al 99% nell'individuare una certa malattia quando essa è presente. Si possono però anche verificare dei "falsi positivi" con probabilità dell'1% (ovvero una persona sana che si sottoponga al test, con una probabilità di 0.01 risulta erroneamente affetta dalla malattia in questione). Se l'incidenza di questo male sulla popolazione è dello 0.5%, qual è la probabilità che un soggetto sia malato, condizionata al fatto che le analisi abbiano dato esito positivo?

Sia M l'evento "il soggetto è malato" ed E l'evento "il risultato dell'analisi è positivo". Allora $\mathbb{P}(M|E)$ si trova tramite:

$$\begin{aligned}\mathbb{P}(M|E) &= \frac{\mathbb{P}(M \cap E)}{\mathbb{P}(E)} \\ &= \frac{\mathbb{P}(E|M)\mathbb{P}(M)}{\mathbb{P}(E|M)\mathbb{P}(M) + \mathbb{P}(E|M^c)\mathbb{P}(M^c)} \\ &= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \approx 0.3322\end{aligned}$$

Perciò solo il 33% delle persone che risultano positive alle analisi sono realmente affette dalla malattia. Siccome molti studenti si stupiscono di questo risultato (infatti le caratteristiche del test sembrano buone e ci si aspetterebbe un valore più elevato), vale forse la pena di presentare una seconda argomentazione che anche se meno rigorosa può aiutare a chiarirsi le idee.

Se lo $0.5\% = 1/200$ della popolazione soffre di questo male, in media su 200 persone vi sarà un solo malato. Se egli si sottopone alle analisi, verrà trovato positivo quasi certamente (con probabilità 0.99), così anche su 200 individui testati ve ne saranno in media 0.99 che saranno correttamente individuati come malati. D'altro canto le (in media) 199 persone sane hanno una probabilità di 0.01 di risultare positive, e quindi in media su 200 analisi vi saranno $199 \times 0.01 = 1.99$ falsi positivi. Se consideriamo che ogni 0.99 positivi veri vi sono in media 1.99 positivi falsi, ricaviamo nuovamente che la frazione di malati reali tra i soggetti positivi alle analisi è di:

$$\frac{0.99}{0.99 + 1.99} \approx 0.3322$$

Esempio 1.7.5

Ad un certo stadio delle indagini su un crimine, l'investigatore capo è convinto al 60% della colpevolezza di un certo sospetto. Supponiamo che si scopra un nuovo indizio che mostra che il colpevole deve possedere una certa caratteristica distintiva (come ad esempio essere mancino, calvo, o avere i capelli castani); inoltre anche il sospettato la possiede. Se tale particolarità interessa il 20% della popolazione, quanto sicuro deve essere l'investigatore della colpevolezza del sospettato?

Denotiamo con G e con C i due eventi "il sospetto è colpevole" e "il sospetto possiede il tratto distintivo del colpevole". Abbiamo:

$$\mathbb{P}(G|C) = \frac{\mathbb{P}(G \cap C)}{\mathbb{P}(C)}$$

dove:

$$\mathbb{P}(G \cap C) = \mathbb{P}(C|G)\mathbb{P}(G) = 1 \times 0.6 = 0.6$$

e dove la probabilità di C si trova condizionando alla colpevolezza o meno del sospetto, nel modo seguente:

$$\begin{aligned}\mathbb{P}(C) &= \mathbb{P}(C|G)\mathbb{P}(G) + \mathbb{P}(C|G^c)\mathbb{P}(G^c) \\ &= 1 \times 0.6 + 0.2 \times 0.4 = 0.68\end{aligned}$$

Qui abbiamo stabilito che la probabilità che il sospetto abbia la caratteristica rilevante se non è colpevole sia quella generale della popolazione, 0.2. Concludendo:

$$\mathbb{P}(G|C) = 0.6/0.68 \approx 0.882$$

e l'ispettore dovrebbe alzare all'88% la sua confidenza sulla colpevolezza del sospetto.

Cosa fare invece se l'indizio rinvenuto non è univoco? Supponiamo ad esempio che esso dica che non è certo, ma vi è il 90% di probabilità che il colpevole possieda questa caratteristica. Come si modifica la risoluzione per tenere conto di questa complicazione?

In questo caso, la probabilità che il sospetto possieda la caratteristica rilevante, supponendo che sia colpevole è di 0.9, mentre prima era pari a 1. Allora,

$$\begin{aligned}\mathbb{P}(C|G) &= \frac{\mathbb{P}(G \cap C)}{\mathbb{P}(C)} \\ &= \frac{\mathbb{P}(C|G)\mathbb{P}(G)}{\mathbb{P}(C|G)\mathbb{P}(G) + \mathbb{P}(C|G^c)\mathbb{P}(G^c)} \\ &= \frac{0.9 \times 0.6}{0.9 \times 0.6 + 0.2 \times 0.4} = \frac{0.54}{0.62} \approx 0.871\end{aligned}$$

che è un valore un po' più inferiore a quello ottenuto precedentemente.

1.7.1 Sistema completo di alternative e formula delle alternative generalizzata

Sia $(H_n)_{n \in \mathbb{N}}$ un insieme finito o numerabile di eventi. Se tale successione di eventi rispetta:

$$(H_n)_{n \in \mathbb{N}}, \quad \mathbb{P}(H_n) > 0 \tag{1}$$

$$i \neq j \in \mathbb{N}, \quad H_i \cap H_j = \emptyset \tag{2}$$

$$\bigcup_{n \in \mathbb{N}} H_n = \Omega \tag{3}$$

allora la successione costituisce un **sistema completo di alternative**. La proprietà (2) si cita dicendo che gli eventi H_i *ricoprono* Ω e significa che si verifica sempre almeno uno di essi (esattamente uno, se - come nel nostro caso - sono anche disgiunti).

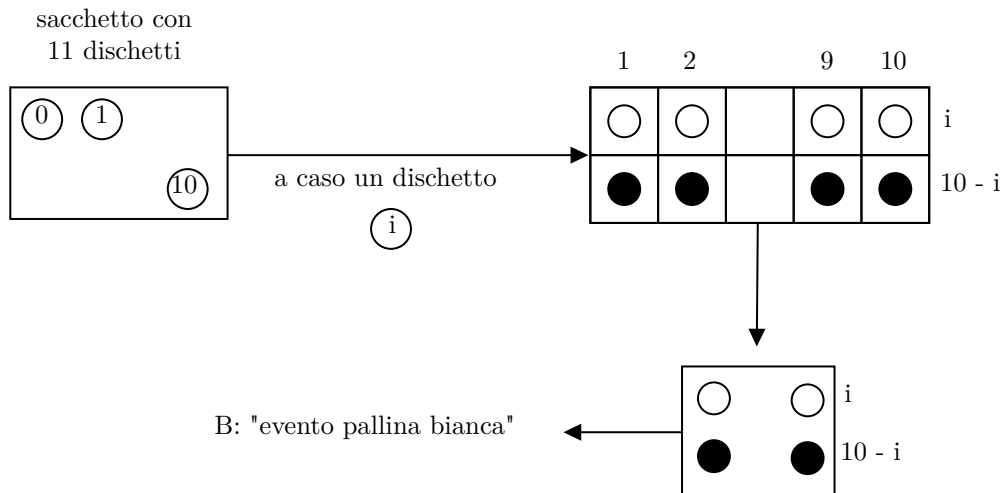
Consideriamo adesso un generico evento $A \in \mathcal{F}$. Possiamo scrivere la probabilità dell'evento A in questo modo:

$$\begin{aligned}
 \mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) \\
 (3) \quad &= \mathbb{P}\left(A \cap \bigcup_{n \in \mathbb{N}} H_n\right) \\
 (\text{proprietà distributiva}) \quad &= \mathbb{P}\left[\bigcup_{n \in \mathbb{N}} (A \cap H_n)\right] \\
 (2) \text{ e } \boxed{M_3} \quad &= \sum_{n \in \mathbb{N}} \mathbb{P}(A \cap H_n) \\
 (\text{teorema della probabilità composta}) \quad &= \sum_{n \in \mathbb{N}} \mathbb{P}(H_n) \cdot \mathbb{P}(A|H_n)
 \end{aligned}$$

La formula così ottenuta viene chiamata **Formula delle alternative** per un sistema completo finito o numerabile.

Esempio 1.7.6

Consideriamo un sacchetto con 11 dischetti numerati da 0 a 10. La prima parte dell'esperimento consiste nel prendere dal sacchetto in maniera casuale un dischetto i dove $i = 0, 1, \dots, 10$. Dopo di ciò, una volta ottenuta questa informazione, andiamo a prendere i palline bianche da un'armadietto che contiene 10 palline bianche e 10 palline nere e le andiamo a mettere in una scatola. Vogliamo che nella scatola ci siano esattamente 10 palline. Per arrivare a questo risultato, dopo aver aggiunto le palline bianche, aggiungiamo anche $10 - i$ palline nere nella scatola. Vogliamo determinare la probabilità di pescare una pallina bianca dalla scatola finale. Chiamiamo l'evento B : "evento pallina bianca".



Non si può determinare la probabilità dell'evento B se prima non abbiamo pescato un dischetto. Possiamo usare la formula delle alternative. Le alternative in questo caso ne sono 11 (perché 11 ne sono i dischetti). Chiamiamo con H_i l'evento "esce il dischetto i ".

$$\begin{aligned}
 \mathbb{P}(B) &= \sum_{i=0}^{10} \mathbb{P}(H_i) \cdot \mathbb{P}(B|H_i) \\
 &= \sum_{i=0}^{10} \frac{1}{11} \cdot \frac{i}{10} \\
 &= \frac{1}{110} \sum_{i=0}^{10} i \\
 &= \frac{1}{110} \cdot \frac{10 \cdot 11}{2} \\
 &= \frac{55}{110} = \frac{1}{2}
 \end{aligned}$$

Il risultato è $\frac{1}{2}$. Significa che questo esperimento non favorisce né la pallina bianca né la pallina nera.

Per permettere all'esperimento di favorire la pallina bianca bisogna:

1. Favorire in qualche modo la probabilità di pescare dischetti con un numero maggiore di 5. In questo modo perdiamo però l'equiprobabilità.
2. Eliminare almeno un dischetto con numero minore di 5.

Sia $A \in \mathcal{F}$ un evento generico e siano $E, E^c \in \mathcal{F}$ due eventi mutuamente esclusivi che formano un *sistema completo di alternative*. Allora possiamo esprimere A come segue:

$$\mathbb{P}(A) = \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)\mathbb{P}(E^c)$$

Questa equazione è proprio la **formule delle alternative** (ma con solo *due alternative*). Possiamo esprimere la probabilità dell'evento E dato A come:

$$\mathbb{P}(E|A) = \frac{\mathbb{P}(A|E)\mathbb{P}(E)}{\mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)\mathbb{P}(E^c)}$$

Possiamo **generalizzare la formula delle alternative**. Sia $(H_n)_{n \in \mathbb{N}}$ un successione di eventi che forma un *sistema completo di alternative* e sia $A \in \mathcal{F}$. Possiamo esprimere l'evento A come:

$$\mathbb{P}(A) = \sum_{n \in \mathbb{N}} \mathbb{P}(A|H_n)\mathbb{P}(H_n)$$

e per ogni $i = 1, 2, \dots, n$ possiamo esprimere H_i come:

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(A|H_i)\mathbb{P}(H_i)}{\sum_{n \in \mathbb{N}} \mathbb{P}(A|H_n)\mathbb{P}(H_n)}$$

Esempio 1.7.7

Due compagni E_1 ed E_2 scommettono un'aperitivo giocando a carte. E_2 mischia le carte e dà una carta ad E_1 e una carta a sé stesso: vince chi ha la carta più alta (in caso di parità si ripete).

Denotiamo con O l'evento *onesto* mentre con B l'evento *baro*. La **probabilità a priori** che E_2 sia onesto è $\mathbb{P}(O) = 0.90$ quindi $\mathbb{P}(B) = 0.10$.

Vogliamo determinare di quanto bisogna aumentare la probabilità che E_2 stia barando nel caso vincessse il 1° aperitivo (denotiamo V_1 l'evento " E_2 ha vinto il 1° aperitivo").

$$\mathbb{P}(B|V_1) = \frac{\mathbb{P}(V_1|B)\mathbb{P}(B)}{\mathbb{P}(V_1|B)\mathbb{P}(B) + \mathbb{P}(V_1|O)\mathbb{P}(O)}$$

E' chiaro che $\mathbb{P}(V_1|B) = 1$ mentre la probabilità $\mathbb{P}(V_1|O) = \frac{1}{2}$. Dunque, sostituendo:

$$\mathbb{P}(B|V_1) = \frac{1 \cdot 0.10}{1 \cdot 0.10 + \frac{1}{2} \cdot 0.90} = \frac{2}{11} = 0.18$$

Vogliamo adesso determinare di quanto bisogna aumentare la probabilità che E_2 stia barando nel caso vincessse il 1° e il 2° aperitivo (denotiamo questo evento come V_2) partendo sempre dalle stesse probabilità a priori:

$$\mathbb{P}(B|V_2) = \frac{\mathbb{P}(V_2|B)\mathbb{P}(B)}{\mathbb{P}(V_2|B)\mathbb{P}(B) + \mathbb{P}(V_2|O)\mathbb{P}(O)}$$

La probabilità $\mathbb{P}(V_2|B) = 1$ rimane la stessa, mentre la probabilità $\mathbb{P}(V_2|O) = (\frac{1}{2})^2 = \frac{1}{4}$ (capiremo meglio il perché quando parleremo di *eventi indipendenti*).

Dunque, sostituendo:

$$\mathbb{P}(B|V_2) = \frac{\frac{1}{10}}{\frac{1}{10} \cdot \frac{1}{4} + \frac{9}{10}} = \frac{4}{13} = 0.31$$

Le probabilità ottenute in base alle probabilità a priori sono dette **probabilità a posteriori**.

Determiniamo adesso la probabilità che E_2 stia barando dopo che ha vinto il 6° aperitivo di fila (denotiamo questo evento con V_6) mantenendo le stesse probabilità a priori.

$$\mathbb{P}(B|V_6) = \frac{\mathbb{P}(V_6|B)\mathbb{P}(B)}{\mathbb{P}(V_6|B)\mathbb{P}(B) + \mathbb{P}(V_6|O)\mathbb{P}(O)}$$

La probabilità $\mathbb{P}(V_6|B) = 1$ rimane uguale mentre la probabilità $\mathbb{P}(V_6|O) = (\frac{1}{2})^6 = \frac{1}{2^6}$. Dunque, sostituendo:

$$\mathbb{P}(B|V_6) = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{1}{2^6} \cdot \frac{9}{10}} = \frac{64}{73} \approx 0.88$$

Questo modo di procedere (cioè quello di cambiare l'evento V ma rinomare uguali le probabilità di O e B) viene detto **inferenza diretta**.

Procediamo adesso in maniera leggermente differente ma che comunque ci porta allo stesso risultato. Procediamo con il metodo dell'**inferenza iterativa** che consiste di cambiare ad ogni passo le probabilità a priori con le probabilità a posteriori calcolate nel passo precedente.

Ad esempio, se nel 1° passo abbiamo ottenuto queste probabilità posteriori, cioè $\mathbb{P}(B|V_1) = \frac{2}{11}$ e $\mathbb{P}(O|V_1) = \frac{9}{11}$, queste diventeranno le nuove probabilità a priori nel 2° passo, quindi (chiamando adesso l'evento della vincita V e non più V_i):

$$\mathbb{P}(B|V) = \frac{\frac{2}{11}}{\frac{2}{11} + \frac{1}{2} \cdot \frac{9}{11}} = \frac{4}{13}$$

Notiamo che non fa differenza se usiamo il metodo dell'inferenza diretta o il metodo dell'inferenza iterativa, il risultato rimane uguale (infatti questo risultato coincide con quello di prima).

Esempio 1.7.8

Un aereo è scomparso, e si suppone che possa essere caduto in una qualsiasi di tre regioni, con uguale probabilità. Per $i = 1, 2, 3$, sia $1 - \alpha_i$ la probabilità di rintracciare il velivolo caduto nella regione i -esima (le costanti α_i rappresentano la probabilità di non rinvenire il velivolo; sono normalmente dovute alle condizioni geografiche e ambientali delle regioni). Qual'è la probabilità che l'aereo si trovi in ciascuna delle tre regioni se una ricerca della regione 1 ha dato esito negativo?

Per $i = 1, 2, 3$, denotiamo con R_i l'evento "*il velivolo si trova nella regione i -esima*"; sia E l'evento "*la ricerca nella regione 1 non ha successo*". Dalla formula di Bayes otteniamo per R_1 :

$$\begin{aligned} \mathbb{P}(R_1|E) &= \frac{\mathbb{P}(E|R_1)\mathbb{P}(R_1)}{\sum_{i=1}^3 \mathbb{P}(E|R_i)\mathbb{P}(R_i)} \\ &= \frac{\frac{\alpha_1}{3}}{\frac{\alpha_1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{\alpha_1}{\alpha_1 + 3} \end{aligned}$$

mentre per $j = 2, 3$ abbiamo:

$$\begin{aligned} \mathbb{P}(R_j|E) &= \frac{\mathbb{P}(E|R_j)\mathbb{P}(R_j)}{\sum_{i=1}^3 \mathbb{P}(E|R_i)\mathbb{P}(R_i)} \\ &= \frac{\frac{1}{3}}{\frac{\alpha_1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{1}{\alpha_1 + 2} \end{aligned}$$

Quindi se ad esempio fosse $\alpha_1 = 0.4$, la probabilità che il velivolo sia nella prima regione nonostante cercandolo lì non sia stato trovato sarebbe di $\frac{1}{6}$.

1.7.2 Eventi indipendenti

Siano $E, F \in \mathcal{F}$ due eventi. La probabilità di E condizionata ad F è generalmente diversa dalla probabilità non condizionata $\mathbb{P}(E)$: sapere che l'evento F si è verificato, modifica di solito la probabilità che si sia verificato E . Nel caso particolare in cui invece $\mathbb{P}(E|F)$ e $\mathbb{P}(E)$ siano uguali, diciamo che E è **indipendente da F** . Quindi E è indipendente da F se la conoscenza che F si è avverato non cambia la probabilità di E . Se vale questo allora possiamo scrivere:

$$\mathbb{P}(E|F) = \mathbb{P}(E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Quindi possiamo scrivere (la probabilità dell'intersezione si fattorizza nel prodotto di A e B):

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$$

e vale anche:

$$\mathbb{P}(E) > 0 \quad \mathbb{P}(E \cap F) = \mathbb{P}(F|E)\mathbb{P}(E) = \mathbb{P}(F)\mathbb{P}(E)$$

Definizione 1.7.1 (Due eventi indipendenti - Formula di fattorizzazione)

Siano $A, B \in \mathcal{F}$ due eventi. A e B sono indipendenti se e soltanto se vale questa equazione:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Da questa definizione si capisce che **l'indipendenza degli eventi riguarda la misura della probabilità**.

Proposizione 1.7.1 (Se A e B sono indipendenti, lo sono anche A e B^c)

Siano $A, B \in \mathcal{F}$ due eventi indipendenti.

Allora lo sono anche A e B^c .

Dimostrazione.

L'evento A può essere scritto come $A = (A \cap B) \dot{\cup} (A \cap B^c)$. Quindi, essendo l'unione di eventi disgiunti, possiamo scrivere:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \\ (\text{A e B indipendenti}) \quad &= \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A \cap B^c) \end{aligned}$$

E quindi possiamo scrivere:

$$\begin{aligned} \mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)[1 - \mathbb{P}(B)] \\ &= \mathbb{P}(A)\mathbb{P}(B^c). \end{aligned}$$

□

Esempio 1.7.9 (Esempio indipendenza eventi zara)

Consideriamo il lancio di due dadi onesti.

L'evento $Z_2 = 7$ e l'evento $\{Z_1^{(1)} = 4\} \cap \{Z_1^{(2)} = 3\}$ NON sono indipendenti. Infatti:

$$\mathbb{P}(Z_2 = 7 | \{Z_1^{(1)} = 4\} \cap \{Z_1^{(2)} = 3\}) = 1$$

L'evento $Z_2 = 7$ e l'evento $Z_1^{(1)} = 4$ sono indipendenti. Infatti:

$$\begin{aligned} \mathbb{P}(Z_2 = 7 | Z_1^{(1)} = 4) &= \frac{1}{6} \\ \mathbb{P}(Z_2 = 7) &= \frac{6}{36} = \frac{1}{6} \end{aligned}$$

L'evento $Z_2 = 7$ e l'evento $Z_1^{(2)} = 3$ sono indipendenti. Infatti:

$$\begin{aligned} \mathbb{P}(Z_2 = 7 | Z_1^{(2)} = 3) &= \frac{1}{6} \\ \mathbb{P}(Z_2 = 7) &= \frac{6}{36} = \frac{1}{6} \end{aligned}$$

Svolgere adesso il problema 36.

Definizione 1.7.2 (Tre eventi indipendenti)

Siano $A, B, C \in \mathcal{F}$ tre eventi. A, B, C sono indipendenti tra loro se valgono queste quattro equazioni contemporaneamente:

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) \\ \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C) \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B) \cdot \mathbb{P}(C)\end{aligned}$$

Proposizione 1.7.2

Siano $A, B, C \in \mathcal{F}$ tre eventi indipendenti.

Allora $\mathbb{P}[A \cap (B \cup C)] = \mathbb{P}(A) \cdot \mathbb{P}(B \cup C)$

Dimostrazione.

$$\begin{aligned}(\text{proprietà distributiva}) \quad \mathbb{P}[A \cap (B \cup C)] &= \mathbb{P}[(A \cap B) \cup (A \cap C)] \\ (\text{formula inclusione-esclusione}) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) - \mathbb{P}(A \cap B \cap C) \\ (\text{A, B e C indipendenti tra loro}) &= \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B \cap C) \\ &= \mathbb{P}(A)[\mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C)] \\ (\text{formula inclusione-esclusione inversa}) &= \mathbb{P}(A)\mathbb{P}(B \cup C).\end{aligned}$$

□

Definizione 1.7.3 (Indipendenza numero finito arbitrario di eventi)

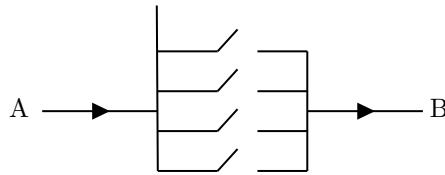
Gli eventi E_1, E_2, \dots, E_n con $n \geq 2$ sono indipendenti se per ogni loro sottogruppo $E_{\alpha_1}, E_{\alpha_2}, \dots, E_{\alpha_r}$ con $1 \leq \alpha_1 < \dots < \alpha_r \leq n$, vale l'equazione:

$$\mathbb{P}\left(\bigcap_{i=1}^r E_{\alpha_i}\right) = \prod_{i=1}^r \mathbb{P}(E_{\alpha_i})$$

In altre parole, $n \geq 2$ eventi sono indipendenti tra loro se *comunque* si sceglie un sottoinsieme di essi, per tali eventi vale la formula di fattorizzazione.

Esempio 1.7.10 (Cicuito parallelo)

Consideriamo un circuito elettrico collegato in parallelo. Abbiamo quattro componenti. Denotiamo con E_i



l'evento: "*il componente i funziona*" e sia $\mathbb{P}(E_i) = p_i$ con $i = 1, 2, 3, 4$. I quattro eventi sono *indipendenti* tra di loro. Vogliamo calcolare la probabilità che il sistema funzioni correttamente.

Innanzitutto notiamo che:

$$\mathbb{P}(\text{il sistema funziona}) = 1 - \mathbb{P}(\text{il sistema non funziona})$$

La frase *il sistema funziona* è chiaramente l'unione di tutti gli E_i . Possiamo esprimere l'unione in questa maniera e risolvere il problema:

$$\begin{aligned}
 (\text{Teorema di De Morgan}) \quad \mathbb{P}\left(\bigcup_{i=1}^4 E_i\right) &= \mathbb{P}\left[\left(\bigcap_{i=1}^4 E_i^c\right)^c\right] \\
 &= 1 - \mathbb{P}\left(\bigcap_{i=1}^4 E_i^c\right) \\
 &= 1 - \prod_{i=1}^4 \mathbb{P}(E_i^c) \\
 &= 1 - \prod_{i=1}^4 (1 - p_i).
 \end{aligned}$$

Variabili aleatorie e valore atteso

2 Variabili aleatorie e valore atteso

2.1 Variabili aleatorie

Quando si realizza un esperimento casuale, non sempre si è interessati in ugual modo a tutte le informazioni ricavabili dal suo esito. Spesso si può individuare una singola quantità numerica (ricavabile dall'esito stesso) che racchiude tutto ciò che in realtà vogliamo sapere. Se tiriamo due dadi, ad esempio, può accadere che ci interessi solamente il valore della loro somma, e non ciascuno dei punteggi. Potremmo volere registrare che il totale realizzato è 7, senza dare importanza a quale sia l'esito vero e proprio dell'esperimento, tra i sei possibili, che sono (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) e (6, 1).

Quantità di interesse che, come queste, sono determinate dal risultato di un esperimento casuale sono dette *variabili aleatorie*. Siccome il valore di una variabile aleatoria è determinato dall'esito dell'esperimento, possiamo assegnare delle probabilità ai suoi valori possibili.

Esempio 2.1.1

Si tirano due dadi indipendenti e non truccati, e si denota con la lettera X la *variabile aleatoria definita dalla loro somma*. Ha senso domandarsi quanto vale la probabilità che $X = 3$, ovvero che la probabilità dell'evento $\{s \in \Omega \mid X(s) = 3\}$ dove $\Omega = \{(i, j, k) \mid i, j, k = 1, 2, \dots, 6\}$ è lo spazio degli esiti. Vi sono due elementi dello spazio degli esiti di questo esperimento che danno ad X il valore 3. Essi sono (1, 2) e (2, 1). Perciò, con una notazione più leggera:

$$\{X = 3\} = \{s \in \Omega \mid X(s) = 3\} = \{(1, 2), (2, 1)\}$$

e di conseguenza la probabilità che $X = 3$ è pari a $2/36$ perché abbiamo a che fare con *esiti equiprobabili*. Il modo corretto di scrivere questo risultato sarebbe $P(\{X = 3\}) = 2/36$, ma è invalso l'uso di scrivere, con leggero abuso di notazione $P(X = 3) = 2/36$. Notiamo che ogni evento A appartiene alle parti dello spazio degli eventi, cioè ogni evento A è un sottoinsieme dello spazio degli eventi Ω , dunque $A \in P(\Omega)$. Ricorrendo a questa convenzione elenchiamo le probabilità per tutti i valori possibili di X .

$$\begin{aligned}P(X = 2) &= P\{(1, 1)\} = 1/36 \\P(X = 3) &= P\{(1, 2), (2, 1)\} = 2/36 \\P(X = 4) &= P\{(2, 2), (1, 3), (3, 1)\} = 3/36 \\P(X = 5) &= P\{(4, 1), (1, 4), (3, 2), (2, 3)\} = 4/36 \\P(X = 6) &= P\{(5, 1), (1, 5), (3, 3), (4, 2), (2, 4)\} = 5/36 \\P(X = 7) &= P\{(6, 1), (1, 6), (5, 2), (2, 5), (4, 3), (3, 4)\} = 6/36 \\P(X = 8) &= P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = 5/36 \\P(X = 9) &= P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = 4/36 \\P(X = 10) &= P\{(4, 6), (5, 5), (6, 4)\} = 3/36 \\P(X = 11) &= P\{(5, 6), (6, 5)\} = 2/36 \\P(X = 12) &= P\{(6, 6)\} = 1/36\end{aligned}$$

La variabile aleatoria può assumere tutti i valori interi che vanno da 2 a 12 (insieme chiamato *spettro della trasformazione* - trasformazione che ad ogni coppia associa la sua somma), con probabilità specificate sopra.

Siccome X deve assumere uno di questi valori, ne segue che $\Omega = \bigcup_{i=2}^{12} \{X = i\}$, e di conseguenza:

$$1 = P(\Omega) = P\left(\bigcup_{i=2}^{12} \{X = i\}\right) = \sum_{i=2}^{12} P(X = i)$$

come si verifica facilmente dalle equazioni sopra.

Un'altra variabile aleatoria di possibile interesse all'interno di questo esperimento è il valore del primo dado. La denotiamo con Y e notiamo che:

$$P(Y = i) = 1/6, \quad i = 1, 2, 3, 4, 5, 6$$

Ovvero Y può assumere ciascuno dei valori interi da 1 a 6 con la stessa probabilità.

Esempio 2.1.2

Un tizio acquista due componenti elettronici, ciascuno dei quali può essere *accettabile* o *difettoso*. Supponiamo che le probabilità dei 4 esiti possibili - (d,d), (d,a), (a,d), (a,a) - siano rispettivamente 0.09, 0.21, 0.21 e 0.49.

Sia X il numero di componenti accettabili; allora X è una variabile aleatoria che può assumere i valori 0, 1 o 2 con probabilità:

$$\begin{aligned}P(X = 0) &= 0.09 \\P(X = 1) &= 0.42 \\P(X = 2) &= 0.49\end{aligned}$$

Se vogliamo limitarci a registrare se vi sia almeno un componente accettabile, possiamo definire una variabile aleatoria I come segue,

$$I := \begin{cases} 1 & \text{se } X = 1 \text{ o } 2 \\ 0 & \text{se } X = 0 \end{cases}$$

Se con A si denota l'evento che vi sia almeno un componente accettabile, allora I è detta la *funzione indicatrice* dell'evento A , infatti I assume i valori 1 o 0 a seconda se l'evento A si verifica o meno. Le probabilità corrispondenti ai valori possibili di I sono:

$$\begin{aligned}P(I = 1) &= 0.91 \\P(I = 0) &= 0.09\end{aligned}$$

Esempio 2.1.3 (Estensione al caso di tre dadi)

Si consideri l'esperimento ξ_3 che consiste nel lanciare tre dadi onesti di colore differente in maniera tale che si possa parlare di primo, secondo e terzo dado. Lo spazio campionario, ovvero l'insieme dei possibili risultati di ξ_3 , è:

$$\Omega_3 := \{(i, j, k) : i, j, k = 1, 2, 3, \dots, 6\}.$$

Ne risulta che $|\Omega_3| = 6^3$, ovvero che in Ω_3 ci sono 216 coppie ordinate ognuna delle quali è un punto campionario. D'altra parte, l'ipotesi dell'onestà dei tre dadi porta a ritenere che:

$$i, j, k = 1, 2, \dots, 6 \quad \mathbb{P}_3((i, j, k)) = \frac{1}{6^3} = \frac{1}{216}$$

ovvero che è possibile utilizzare la definizione di probabilità introdotta da Laplace. Allora, considerando evento ogni sottoinsieme di Ω_3 , si ha:

$$A \in P(\Omega_3), \quad \mathbb{P}_3(A) = \frac{|A|}{216}.$$

In definitiva, la terna $(\Omega_3, P(\Omega_3), \mathbb{P}_3)$ descrive completamente e in maniera formale l'esperimento ξ_3 .

In questo contesto, ad esempio, ha senso considerare la somma dei punteggi dei tre dadi:

$$Z_3 : (i, j, k) \in \Omega_3 \longrightarrow (i + j + k) \in \mathbb{R}$$

La funzione Z_3 assume con probabilità positiva i seguenti valori: 3, 4, ..., 17, 18. Per ognuno di tali valori è possibile determinare agevolmente la sua controimmagine tramite la funzione Z_3 . A tale scopo, per sintetizzare le formule, è utile individuare le classi di terne compatibili con la somma in questione e che sono diverse per almeno un punteggio con le rispettive dimensioni. Ad esempio, per $Z_3 = 9$ ci sono le seguenti 6 classi terne:

$$[(1, 2, 6); \textcolor{red}{6}], [(1, 3, 5); \textcolor{red}{6}], [(1, 4, 4); \textcolor{red}{3}], [(2, 2, 5); \textcolor{red}{3}], [(2, 3, 4); \textcolor{red}{6}], [(3, 3, 3); \textcolor{red}{1}].^2$$

Dopo di ciò, si ha:

$$\mathbb{P}_3(Z_3) = \frac{\textcolor{red}{6} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{3} + \textcolor{red}{6} + \textcolor{red}{1}}{6^3} = \frac{25}{216}$$

Allo stesso modo, in $\{Z_3 = 10\}$ ci sono le seguenti 6 classi terne:

$$[(1, 3, 6); \textcolor{red}{6}], [(1, 4, 5); \textcolor{red}{6}], [(2, 2, 6); \textcolor{red}{3}], [(2, 3, 5); \textcolor{red}{6}], [(2, 4, 4); \textcolor{red}{3}], [(3, 3, 4); \textcolor{red}{3}]$$

e, pertanto, risulta:

$$\mathbb{P}_3(Z_3 = 10) = \frac{\textcolor{red}{6} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{6} + \textcolor{red}{3} + \textcolor{red}{3}}{6^3} = \frac{27}{216}$$

Queste formule forniscono la soluzione al problema posto dai gentiluomini fiorentini a Galileo Galilei: in totale, ci sono 25 terne di punteggi che fanno realizzare il 9 e 27 terne di punteggi che fanno realizzare il 10.

Notiamo che:

$$\sum_{s=3}^{18} \mathbb{P}_3(Z_3 = s) = 1$$

²La scrittura $[(1, 2, 6); \textcolor{red}{6}]$ indica le sei terne che si ottengono permutando i punteggi 1, 2 e 6 in tutti i modi possibili; analogamente, la scrittura $[(1, 4, 4); \textcolor{red}{3}]$ indica le tre terne che si ottengono permutando i punteggi 1, 4 e 4 in tutti i modi possibili.

Osservazione 2.1.1

Il numero aleatorio Z_2 può essere considerato anche nello spazio di probabilità $(\Omega_3, P(\Omega_3), \mathbb{P}_3)$:

$$Z_2 : (i, j, k) \in \Omega \longrightarrow (i + j + 0 \cdot k) \in \mathbb{R}$$

Ovviamente, questo comporta un aggravio nelle formule in quanto, ad esempio,

$$\{Z_2 = 2\} = \{(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 1, 4), (1, 1, 5), (1, 1, 6)\} \implies |Z_2 = 2| = 6$$

Questo fatto non altera i risultati individuati nel gioco della zara a due dadi; infatti, relativamente all'evento $Z_2 = 2$, si ha:

$$\mathbb{P}_3(Z_2 = 2) = \frac{6}{216} = \frac{1}{36} = \mathbb{P}_2(Z_2 = 2).$$

Variabili aleatorie discrete e continue

Negli esempi precedenti tutte le variabili aleatorie disponevano di un insieme finito di valori possibili. Variabili aleatorie con un numero finito o numerabile di valori possibili sono dette **discrete**. Esistono comunque anche variabili aleatorie dette appunto **continue**, che possono assumere un insieme continuo di valori possibili, come può essere un intervallo di numeri reali.

2.2 Continuità e misurabilità

In questa sezione capiremo il concetto di **continuità** (presentato in maniera molto astratta) che ci permetterà di definire la **misurabilità** e i **numeri aleatori** completando la *terna di probabilità di un numero aleatorio* (o variabile aleatoria).

Definizione 2.2.1 (Topologia)

Si definisce **topologia** una collezione \mathcal{T} di sottoinsiemi di un insieme S (detto anche *insieme sostegno*) tali che:

- L'insieme vuoto e S appartengono a \mathcal{T} : $\emptyset \in \mathcal{T}$ e $S \in \mathcal{T}$.
- L'unione di una quantità arbitraria di insiemi appartenenti a \mathcal{T} appartiene a \mathcal{T} : $\bigcup Z \in \mathcal{T}, \forall Z \in \mathcal{T}$.
- L'intersezione di due insiemi appartenenti a \mathcal{T} appartiene a \mathcal{T} : $Y \cap W \in \mathcal{T}, \forall Y, W \in \mathcal{T}$.

Definizione 2.2.2 (Spazio topologico)

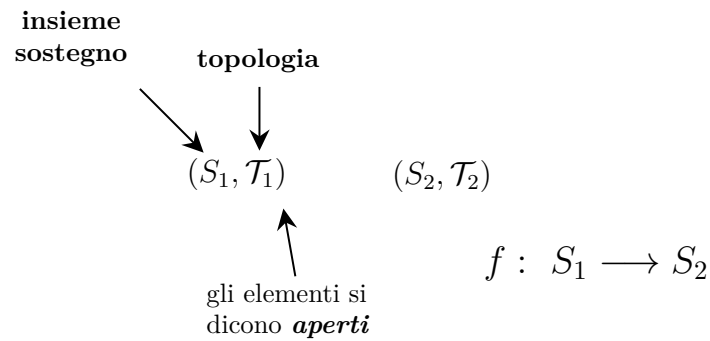
Uno **spazio topologico** è una coppia (S, \mathcal{T}) dove S è un insieme e \mathcal{T} una topologia.

In uno spazio topologico gli insiemi che costituiscono \mathcal{T} si dicono *aperti* in S . Si dice che la collezione \mathcal{T} di aperti è una topologia per S . Se dal contesto è chiaro di che topologia si sta parlando, per brevità si indica lo spazio topologico solo con il nome S dell'insieme.

La **continuità** riguarda la **topologia**.

Definizione 2.2.3 (Funzione f continua)

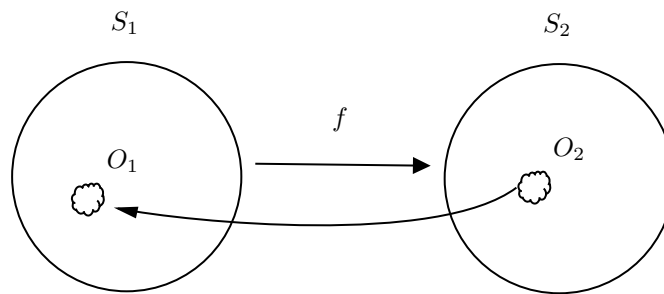
Siano \mathcal{T}_1 e \mathcal{T}_2 due topologie rispettivamente per gli insiemi sostegno S_1 e S_2 (quindi due spazi topologici) e sia f una funzione da S_1 ad S_2 :



Diciamo che f è **continua** se e soltanto se:

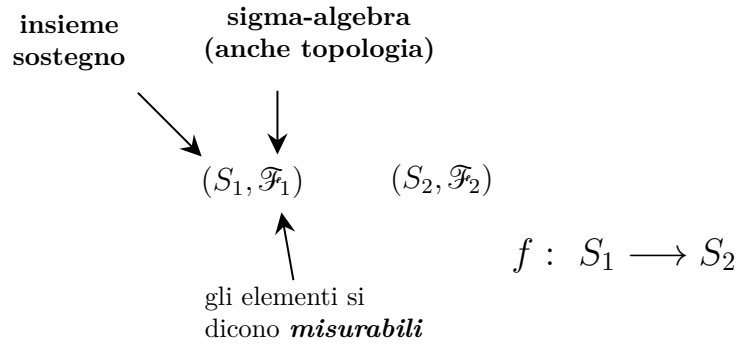
$$\forall O_2 \in \mathcal{T}_2, \{f \in O_2\} \in \mathcal{T}_1$$

A parole: *comunque* preso un aperto O_2 appartenente a \mathcal{T}_2 , la controimmagine di O_2 (indicata con $\{f \in O_2\}$) deve appartenere a \mathcal{T}_1 .



Definizione 2.2.4 (Funzione f misurabile)

Siano \mathcal{F}_1 e \mathcal{F}_2 due σ -algebre (una σ -algebra è una topologia) rispettivamente per gli insiemi sostegno S_1 e S_2 (quindi due spazi topologici) e sia f una funzione da S_1 ad S_2 :



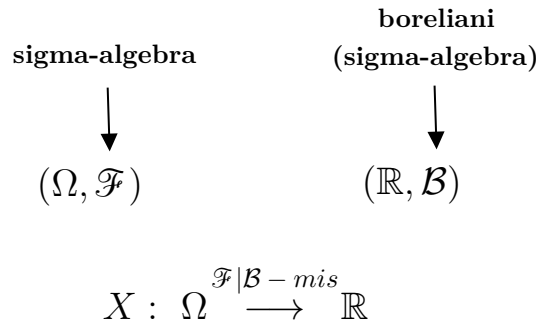
Diciamo che f è **misurabile** se e solo se:

$$\forall B_2 \in \mathcal{F}_2, \{f \in B_2\} \in \mathcal{F}_1$$

e si indica con:

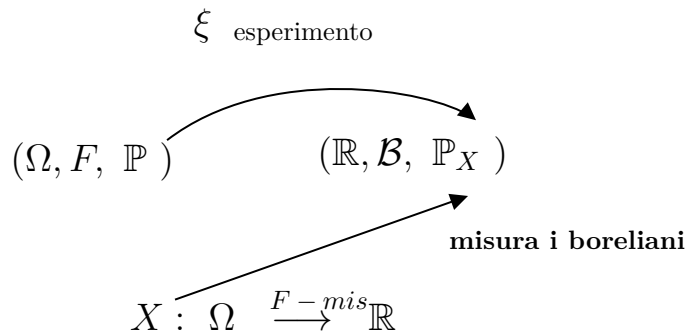
$$f : S_1 \xrightarrow{\mathcal{F}_1 | \mathcal{F}_2 - mis} S_2$$

Definizione terna probabilità numero aleatorio



X è un *numero aleatorio misurabile* (possiamo omettere \mathcal{B} sopra la funzione perché è scontato).

Nel caso di un esperimento generico aleatorio ξ dobbiamo essere in grado di definire anche *la funzione di probabilità* \mathbb{P}_X e questo ci è permesso farlo proprio grazie alla **misurabilità del numero aleatorio X** .



Definiamo \mathbb{P}_X in questo modo (le parentesi graffe possono essere omesse):

$$B \in \mathcal{B} \quad \mathbb{P}_X(B) := \mathbb{P}(\{X \in B\})$$

Assiomi della probabilità

$$B \in \mathcal{B}, \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B) \geq 0 \quad (1)$$

$$\mathbb{P}_X(\{\mathbb{R}\}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1 \quad (2)$$

$$(B_n)_{n \in \mathbb{N}} \subseteq \mathcal{B} \text{ successione di boreali tali che } B_i \cap B_j = \emptyset \text{ con } i \neq j \quad (3)$$

$$\mathbb{P}_X\left(\bigcup B_n\right) = \mathbb{P}\left(X \in \bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \{X \in B_n\}\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(X \in B_n) = \sum_{n \in \mathbb{N}} \mathbb{P}_X(B_n)$$

2.3 Definizione di valore atteso (o media) di un numero aleatorio discreto

Uno dei concetti più importanti in tutta la teoria della probabilità è quello di valore atteso.

Definizione 2.3.1 (Valore atteso)

Sia X una variabile aleatoria discreta che può assumere i valori x_1, x_2, \dots ; il **valore atteso** di X , che si indica con $\mathbb{E}(X)$, è (se esiste) il numero:

$$\mathbb{E}(X) := \sum_i x_i \mathbb{P}(X = x_i)$$

In altri termini, si tratta della media pesata dei valori possibili di X , usando come pesi le probabilità che tali valori vengano assunti da X . Per questo $\mathbb{E}(X)$ è anche detta **media** di X , oppure *aspettazione* (dal termine inglese *expectation*).

Per illustrare il concetto di media pesata, facciamo un semplice esempio. Se X è una variabile aleatoria con funzione di probabilità:

$$p_0 := \mathbb{P}(X = 0) := \frac{1}{2} =: \mathbb{P}(X = 1) =: p_1$$

allora,

$$\mathbb{E}(X) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{0+1}{2} = \frac{1}{2}$$

è in questo caso semplicemente la media aritmetica dei valori che X può assumere. Però, se:

$$p_0 = \frac{1}{3}, \quad p_1 = \frac{2}{3}$$

allora,

$$\mathbb{E}(X) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{0+1 \times 2}{3} = \frac{2}{3}$$

è una media pesata degli stessi valori 0 e 1, dove al secondo è stato dato un peso che è il doppio di quello del primo.

L'interpretazione frequentistica della probabilità fornisce una importante giustificazione del concetto di valore atteso. Da tale punto di vista la probabilità di un evento è definita come il limite a cui tende - empiricamente - il rapporto tra il numero di ripetizioni in cui si è realizzato l'evento e il numero totale di ripetizioni di un esperimento. Consideriamo una variabile aleatoria X che può assumere i valori x_1, x_2, \dots, x_n con funzione di massa di probabilità p . Immaginando che X sia la vincita in una singola mano di un gioco casuale, qual'è la vincita media (nel senso comune del termine) se giochiamo molte mani? Su un numero N di ripetizioni dell'esperimento, ciascuno dei valori x_i si verificherà un certo numero N_i di volte. L'interpretazione frequentista afferma che se N è molto grande, $N_i \approx N\mathbb{P}(X = x_i)$. D'altronde ci si convince facilmente che la vincita media è data da:

$$\frac{x_1 N_1 + x_2 N_2 + \dots + x_n N_n}{N} = \sum_{i=1}^n x_i \frac{N_i}{N} \approx \sum_{i=1}^n x_i \mathbb{P}(X = x_i) =: \mathbb{E}(X)$$

e quindi coincide approssimativamente con la definizione di valore atteso di X .

Esempio 2.3.1

Sia X il punteggio che si ottiene lanciando un dado non truccato. Quanto vale $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{42}{6} = \frac{7}{2} = 3.5$$

E' utile notare che in questo esempio, il valore atteso di X non è uno dei valori che X può assumere (tirando un dado non c'è modo di ottenere un punteggio di 3.5). Perciò, anche se $\mathbb{E}(X)$ è chiamato *valore atteso* di X , non vuole affatto dire che noi ci attendiamo di vedere questo valore, ma piuttosto che ci aspettiamo che sia il limite a cui tende il punteggio medio del dado su un numero crescente di ripetizioni. In effetti su molti lanci di dado la media aritmetica di tutti i valori ottenuti tende a $7/2$.

Esempio 2.3.2

Se I è la funzione indicatrice di un evento A , ovvero se:

$$I := \begin{cases} 1, & \text{se } A \text{ si verifica} \\ 0, & \text{se } A \text{ non si verifica} \end{cases}$$

allora:

$$\mathbb{E}(I) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = \mathbb{P}(X = 1) = \mathbb{P}(A)$$

2.4 La funzione generatrice dei momenti

La *funzione generatrice dei momenti*, o più semplicemente *funzione generatrice* ϕ , di una variabile aleatoria X , è definita, per tutti i t reali per i quali il valore atteso e^{tX} ha senso, dall'espressione:

$$\phi(t) := \mathbb{E}(e^{tX}) = \sum_x e^{tx} \mathbb{P}(X = x)$$

Il nome adottato deriva dal fatto che tutti i momenti di cui è dotata X possono essere ottenuti derivando più volte nell'origine la funzione $\phi(t)$. Ad esempio,

$$\phi'(t) = \frac{d}{dt} \mathbb{E}(e^{tX}) = \mathbb{E}\left[\frac{d}{dt} e^{tX}\right] = \mathbb{E}[X e^{tX}]$$

da cui $\phi'(0) = \mathbb{E}(X)$. Analogamente:

$$\phi''(t) = \frac{d^2}{dt^2} \mathbb{E}(e^{tX}) = \mathbb{E}\left[\frac{d^2}{dt^2} e^{tX}\right] = \mathbb{E}[X^2 e^{tX}]$$

da cui $\phi''(0) = \mathbb{E}(X^2)$, è il momento secondo di X . Più in generale, la derivata n -esima di $\phi(t)$ calcolata in 0 fornisce il momento n -esimo di X :

$$\phi^{(n)}(0) = \mathbb{E}(X^n), \quad n \geq 1$$

2.4.1 Proprietà funzione generatrice dei momenti

Elenchiamo innanzitutto due semplici proprietà della funzione generatrice dei momenti:

1. $\mu \in \mathbb{R}$, $\phi_\mu(t) = \mathbb{E}(e^{t\mu}) = e^{t\mu}$.
2. $a \in \mathbb{R} - \{0\}$, $\phi_{aX}(t) = \mathbb{E}(e^{taX}) = \phi_X(at)$

Proposizione 2.4.1 (Se X e Y sono indipendenti allora $\phi_X(t) \cdot \phi_Y(t) = \phi_{X+Y}(t)$)

La funzione generatrice dei momenti della somma di variabili aleatorie indipendenti è il prodotto delle funzioni generatrici delle singole variabili aleatorie.

Dimostrazione.

$$\begin{aligned} \phi_{X+Y}(t) &= \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left(e^{tX} \cdot e^{tY}\right) \\ (X \text{ e } Y \text{ indipendenti per HP}) &= \mathbb{E}(e^{tX}) \cdot \mathbb{E}(e^{tY}) \\ &= \phi_X(t) \cdot \phi_Y(t) \end{aligned}$$

□

Proposizione 2.4.2 (X_1, X_2, \dots, X_n indipendenti allora $\phi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \phi_{X_i}(t)$)

Un'altra importante proprietà di ϕ è che la funzione generatrice dei momenti della somma di variabili aleatorie indipendenti è il prodotto delle funzioni generatrici delle singole variabili aleatorie.

Se X è una variabile aleatoria definita come:

$$X = \sum_{i=1}^n X_i \quad \text{dove ogni variabile } X_i \text{ è indipendente dalle altre variabili}$$

allora:

$$\phi_X(t) = \prod_{i=1}^n \phi_{X_i}(t)$$

Dimostrazione.

$$\begin{aligned} \phi_{\sum_{i=1}^n X_i}(t) &= \mathbb{E} \left(\prod_{i=1}^n e^{tX_i} \right) \\ (\text{indipendenza HP}) \quad &= \prod_{i=1}^n \mathbb{E}(e^{tX_i}) \\ &= \prod_{i=1}^n \phi_{X_i}(t) \end{aligned}$$

□

2.5 Funzione di distribuzione

Consideriamo X un numero aleatorio discreto con spettro $S_X = \{x_1, \dots, x_m, \dots\}$ e valori di probabilità $(p_k)_{k \in S_X}$:

$$(\mathbb{R}, \mathcal{B}, \mathbb{P}_X) \quad p_k = \mathbb{P}(X = k) = \mathbb{P}_X(\{k\})$$

$$B \in \mathcal{B} \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B) = \sum_{k: x_k \in B} 1_{\{x_k \in B\}} \cdot p_k$$

dove l'ultimo passaggio è proprio l'intersezione tra il boreliano B e lo spettro S_X .

Poniamo $X \sim \Pi(\lambda)$ con $\lambda > 0$. Il suo spettro è $S_X = \{0, 1, 2, \dots\}$ e la probabilità è $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ (variabile aleatoria di Poisson trattata nel Capitolo 3 paragrafo 3.5). Valutiamo le seguenti probabilità (intersezioni tra il boreliano dato e lo spettro della variabile aleatoria X):

$$\begin{aligned} \mathbb{P}(1 < X < \sqrt{2}) &= \mathbb{P}_X(\emptyset) = 0 \quad (\text{intersezione vuota}) \\ \mathbb{P}(1 \leq X < \sqrt{2}) &= \mathbb{P}_X(\{1\}) \\ \mathbb{P}(1 \leq X \leq \sqrt{5}) &= \mathbb{P}_X(\{1\}) + \mathbb{P}_X(\{2\}) \\ \mathbb{P}_X([0, +\infty[) &= 1 \\ \mathbb{P}_X([0, +\infty]) &= 1 - \mathbb{P}_X(\{0\}) \end{aligned}$$

Tra tutti i boreliani quelli che sono di una fondamentale importanza nel calcolo delle probabilità sono le **semirette sinistre chiuse**:

$$(\quad - \infty, x]_{x \in \mathbb{R}}$$

x è l'origine della semiretta chiusa. Questa scrittura è dunque la **famiglia delle semirette sinistre chiuse**. Se poniamo $B =] - \infty, x]$:

$$B =] - \infty, x], \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X \leq x) := F_X(x)$$

Abbiamo così ottenuto una funzione reale di variabile reale (perché facciamo variare x). Tale funzione F_X è chiamata **funzione di distribuzione** (in inglese *distribution function*). Con questa funzione passiamo da una funzione \mathbb{P}_X che vuole un sottoinsieme a una funzione F_X che vuole un numero reale (più facile da maneggiare

$$F_X : \mathbb{R} \rightarrow [0, 1]$$

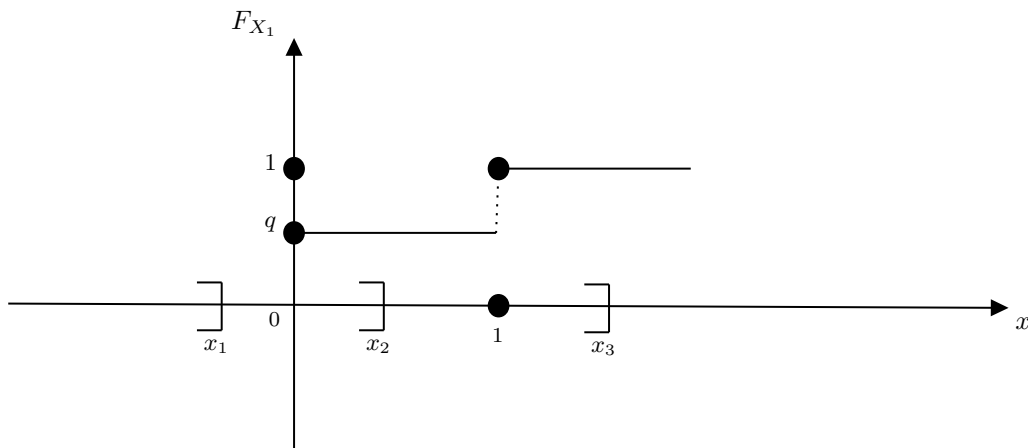
$$x \mapsto F_X(x) := \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x)$$

- valutare monotonia più semplicemente ecc...). La funzione F_X opera una restrizione della funzione \mathbb{P}_X perché non considera tutti i boreliani ma solo l'intervallo $]-\infty, x]$. Di conseguenza risulta essere più completa la funzione \mathbb{P}_X perché permette di misurare tutti i boreliani e non solo $]-\infty, x]$ ma F_X è molto più maneggevole. Dalla funzione F_X è però possibile ricostruire \mathbb{P}_X . Ad esempio:

$$\mathbb{P}(X > a) = 1 - F_X(a)$$

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$$

Facciamo qualche esempio. Sia X_1 un numero aleatorio di Bernoulli ($X_1 \sim B(1, p) \sim \text{Bernoulli}(p)$) con probabilità $\mathbb{P}(X_1 = 1) = p$ e $\mathbb{P}(X_1 = 0) = 1 - p = q$. Lo spettro è $S_{X_1} = \{0, 1\}$. Com'è fatta F_{X_1} ?



$$F_{X_1}(x_1) = \mathbb{P}(X_1 \leq x_1) = 0$$

$$F_{X_1}(x_2) = \mathbb{P}(X_1 \leq x_2) = \mathbb{P}(X_1 = 0) = q$$

$$F_{X_1}(x_3) = \mathbb{P}(X_1 \leq x_3) = \mathbb{P}(\{X_1 = 0\} \cup \{X_1 = 1\}) = \mathbb{P}(X_1 = 0) + \mathbb{P}(X_1 = 1) = p + q = p + (1 - p) = 1$$

$$\lim_{x \rightarrow +\infty} F_{X_1}(x) = 1$$

$$\lim_{x \rightarrow -\infty} F_{X_1}(x) = 0$$

$$F_X(0^+) - F_X(0^-) = F_X(0) - F(0^-) = q - 0 > 0$$

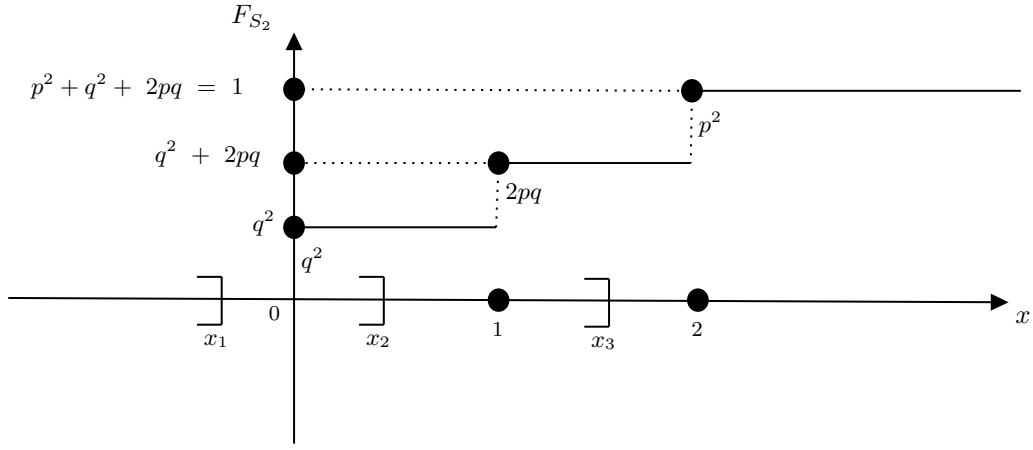
I numeri aleatori che stiamo studiando hanno una particolarità: $F_X(x^+) - F_X(x^-)$ fa sempre zero *tranne* su tutti i valori dello spettro:

$$F_X(x) - F_X(x^-) = \begin{cases} 0, & \text{se } x \notin S_X \\ p_k, & \text{se } x = x_k \in S_k \end{cases}$$

Con la funzione F non abbiamo perso informazioni in termini probabilistici. Infatti possiamo ricostruire dalla funzione di distribuzione la probabilità di tutti i boreliani, in particolare i singoletti perché valgono sempre zero tranne quelli che appartengono allo spettro. Possiamo riottenere le probabilità dei singoletti tramite l'operazione descritta appena sopra. Possiamo quindi ricostruire la distribuzione di probabilità, ricostruiamo cioè \mathbb{P}_X con questa operazione.

Facciamo adesso un altro esempio. Consideriamo il numero aleatorio $S_2 = X_1 + X_2$ (distribuzione binomiale $S_2 \sim B(2, p)$). Lo spettro di S_2 è $S_{S_2} = \{0, 1, 2\}$ dove i valori assumono probabilità rispettivamente $(q^2, 2pq, p^2)$:

Dal grafico notiamo che le alzate sono le probabilità dei singoletti che si ottengono facendo la differenza. Ci sono tre punti di discontinuità di 1° specie. E' una funzione monotona crescente.



Consideriamo adesso il numero aleatorio $W_1 = \mathcal{T} \sim G(p)$. Il suo spettro è uguale all'insieme dei numeri naturali \mathbb{N} e la probabilità è $\mathbb{P}(W_1 = n) = pq^{n-1}$ con $n \in \mathbb{N}$. Se analizziamo la funzione di distribuzione F_{W_1} per arrivare ad uno nel grafico dobbiamo per forza considerare il limite:

$$\lim_{x \rightarrow +\infty} F_{W_1}(x) = 1$$

Questa funzione ha un numero infinito numerabile di punti di discontinuità di prima specie.

2.5.1 Successione di eventi monotona crescente/decrescente

Consideriamo $(E_n)_{n \in \mathbb{N}}$ una successione di eventi.

Tale successione $(E_n)_{n \in \mathbb{N}}$ si dice **crescente** se e soltanto se:

$$\forall n \in \mathbb{N} \ E_n \subseteq E_{n+1} \text{ e si scrive } E_n \uparrow \text{ e il suo limite è } \bigcup_{m \in \mathbb{N}} E_m$$

mentre si dice **decrescente** se e soltanto se:

$$\forall n \in \mathbb{N} \ E_{n+1} \subseteq E_n \text{ e si scrive } E_n \downarrow \text{ e il suo limite è } \bigcap_{m \in \mathbb{N}} E_m$$

Teorema 12 (Il limite della probabilità di una successione di eventi monotona è la probabilità del limite)

Sia $(E_n)_{n \in \mathbb{N}}$ una successione monotona. Allora:

$$\lim_{n \rightarrow \infty} \mathbb{P}(E_n) = \mathbb{P}(\lim_{n \rightarrow \infty} E_n)$$

Questo teorema proviene dall'analisi matematica:

$$\lim_{x \rightarrow x_0} f(x) = f(x_0) = f(\lim_{x \rightarrow x_0} x)$$

Operativamente con questa proprietà calcoliamo il limite con valori provenienti dal dominio e alla fine calcoliamo una sola volta la funzione f . Quindi se dobbiamo fare il limite di una successione degli eventi, se questa è crescente facciamo l'unione mentre se è decrescente facciamo l'intersezione e poi calcoliamo una sola volta la probabilità del risultato dell'operazione insiemistica.

2.5.2 Proprietà della funzione di distribuzione

Sia F_X la funzione di distribuzione, per ogni $x \in \mathbb{R}$ $F_X(x) = \mathbb{P}(X \leq x)$.

1. F_X è non decrescente
2. F_X è continua destra in \mathbb{R}
3. $\lim_{x \rightarrow +\infty} F_X(x) = 1$ e $\lim_{x \rightarrow -\infty} F_X(x) = 0$

Dimostrazione 1 (F_X è non decrescente):

$$\begin{aligned} x_1 < x_2 &\implies]-\infty, x_1] \cup]x_1, x_2] \\ F_X(x_2) &= \mathbb{P}(X \leq x_2) = \mathbb{P}(X \leq x_1) + \mathbb{P}(x_1 < X \leq x_2) = F_X(x_1) + \mathbb{P}(x_1 < X \leq x_2) \\ \text{siccome la quantità } \mathbb{P}(x_1 < X \leq x_2) &\geq 0 \text{ possiamo concludere che } F_X(x_2) \geq F_X(x_1) \end{aligned}$$

□

Dimostrazione 2 (F_X è continua destra in \mathbb{R})

F_X è continua a destra in \mathbb{R} vuole dire che

$$\forall x \in \mathbb{R} \quad \lim_{\epsilon \rightarrow 0^+} F_X(x + \epsilon) = F_X(x)$$

Per dimostrarlo ci serviamo del **teorema Ponte**. Consideriamo questo limite:

$$\lim_{x \rightarrow x_0} f(x) = l, \quad l \in \mathbb{R}$$

Il risultato del limite è un numero reale $l \in \mathbb{R}$. L'operazione di questo limite consiste nell'avvicinarsi in continuità da x a x_0 , senza "saltelli". E' possibile anche però avvicinarsi a x_0 da destra e da sinistra non in continuità, ma a "saltelli" numerabili. Quello che vogliamo fare è estrarre una successione (numerabile) di valori $(x_n)_{n \in \mathbb{N}}$ nell'intorno del punto x_0 dove si vuole il limite. Quindi, se consideriamo questo limite invece dell'altro:

$$\lim_{n \rightarrow \infty} f(x_n) = l \in \mathbb{R}$$

quello che otteniamo è sempre lo stesso numero reale $l \in \mathbb{R}$. Il teorema Ponte dice che (x_0 può essere anche $+\infty, -\infty$):

$$\lim_{x \rightarrow x_0} f(x) = l \in \mathbb{R}$$



per ogni successione $x_n \rightarrow x_0$ nel dominio di f tale che $x_n \neq x_0$

$$\lim_{n \rightarrow \infty} f(x_n) = l \in \mathbb{R}$$

dove $(x_n)_{n \in \mathbb{N}}$ è una successione numerabile di valori che appartengono all'intorno di x_0 . Sostanzialmente ci stiamo avvicinando a x con una successione: comunque ci avviciniamo ad x il risultato del limite sarà sempre $l \in \mathbb{R}$.

Per dimostrare che:

$$\forall x \in \mathbb{R} \quad \lim_{\epsilon \rightarrow 0^+} F_X(x + \epsilon) = F_X(x)$$

trasformiamo questo limite in un limite di una successione numerabile di valori appartenenti nell'intorno di x :

$$\begin{aligned} \lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(X \leq x + \frac{1}{n}\right) = \lim_{n \rightarrow \infty} \left(\mathbb{P}(X \leq x) + \mathbb{P}(x < X \leq x + \frac{1}{n})\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x) + \lim_{n \rightarrow \infty} \mathbb{P}(x < X \leq x + \frac{1}{n}) \\ &= \lim_{n \rightarrow \infty} F_X(x) + \lim_{n \rightarrow \infty} \mathbb{P}(x < X \leq x + \frac{1}{n}) \\ (\text{il limite di una costante è una costante}) &= F_X(x) + \lim_{n \rightarrow \infty} \mathbb{P}(x < X \leq x + \frac{1}{n}) \\ (\text{Teorema 12}) &= F_X(x) + \mathbb{P}\left(\lim_{n \rightarrow \infty} \{x < X \leq x + \frac{1}{n}\}\right) \\ (\text{la controimmagine è il vuoto}) &= F_X(x) + \mathbb{P}(\emptyset) \\ (\text{la probabilità del vuoto è 0}) &= F_X(x) \end{aligned}$$

Siamo quindi arrivati a questo punto:

$$\lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) = F_X(x)$$

Siccome F_X è una funzione non decrescente (e quindi NON è una funzione che oscilla), allora possiamo affermare che comunque prendiamo una successione $(x_n)_{n \in \mathbb{N}}$ che tende a x con $n \rightarrow \infty$ il suo risultato sarà sempre $F_X(x)$. Utilizzando adesso il teorema ponte abbiamo quindi dimostrato la tesi:

$$\lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) = F_X(x) \xrightarrow[\text{Teorema ponte}]{F_X \text{ non decrescente}} \lim_{\epsilon \rightarrow 0^+} F_X(x + \epsilon) = F_X(x)$$

□

Dimostrazione 3 ($\lim_{x \rightarrow +\infty} F_X(x) = 1$)

Partiamo da questo limite ($n \in \mathbb{N}$):

$$\lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} \{X \leq n\}\right) = \mathbb{P}(\Omega) = 1$$

Ricordiamo che la scrittura $\{X \leq n\} = \{\omega \in \Omega \mid X(\omega) \leq n\}$ identifica degli eventi e con $n \rightarrow \infty$ corrisponde proprio a Ω : la probabilità di Ω è uno.

Adesso applichiamo il fatto che F_X è non decrescente e applichiamo il teorema Ponte (esattamente lo stesso ragionamento della successione di prima):

$$\lim_{n \rightarrow \infty} F_X(n) = 1 \xrightarrow[\text{Teorema ponte}]{F_X \text{ non decrescente}} \lim_{x \rightarrow +\infty} F_X(x) = 1$$

Allo stesso modo dimostriamo che:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0$$

Prendiamo questo limite:

$$\lim_{n \rightarrow +\infty} F_X(-n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq -n) = \mathbb{P}\left(\lim_{n \rightarrow +\infty} \{X \leq -n\}\right) = \mathbb{P}(\emptyset) = 0$$

Quindi sapendo che F_X è non decrescente, applichiamo il teorema Ponte:

$$\lim_{n \rightarrow \infty} F_X(-n) = 0 \xrightarrow[\text{Teorema ponte}]{F_X \text{ non decrescente}} \lim_{x \rightarrow -\infty} F_X(x) = 0$$

□

Sostanzialmente il teorema Ponte ci permette di passare da un dominio di valori reali a un dominio di valori naturali (e quindi proprio i valori che assumono le variabili aleatorie che abbiamo studiato fino ad adesso). Siccome il teorema Ponte lega i limiti di funzioni reali con i limiti di successioni numerabili, dimostrare queste proprietà studiando i limiti di successioni monotone implica dimostrarle per i limiti di funzioni reali.

Vediamo adesso alcuni esempi su come possiamo ricostruire la funzione \mathbb{P} utilizzando la funzione di distribuzione:

$$\begin{aligned} \mathbb{P}(a < X \leq b) &= F_X(b) - F_X(a) \\ \mathbb{P}(X = a) &= F_X(a) - F_X(a^-) \\ \mathbb{P}(a < X < b) &= F_X(b) - F_X(a) - [F_X(a) - F_X(a^-)] \\ \mathbb{P}(a \leq X < b) &= F_X(b) - F_X(a) - [F_X(a) - F_X(a^-)] + [F_X(a) - F_X(a^-)] \\ \mathbb{P}(X > a) &= 1 - F_X(a) \\ \mathbb{P}(X \geq a) &= 1 - F_X(a) + [F_X(a) - F_X(a^-)] = 1 - F_X(a^-) \end{aligned}$$

Modelli di variabili aleatorie discrete

3 Modelli di variabili aleatorie discrete

3.1 Variabili aleatorie di Bernoulli e binomiali

Consideriamo l'esperimento ξ : lancio di una moneta truccata un numero indefinito di volte. Lo spazio campione è $\Omega = \{T, C\}^\infty$ dove T sta per testa e C sta per croce. Un elemento $\omega \in \Omega$ è del tipo $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$ dove $\omega_n \in \{T, C\}$. La famiglia degli eventi generata dalla famiglia degli eventi generatori è $\mathcal{F} = \sigma(\mathcal{G})$ dove la famiglia degli eventi generatori è $\mathcal{G} = (T_n)_{n \in \mathbb{N}}$. Ad esempio $T_1 \subseteq \Omega$ sono tutte le successioni con T al primo posto, cioè:

$$T_1 = \{(T, \omega_2, \omega_3, \dots, \omega_n, \dots) \mid \text{per ogni } i = 2, \dots \quad \omega_i = T \text{ oppure } \omega_i = C\}$$

Consideriamo adesso il numero aleatorio X_1 così definito:

$$\begin{aligned} X_1 : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longrightarrow X_1(\omega) = \begin{cases} 1, & \text{se } \omega_1 = T \\ 0, & \text{se } \omega_1 = C \end{cases} \end{aligned}$$

Possiamo dire che X_1 è *misurabile*?

- Se B non contiene 0 e non contiene 1 allora $\{X_1 \in B\} = \emptyset \in \mathcal{F}$.
- Se B contiene 1 e non contiene 0 allora $\{X_1 \in B\} = T \in \mathcal{F}$.
- Se B contiene 0 e non contiene 1 allora $\{X_1 \in B\} = T^c \in \mathcal{F}$.
- Se B contiene 0 e contiene 1 $\{X_1 \in B\} = \Omega \in \mathcal{F}$

Possiamo affermare che X_1 è un numero aleatorio *discreto* dato che l'insieme dei valori che esso può assumere è finito e lo denotiamo con $S_{X_1} = \{0, 1\}$ (chiamato anche *spettro della trasformazione*).

Adesso *definiamo le probabilità* \mathbb{P} . Diciamo che $p \in \{0, 1\}$ è la probabilità che il primo lancio sia testa. Diciamo che $1 - p := q$ è la probabilità che non si verifichi testa al primo lancio:

$$\begin{aligned} \mathbb{P}_{X_1}(\{1\}) &= \mathbb{P}(X_1 = 1) = p \\ \mathbb{P}_{X_1}(\{0\}) &= \mathbb{P}(X_1 = 0) = 1 - p := q \\ \mathbb{P}_{X_1}(\mathbb{R}) &= 1 \end{aligned}$$

Possiamo adesso allora definire la probabilità \mathbb{P}_{X_1} :

$$B \in \mathcal{B}, \quad \mathbb{P}_{X_1}(B) = \begin{cases} 0, & \text{se } B \cap \{0, 1\} = \emptyset \\ p, & \text{se } B \cap \{0, 1\} = \{1\} \\ q, & \text{se } B \cap \{0, 1\} = \{0\} \\ 1, & \text{se } B \cap \{0, 1\} = \{0, 1\} \end{cases}$$

Abbiamo dunque completato la descrizione di un numero aleatorio: abbiamo cioè determinato tutti i suoi valori possibili e abbiamo definito la sua funzione di probabilità \mathbb{P}_X (una volta assegnati i valori di probabilità alla funzione \mathbb{P}).

Genericamente parlando, il numero aleatorio X_n è un **numero di Bernoulli** e si denota così:

$$n \in \mathbb{N}, \quad X_n \sim (1, p)$$

dove 1 denota il fatto che stiamo parlando di un solo esperimento e con p indichiamo la probabilità che l'esperimento si concluda con esito "croce". In altre parole, una variabile aleatoria è un numero di Bernoulli se può assumere solo i valori 0 e 1.

Vogliamo adesso verificare se la somma $X_1 + X_2$ (cioè la somma di due numeri aleatori) è misurabile. Denotiamo questa somma come $S_2 := X_1 + X_2$ e notiamo subito che questa variabile aleatoria è un **numero di Bernoulli binomiale** (dato che adesso ci sono in gioco due esperimenti e non più uno) e cioè:

$$S_2 := X_1 + X_2 \sim B(2, p)$$

Lo spettro di S_2 è:

$$S_{S_2} = \{0, 1, 2\}$$

quindi tale numero aleatorio può assumere tre valori.

Cerchiamo adesso di capire quali sono le contro-immagini di ogni possibile valore di S_2 :

$$B \in \mathcal{B}, \{S_2 \in B\} = \begin{cases} \emptyset, & \text{se } S_{S_2} \cap B = \emptyset \\ C_1 \cap C_2, & \text{se } S_{S_2} \cap B = \{0\} \\ (T_1 \cap C_2) \dot{\cup} (C_1 \cap T_2), & \text{se } S_{S_2} \cap B = \{1\} \\ T_1 \cap T_2, & \text{se } S_{S_2} \cap B = \{2\} \\ (C_1 \cap C_2) \dot{\cup} (C_1 \cap T_2) \dot{\cup} (T_1 \cap C_2), & \text{se } S_{S_2} \cap B = \{0, 1\} \\ (C_1 \cap C_2) \dot{\cup} (T_1 \cap T_2), & \text{se } S_{S_2} \cap B = \{0, 2\} \\ (C_1 \cap T_2) \dot{\cup} (T_1 \cap C_1) \dot{\cup} (T_1 \cap T_2), & \text{se } S_{S_2} \cap B = \{1, 2\} \\ \Omega, & \text{se } S_{S_2} \cap B = \{0, 1, 2\} \end{cases}$$

Tutto dipende dall'intersezione di un boreliano con lo spettro.

In generale, possiamo dire che:

La somma di due funzioni misurabili (continue) è una funzione misurabile (continua)

Non ci resta adesso che definire la funzione di probabilità \mathbb{P} (e di conseguenza definiamo anche \mathbb{P}_{S_2}) per i tre valori possibili dello spettro di S_2 .

$$p_0 := \mathbb{P}(S_2 = 0) = \mathbb{P}(C_1 \cap C_2) = \mathbb{P}(C_1)\mathbb{P}(C_2) = (1 - p)^2$$

$$\begin{aligned} p_1 := \mathbb{P}(S_2 = 1) &= \mathbb{P}((C_1 \cap T_2) \dot{\cup} (T_1 \cap C_2)) = \mathbb{P}(C_1 \cap T_2) + \mathbb{P}(T_1 \cap C_2) = \mathbb{P}(C_1)\mathbb{P}(T_2) + \mathbb{P}(T_1)\mathbb{P}(C_2) \\ &= (1 - p)p + p(1 - p) \\ &= 2(1 - p)p \end{aligned}$$

$$p_2 := \mathbb{P}(S_2 = 2) = \mathbb{P}(T_1 \cap T_2) = \mathbb{P}(T_1)\mathbb{P}(T_2) = p^2$$

Abbiamo così definito la **distribuzione di probabilità** di S_2 :

$$S_2 \sim B(2, p) \text{ (numero aleatorio con distribuzione binomiale)}$$

$$S_{S_2} = \{0, 1, 2\}$$

$$((1 - p)^2, 2p(1 - p), p^2)$$

Così come in precedenza abbiamo definito la distribuzione di probabilità di X_1 :

$$X_1 \sim B(1, p) \equiv B(p) \text{ (numero aleatorio con distribuzione di Bernoulli)}$$

$$S_{X_1} = \{0, 1\}$$

$$(1 - p, p)$$

Consideriamo $S_3 := X_1 + X_2 + X_3$. Anche questo numero aleatorio ha una distribuzione binomiale. Per semplicità poniamo $q := 1 - p \in (0, 1)$:

$$S_3 \sim B(3, p) \text{ (numero aleatorio con distribuzione binomiale)}$$

$$S_{S_3} = \{0, 1, 2, 3\}$$

$$(q^3, 3q^2p, 3qp^2, p^3)$$

Calcoliamo adesso i valori attesi (o medie) dei numeri aleatori X_1 , S_2 e S_3 .

$$\mathbb{E}(X_1) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$\mathbb{E}(S_2) = 0 \cdot (1 - p)^2 + 1 \cdot 2(1 - p)p + 2 \cdot p^2 = 2p((1 - p) + p) = 2p$$

$$\mathbb{E}(S_3) = 0 \cdot p^0 q^3 + 1 \cdot 3q^2 p + 2 \cdot 3p^2 q + 3p^3 q^0 = 3p$$

Osservando questi risultati si intuisce che **la media è il prodotto dei parametri** (in una distribuzione di Bernoulli o binomiale).

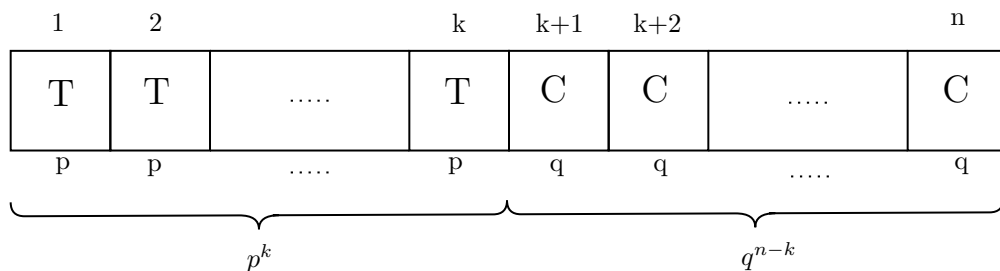
3.1.1 Distribuzione di probabilità binomiale con parametri n e p

Supponiamo di realizzare n ripetizioni indipendenti di un esperimento, ciascuna delle quali può concludersi in un "successo" (oppure valore 1 oppure valore *testa*) con probabilità p , o in un "fallimento" (oppure valore 0 oppure *croce*) con probabilità $1 - p$. Se X denota il numero totale di successi, X si dice **variabile aleatoria binomiale di parametri** (n, p) .

Consideriamo il numero aleatorio $S_n = X_1 + X_2 + \dots + X_n$. Il suo spettro è $S_{S_n} = \{0, 1, 2, \dots, n\}$. La distribuzione binomiale $B(n, p)$ descrive una variabile aleatoria S_n definita come la somma di n variabili aleatorie indipendenti X_i di uguale legge di Bernoulli $B(p)$. La funzione di probabilità per una variabile aleatoria binomiale di parametri (n, p) è data (ricordiamo che $q := 1 - p$):

$$p_k := \mathbb{P}(S_n = k) = p^k \cdot q^{n-k} \cdot \binom{n}{k}$$

dove il coefficiente binomiale $\binom{n}{k}$ conta tutte le permutazioni lunghe n con ripetizioni (lettere che si ripetono k volte e $n - k$ volte).



$$\mathbb{P}(S_n \in \mathbb{R}) = \sum_{k=0}^n p_k = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1^n = 1$$

L'ultimo risultato è stato ottenuto grazie al *binomio di Newton*³

³Tale formula afferma che:

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

La media di S_n è:

$$\begin{aligned}\mathbb{E}(S_n) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{(n-k)} \\ &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k q^{(n-k)} \\ &= \sum_{k=1}^n k \frac{n \cdot (n-1)!}{k \cdot (k-1)!(n-1-(k-1))!} p^k q^{(n-k)} \\ &= np \sum_{k=1}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{(n-1-y)}\end{aligned}$$

dove nell'ultimo passaggio si è posto $y = k - 1$. Tenendo presente che:

$$\sum_{k=1}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{(n-1-y)} = 1$$

poiché si tratta della somma delle probabilità di una distribuzione binomiale con parametri $n - 1$ e p si ottiene:

$$\mathbb{E}(S_n) = n \cdot p$$

e quindi abbiamo dimostrato che la media di un numero aleatorio con distribuzione binomiale è il prodotto dei suoi parametri.

Questo risultato lo si poteva ottenere notando il fatto che **la media della somma è uguale alla somma delle medie**:

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \mathbb{E}(X_k) = \sum_{k=1}^n p = p \sum_{k=1}^n 1 = p \cdot n \text{ (proprietà di linearità della media)}$$

Esempio 3.1.1

Una azienda produce dischetti per PC che sono difettosi con probabilità 0.01, indipendentemente l'uno dall'altro. Questi dischetti sono poi venduti in confezioni da 10 pezzi, con la garanzia di rimborso in caso vi sia più di un pezzo difettoso. Che percentuale delle confezioni viene ritornata? Se si comprano tre confezioni, qual'è la probabilità di ritornarne esattamente una?

Se X è il numero di pezzi difettosi in una scatola da 10 dischetti, X è una variabile aleatoria binomiale di parametri (10,0.01). Perciò, assumendo che tutti i clienti che ne hanno la possibilità sfruttino la garanzia, la probabilità che una scatola sia ritornata è pari a:

$$\begin{aligned}\mathbb{P}(X > 1) &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \\ &= 1 - \binom{10}{0} \cdot 0.01^0 \cdot 0.99^{10} - \binom{10}{1} \cdot 0.01^1 \cdot 0.99^9 \approx 0.0043\end{aligned}$$

Poiché ogni scatola - indipendentemente dalle altre - viene resa con probabilità di circa 0.43%, a lungo andare sarà reso circa lo 0.43% delle confezioni. Da quanto detto segue inoltre, che acquistando 3 scatole, il numero di quelle che verranno rese è una variabile aleatoria binomiale di parametri (3,0.0043), quindi la probabilità richiesta è:

$$\binom{3}{1} \cdot 0.0043^1 \cdot 0.9957^2 \approx 0.013$$

Esempio 3.1.2

Supponiamo per semplicità che il colore degli occhi di ogni persona sia determinato da una sola coppia di geni, con il fenotipo "occhi castani" dominante rispetto a quello "occhi azzurri". Ciò significa che un individuo con due geni per gli occhi azzurri presenta occhi azzurri, mentre uno che abbia almeno un gene per gli occhi castani ce li avrà di quel colore. Quando due individui procreano, ciascuno dei figli prende a caso uno dei due geni da ciascuno dei due genitori. Se il figlio maggiore di una coppia di persone con gli occhi castani ha gli occhi azzurri, qual è la probabilità che esattamente 2 degli altri 4 figli (che non comprendono gemelli) abbiano gli occhi azzurri?

Il fatto che il figlio maggiore abbia gli occhi azzurri e i genitori hanno gli occhi castani ci assicura che la coppia ha un gene per gli occhi azzurri e un gene per gli occhi castani rispettivamente. Un gene viene attribuito

al figlio dal genitore a caso dunque siamo in una situazione di equiprobabilità: $\frac{1}{2}$ è la probabilità di avere un gene per gli occhi castani e $\frac{1}{2}$ è la probabilità di avere un gene per gli occhi azzurri. La probabilità che un figlio abbia gli occhi azzurri è dunque di $\frac{1}{4}$. Costruiamo adesso un numero aleatorio $Y \sim B(4, \frac{1}{4})$ che ha distribuzione binomiale di parametri $(4, \frac{1}{4})$ perché 4 sono gli altri figli e $\frac{1}{4}$ è la probabilità di avere un figlio con gli occhi azzurri. Quindi il numero aleatorio Y indica quanti figli con gli occhi azzurri sono nati. Dunque, la probabilità cercata è di (vogliamo esattamente 2 figli con gli occhi azzurri):

$$\binom{4}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = \frac{27}{128} \approx 0.2109$$

3.1.2 Calcolo esplicito della distribuzione binomiale

Sia S_n un numero aleatorio con distribuzione binomiale di parametri (n, p) . Per calcolare operativamente la funzione di massa:

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}$$

è molto utile la seguente relazione tra $\mathbb{P}(S_n = k + 1)$ e $\mathbb{P}(S_n = k)$:

$$\mathbb{P}(S_n = k + 1) = \frac{p}{q} \cdot \frac{n - k}{k + 1} \cdot \mathbb{P}(S_n = k)$$

Dimostrazione.

$$\begin{aligned} \mathbb{P}(S_n = k + 1) &= \binom{n}{k + 1} p^{k+1} q^{n-k-1} = \frac{p}{q} \binom{n}{k + 1} p^k q^{n-k} \\ &= \frac{p}{q} p^k q^{n-k} \cdot \frac{n!}{(k + 1)!(n - k - 1)!} \\ &= \frac{p}{q} p^k q^{n-k} \cdot \frac{1}{(k + 1)} \cdot \frac{n!}{k!(n - k - 1)!} \\ &= \frac{p}{q} \cdot \frac{1}{k + 1} \cdot p^k \cdot q^{n-k} \cdot (n - k) \cdot \frac{n!}{k!(n - k)!} \\ &= \frac{p}{q} \cdot \frac{n - k}{k + 1} \cdot \mathbb{P}(S_n = k) \end{aligned}$$

□

Con questa formula non siamo costretti a calcolare i fattoriali. Si parte con $\mathbb{P}(S_n = 0)$:

$$\begin{aligned} P(S_n = 0) &= \binom{n}{0} p^0 q^n = q^n \\ P(S_n = 1) &= \frac{p}{q} \cdot \frac{n - 1}{2} \cdot \mathbb{P}(S_n = 0). \end{aligned}$$

3.2 Distribuzione geometrica - legge geometrica

Ci troviamo in questa situazione $(\{T, C\}^\infty, \mathcal{F} = \sigma[(T_n)_{n \in \mathbb{N}}], \mathbb{P})$ (lancio di una moneta *truccata* un numero indefinito di volte). Definiamo il numero aleatorio $\mathcal{T} \geq 1 : X_{\mathcal{T}}$ che indica in che posizione si presenta una T per la prima volta nella successione. Ad esempio $X_{\mathcal{T}} = 1$ indica che T si è presentato in prima posizione. Sia $p := \mathbb{P}(T)$ la probabilità che in un lancio esce T (testa) e $q := \mathbb{P}(T^c) = \mathbb{P}(C)$ la probabilità che in un lancio esce C (croce). Il numero aleatorio \mathcal{T} ha i seguenti valori $S_{\mathcal{T}} = \{1, 2, 3, \dots\} = \mathbb{N}$ con $n \in \mathbb{N}$ (coincide con l'insieme dei numeri naturali). La probabilità che $\mathbb{P}(\mathcal{T} = n)$ è:

$$\mathbb{P}(\mathcal{T} = n) = \mathbb{P}(C_1 \cap C_2 \cap \dots \cap C_{n-1} \cap T_n) = \mathbb{P}(C_1) \cdot \mathbb{P}(C_2) \cdot \dots \cdot \mathbb{P}(C_{n-1}) \cdot \mathbb{P}(T_n) = q^{n-1} \cdot p$$

e viene chiamata **legge geometrica** (o **distribuzione geometrica**) e si indica con $\mathcal{T} \sim G(p)$. Il numero aleatorio \mathcal{T} ha legge geometrica.

Un numero aleatorio \mathcal{T} descritto da una qualsiasi distribuzione deve rispettare:

1. $n \in \mathbb{N}, \mathbb{P}(\mathcal{T} = n) > 0$
2. $\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{T} = n) = 1$ (converge a 1)

Dimostriamo il punto 2. (dato che non sempre è così) per la distribuzione geometrica:

$$\begin{aligned} \sum_{n=1}^{\infty} p \cdot q^{n-1} &= p \sum_{n=1}^{\infty} q^{n-1} = p \sum_{n=0}^{\infty} q^n \\ \text{(serie geometrica di ragione } q) &= p \cdot \frac{1}{1-q} = p \cdot \frac{1}{p} = 1 \end{aligned}$$

Calcoliamo la media del numero aleatorio \mathcal{T} :

$$\mathbb{E}(\mathcal{T}) = \sum_{n=1}^{\infty} n \cdot p \cdot q^{n-1} = p + 2pq + 3pq^2 + 4pq^3 + \dots = p \cdot (1 + 2q + 3q^2 + 4q^3 + \dots)$$

Il secondo fattore (quello a sinistra di p nell'ultimo passaggio) può essere visto in questa maniera:

$$\begin{aligned} 1 + q + q^2 + q^3 + \dots &= \frac{1}{p} \\ + q + q^2 + q^3 + \dots &= q \cdot (1 + q + q^2 + \dots) = \frac{q}{p} \\ + q^2 + q^3 + \dots &= q^2 \cdot (1 + q + q^2 + \dots) = \frac{q^2}{p} \\ + q^3 + \dots &= q^3 \cdot (1 + q + q^2 + \dots) = \frac{q^3}{p} \\ &\dots \end{aligned}$$

Quindi:

$$\mathbb{E}(\mathcal{T}) = p \cdot \left(\frac{1}{p} + \frac{q}{p} + \frac{q^2}{p} + \frac{q^3}{p} + \dots \right) = \frac{1}{p}$$

Riepilogo serie geometrica

$$h \in \mathbb{R}, \quad 1 + h + h^2 + \dots + h^n = \frac{1 - h^{n+1}}{1 - h} \quad (\text{progressione geometrica di ragione } h - \text{somma parziale}).$$

$$|h| < 1 \quad S = \lim_{n \rightarrow \infty} \frac{(1 - h)^{n+1}}{1 - h} = \frac{1}{1 - h} \quad (\text{è il limite della somma parziale})$$

3.3 Distribuzione di Pascal

Sia W_k un numero aleatorio che rappresenta *il numero delle prove per vedere la k -esima testa* con $k \in \mathbb{N}$. E' chiaro che $W_1 = \mathcal{T} \sim G(p)$ dato che \mathcal{T} rappresenta proprio il numero delle prove per vedere la prima testa nella successione. Inoltre, $\mathbb{P}(W_k = k - 1) = 0$ dato che abbiamo bisogno di almeno k prove. Dunque, lo spettro del numero aleatorio W_k è uguale a $S_{W_k} = \{k, k + 1, k + 2, \dots\}$.

Vogliamo adesso determinare la probabilità di $W_k = n$ con $n \geq k$:

$$\mathbb{P}(W_k = n) = \mathbb{P}(S_{n-1} = k - 1, X_n = 1) = \mathbb{P}(S_{n-1} = k - 1) \cdot \mathbb{P}(X_n = 1) = \binom{n-1}{k-1} p^{k-1} q^{n-k} p = \binom{n-1}{k-1} p^k q^{n-k}$$

Il numero aleatorio W_k ha **distribuzione di Pascal** e si indica con $W_k \sim \text{Pascal}(p, k)$

Possiamo esprimere W_k utilizzando il numero aleatorio \mathcal{T} . Osserviamo che:

$$\mathcal{T}^{(1)} \sim \mathcal{T}^{(2)} \sim \dots \mathcal{T}^{(k-1)} \sim \mathcal{T}^{(k)} \sim G(p)$$

dove con $\mathcal{T}^{(1)}$ stiamo dicendo "quante prove per osservare la prima T ". Poi si azzerà il contatore delle prove e con $\mathcal{T}^{(2)}$ indichiamo "quante prove per osservare la seconda T ", poi si azzerà il contatore delle prove e così via... Tutte queste $\mathcal{T}^{(i)}$ hanno la stessa distribuzione (la legge geometrica).

Dunque, la media di W_k può essere espressa così. Innanzitutto osserviamo che W_k è uguale a:

$$W_k = \sum_{i=1}^k \mathcal{T}^{(i)}$$

E quindi la media di W_k può essere calcolata così (utilizzando la *proprietà di linearità della media*):

$$\mathbb{E}(W_k) = \mathbb{E}\left(\sum_{i=1}^k \mathcal{T}^{(i)}\right) = \sum_{i=1}^k \mathbb{E}(\mathcal{T}^{(i)}) = \sum_{i=1}^k \mathbb{E}(T) = \mathbb{E}(T) \cdot \sum_{i=1}^k 1 = \frac{k}{p}$$

3.4 Distribuzione di probabilità uniforme e discreta

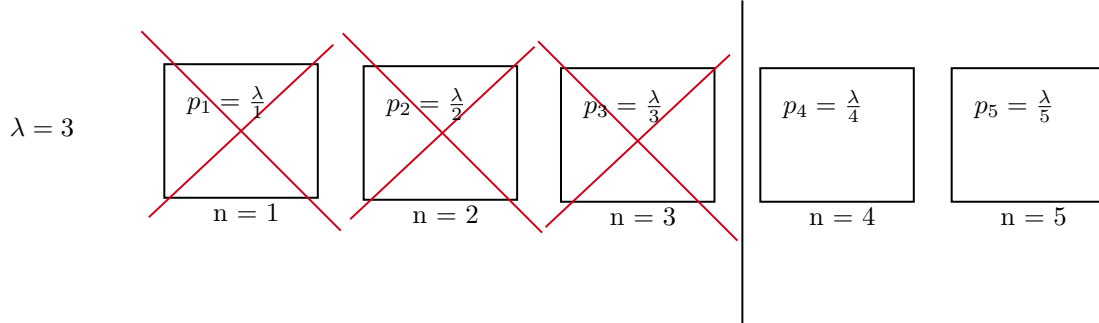
Sia ξ il lancio di un dado *onesto*. Sia Z_1 la variabile aleatoria che rappresenta il punteggio ottenuto. Lo spettro di Z_1 è $S_{Z_1} = \{1, 2, \dots, 6\}$ e la probabilità di $\mathbb{P}(Z_1 = i) = \frac{1}{6}$ con $i = 1, 2, \dots, 6$. Questo è un numero aleatorio **casuale**: tutti i valori finiti dello spettro hanno la stessa probabilità di verificarsi. Un numero aleatorio con queste proprietà si dice che ha **distribuzione uniforme discreta** e si denota con:

$$Z_1 \sim U_\alpha(1, 2, 3, 4, 5, 6)$$

3.5 Distribuzione di Poisson

Proseguiamo la panoramica con un'altra importante variabile aleatoria discreta che assume valori interi non negativi. Ci troviamo in questa situazione $(\{T, C\}^\infty, \mathcal{F} = \sigma[(T_n)_{n \in \mathbb{N}}], \mathbb{P})$ (lancio di una moneta *truccata* un numero indefinito di volte). Facciamo adesso assumere alla variabile aleatoria S_n valori interi non negativi, cioè il suo spettro sarà $S_{S_n} = \mathbb{N}$. La sua funzione di massa di probabilità sarà $\mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}$ con $k \in \mathbb{N}$.

Scegliamo però la probabilità che si verifichi una croce in un modo particolare facendola dipendere da un parametro $\lambda > 0$. Ad esempio:



Fissato $\lambda = 3$ è chiaro che bisogna scartare 1, 2, 3 perché altrimenti la probabilità assume rispettivamente i valori 3, $\frac{3}{2}$ e $\frac{3}{3}$.

La regola è che il prodotto tra n e la probabilità p_n è uguale a λ :

$$n \cdot p_n = \lambda$$

In questa maniera, se si manda $n \rightarrow \infty$, la probabilità p tende a zero $p_n \rightarrow 0$. Aumentando il numero delle prove la probabilità della testa diminuisce. Questo ci permette di calcolare la probabilità di S_n quando $n \in \mathbb{N}$:

$$\begin{aligned} \mathbb{P}(S_n = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p^k q^{n-k} = \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^k} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)(n-k)!}{n \cdot n \cdot n \dots n \cdot (n-k)!} \cdot \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^k} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{n \cdot n \cdot n \dots n} \cdot \frac{\lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^n}{\lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^k} \\ (\text{il limite al denominatore è } 1) &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n-1}{n} \cdot \lim_{n \rightarrow \infty} \frac{n-2}{n} \dots \lim_{n \rightarrow \infty} \frac{n-k+1}{n} \cdot \lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^n \\ (\text{tutti i limiti sono uguali a } 1) &= \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^n \\ &= \frac{\lambda^k}{k!} \cdot e^{-\lambda} \end{aligned}$$

L'ultimo passaggio è stato ottenuto dal numero di nepero⁴.

Definizione 3.5.1 (Numero aleatorio di Poisson)

Una variabile aleatoria y che assuma i valori $0, 1, 2, \dots$, è una variabile aleatoria *di Poisson o poissoniana* di parametro λ , $\lambda > 0$, se la sua funzione di massa di probabilità è data da:

$$\mathbb{P}(y = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

e si denota $y \sim \text{Poisson}(\lambda) \sim \Pi(\lambda)$

⁴ $e = \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$

E' immediato verificare che la funzione di massa di probabilità di un numero aleatorio di Poisson y è accettabile, infatti:

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

Ricordando che:

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda} \text{ (sviluppo in serie di potenze dell'esponenziale per tutti i numeri } \lambda \text{)}$$

Calcoliamo adesso la media di $y \sim \Pi(\lambda)$:

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

Quindi λ è la media.

3.5.1 Calcolo esplicito della distribuzione di Poisson

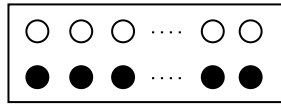
Se y è un numero aleatorio di Poisson $y \sim \Pi(\lambda)$ di media λ , allora:

$$\mathbb{P}(y = k+1) = e^{-\lambda} \cdot \frac{\lambda^{k+1}}{(k+1)!} = \frac{\lambda}{k+1} \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \frac{\lambda}{k+1} \cdot \mathbb{P}(y = k)$$

Si parte da $\mathbb{P}(y = 0) = e^{-\lambda} \cdot \frac{\lambda^0}{0!} = e^{-\lambda}$

3.6 Numeri aleatori ipergeometrici

Consideriamo una scatola con $|B|$ palline bianche e $|N|$ palline nere. Il numero totale di palline è $|T|$.



Consideriamo adesso il numero aleatorio X_1 :

$$X_1 := \begin{cases} 1, & \text{se esce Bianca} \\ 0, & \text{altrimenti} \end{cases}$$

Supponendo che ad ogni estrazione la pallina pescata viene rimessa nella scatola scutendola (*con rimpiazzo*), il numero aleatorio X_1 è chiaramente un numero aleatorio di Bernoulli $X_1 \sim \text{Bernoulli}(p) \sim B(1, p)$. In questo caso però la probabilità p non è un più un numero reale nell'intervallo $[0, 1]$ ma *un numero razionale*. Infatti la probabilità che $X_1 = 1$ e cioè che esca la pallina bianca è:

$$p = \frac{|B|}{|T|}$$

Consideriamo adesso le *estrazioni senza rimpiazzamento o in blocco* (equiprobabili). Definiamo X come il numero delle biglie bianche estratte. Poniamo $n \in \mathbb{N}$ uguale alla dimensione del blocco (il numero totale di estrazioni fatte). Poniamo uguale ad a il numero totale di palline bianche presenti nella scatola, con b il numero totale di palline nere presenti nella scatola e con $a + b$ il numero totale di palline presenti nella scatola. Vogliamo calcolare la probabilità in generale $\mathbb{P}(X = k)$ cioè la probabilità che tra le palline pescate a caso in blocco (blocco da n palline) ci siano esattamente k palline bianche:

$$\mathbb{P}(X = k) = \frac{\binom{a}{k} \cdot \binom{b}{n-k}}{\binom{a+b}{n}}$$

E' chiaro che se ci sono 5 palline bianche, non si può calcolare il coefficiente binomiale $\binom{5}{6}$ e quindi per convenzione poniamo:

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!}, & \text{se } k \leq n \\ 0, & \text{se } k > n \end{cases}$$

Lo spettro del numero aleatorio X (ovvero i valori che può assumere k) è definito da questo insieme:

$$S_X = \{\max(0, n - b), \dots, \min(n, a)\}$$

Esempio 3.6.1

Consideriamo una scatola con N batterie accettabili e M batterie difettose. Estraiamo un blocco di n batterie. La variabile X conta il numero delle batterie accettabili.

$$k \in \mathbb{N} \quad \mathbb{P}(X = k) = \frac{\binom{N}{k} \cdot \binom{M}{n-k}}{\binom{N+M}{n}}$$

Supponiamo che ci siano 15 batterie accettabili e 5 batterie guaste. Peschiamo a caso 6 batterie. Qual'è la probabilità che tra le 6 batterie ce ne siano *almeno* 4?

Il valore k deve essere compreso tra 4 e 6 (estremi inclusi) per far sì che l'evento "almeno 4 batterie accettabili" si avveri. Dunque:

$$\begin{aligned} \mathbb{P}(4 \text{ batterie accettabili}) &= \mathbb{P}(X = 4) \cup \mathbb{P}(X = 5) \cup \mathbb{P}(X = 6) \\ &= \sum_{k=4}^6 \frac{\binom{15}{k} \cdot \binom{5}{6-k}}{\binom{20}{6}} \\ &= \frac{\binom{15}{4} \cdot \binom{5}{2} + \binom{15}{5} \cdot \binom{5}{1} + \binom{15}{6} \cdot \binom{5}{0}}{\binom{20}{6}} \end{aligned}$$

Le variabili X così descritte sono **variabili aleatorie ipergeometriche** e si indicano così:

$$X \sim I_G(a, b, n)$$

Consideriamo X_k il numero aleatorio definito così:

$$X_k := \begin{cases} 1, & \text{se esce bianca con probabilità } \frac{a}{a+b} \\ 0, & \text{se esce nera con probabilità } \frac{b}{a+b} \end{cases}$$

Dove $k = 1, \dots, n$ ed indica la k -esima pallina. Ad esempio:

$$\begin{aligned} \mathbb{P}(X_1 = 1) &= \frac{a}{a+b} \quad (\text{la 1}^\circ \text{ pallina è bianca}) \\ \mathbb{P}(X_2 = 1) &= \frac{a}{a+b} \quad (\text{la 2}^\circ \text{ pallina è bianca}) \end{aligned}$$

Vogliamo dimostrare che la probabilità rimane sempre la stessa indipendentemente dal numero di palline già estratte. Dimostriamolo solo per X_2 (ma vale in generale):

$$\begin{aligned} \mathbb{P}(X_2 = 1) &= \mathbb{P}(X_2 = 1 \cap [\{X_1 = 0\} \cup \{X_1 = 1\}]) \\ &= \mathbb{P}(\{X_2 = 1\} \cap \{X_1 = 0\} \cup \{X_2 = 1\} \cap \{X_1 = 1\}) \\ &= \mathbb{P}(\{X_2 = 1\} \cap \{X_1 = 0\}) + \mathbb{P}(\{X_2 = 1\} \cap \{X_1 = 1\}) \\ &= \mathbb{P}(X_1 = 0) \cdot \mathbb{P}(X_2 = 1 | X_1 = 0) + \mathbb{P}(X_1 = 1) \cdot \mathbb{P}(X_2 = 1 | X_1 = 1) \\ &= \frac{b}{a+b} \cdot \frac{a}{a+b-1} + \frac{a}{a+b} \cdot \frac{a-1}{a+b-1} \\ &= \frac{a(b+a-1)}{(a+b) \cdot (a+b-1)} = \frac{a}{a+b} \end{aligned}$$

In generale, per ogni $k = 1, \dots, n$ $\mathbb{P}(X_k = 1) = \frac{a}{a+b}$. Non ci conosce quello che è successo prima.

La media $\mathbb{E}(X_k)$:

$$\mathbb{E}(X_k) = 1 \cdot \frac{a}{a+b} + 0 \cdot \frac{b}{a+b} = \frac{a}{a+b}$$

Proprio come la media di un numero aleatorio di Bernoulli è la probabilità p . Non a caso infatti:

$$\begin{aligned} X &\sim B(n, p) & \mathbb{E}(X) &= n \cdot p \\ X &\sim I_G(n, a, b) & \mathbb{E}(X) &= n \cdot p = n \cdot \frac{a}{a+b} \end{aligned}$$

I due schemi si differenziano sulla *varianza*. I valori che può assumere k in uno schema ipergeometrico generalmente sono di meno rispetto ad uno schema binomiale. I due schemi però tendono a coincidere se a e b sono grandi.

Modelli di variabili aleatorie continue

4 Modelli di variabili continue

Fino ad adesso abbiamo trattato solo variabili aleatorie discrete la quale legge o distribuzione di probabilità \mathbb{P}_X (dove X è una variabile aleatoria) è ricostruita mediante il suo spettro S_X dove la probabilità dei suoi elementi è già fissata: $x \in S_X$, $\mathbb{P}(X = x) > 0$. Dunque la legge \mathbb{P}_X può essere ricostruita tramite queste due informazioni: se prendiamo un boreliano (perché \mathbb{P}_X misura proprio i boreliani), il valore di probabilità di tale boreliano è la sommatoria delle probabilità degli elementi contenuti nell'intersezione del boreliano con lo spettro S_X

$$B \in \mathcal{B}, \quad \mathbb{P}_X(B) = \sum_{x \in B \cap S_X} \mathbb{P}(X = x).$$

Quindi la ricostruzione avviene in questo modo, in particolare si ricostruisce la funzione di distribuzione F_X : basta scegliere un B uguale ad una semi-retta. Questo è un caso particolare, perché la formula che abbiamo visto sopra lavora su tutti i boreliani, mentre F_X si riferisce ai boreliani del tipo $] - \infty, x]$ dove x è l'origine.

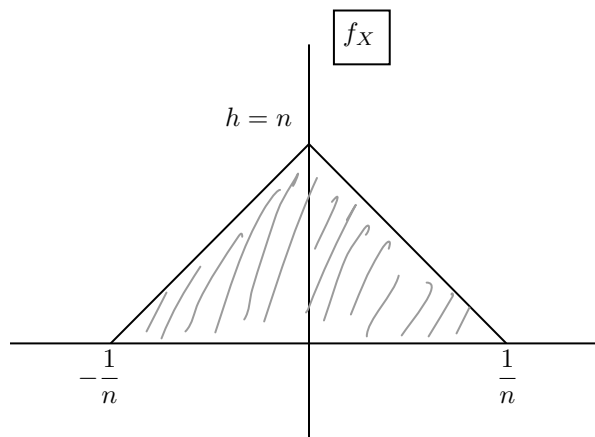
Per quanto riguarda le variabili aleatorie **continue**, la legge \mathbb{P}_X è ricostruita per il tramite di una funzione chiamata **funzione di densità di probabilità** f_X (spesso abbreviata con **f.d.p.**). Siccome ci stiamo muovendo nel continuo, la probabilità che X appartiene a B (dove X è una variabile aleatoria continua e B un boreliano) si misura non più con una sommatoria ma con un integrale:

$$\mathbb{P}_X(B) = \int_B f_X(x) dx$$

Non bisogna interpretare la funzione di densità di probabilità f_X come una probabilità. A differenza delle variabili aleatorie che offrono un insieme di punti finiti o numerabili dotati di massa di probabilità maggiore di zero (cioè con probabilità maggiore di zero), in una variabile aleatoria continua nessun punto dell'asse reale è dotato di massa di probabilità maggiore di zero: la probabilità in questo caso si trova in un intervallo o nell'unione di più intervalli (può essere anche tutto l'asse reale oppure l'intervallo $[0, 1]$ oppure in un intervallo $[0, +\infty]$). Quindi la probabilità non si distribuisce sui singoli punti (come nelle variabili aleatorie discrete) ma su un intervallo reale. Non è possibile lavorare sui singoletti ma bisogna lavorare sugli intervalli.

Esistono molti tipi di variabili aleatorie oltre a quelle discrete e continue. Esistono variabili aleatorie che sono anche un misto tra discreto e continuo (divisa in parti discrete e parti continue). Esiste anche una particolare variabile aleatoria chiamata **variabile aleatoria singolare**: in teoria della probabilità, una variabile casuale singolare (o continua singolare) è una variabile casuale molto particolare, che è raro incontrare negli studi pratici, in quanto ha un comportamento abbastanza "patologico" (introdotta dal professore Buonocore per curiosità, non obbligatorio lo studio).

Consideriamo ad esempio il seguente triangolo:



Osserviamo che l'altezza h è proprio uguale ad n perché deve sicuramente accadere che:

$$\mathbb{P}_X \left(\left[-\frac{1}{n}, \frac{1}{n} \right] \right) = \int_{-\frac{1}{n}}^{\frac{1}{n}} f_X(x) dx = 1$$

Siccome l'integrale misura proprio l'area sottostante alla funzione in un determinato intervallo, allora l'area del

triangolo deve essere 1. Quindi l'altezza si ricava così (la base è $\frac{2}{n}$):

$$\begin{aligned}\frac{b \cdot h}{2} &= 1 \\ h &= \frac{2}{b} \\ h &= \frac{2}{\frac{2}{n}} = n\end{aligned}$$

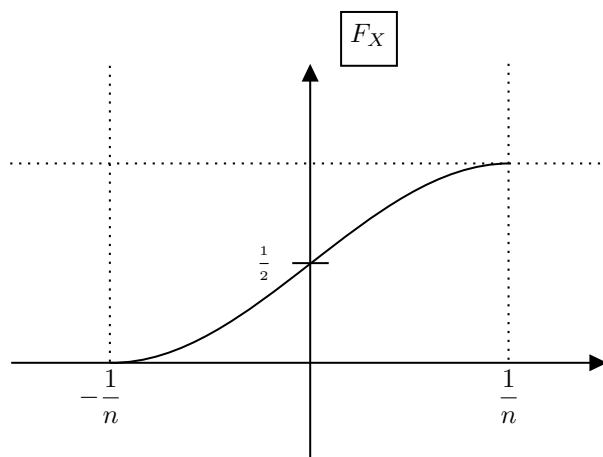
Adesso se facciamo l'operazione di limite per $n \rightarrow \infty$, l'intervallo $[-\frac{1}{n}, \frac{1}{n}]$ si restringe sempre di più e l'altezza del triangolo aumenta. L'intervallo si avvicina sempre più a 0 e l'area (che deve sempre valere 1) diventa una freccia infinitamente lunga. Queste sono le variabili aleatorie singolari.

Torniamo adesso alle variabili aleatorie continue. Vediamo come **la funzione di distribuzione di una variabile aleatoria continua si ricostruisce attraverso un'integrazione**:

$$B =] - \infty, x], \quad \mathbb{P}_X(] - \infty, x]) = \mathbb{P}(X \leq x) = F_X(x) = \int_{-\infty}^x f_X(y) dy$$

Da questo risultato si deduce che la funzione di distribuzione F_X è la funzione integranda di f_X (F_X è una primitiva della funzione di densità di probabilità f_X). Mentre f_X rispetto a F_X è la derivata.

Ritorniamo all'esempio del triangolo e proviamo a fare l'operazione di funzione integranda sulla funzione di densità di probabilità che descrive il triangolo. Il risultato è questo:



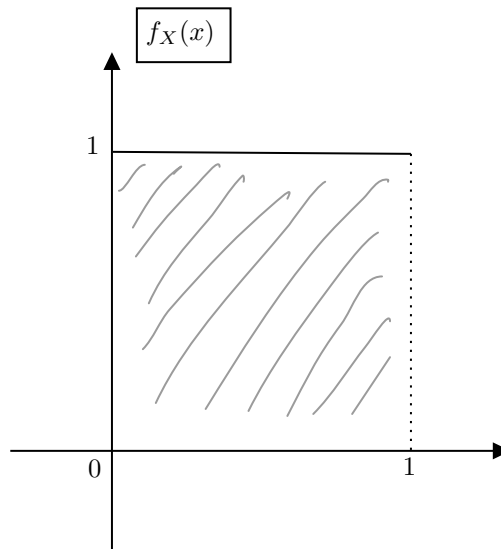
Tornando al discorso della variabile aleatoria singolare, se facciamo tendere $n \rightarrow \infty$, l'intervallo $[-\frac{1}{n}, \frac{1}{n}]$ si stringe sempre di più fino ad arrivare ad un intervallo da 0 a 0 e "l'escursione" da 0 a 1 (sulle ordinate) avviene solo nel punto di ascissa 0 (ed ecco la particolarità e la stranezza di una variabile aleatoria singolare).

4.1 Variabile aleatoria continua con legge uniforme

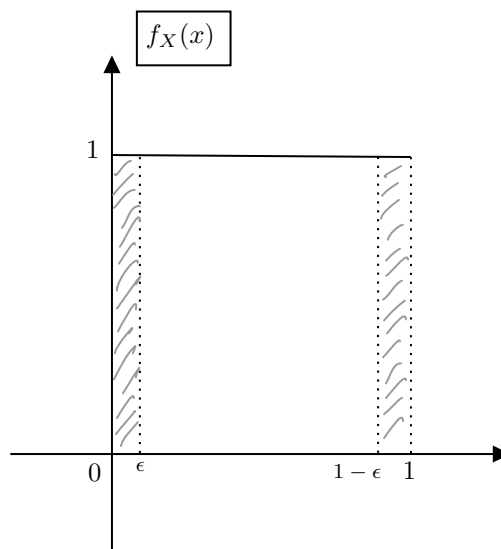
Su questo tipo di variabile non è possibile fare nessuna previsione. Una variabile aleatoria con legge uniforme si indica con $X \sim U(0, 1)$ se l'intervallo considerato è $[0, 1]$. Per assegnare una legge continua, bisogna ovviamente determinare la funzione di densità di probabilità:

$$f_X(x) = \begin{cases} 1, & \text{se } x \in (0, 1) \\ 0, & \text{altrimenti} \end{cases}$$

Il grafico di tale funzione $f(x)$ è:



Non possiamo fare nessuna previsione. L'area del quadrato è $b \cdot h = 1$. Quello che possiamo osservare è che la variabile uniforme non favorisce alcun intervallo di medesima ampiezza. Ad esempio gli intervalli $]0, \epsilon[$ e $]1 - \epsilon, 1[$ hanno la stessa probabilità:



Infatti:

$$\begin{aligned} \mathbb{P}_X(]0, \epsilon[) &= \int_0^\epsilon 1 \cdot dx = \epsilon \\ \mathbb{P}_X(]1 - \epsilon, 1[) &= \int_{1-\epsilon}^1 1 \cdot dx = 1 \cdot (1 - 1 + \epsilon) = \epsilon \end{aligned}$$

Ecco perché si chiama uniforme: perché la sua funzione di densità di probabilità è piatta.

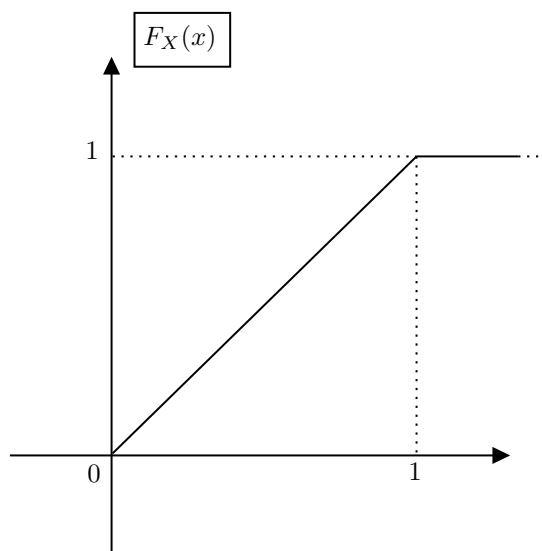
Adesso facciamo l'operazione di funzione integranda su tale funzione $f_X(x)$:

$$x < 0, \quad F_X(x) = \int_{-\infty}^x f_X(y) \cdot dy = \int_{-\infty}^x 0 \cdot dy = 0$$

$$0 < x < 1, \quad F_X(x) = \int_{-\infty}^x f_X(y) \cdot dy = \int_{-\infty}^0 f_X(y) \cdot dy + \int_0^x f_X(y) \cdot dy = 0 + (1 \cdot x) = x$$

$$x \geq 1, \quad F_X(x) = 1$$

Dunque il corrispondente grafico di F_X è il seguente:



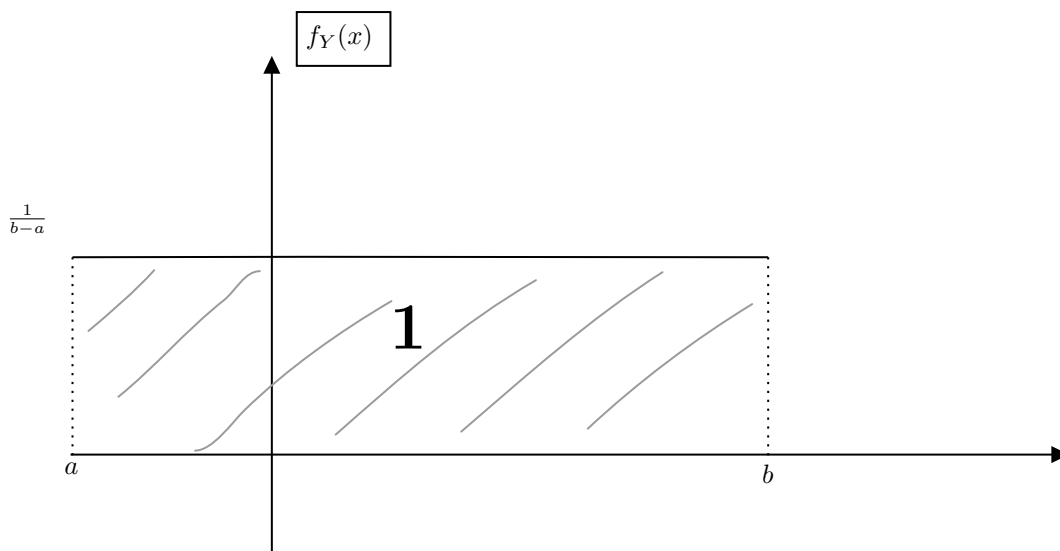
Possiamo anche considerare la legge uniforme in un **intervallo generico** $]a, b[$. Chiamiamo tale variabile $Y \sim U(a, b)$ con $a < b$. Essendo un intervallo generico, la funzione di densità di probabilità $f_Y(x)$ formerà molto probabilmente un rettangolo. L'altezza di questo rettangolo sarà:

$$h = \frac{1}{\text{base}} = \frac{1}{b - a}$$

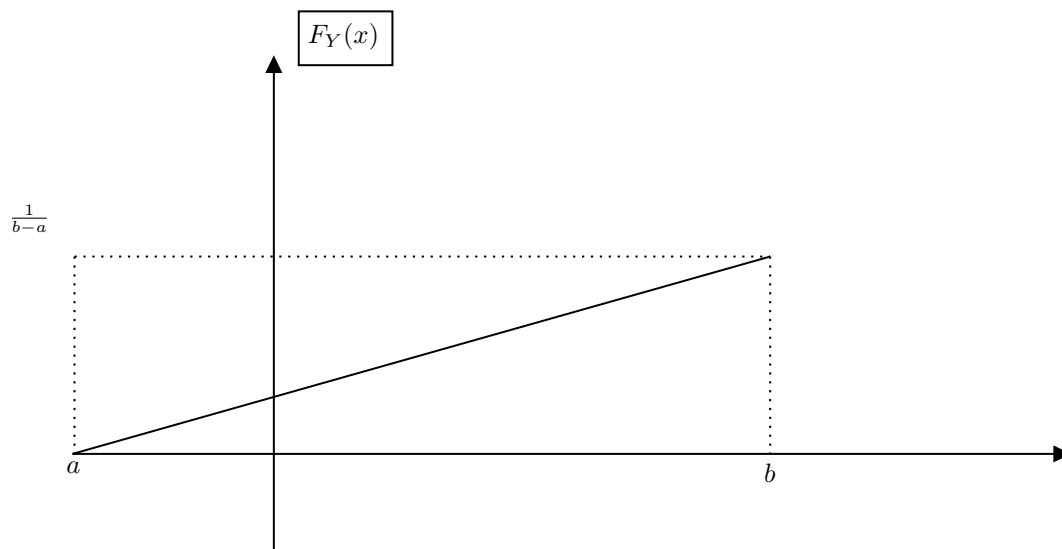
quindi la funzione $f_Y(x)$ sarà:

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

Il corrispondente grafico sarà:



La funzione integranda F_Y di $f_Y(x)$ sarà questa:



Ma è davvero necessario utilizzare la variabile aleatoria Y ? Possiamo ricavare **tutte le leggi uniformi servendoci solo della variabile aleatoria X con questa formula:**

$$Y = a + (b - a)X \quad \text{con } a < b$$

Parliamo adesso della media. Quando stavamo nel caso discreto avevamo la somma dei prodotti valore dello spettro per la rispettiva probabilità. Nel caso continuo dobbiamo calcolare un integrale. Facciamo una piccola osservazione: abbiamo detto che tutti i punti dell'asse reale hanno probabilità zero, infatti:

$$\mathbb{P}(a \leq X \leq a) = 0 \quad \text{perché l'area è 0 ovviamente}$$

quindi questa è una dimostrazione matematica di questa affermazione. Allora per ottenere la probabilità di un singolo punto dobbiamo considerare necessariamente un piccolo intervallo intorno ad esso. Quindi la media di un numero aleatorio continuo uniforme sarà $(f_X(x) \cdot dx)$ non è zero perché dx è un intervallo infinitesimo, rappresenta la probabilità dell'intervallino attorno ad x :

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} x \cdot f_X(x) \cdot dx = \int_0^1 x \cdot 1 \cdot dx \\ &= \int_0^1 x \cdot dx \\ &= \frac{x^2}{2} \Big|_0^1 \\ &= \frac{1}{2}(1^2 - 0^2) = \frac{1}{2} \end{aligned}$$

La media di una variabile uniforme vale $\frac{1}{2}$ proprio perché non favorisce nessuno.

Qual'è la **media della variabile aleatoria** Y ? Possiamo sfruttare l'equazione che mette in relazione X con Y per calcolarla:

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}[a + (b - a)X] \\ \text{(l'integrale di una somma è la somma degli integrali)} &= \mathbb{E}(a) + \mathbb{E}[(b - a)X] \\ \text{(la media di una costante è una costante)} &= a + \mathbb{E}[(b - a)X] \\ &= a + (b - a)\mathbb{E}(X) \\ &= a + \frac{b - a}{2} = \frac{a + b}{2}\end{aligned}$$

Calcoliamo la media utilizzando adesso la **funzione generatrice dei momenti**. Trasformiamo X in e^{tX} . Dunque:

$$\begin{aligned}t \in \mathbb{R} - \{0\}, \quad \phi_X(t) = \mathbb{E}(e^{tX}) &= \int_{\mathbb{R}} e^{tx} \cdot f_X(x) \cdot dx = \int_0^1 e^{tx} \cdot f_X(x) \cdot dx \\ &= \int_0^1 e^{tx} \cdot 1 \cdot dx \\ &= \frac{e^{tx}}{t} \Big|_0^1 = \frac{1}{t}(e^t - 1)\end{aligned}$$

La funzione generatrice dei momenti serve per calcolare la media $\mathbb{E}(X)$ e anche la media $\mathbb{E}(X^2)$. Per fare ciò, dobbiamo calcolare tale funzione nel punto $t = 0$ ma in questo caso abbiamo escluso questo valore (perché non può stare al denominatore). Allora quello che dobbiamo fare è calcolare la funzione generatrice dei momenti con $t = 0$ per continuità (sfruttando il teorema di De Hopital):

$$\lim_{t \rightarrow 0} \frac{e^t - 1}{t} = \lim_{t \rightarrow 0} \frac{e^t}{1} = 1$$

Stessa cosa per $\phi'(t)$:

$$\phi'(t) = \frac{e^t \cdot t - (e^t - 1)}{t^2} = \frac{te^t - e^t + 1}{t^2}$$

Quindi $\phi'(0)$ sarà (sfruttando sempre De Hopital):

$$\phi'(0) = \lim_{t \rightarrow 0} \frac{te^t + 1 - e^t}{t^2} = \lim_{t \rightarrow 0} \frac{e^t + te^t - e^t}{2t} = \frac{1}{2} = \mathbb{E}(X)$$

Calcoliamo adesso $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \int_0^1 x^2 \cdot f_X(x) \cdot dx = \int_0^1 x^2 \cdot 1 \cdot dx = \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}(1-0) = \frac{1}{3}$$

Quanto vale adesso la media di Y^2 ?

$$\begin{aligned}\mathbb{E}(Y^2) &= \mathbb{E}\{[a + (b-a)X]^2\} = \mathbb{E}\{[a^2 + (b-a)^2X^2 + 2a(b-a)X]\} = a^2 + (b-a)^2\mathbb{E}(X^2) + 2a(b-a)\mathbb{E}(X) \\ &= a^2 + \frac{(b-a)^2}{3} + a(b-a) \\ &= \frac{3a^2 + b^2 + a^2 - 2ab + 3ab - 3a^2}{3} \\ &= \frac{a^2 + ab + b^2}{3}\end{aligned}$$

Abbiamo trovato la media di un quadrato di una variabile uniforme generica nell'intervallo (a, b) .

4.1.1 Definizione varianza - varianza legge uniforme

Ci servirebbe capire adesso **quant'è affidabile la media di X** (di quanto è un buon previsore del risultato). Il risultato $\frac{1}{2}$ è un buon esito su cui scommettere? Questa preziosa informazione viene chiamata **varianza di X** e si denota con $\mathbb{D}^2(X)$ (questa informazione ha senso solo se la media della variabile aleatoria presa in considerazione esiste). Essa fornisce una misura della variabilità dei valori assunti dalla variabile stessa. Nello specifico, la misura di quanto essi si discostino quadraticamente rispettivamente dalla media aritmetica o dal valore atteso. La varianza è definita così:

$$\mathbb{D}^2(X) := \mathbb{E}\{[X - \mathbb{E}(X)]^2\}$$

dove l'argomento viene chiamato **scarto**. Dunque facciamo il quadrato dello scarto (è quindi un'altra variabile aleatoria). Facciamo la media del quadrato dello scarto. Tale valore più è vicino a 0 più la media prevede bene il risultato. Tale valore **misura la variabilità della legge della variabile aleatoria**.

Se andiamo a fare i calcoli scopriamo che $\mathbb{D}^2(X)$ è uguale a:

$$\mathbb{D}^2(X) = \mathbb{E}\{[X^2 - 2\mathbb{E}(X)X + \mathbb{E}^2(X)]\} = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

dunque:

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

quindi è la media del secondo ordine meno il quadrato della media del primo ordine. Nella legge uniforme, la varianza di X vale:

$$\mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Il valore della varianza di X ci permette di calcolare quella che viene chiamata **deviazione standard** (o scarto quadratico medio, o scarto tipo, o scostamento quadratico medio) che è uno dei modi per esprimere la dispersione dei dati intorno ad un indice di posizione, quale può essere, ad esempio, la media aritmetica o una sua stima. **Lo scarto quadratico medio è la radice quadrata della varianza**, quindi è definita come:

$$\sigma_X := \sqrt{\mathbb{D}^2(X)}$$

La deviazione standard della variabile aleatoria con legge uniforme X è:

$$\sigma_X = \sqrt{\frac{1}{12}} = \frac{1}{2\sqrt{3}} \approx 0.3$$

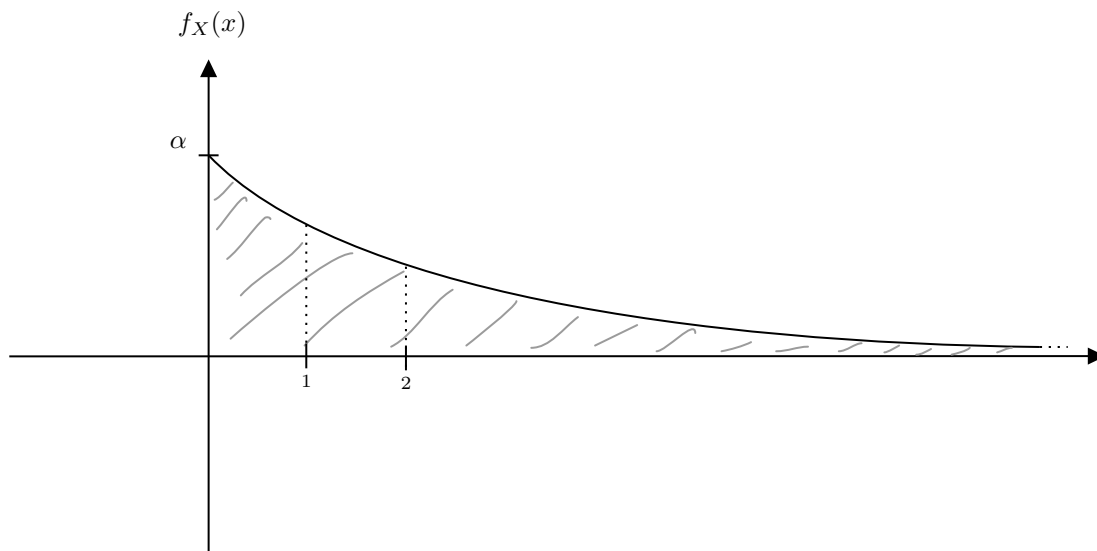
Questa informazione dice che i valori probabili sono $]\frac{1}{2} - 0.3, \frac{1}{2} + 0.3[$ o meglio quelli che appartengono a $]0.2, 0.8[$. Tale intervallo è molto ampio (basta paragonarlo all'intervallo $]0, 1[$, è l'80%). Questo ci dice che la media $\frac{1}{2}$ NON è una buona previsione.

4.2 Variabile aleatoria esponenziale

Una variabile aleatoria continua la cui funzione di densità di probabilità è data da:

$$f_X(x) = \begin{cases} \alpha \cdot e^{-\alpha x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$$

per un opportuno valore $\alpha > 0$, si dice **esponenziale** con parametro (o *intensità*) α . Una variabile aleatoria che ha questa legge viene comunemente indicata con $X \sim \text{Esp}(\alpha)$ con $\alpha > 0$. Il supporto della funzione di densità di probabilità di una variabile con legge esponenziale è $]0, +\infty[$. Il suo grafico è:



Calcoliamo la probabilità che il numero aleatorio X con questa legge assuma valore maggiore o uguale di 2:

$$\mathbb{P}(X \geq 2) = \int_2^{+\infty} \alpha e^{-\alpha x} dx = -e^{-\alpha x} \Big|_2^{+\infty} = e^{-\alpha x} \Big|_{+\infty}^2 = e^{-2\alpha} - \lim_{x \rightarrow +\infty} e^{-\alpha x} = e^{-2\alpha}$$

Sostanzialmente abbiamo calcolato l'area sottostante al grafico da nell'intervallo $]2, +\infty[$ (ricordiamo inoltre che per essere una distribuzione, tutta l'area deve essere uguale a 1).

Adesso vogliamo vedere come si comporta la **funzione di distribuzione** di un numero aleatorio con legge esponenziale. Ricordiamo che $F_X(x)$ è uguale a:

$$F_X(x) = \mathbb{P}_X(X \leq x) = \int_{-\infty}^x f_X(y) dy$$

E' chiaro che se $x \leq 0$, non esiste area da calcolare nel grafico quindi l'integrale vale 0. Mentre se $x > 0$ (eliminiamo tutto quello che viene prima dello zero perché senza area):

$$F_X(x) = \int_0^x \alpha e^{-\alpha y} dy = e^{-\alpha y} \Big|_x^0 = 1 - e^{-\alpha x}$$

Concludiamo che la funzione di distribuzione di un numero aleatorio X con legge esponenziale ha questo comportamento:

$$F_X(x) = \begin{cases} 1 - e^{-\alpha x}, & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Calcoliamo adesso la **media** di un numero esponenziale:

$$\begin{aligned}
 \mathbb{E}(X) &= \int_{\mathbb{R}} x f_X(x) dx = \int_0^{+\infty} x \cdot \alpha e^{-\alpha x} dx \\
 (\text{integrazione per parti per integrali definiti}) &= -e^{-\alpha x} \cdot x \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\alpha x} dx \\
 (\text{l'addendo sinistro vale zero perché tende a } 0) &= \int_0^{+\infty} e^{-\alpha x} dx \\
 (\text{moltiplico e divido per } \alpha) &= \frac{1}{\alpha} \cdot \int_0^{+\infty} \alpha e^{-\alpha x} dx \\
 (\text{il moltiplicatore vale } 1 \text{ perché tutta l'area della funzione } f_X(x)) &= \frac{1}{\alpha}
 \end{aligned}$$

Un passaggio è stato ottenuto utilizzando l'integrazione per parti per integrali definiti. Siano $f, g : [a, b] \rightarrow \mathbb{R}$ due funzioni continue e supponiamo che le loro derivate siano pure continue su $[a, b]$. Allora:

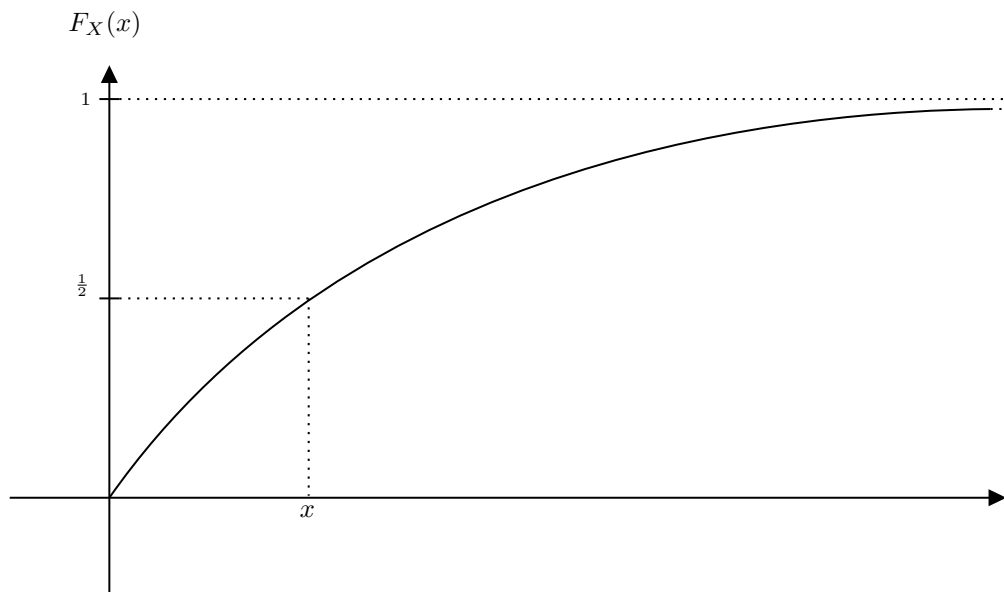
$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)dx$$

In conclusione, possiamo dire che:

$$\mathbb{E}(X) = \frac{1}{\alpha}$$

la **media** è il **reciproco del parametro** α .

Qual'è il grafico della funzione di distribuzione $F_X(x)$ di un numero aleatorio X con legge esponenziale?



Qual'è il valore x tale che $F_X(x) = \frac{1}{2}$ (e quindi probabilità di $\frac{1}{2}$ che corrisponde a metà area della funzione f_X)?

$$\begin{aligned}
 F_X(x) &= \frac{1}{2} \\
 1 - e^{-\alpha x} &= \frac{1}{2} \\
 e^{-\alpha x} &= \frac{1}{2} \\
 \ln e^{-\alpha x} &= -\ln \frac{1}{2} \\
 \alpha x &= \ln 2 \\
 (\text{questo punto viene chiamato } \mathbf{mediana} \text{ di } F_X) & \quad x = \frac{1}{\alpha} \ln 2
 \end{aligned}$$

Esaminiamo adesso la **funzione generatrice dei momenti** di un numero aleatorio X con legge esponenziale:

$$\begin{aligned}\phi_X(t) &:= \mathbb{E}(e^{tX}) = \int_{\mathbb{R}} e^{tx} f_X(x) dx \\ &= \int_0^{+\infty} e^{tx} \cdot \alpha e^{-\alpha x} dx \\ &= \int_0^{+\infty} \alpha e^{-x(\alpha-t)} dx \\ (\text{moltiplico e divido per } \alpha - t) &= \frac{\alpha}{\alpha - t} \int_0^{+\infty} (\alpha - t) e^{-(\alpha-t)x} dx \\ (\text{il moltiplicatore vale 1}) &= \frac{\alpha}{\alpha - t}\end{aligned}$$

Il denominatore deve essere maggiore di zero, cioè $\alpha - t > 0 \rightarrow t < \alpha$, quindi la funzione generatrice dei momenti di un numero aleatorio X con legge esponenziale è definita nell'intervallo $] -\infty, \alpha[$:

$$\phi_X(t) = \frac{\alpha}{\alpha - t} \text{ con } t < \alpha$$

Facciamo qualche verifica (ricordiamo che la funzione generatrice dei momenti serve per calcolare facilmente la media di qualsiasi ordine):

$$\begin{aligned}\mathbb{E}(X^0) &= \mathbb{E}(1) = 1 \\ \phi_X(0) &= \frac{\alpha}{\alpha - 0} = 1\end{aligned}$$

Calcoliamo la derivata prima di ϕ_X che deve corrispondere alla media $\mathbb{E}(X)$:

$$\phi'_X(t) = \left(\frac{\alpha}{\alpha - t} \right)' = \frac{\alpha' \cdot (\alpha - t) - \alpha \cdot (\alpha - t)'}{(\alpha - t)^2} = \frac{0 \cdot (\alpha - t) - \alpha \cdot (-1)}{(\alpha - t)^2} = \frac{\alpha}{(\alpha - t)^2}$$

Infatti se calcoliamo $\phi'_X(0)$ otteniamo proprio $\mathbb{E}(X)$:

$$\phi'_X(0) = \frac{\alpha}{\alpha^2} = \frac{1}{\alpha} = \mathbb{E}(X)$$

Calcoliamo la derivata seconda di ϕ_X che deve corrispondere alla media del secondo ordine $\mathbb{E}(X^2)$:

$$\begin{aligned}\phi''_X(t) &= \left(\frac{\alpha}{(\alpha - t)^2} \right)' = \frac{\alpha' \cdot [(\alpha - t)^2] - \alpha \cdot [(\alpha - t)^2]'}{\alpha^4} = \frac{0 \cdot [(\alpha - t)^2] - \alpha \cdot (2t - 2\alpha)}{(\alpha - t)^4} = \frac{-2\alpha t + 2\alpha^2}{(\alpha - t)^4} = \frac{2\alpha(\alpha - t)}{(\alpha - t)^4} \\ &= \frac{2\alpha}{(\alpha - t)^3}\end{aligned}$$

Infatti se calcoliamo $\phi''_X(0)$ otteniamo proprio $\mathbb{E}(X^2)$:

$$\phi''_X(0) = \frac{2}{\alpha^2} = \mathbb{E}(X^2)$$

Calcoliamo adesso la **varianza di X** con legge esponenziale. Ricordiamo che la varianza di un numero aleatorio si calcola come la differenza della media del secondo ordine e la media del primo ordine al quadrato:

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{2}{\alpha^2} - \frac{1}{\alpha} = \frac{1}{\alpha^2}$$

quindi otteniamo che la varianza di un numero aleatorio X con legge esponenziale è uguale a:

$$\mathbb{D}^2(X) = \frac{1}{\alpha^2}$$

Calcoliamo adesso la **deviazione standard** di X , otteniamo:

$$\sqrt{\mathbb{D}^2(X)} = \frac{1}{\alpha}$$

Quindi la **deviazione standard di un numero aleatorio con legge esponenziale è uguale alla sua media**.

Quindi, quando il rapporto tra la deviazione standard e la media di un numero aleatorio è circa 1, potrebbe accadere che stiamo avendo a che fare con un numero aleatorio con legge esponenziale:

$$\frac{\sqrt{\mathbb{D}^2(X)}}{\mathbb{E}(X)} = 1$$

4.2.1 Proprietà di assenza di usura

Un numero aleatorio X con legge esponenziale gode di una proprietà esclusiva chiamata **assenza di usura** (o assenza di memoria). Con questa espressione, si intende che:

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s) \quad \forall s, t \geq 0$$

Dimostrazione.

$$\begin{aligned} \mathbb{P}(X > s + t | X > t) &= \frac{\mathbb{P}(X > s + t, X > t)}{\mathbb{P}(X > t)} \\ &= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)} \\ &= \frac{1 - \mathbb{P}(X \leq s + t)}{1 - \mathbb{P}(X \leq t)} \\ &= \frac{1 - F_X(s + t)}{1 - F_X(t)} \\ &= \frac{e^{-\alpha(s+t)}}{e^{-\alpha t}} \\ &= \frac{e^{-\alpha s} \cdot e^{-\alpha t}}{e^{-\alpha t}} = e^{-\alpha s} = 1 - F_X(s) = \mathbb{P}(X > s) \end{aligned}$$

□

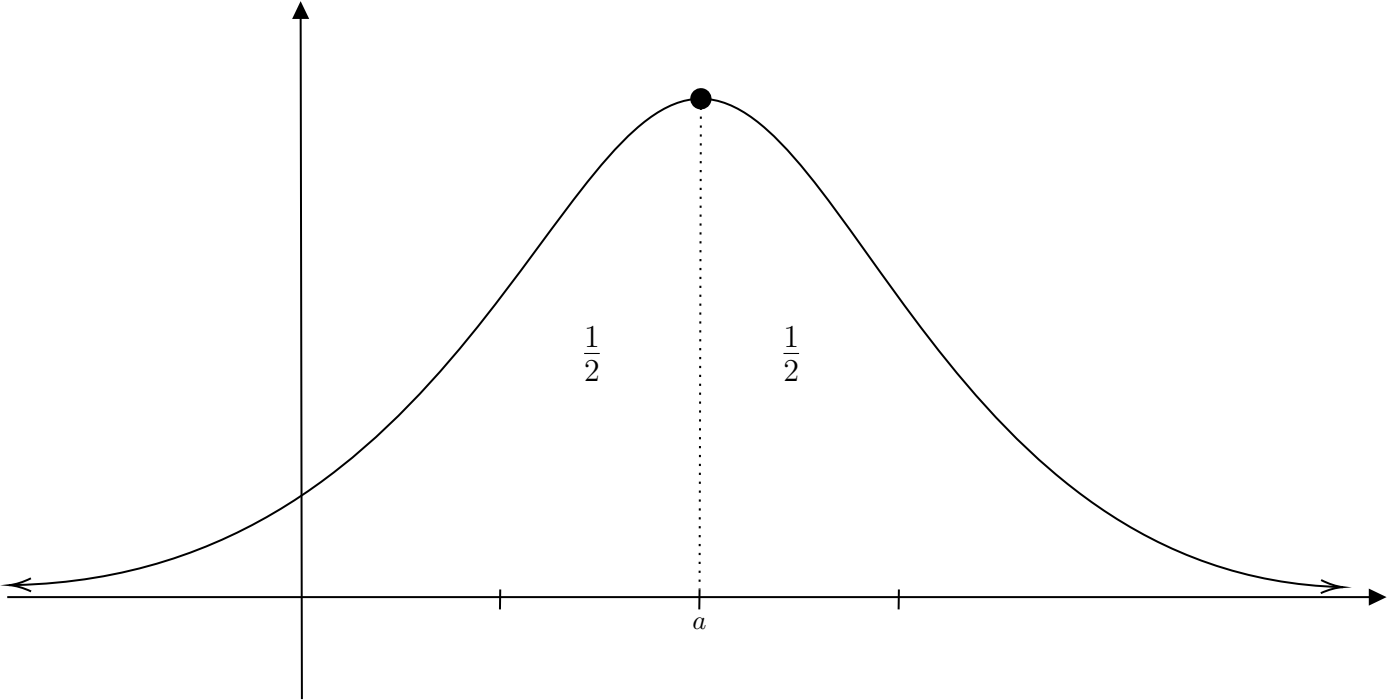
Per capire questa proprietà, si immagini che X rappresenti il tempo di vita di un certo oggetto prima di guastarsi. Sapendo che tale oggetto è già in funzione da un tempo t e non si è ancora rotto, qual'è la probabilità che esso continui a funzionare almeno per un ulteriore intervallo di tempo s ? Chiaramente la probabilità richiesta è quella espressa dal membro sinistro della proprietà di sopra, e cioè $\mathbb{P}(X > s + t | X > t)$. Infatti dire che l'oggetto non si è ancora guastato al tempo t equivale a dire che il tempo in cui avverrà la rottura (X), è superiore a t , mentre affermare che l'oggetto funzionerà per un ulteriore tempo s a partire dal tempo t , significa che il tempo X dovrà essere maggiore di $t + s$. In questo senso, la proprietà di assenza di usura afferma che la distribuzione del tempo di vita rimanente dell'oggetto considerato, è la medesima sia nel caso in cui esso stia funzionando da un tempo t , sia nel caso in cui esso sia nuovo o, in altri termini, se la proprietà di assenza di usura è soddisfatta, non vi è alcun bisogno di tenere presente l'età dell'oggetto, perché fino a che esso funziona, si comporta esattamente come se fosse "nuovo di zecca".

4.3 Variabile aleatoria normale o gaussiana

Una variabile aleatoria X si dice *normale* o *gaussiana* di parametri a e b^2 e si denota con $X \sim \mathcal{N}(a, b^2)$, se X ha funzione di densità di probabilità data da:

$$f_X(x, a, b^2) = \frac{1}{\sqrt{2\pi b^2}} \cdot e^{-\frac{(x-a)^2}{2b^2}} \text{ con } a \in \mathbb{R}, b^2 > 0 \text{ e } \forall x \in \mathbb{R}$$

Il grafico della funzione f_X (anche detta *curva a campana*) si presenta così:



La funzione di densità di probabilità di un numero aleatoria con distribuzione normale è una curva a campana simmetrica rispetto all'asse $x = a$ (**asse di simmetria**). Infatti, l'area che si trova a sinistra del punto a è $\frac{1}{2}$ e l'area che si trova a destra del punto a è $\frac{1}{2}$. Inoltre, tale parametro a rappresenta sia la **moda** (*Mo*) che la **mediana** (*Me*). Il supporto della funzione f_X è l'insieme dei numeri reali \mathbb{R} .

Quando poniamo i parametri $a = 0$ e $b^2 = 1$ diventa un **numero gaussiano standard** (o *normale standard*) e si denota con $Z \sim \mathcal{N}(0, 1)$.

4.3.1 Funzione generatrice dei momenti di un numero aleatorio gaussiano standard

Determiniamo la funzione generatrice dei momenti del numero gaussiano standard:

$$\begin{aligned} \phi_Z(t) &= \mathbb{E}(e^{tZ}) = \int_{\mathbb{R}} e^{tx} f_Z(x) dx \\ &= \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2tx)} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[(x^2 - 2tx + t^2) - t^2]} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[(x-t)^2 - t^2]} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} \cdot e^{\frac{t^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

(il moltiplicando a destra è 1 - f.d.p. di $\mathcal{N}(t, 1)$)

Otteniamo dunque che la funzione generatrice dei momenti di un numero aleatorio gaussiano standard è:

$$\phi_Z(t) = e^{\frac{t^2}{2}}$$

4.3.2 Media e varianza numero aleatorio gaussiano standard

Calcoliamo la media $\mathbb{E}(Z)$ del numero aleatorio gaussiano standard:

$$\mathbb{E}(Z) = \phi'_Z(t) = e^{\frac{1}{2}t^2} \cdot t|_{t=0} = 1 \cdot 0 = 0.$$

Otteniamo che **la media è uguale al parametro a** (questo vale in generale).

Calcoliamo la media del secondo ordine:

$$\mathbb{E}(Z^2) = \phi''_Z(t) = (\phi_Z(t) \cdot t)' = \phi'_Z(t) \cdot t + \phi_Z(t) \cdot 1 = \phi_Z(0) = 1$$

Quindi la media del secondo ordine è uguale a 1 (in generale vale che è uguale a $a^2 + b^2$).

Calcoliamo adesso la varianza di un numero aleatorio gaussiano standard:

$$\mathbb{D}^2(Z) = \mathbb{E}(Z^2) - \mathbb{E}^2(Z) = 1 - 0 = 1$$

Otteniamo che **la varianza è uguale al parametro b** (questo vale in generale). Quindi vale questa proprietà:

$$\begin{array}{ccc} & \mathbb{E}(Z) & \mathbb{D}^2(Z) \\ & \nwarrow & \nearrow \\ Z \sim \mathcal{N}(0, 1) & & \end{array}$$

4.3.3 Proprietà varianza (in generale)

Mettiamo adesso in evidenza alcune proprietà che la varianza gode in generale (nel senso che tali proprietà non si limitano solo ai numeri aleatori gaussiani ma valgono in generale).

Proposizione 4.3.1 (Proprietà varianza 1)

Sia X un numero aleatorio qualsiasi che ammette finita la varianza e Y un altro numero aleatorio definito come:

$$c \in \mathbb{R} \quad Y = X + c$$

Allora possiamo dire che:

$$\mathbb{D}^2(Y) = \mathbb{D}^2(X)$$

Dimostrazione.

$$\begin{aligned} \mathbb{D}^2(Y) &= \mathbb{E}\{[Y - \mathbb{E}(Y)]^2\} \\ &= \mathbb{E}[(X + c - \mathbb{E}(X) - c)^2] \\ &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} \\ &= \mathbb{D}^2(X). \end{aligned}$$

□

Proposizione 4.3.2 (Proprietà varianza 2)

Sia X un numero aleatorio qualsiasi che ammette finita la varianza e T un altro numero aleatorio definito come:

$$d \in \mathbb{R} \quad T = \frac{X}{d}$$

Allora possiamo dire che:

$$\mathbb{D}^2(T) = \mathbb{D}^2\left(\frac{X}{d}\right) = \frac{1}{d^2} \cdot \mathbb{D}^2(X)$$

Dimostrazione.

$$\begin{aligned}
\mathbb{D}^2(T) &= \mathbb{E}\{[T - \mathbb{E}(T)]^2\} \\
&= \mathbb{E}\left\{\left[\frac{X}{d} - \frac{\mathbb{E}(X)}{d}\right]^2\right\} \\
&= \mathbb{E}\left\{\left[\frac{X - \mathbb{E}(X)}{d}\right]^2\right\} \\
&= \mathbb{E}\left\{\frac{[X - \mathbb{E}(X)]^2}{d^2}\right\} \\
&= \frac{1}{d^2} \cdot \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \frac{1}{d^2} \cdot \mathbb{D}^2(X).
\end{aligned}$$

□

4.3.4 Operazione di standardizzazione

Consideriamo adesso questa operazione:

$$\frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{D}^2(X)}}$$

Si tratta dello scarto diviso la deviazione standard. Questa operazione viene chiamata **operazione di standardizzazione**. Calcoliamo adesso la media e la deviazione dell'operazione di standardizzazione:

$$\begin{aligned}
\mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{D}^2(X)}}\right) &= \frac{1}{\sqrt{\mathbb{D}^2(X)}} \cdot [\mathbb{E}(X) - \mathbb{E}(X)] = 0 \\
\mathbb{D}^2\left(\frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{D}^2(X)}}\right) &= \frac{1}{\mathbb{D}^2(X)} \cdot \mathbb{D}^2[X - \mathbb{E}(X)] = \frac{1}{\mathbb{D}^2(X)} \cdot \mathbb{D}^2(X) = 1
\end{aligned}$$

dove il secondo risultato è stato ottenuto sfruttando le proprietà 4.3.1 e 4.3.2 della varianza.

Se X è un numero aleatorio gaussiano allora l'operazione di standardizzazione su X genera il numero aleatorio Y gaussiano standard:

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{D}^2(X)}} \sim \mathcal{N}(0, 1)$$

Un risultato importante riguardo i numeri aleatori gaussiani è che se X è gaussiana e Y è una trasformazione lineare di X , allora Y è a sua volta gaussiana. In particolare, se prendiamo il numero aleatorio $Z \sim \mathcal{N}(0, 1)$ gaussiano standardizzato e consideriamo quest altro numero aleatorio:

$$a + bZ$$

e poi calcoliamo la sua media e la sua varianza:

$$\begin{aligned}
\mathbb{E}(a + bZ) &= a + b\mathbb{E}(Z) = a \\
\mathbb{D}^2(a + bZ) &= \mathbb{D}^2(bZ) = b^2\mathbb{D}^2(Z) = b^2
\end{aligned}$$

4.3.5 Funzione generatrice dei momenti di un numero aleatorio normale

Sia $X \sim (\mu, \sigma^2)$ un numero aleatorio con distribuzione normale con media μ e varianza σ^2 . Vogliamo determinare la sua funzione generatrice dei momenti. Notiamo innanzitutto che, usando l'operazione di standardizzazione otteniamo questo risultato:

$$\frac{X - \mu}{\sigma} \sim Z \iff X - \mu \sim Z\sigma \iff X \sim \mu + \sigma Z$$

Quindi:

$$X \sim \mu + \sigma Z$$

Sfruttando questo risultato e utilizzando le proprietà della funzione generatrice dei momenti, possiamo facilmente ottenere la funzione generatrice dei momenti di un numero aleatorio normale:

$$\begin{aligned}
\phi_X(t) &= \phi_{\mu + \sigma Z}(t) \\
(\text{Proposizione 2.4.1 f.g.m.}) &= \phi_\mu(t) \cdot \phi_{\sigma Z}(t) \\
(\text{Proprietà 1}) &= e^{t\mu} \cdot \phi_Z(\sigma t) = e^{t\mu} \cdot e^{\sigma^2 \frac{t^2}{2}}
\end{aligned}$$

Tutte le leggi gaussiane hanno questa funzione generatrice dei momenti.

4.3.6 Calcolo punto massimo

Vogliamo adesso determinare il **punto massimo** (sull'ascissa) della funzione di densità di probabilità di un numero aleatorio gaussiano generale X . Procediamo calcolando la derivata prima della funzione di densità di probabilità f_X :

$$f'_X(x, a, b) = -f_X(x, a, b) \frac{x - a}{b}$$

Calcolando $f'_X(x, a, b) = 0$ otteniamo un **unico punto stazionario** e cioè il parametro a :

$$f'_X(x, a, b) = 0 \iff x - a = 0 \iff x = a$$

Bisogna adesso capire se tale punto stazionario a è un punto di massimo o di minimo relativo (perché se la derivata in quel punto è 0 non possiamo ancora concludere che è un punto di massimo o di minimo). Procediamo calcolando la derivata seconda di f_X :

$$f''_X(x, a, b) = -\frac{1}{b}[f'_X(x, a, b)(x - a) + f_X(x, a, b)]$$

Calcoliamo adesso la derivata seconda f''_X nel punto a e otteniamo:

$$f''_X(a, a, b) = -\frac{1}{b}f_X(a, a, b)$$

Siccome $b = \sqrt{b^2} > 0$ e $f_X(x, a, b) > 0 \forall x \in \mathbb{R}$ la quantità ottenita è *negativa* e quindi possiamo concludere che il parametro a è un **punto di massimo**.

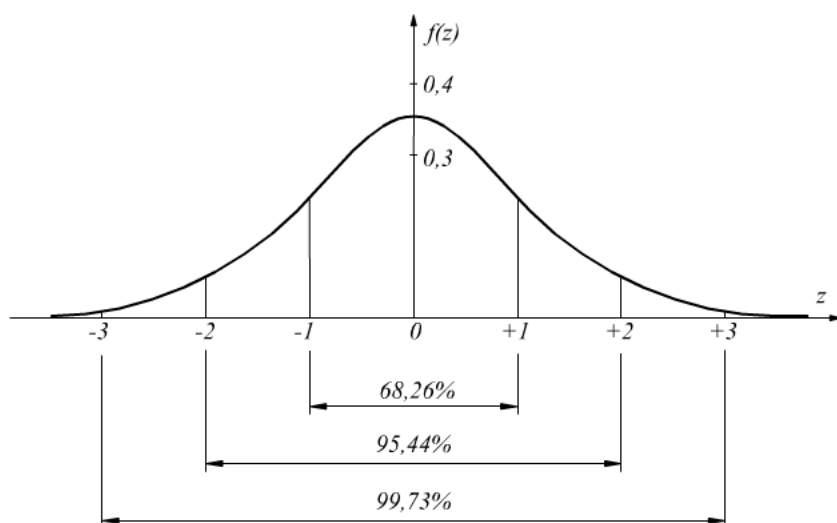
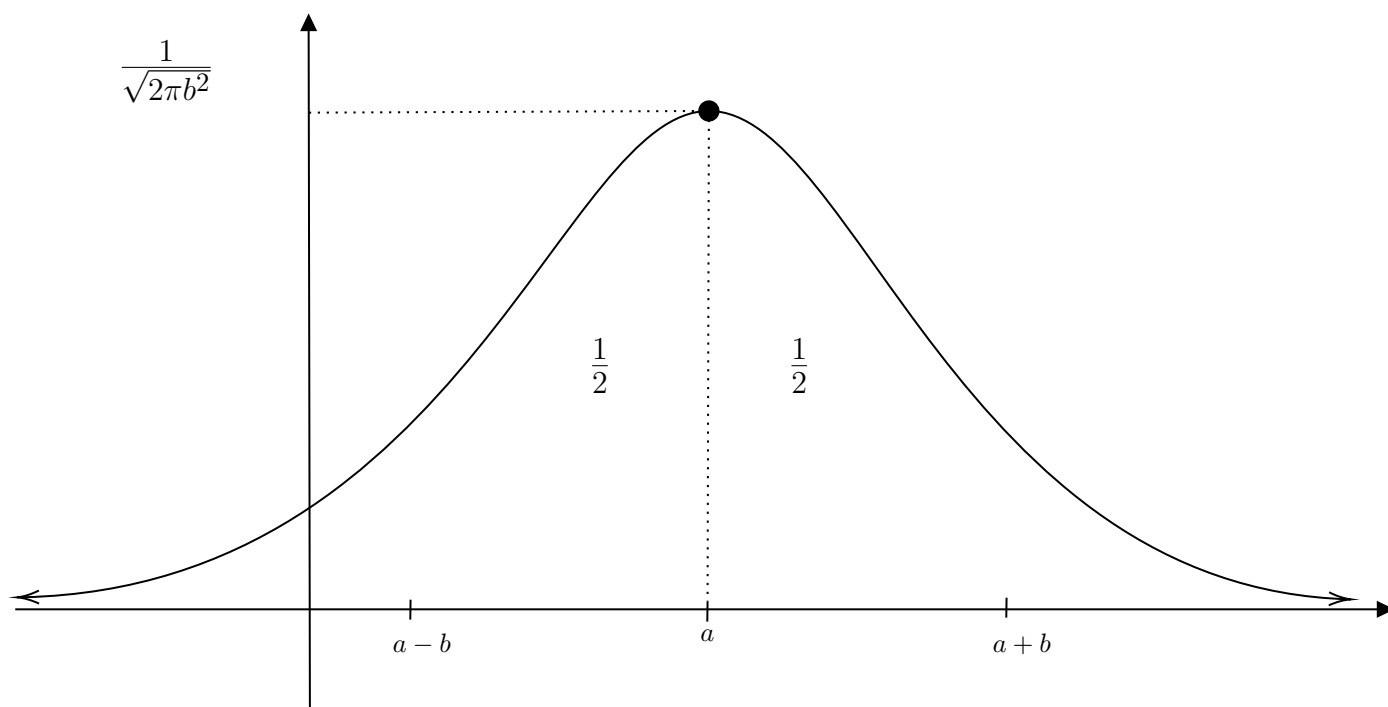
4.3.7 Calcolo punti di flesso

Vogliamo adesso calcolare i **punti di flesso** della funzione di densità di probabilità di un numero aleatorio gaussiano f_X . Un punto di flesso per una curva o funzione è un punto in cui si manifesta un cambiamento di convessità o di segno di curvatura. I punti di flesso si calcolano andando a pescare i punti dove la derivata seconda della funzione si annulla (diventa uguale a 0):

$$\begin{aligned} f''_X(x, a, b) &= -f'_X(x, a, b) \left(\frac{x - a}{b} \right) - f_X(x, a, b) \left(\frac{x - a}{b} \right)' \\ &= f_X(x, a, b) \cdot \frac{(x - a)^2}{b^2} - f_X(x, a, b) \cdot \frac{1}{b} \\ &= \frac{f_X(x, a, b)}{b} \cdot \left[\frac{(x - a)^2}{b} - 1 \right] \end{aligned}$$

Poniamo il risultato uguale a 0:

$$\begin{aligned} \frac{f_X(x, a, b)}{b} \cdot \left[\frac{(x - a)^2}{b} - 1 \right] &= 0 \\ \iff \frac{(x - a)^2}{b} - 1 &= 0 \\ \iff \frac{(x - a)^2 - b}{b} &= 0 \\ \iff (x - a)^2 &= b \\ \iff x = a + b \vee x = a - b \end{aligned}$$



Esempio 4.3.1

Sia X una variabile aleatoria gaussiana con parametri $a = 3$ e $b = 4$ quindi $X \sim \mathcal{N}(3, 16)$. Vogliamo calcolare le seguenti probabilità:

1. $\mathbb{P}(X < 11)$
2. $\mathbb{P}(X > -1)$
3. $\mathbb{P}(2 < X < 7)$

Per calcolare queste probabilità dobbiamo ricorrere a $Z \sim \mathcal{N}(0, 1)$ per rendere i calcoli più semplici ma anche perché la funzione di distribuzione $F_X(x)$ di una variabile aleatoria X gaussiana non è esprimibile in funzioni elementari (non è possibile esprimere una forma chiusa per il risultato di questo integrale). Allora procediamo a **standardizzare** tale variabile X :

$$Y = \frac{X - 3}{4}$$

Dunque calcoliamo il punto 1:

$$\mathbb{P}(X < 11) = \mathbb{P}(X - 3 < 11 - 3) = \mathbb{P}\left(\frac{X - 3}{4} < \frac{11 - 3}{4}\right) = \mathbb{P}(Y < 2) = 0.9772$$

Calcoliamo il punto 2:

$$\mathbb{P}(X > -1) = 1 - \mathbb{P}(X \leq -1) = 1 - \mathbb{P}(Y \leq -1) = 1 - [0.5 - \mathbb{P}(-1 < Y < 0)] = 0.5 + 0.34 = 0.84$$

Calcoliamo il punto 3:

$$\begin{aligned}\mathbb{P}(2 < X < 7) &= \mathbb{P}\left(-\frac{1}{4} < Y < 1\right) = F_Z(1) - F_Z\left(-\frac{1}{4}\right) \\ &= \mathbb{P}\left(\left\{-\frac{1}{4} < Y \leq 0\right\} \cup \{0 < Y < 1\}\right) \\ &= \mathbb{P}\left(-\frac{1}{4} < Y \leq 0\right) + \mathbb{P}(0 < Y < 1) \approx 0.0987 + 0.34\end{aligned}$$

Esempio 4.3.2

La potenza W dissipata da una resistenza è proporzionale al quadrato della differenza di potenziale ai suoi capi, ovvero:

$$W = rV^2 \quad \text{con } r = 3$$

Si supponga che $V \sim \mathcal{N}(6, 1)$. Calcolare:

1. $\mathbb{P}(W > 120)$
2. $\mathbb{E}(W)$

Calcoliamo il punto 1:

$$\begin{aligned}\mathbb{P}(W > 120) &= \mathbb{P}(3 \cdot V^2 > 120) = \mathbb{P}(V^2 > 40) = \mathbb{P}(V > \sqrt{40}) \\ &\quad (\text{standardizziamo}) \quad = \mathbb{P}(V - 6 > \sqrt{40} - 6) \\ &= \mathbb{P}(Y > \sqrt{40} - 6) \\ &= 1 - \mathbb{P}(Y \leq \sqrt{40} - 6) \\ &= 1 - \mathbb{P}(Y \leq 0.3246) \approx 0.3727\end{aligned}$$

Calcoliamo il punto 2. Sappiamo che $V \sim \mathcal{N}(6, 1)$ quindi sappiamo che $\mathbb{E}(V) = 6$ e $\mathbb{D}^2(V) = 1$. Per calcolare $\mathbb{E}(V^2)$ la media del secondo ordine di V , possiamo ricorrere alla stessa definizione di varianza:

$$\mathbb{D}^2(V) = \mathbb{E}(V^2) - \mathbb{E}^2(V)$$

e quindi, possiamo calcolare facilmente $\mathbb{E}(V^2)$:

$$\mathbb{E}(V^2) = \mathbb{D}^2(V) + \mathbb{E}^2(V) = 1 + (6)^2 = 1 + 36 = 37.$$

Dunque, concludiamo che:

$$\mathbb{E}(W) = 3 \cdot \mathbb{E}(V^2) = 3 \cdot 37 = 111 \text{ Watt}$$

□

Consideriamo adesso di nuovo il numero aleatorio gaussiano standard $Z \sim \mathcal{N}(0, 1)$. La sua funzione di distribuzione è uguale a:

$$F_Z(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

Ma cosa dire se eleviamo questo numero aleatorio al quadrato? Se poniamo $Y = Z^2$ questa trasformazione cambia radicalmente la distribuzione di probabilità perché Y non può assumere valori negativi dunque il supporto cambia da $] - \infty, +\infty[$ a $]0, +\infty[$. La funzione di distribuzione di Y sarà:

$$y > 0 \quad F_Y(y) = \begin{cases} 0 & , \text{ se } y \leq 0 \\ 2 \cdot [F_Z(\sqrt{y}) - \frac{1}{2}] & , \text{ se } y > 0 \end{cases}$$

Per capire questo risultato, calcoliamo $\mathbb{P}(Y \leq y)$:

$$\begin{aligned}\mathbb{P}(Y \leq y) &= \mathbb{P}(Z^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ (\text{per simmetria}) \quad &= 2 \cdot \mathbb{P}(0 \leq Z \leq \sqrt{y}) = 2 \cdot \left[F_Z(\sqrt{y}) - \frac{1}{2} \right]\end{aligned}$$

La funzione di distribuzione F_Y può anche essere vista in questo modo:

$$F_Y(y) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y})$$

Sfruttando questa uguaglianza, deriviamo l'espressione e ricaviamo f_Y :

$$\begin{aligned} f_Y(y) &= f_Z(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_Z(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ \text{(i due addendi sono uguali per simmetria)} \quad &= \frac{1}{\sqrt{y}} \cdot f_Z(\sqrt{y}) \\ &= \frac{1}{\sqrt{y}\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}} \end{aligned}$$

Quindi otteniamo che:

$$f_Y(y) = \frac{1}{\sqrt{y}\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}$$

e viene chiamata **legge del quadrato** o **chi-quadrato** (Figura 3).

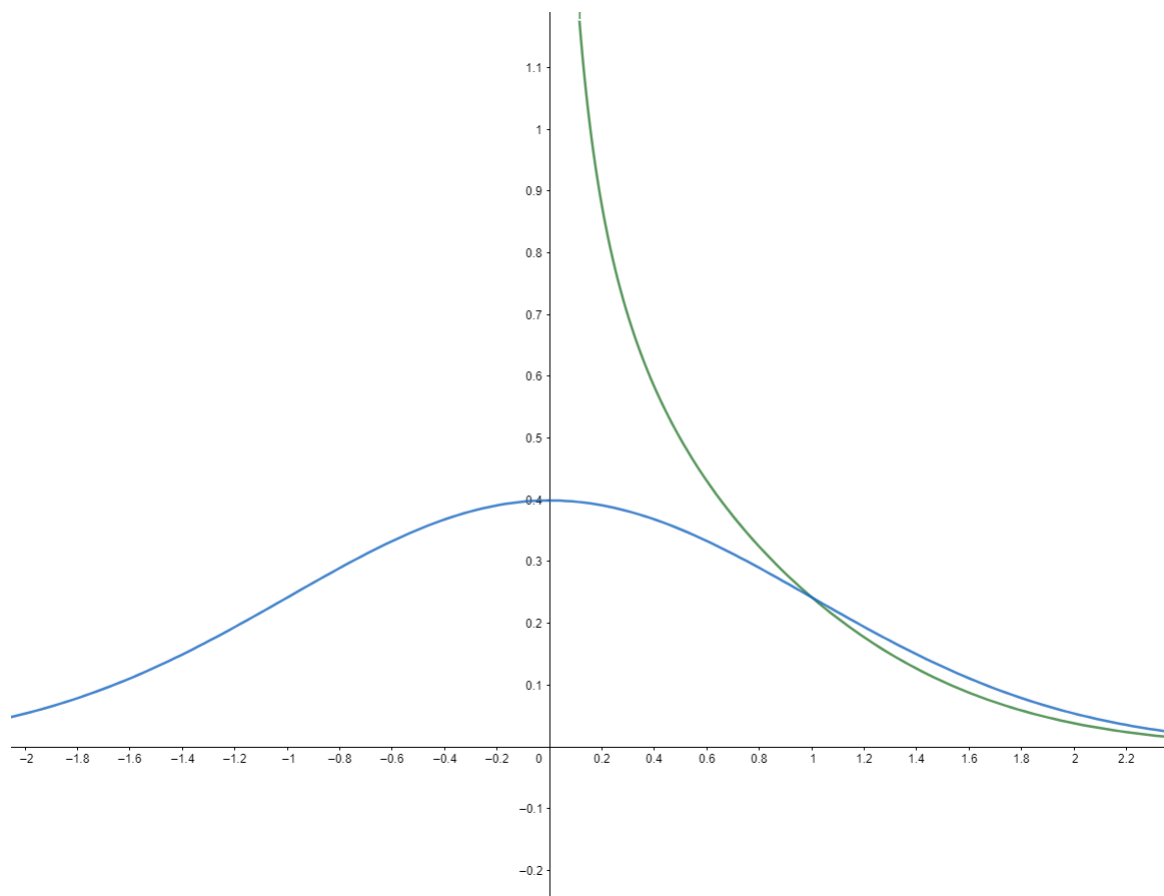


Figura 3: La funzione in blu è normale standard - la funzione in verde è legge del quadrato

Diamo adesso la definizione di **quantile superiore** di ordine $\alpha \in (0, 1)$ di Z che si denota con Z_α (dove Z è il numero aleatorio gaussiano standard $Z \sim \mathcal{N}(0, 1)$). La definizione è:

$$\mathbb{P}(Z_\alpha) = \alpha$$

Il numero α viene scelto arbitrariamente e l'idea è quella di identificare qual'è il numero oltre il quale non bisogna andare perché altrimenti si va a finire in una zona a bassa probabilità.

Ad esempio se scegliamo $\alpha = 0.05$, il corrispondente quantile superiore è $Z_{0.05} \approx 1,645$,

Copie e vettori di variabili aleatorie

5 Copie e vettori di variabili aleatorie

Ci sono situazioni in cui la scelta di ridurre un esperimento casuale allo studio di una sola variabile aleatoria, è destinata a fallire a priori, perché l'oggetto di interesse sono proprio le relazioni presenti tra due o più grandezze numeriche. Ad esempio, in un esperimento sulle possibili cause di tumore, potremmo voler indagare il rapporto tra il numero medio di sigarette fumate quotidianamente e l'età in cui viene riscontrata questa patologia.

Per specificare la relazione tra due variabili aleatorie X e Y , il punto di partenza è estendere il concetto di funzione di distribuzione.

Definizione 5.0.1 (Funzione di distribuzione congiunta di X e Y)

Siano X e Y due variabili aleatorie che riguardano lo stesso esperimento casuale. Si dice *funzione di distribuzione congiunta* di X e Y - e si indica normalmente con la lettera F - la funzione di due variabili seguente:

$$F(x, y) := \mathbb{P}(X \leq x, Y \leq y)$$

dove la virgola nell'argomento $\mathbb{P}()$ denota l'intersezione tra eventi.

La conoscenza di questa funzione permette, almeno in teoria, di calcolare le probabilità di tutti gli eventi che dipendono, singolarmente o congiuntamente, da X e Y . Ad esempio la funzione di distribuzione di X - che denotiamo questa volta con F_X - può essere ottenuta dalla funzione di ripartizione congiunta F così:

$$\begin{aligned} F_X(x) &:= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(X \leq x, Y < \infty) \\ &= F(x, \infty) \end{aligned}$$

E analogamente la funzione di ripartizione di Y :

$$F_Y(y) = F(\infty, y)$$

5.1 Distribuzione congiunta per variabili aleatorie discrete

Se sappiamo che un vettore aleatorio è di tipo discreto, possiamo definire e utilizzare la funzione di massa di probabilità congiunta.

Definizione 5.1.1 (Funzione di massa di probabilità congiunta)

Se X e Y sono variabili aleatorie discrete che assumono i valori x_1, x_2, \dots e y_1, y_2, \dots rispettivamente, la funzione:

$$p(x_i, y_j) := \mathbb{P}(X = x_i, Y = y_j), \quad i = 1, 2, \dots, \quad j = 1, 2, \dots$$

è la loro *funzione di massa di probabilità congiunta*.

Le funzioni di massa individuale di X e Y si possono ricavare da quella congiunta notando che, siccome Y deve assumere uno dei valori y_i , l'evento $\{X = x_i\}$ può essere visto come l'unione al variare di j degli eventi $\{X = x_i, Y = y_j\}$, che sono mutuamente esclusivi; in formule:

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

da cui, grazie all'Assioma 3:

$$\begin{aligned} p_X(x_i) &:= \mathbb{P}(X = x_i) \\ &= \mathbb{P}\left(\bigcup_j \{X = x_i, Y = y_j\}\right) \\ &= \sum_j \mathbb{P}(X = x_i, Y = y_j) \\ &=: \sum_j p(x_i, y_j) \end{aligned}$$

Analogamente per p_Y :

$$p_Y(y_j) = \sum_i p(x_i, y_j)$$

Esempio 5.1.1

Da un gruppo di 12 batterie - di cui 3 nuove, 4 usate e 5 difettose - ne vengono scelte tre a caso. Siano X e Y rispettivamente il numero di batterie nuove e usate tra quelle scelte. La funzione di massa di probabilità congiunta, $p(x_i, y_j)$ è data dai valori seguenti, come il lettore può verificare facilmente:

$$\begin{aligned} p(0,0) &= \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220} = 0.0454 & p(0,1) &= \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220} = 0.1818 \\ p(0,2) &= \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220} = 0.1363 & p(0,3) &= \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220} = 0.0181 \\ p(1,0) &= \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220} = 0.1363 & p(1,1) &= \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220} = 0.2727 \\ p(1,2) &= \frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}} = \frac{18}{220} = 0.0818 & p(2,0) &= \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220} = 0.068 \\ p(2,1) &= \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{12}{220} = 0.0545 & p(3,0) &= \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220} = 0.0045 \end{aligned}$$

Queste probabilità possono essere convenientemente presentate in forma tabellare:

		Y				
		0	1	2	3	
X	0	p(0,0)	p(0,1)	p(0,2)	p(0,3)	p(0,.)
	1	p(1,0)	p(1,1)	p(1,2)	p(1,3)	p(1,.)
	2	p(2,0)	p(2,1)	p(2,2)	p(2,3)	p(2,.)
	3	p(3,0)	p(3,1)	p(3,2)	p(3,3)	p(3,.)
		p(.,0)	p(.,1)	p(.,2)	p(.,3)	1

		Y				
		0	1	2	3	
X	0	0.0454	0.1818	0.1363	0.0181	p(0,.)
	1	0.1363	0.2727	0.0818	0	p(1,.)
	2	0.068	0.0545	0	0	p(2,.)
	3	0.0545	0	0	0	p(3,.)
		p(.,0)	p(.,1)	p(.,2)	p(.,3)	1

Se vogliamo lo spettro delle variabili aleatorie X e Y è $S_{X,Y} = \{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (2,0), (2,1), (3,0)\}$. Inoltre, per quello che abbiamo detto prima, $p(0, \cdot)$ rappresenta la propabilità $\mathbb{P}(X = 0)$. Infatti:

$$\begin{aligned} \mathbb{P}(X = 0) &= \mathbb{P}(\{X = 0\} \cap \Omega) = \mathbb{P}(\{X = 0\} \cap [\{Y = 0\} \cup \{Y = 1\} \cup \{Y = 2\} \cup \{Y = 3\}]) \\ &= \mathbb{P}(\{X = 0\} \cap \{Y = 0\} \cup \{X = 0\} \cap \{Y = 1\} \cup \{X = 0\} \cap \{Y = 2\} \cup \{X = 0\} \cap \{Y = 3\}) \\ &= \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 0, Y = 1) + \mathbb{P}(X = 0, Y = 2) + \mathbb{P}(X = 0, Y = 3) = p(0, \cdot) \end{aligned}$$

La distribuzione a destra viene chiamata **distribuzione marginale di X** mentre la distribuzione in basso viene chiamata **distribuzione marginale di Y** .

5.2 Distribuzione congiunta per variabili aleatorie continue

Due variabili aleatorie X e Y sono *congiuntamente continue* se esiste una funzione non negativa $f_{(X,Y)}(x,y)$ definita per tutti gli $x \in \mathbb{R}$ e $y \in \mathbb{R}$ tale che:

$$B \in \mathcal{B}^2 \quad \mathbb{P}_{(X,Y)}(B) = \mathbb{P}[(X,Y) \in B] = \int \int_B f_{(X,Y)}(x,y) dx dy$$

e viene chiamata **funzione di densità di probabilità congiunta**. I boreliani B in questo caso appartengono al prodotto cartesiano $] -\infty, x] \times] -\infty, y]$.

La funzione di distribuzione congiunta di due variabili X e Y congiunte continuamente è quindi definita come:

$$F_{(X,Y)}(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(x,y) dx dy$$

mentre la funzione di densità di probabilità congiunta può essere ottenuta derivando due volte $F_{(X,Y)}$:

$$f_{(X,Y)}(x,y) = \frac{d^2 F_{(X,Y)}(x,y)}{dx dy}$$

Le funzioni di densità di probabilità di X e Y possono essere ricavate stesso dalla funzione di densità di probabilità congiunta:

$$\begin{aligned} x \in \mathbb{R}, \quad f_X(x) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y) dy \\ y \in \mathbb{R}, \quad f_Y(y) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y) dx \end{aligned}$$

Nel caso di $f_X(x)$, fissiamo un x , lasciamo variare y e calcoliamo l'integrale (dualmente con $f_Y(y)$).

Esempio 5.2.1

X e Y sono congiuntamente distribuite con:

$$f_{(X,Y)}(x,y) = \begin{cases} 2e^{-x}e^{-2y} & x > 0, y > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Calcoliamo la probabilità $\mathbb{P}(X > 1, Y < 1)$:

$$\begin{aligned} \mathbb{P}(X > 1, Y < 1) &= \int_0^1 dy \int_1^{+\infty} 2e^{-x}e^{-2y} dx = 2 \int_0^1 e^{-2y} dy \int_1^{+\infty} e^{-x} dx \\ &= 2 \int_0^1 e^{-2y} dy \cdot e^{-x}|_{+\infty}^1 \\ &= 2 \int_0^1 e^{-2y} dy \cdot e^{-1} \\ &= 2 \cdot e^{-1} \int_0^1 e^{-2y} dy \\ &= 2e^{-1} \frac{1}{2} e^{-2y}|_1^0 = e^{-1} [1 - e^{-2}] \end{aligned}$$

Calcoliamo adesso la probabilità che $\mathbb{P}(X < a)$ con $a > 0$ qualsiasi:

$$\mathbb{P}(X < a) = \int_0^a \int_0^{+\infty} 2e^{-x}e^{-2y} dx dy = 2 \int_0^a e^{-x} dx \int_0^{+\infty} e^{-2y} dy = 1 - e^{-a}$$

Calcoliamo adesso la probabilità $\mathbb{P}(X < Y)$. In questo caso la regione su cui integrare è quella dove $x < y$. Gli estremi di integrazione che corrispondono a questo dominio possono essere scelti in due modi:

1. o si integra internamente in dx tra gli estremi 0 e y (infatti $x > 0$ altrimenti f è nulla, mentre $x < y$ è la definizione della regione che stiamo considerando), ed esternamente in dy tra 0 e $+\infty$ (infatti basta porre la condizione $x < y$ sull'integrale interno;
2. o si integra internamente in dy tra x e $+\infty$ (per rispettare $x < y$), ed esternamente in dx tra 0 e $+\infty$.

Svolgere il problema 10.

5.3 Distribuzione congiunta quando le variabili sono indipendenti

Quando due variabili aleatorie X e Y sono congiunte abbiamo una tabella del genere:

$p(1,1)$	$p(1,2)$	$p(1,3)$	\dots	$p(1,n)$	$p(1,\cdot)$
$p(2,1)$	$p(2,2)$	$p(2,3)$	\dots	$p(2,n)$	$p(2,\cdot)$
\dots	\dots	\dots	\dots	\dots	\dots
$p(m,1)$	$p(m,2)$	$p(m,3)$	\dots	$p(m,n)$	$p(m,\cdot)$
$p(\cdot,1)$	$p(\cdot,2)$	$p(\cdot,3)$	\dots	$p(\cdot,n)$	1

dove vale la seguente proprietà:

$$\forall i = 1, \dots, n \quad \forall j = 1, \dots, m \quad p(i, j) = p(i, \cdot) \times p(\cdot, j)$$

e si dice che X e Y sono **indipendenti**.

Ad esempio:

$1/4$	$1/4$	$1/2$
$1/4$	$1/4$	$1/2$
$1/2$	$1/2$	1

Se X e Y sono indipendenti, nel **caso discreto** abbiamo che:

$$(x, y) \in S_{(X,Y)} \quad p_{(X,Y)}(x, y) = p_X(x) \cdot p_Y(y)$$

mentre nel **caso continuo** abbiamo che:

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y) \iff F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y)$$

Esempio 5.3.1

Ritornando all'esempio 5.2.1, la seguente funzione:

$$f_{(X,Y)}(x, y) = \begin{cases} 2e^{-x}e^{-2y} & x > 0, y > 0 \\ 0 & \text{altrimenti} \end{cases}$$

presenta due variabili aleatorie $X \sim Esp(1) = e^{-x}$ e $Y \sim Esp(2) = 2e^{-2y}$ che sono indipendenti tra di loro perché le funzioni di densità di probabilità di X e Y sono separate.

5.4 Distribuzioni condizionali

5.4.1 Distribuzioni condizionali caso discreto

Le relazioni esistenti tra due variabili aleatorie possono essere chiarite dallo studio della distribuzione condizionale di una delle due, dato il valore dell'altra. Si ricorda che presi comunque due eventi E e F con $\mathbb{P}(F) > 0$, la probabilità di E condizionata a F è data dall'espressione:

$$\mathbb{P}(E|F) := \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

E' naturale applicare questo schema alle variabili aleatorie discrete.

Esempio 5.4.1

All'interno di una certa popolazione, il 15% delle coppie non ha figli, il 20% ne ha uno, il 35% ne ha due e il 30% ne ha tre. Inoltre ogni bambino, indipendentemente da tutti gli altri, può essere maschio o femmina con pari probabilità. Se si seleziona una famiglia a caso e si denotano con X il numero di femmine e con Y il numero di maschi presenti tra i figli in tale famiglia, si ottiene la seguente funzione di massa di probabilità:

		Y				
		0	1	2	3	
X	0	0.1500	0.1000	0.0875	0.0375	0.3750
	1	0.1000	0.1750	0.1125	0	0.3875
	2	0.0875	0.1125	0	0	0.2000
	3	0.0375	0	0	0	0.0375
		0.3750	0.3875	0.2000	0.0375	1

Vogliamo determinare la probabilità che in questa famiglia scelta a caso ci siano 0 figli maschi sapendo però che c'è una figlia femmina:

$$\mathbb{P}(Y = 0|X = 1) = \frac{\mathbb{P}(Y = 0, X = 1)}{\mathbb{P}(X = 1)} = \frac{0.1000}{0.3875} = \frac{8}{31} = 0.2581$$

Se confrontiamo $\mathbb{P}(Y = 0|X = 1)$ con $\mathbb{P}(Y = 0)$ ci accorgiamo che ovviamente la probabilità è cambiata. Dunque la probabilità condizionata stravolge la legge di probabilità. Denotiamo questa nuova distribuzione così $(Y|\{X = 1\})$.

Esempio 5.4.2

Siano X e Y due variabile aleatorie discrete con funzione di massa congiunta p data da:

$$p(0, 0) = 0.4 \quad p(0, 1) = 0.2 \quad p(1, 0) = 0.1 \quad p(1, 1) = 0.3$$

Vogliamo determinare la legge $(X|\{Y = 1\})$:

$$\begin{aligned} \mathbb{P}(X = 0|Y = 1) &= \frac{0.2}{0.5} = 0.4 \\ \mathbb{P}(X = 1|Y = 1) &= \frac{0.3}{0.5} = 0.6 \end{aligned}$$

Notiamo che $\mathbb{P}(X = 0|Y = 1) + \mathbb{P}(X = 1|Y = 1) = 1$

5.4.2 Distribuzioni condizionali caso continuo

Siano X e Y due variabili aleatorie con funzione di densità congiunta f . Si dice *densità condizionale* di X rispetto a Y , e si indica con $f_{X|Y}(\cdot|\cdot)$, la funzione di due variabili seguente, che è definita per ogni x e per tutte le y per le quali $f_Y(y) > 0$:

$$f_{X|Y} := \frac{f(x, y)}{f_Y(y)}$$

Esempio 5.4.3

Siano X e Y due variabili aleatorie continue e $f_{(X,Y)}(x, y)$ la loro funzione di densità congiunta data da:

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{12}{5}x(2-x-y) & 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{altrimenti} \end{cases}$$

Vogliamo determinare, fissato $y_0 \in (0, 1)$ la funzione $f_{X|Y}(x|y_0)$. Per farlo, dobbiamo ricorrere alla formula di prima:

$$f_{X|Y} := \frac{f(x, y)}{f_Y(y)}$$

mentre $f_{X,Y}(x, y)$ è data dal problema. Quindi:

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dx = \int_0^1 2.4x(2-x-y) dx \\ &= 2.4 \left[\int_0^1 2x dx - \int_0^1 x^2 dx - y \int_0^1 x dx \right] \\ &= 2.4 \left[x^2 \Big|_0^1 - \frac{x^3}{3} \Big|_0^1 - y \frac{x^2}{2} \Big|_0^1 \right] = 2.4 \left(1 - \frac{1}{3} - \frac{y}{2} \right) = \frac{8}{5} - \frac{6}{5}y \end{aligned}$$

Quindi:

$$f_{X|Y} := \frac{f(x, y)}{f_Y(y)} = \frac{\frac{12}{5}x(2-x-y_0)}{\frac{8}{5} - \frac{6}{5}y_0} = \frac{6x(2-x-y_0)}{4-3y_0}$$

□

5.5 Valore atteso

Nel **caso discreto**, il valore atteso di una variabile aleatoria X discreta è dato da:

$$\mathbb{E}(X) = \sum_{x_i \in S_X} x_i \cdot \mathbb{P}_{\mathbb{X}}(x_i) = \sum_{x_i \in S_X} x_i \cdot \mathbb{P}(X = x_i)$$

con la condizione che esista finito:

$$\mathbb{E}(X) = \sum_{x_i \in S_X} |x_i| \cdot \mathbb{P}(X = x_i) < \infty$$

Nel **caso continuo**, il valore atteso di una variabile aleatoria X continua è dato da:

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_X(x) dx$$

con la condizione che esista finito:

$$\mathbb{E}(|X|) = \int_{\mathbb{R}} |x| \cdot f_X(x) < +\infty$$

5.5.1 Valore atteso di una funzione di variabile aleatoria

Se X è una variabile aleatoria discreta, allora, per ogni funzione reale g :

$$\mathbb{E}[g(X)] = \sum_{x_i \in S_X} g(x_i) \mathbb{P}(X = x_i)$$

Se X è una variabile aleatoria continua con funzione di densità di probabilità $f_X(x)$, allora per ogni funzione reale g :

$$\mathbb{E}[g(x)] = \int_{\mathbb{R}} g(x) f_X(x) dx$$

Proposizione 5.5.1 (Se X ammette la media allora anche $g(X) = aX + b$ ammette la media)

Sia X una variabile aleatoria continua e $g(X)$ una sua funzione:

$$g(X) = ax + b \quad a, b \in \mathbb{R}$$

Allora se X ammette la media allora anche $g(X)$ ammette la media.

Dimostrazione.

$$\begin{aligned} \mathbb{E}(g(X)) &= \mathbb{E}(aX + b) = \int_{\mathbb{R}} (ax + b) \cdot f_X(x) dx \\ &= \int_{\mathbb{R}} [ax f_X(x) + b f_X(x)] dx \\ &= \int_{\mathbb{R}} ax f_X(x) dx + \int_{\mathbb{R}} b f_X(x) dx \\ &= a \int_{\mathbb{R}} x f_X(x) dx + b \int_{\mathbb{R}} f_X(x) dx \\ &= a \mathbb{E}(X) + b \mathbb{E}(1) \\ &= a \mathbb{E}(X) + b \end{aligned}$$

Il caso discreto è uguale (ma si usano le sommatorie).

□

5.5.2 Proprietà valore atteso variabili aleatorie congiunte

Siano X e Y due variabili aleatorie congiunte con funzione di probabilità congiunta $f_{(X,Y)}(x, y)$. Sia poi g una funzione definita come:

$$g(X, Y) := X + Y$$

La media di g è data da:

$$\mathbb{E}[g(X, Y)] = \int_{\mathbb{R}^2} g(X, Y) \cdot f_{X,Y}(x, y) dx dy = \int_{\mathbb{R}^2} (x + y) \cdot f_{X,Y}(x, y) dx dy$$

Proposizione 5.5.2 (La media della somma è uguale alla somma delle medie)

$$\mathbb{E}(g(X, Y)) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Dimostrazione.

$$\begin{aligned} \mathbb{E}(X + Y) &= \int_{\mathbb{R}^2} (x + y) f_{X,Y}(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_{X,Y}(x, y) dx dy \\ \text{(l'integrale di una somma è la somma degli integrali)} &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x f_{X,Y}(x, y) dy \right] dx + \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} y f_{X,Y}(x, y) dx \right] dy \\ &= \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} y f_Y(y) dy = \mathbb{E}(X) + \mathbb{E}(Y) \end{aligned}$$

Ma non possiamo dirlo per la varianza.

□

Se la media esiste (cioè è finita) allora possiamo generalizzare e dire che:

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

Esempio 5.5.1

Un'impresa edile ha recentemente sottoposto i suoi preventivi per tre gare, per degli appalti che le sarebbero profitti per 10.000, 20.000 e 40.000 mila dollari. Se le probabilità di vittoria dei singoli appalti sono rispettivamente 0.2, 0.8 e 0.3, qual'è il profitto totale medio che farà l'azienda?

Siano X_1 , X_2 e X_3 i profitti (in migliaia di dollari) percepiti per i tre lavori. Il profitto totale Y sarà dato da $Y := X_1 + X_2 + X_3$ e quindi:

$$\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3]$$

e quindi:

$$\mathbb{E}(X_1) = 10 \cdot 0.2 + 0 \cdot 0.8 = 2$$

$$\mathbb{E}(X_2) = 16 \cdot 0.8 + 0 \cdot 0.2 = 16$$

$$\mathbb{E}(X_3) = 40 \cdot 0.3 + 0 \cdot 0.7 = 12$$

Perciò:

$$\mathbb{E}(X_1 + X_2 + X_3) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) = 2000 + 16000 + 12000 = 30000$$

Esempio 5.5.2

Una segretaria ha finito di scrivere una pila di N lettere, e ha appena compilato le buste con gli indirizzi, quando tutto il materiale le cade per terra e si mischia. Se si inseriscono le lettere nelle buste in maniera del tutto casuale (nel senso che ciascuna lettera può finire in ogni busta con pari probabilità), qual'è il numero medio di lettere che capitano nella busta corretta?

Sia X il numero di lettere che finiscono nella busta giusta. Il valore atteso $\mathbb{E}(X)$ può essere calcolato molto facilmente notando che $X = X_1 + X_2 + \dots + X_N$, dove:

$$X_i := \begin{cases} 1, & \text{se la lettera } i\text{-esima viene inserita nella busta giusta} \\ 0, & \text{altrimenti} \end{cases}$$

Siccome l' i -esima lettera può finire in una qualunque delle N buste con pari probabilità:

$$\mathbb{P}(X_i = 1) = \frac{1}{N}$$

e quindi:

$$\mathbb{E}(X_i) = 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) = \frac{1}{N}$$

Perciò:

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_N) = N \cdot \frac{1}{N} = 1$$

Quindi, indipendentemente dal numero di lettere presenti, in media vi sarà una sola lettera nella busta giusta.

Esempio 5.5.3

Vogliamo calcolare la probabilità che esca almeno un 6 in quattro lanci di un dado onesto (chiamiamo questo evento A):

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177$$

Vogliamo calcolare la probabilità che esca almeno un doppio 6 in ventiquattro lanci del "doppio dado onesto" (chiamiamo questo evento B):

$$\mathbb{P}(B) = 1 - \mathbb{P}(B^c) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.49$$

Esempio 5.5.4

Ci sono 20 tipi diversi di buoni sconto di cui 1 viene inserito in ogni confezione di biscotti. Se si aprono 10 confezioni, quant'è il valore atteso del numero di buoni differenti che si trovano (chiamiamo questo numero X).

$$i = 1, 2, \dots, 20 \quad X_i = \begin{cases} 1, & \text{se il buono di tipo } i \text{ è presente} \\ 0, & \text{altrimenti} \end{cases}$$

Possiamo dire che:

$$X = \sum_{i=1}^{20} X_i \implies \mathbb{E}(X) = \sum_{i=1}^{20} \mathbb{E}(X_i) = \sum_{i=1}^{20} \mathbb{P}(X_i = 1)$$

Calcoliamo adesso $\mathbb{P}(X_i = 1)$:

$$\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = 1 - \left(\frac{19}{20}\right)^{10}$$

Quindi possiamo concludere:

$$\mathbb{E}(X) = \sum_{i=1}^{20} \left[1 - \left(\frac{19}{20}\right)^{10}\right] = 20 * \left[\left(\frac{19}{20}\right)^{10}\right]$$

5.5.3 La media è il miglior predittore di X rispetto all'errore quadratico medio

La media di una variabile aleatoria X è il **miglior predittore ai fini dell'errore quadratico medio**. L'errore quadratico medio (in inglese *Mean Squared Error*, MSE) indica la discrepanza quadratica media fra i valori dei dati osservati ed i valori dei dati stimati. Con la premessa che la media del numero aleatorio X esiste ed è finita, l'errore quadratico medio è definito come:

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

dove X è il numero da stimare e $\mathbb{E}(X)$ è una sua stima.

Che cosa vuol dire che la media $\mathbb{E}(X)$ è il miglior predittore di X rispetto all'errore quadratico medio? Vuol dire questo:

$$\mathbb{E}(X) = \operatorname{argmin}_{c \in \mathbb{R}} \mathbb{E}[(X - c)^2]$$

cioè il valore $\mathbb{E}(X)$ rende minimo l'errore quadratico medio:

$$\mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \min_{c \in \mathbb{R}} \mathbb{E}[(X - c)^2]$$

Dimostrazione.

$$\begin{aligned} (X - c)^2 &= [X - \mathbb{E}(X) + \mathbb{E}(X) - c]^2 = \{[X - \mathbb{E}(X)] + [\mathbb{E}(X) - c]\}^2 \\ &= [X - \mathbb{E}(X)]^2 + [\mathbb{E}(X) - c]^2 + 2[\mathbb{E}(X) - c][X - \mathbb{E}(X)] \end{aligned}$$

$$\begin{aligned} \mathbb{E}[(X - c)^2] &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} + [\mathbb{E}(X) - c]^2 + \mathbb{E}[X - \mathbb{E}(X)][\mathbb{E}(X) - c] \\ (\text{La quantità } \mathbb{E}[X - \mathbb{E}(X)] &= 0) &= \mathbb{D}^2(X) + [\mathbb{E}(X) - c]^2 \\ (\text{poniamo } c = \mathbb{E}(X)) &= \mathbb{D}^2(X) \end{aligned}$$

Nell'ultimo passaggio, l'unico membro che dipende da c è il secondo addendo. Tale addendo si annulla soltanto con $c = \mathbb{E}(X)$ e questo è proprio il valore che rende minimo l'errore quadratico medio (o lo scarto quadratico medio). □

Nel caso in cui $\mathbb{E}(X^2) = +\infty$ allora bisogna considerare lo **scarto assoluto medio**. L'argomento minimo dello scarto assoluto medio è la **mediana**:

$$\operatorname{argmin}_{c \in \mathbb{R}} \mathbb{E}(|X - c|) = Me \quad (\text{Centro di ordine 1})$$

Mentre la moda in questo contesto è uguale a:

$$\operatorname{argmin}_{c \in \mathbb{R}} \mathbb{E}(|X - c|^0) = Mo \quad (\text{Centro di ordine 0})$$

Mentre la media è uguale a:

$$\mathbb{E}(X) = \operatorname{argmin}_{c \in \mathbb{R}} \mathbb{E}(|X - c|^2) \quad (\text{Centro di ordine 2})$$

Esempio 5.5.5

Siano W, Y, Z tre variabili aleatorie definite in questo modo:

$$\begin{aligned} W &= 0 \text{ con probabilità } 1 && \implies \mathbb{E}(W) = 0 \\ Y &= \begin{cases} -1, & \text{con probabilità } \frac{1}{2} \\ 1, & \text{con probabilità } \frac{1}{2} \end{cases} && \implies \mathbb{E}(Y) = 0 \\ Z &= \begin{cases} -100, & \text{con probabilità } \frac{1}{2} \\ 100, & \text{con probabilità } \frac{1}{2} \end{cases} && \implies \mathbb{E}(Z) = 0 \end{aligned}$$

Se calchiamo la varianza di ogni variabile scopriamo che il numero aleatorio con maggiore variabilità è Z :

$$\begin{aligned} \mathbb{D}^2(W) &= 0 \\ \mathbb{D}^2(Y) &= 1 \\ \mathbb{D}^2(Z) &= 10.000 \end{aligned}$$

Ricordando che:

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

Esempio 5.5.6

Sia $Z_1 \sim U_\alpha(1, 2, 3, 4, 5, 6)$ un numero aleatorio uniforme (Zara 1). Calcoliamo i seguenti valori:

$$\begin{aligned} \mathbb{E}(Z_1) &= \sum_{i=1}^6 i \cdot \frac{1}{6} = \frac{7}{2} \\ \mathbb{E}(Z_1^2) &= \sum_{i=1}^6 i^2 \cdot \frac{1}{6} = \frac{1}{6} \cdot 91 = \frac{91}{6} \\ \mathbb{D}^2(Z_1) &= \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \approx 3 \end{aligned}$$

questo vuol dire che il range è il seguente: $\frac{7}{2} - \sqrt{\frac{35}{12}}, \frac{7}{2} + \sqrt{\frac{35}{12}} \approx [2, 5]$

La media in questo caso **non è un buon predittore**

Ricordiamo che $\mathbb{D}^2(aX + b) = \mathbb{D}^2(aX)$, cioè b non conta, è una semplice traslazione (il fattore b viene chiamato *fattore di posizione* - *location* in inglese) mentre a conta e viene chiamato *fattore di scala*. Ad esempio possiamo esprimere Z utilizzando Y e un fattore di scala 100:

$$Z = 100 \cdot Y$$

5.6 Covarianza

Come abbiamo visto nella sezione precedente, la media della somma di variabili aleatorie coincide con la somma delle loro medie. Per la varianza questo in generale non è vero. Ad esempio:

$$\begin{aligned} \text{Var}(X + X) &= \text{Var}(2X) \\ &= 2^2 \text{Var}(X) \\ &= 4\text{Var}(X) \neq \text{Var}(X) + \text{Var}(X) \end{aligned}$$

Vi è tuttavia un caso importante in cui la varianza della somma di due variabili aleatorie è pari alla somma delle loro varianze, ovvero quando le variabili aleatorie sono indipendenti. Prima di dimostrare questo risultato, dobbiamo definire il concetto di **covarianza** di due variabili aleatorie.

Definizione 5.6.1 (Covarianza)

Siano assegnate due variabili aleatorie X e Y (congiuntamente distribuite) di media μ_X e μ_Y rispettivamente. La loro *covarianza*, che si indica con $Cov(X, Y)$ è (se esiste) la quantità:

$$Cov(X, Y) := \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

Si può ottenere anche una formula alternativa più semplice, analoga a quella della varianza. Si trova espandendo il prodotto al secondo membro.

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\ &= \mathbb{E}[XY] - \mu_X \mathbb{E}[Y] - \mu_Y \mathbb{E}[X] + \mu_X \mu_Y \\ &= \mathbb{E}[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\ &= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] \end{aligned}$$

Dalla definizione si deducono alcune semplici proprietà, quali la *simmetria*:

$$Cov(X, Y) = Cov(Y, X)$$

e il fatto che la covarianza *generalizza il concetto di varianza*:

$$Cov(X, X) = \mathbb{E}(X^2) - \mu_X^2 = \mathbb{D}^2(X)$$

Un'altro enunciato interessante, la cui semplice dimostrazione lasciamo al lettore, è che la covarianza è *bilineare*.

Siano $X = \sum_{i=1}^n \alpha_i X_i$ e $Y = \sum_{j=1}^m \beta_j Y_j$ con α_i e β_j costanti qualsiasi, allora:

$$Cov(X, Y) = Cov\left(\sum_{i=1}^n \alpha_i X_i, \sum_{j=1}^m \beta_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j Cov(X_i, Y_j)$$

Consideriamo adesso un **caso particolare** quando $Y = X$, più precisamente quando:

$$i = 1, 2, \dots, n \quad \alpha_i = 1 \quad Y = X = \sum_{i=1}^n X_i$$

Allora scopriamo questo risultato:

$$\begin{aligned} \mathbb{D}^2(X) &= \mathbb{D}^2\left(\sum_{i=1}^n X_i\right) = Cov(X, X) = Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) \\ (\text{simmetria della covarianza}) \quad &= \sum_{i=1}^n \mathbb{D}^2(X_i) + 2 \sum_{i=1}^n \sum_{j>i}^n Cov(X_i, X_j) \end{aligned}$$

*In generale possiamo affermare quindi che la varianza della somma NON è uguale alla somma delle varianze. Esiste un solo caso in cui questo è effettivamente vero, quando la covarianza di due variabili aleatorie U e V è zero $Cov(U, V) = 0$ e si dice che U e V sono **variabili aleatorie non correlate**. Questo succede quando $\mathbb{E}(U)\mathbb{E}(V) = \mathbb{E}(UV)$ e questo succede quando U e V sono indipendenti. Quindi l'unico caso in cui la varianza della somma è uguale alla somma delle varianze è quando le variabili aleatorie non sono correlate e ciò accade quando:*

$$Cov(U, V) = 0 \iff \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V) = 0 \iff \mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V)$$

Teorema 13 (Due variabili se sono indipendenti allora sono anche non correlate)

Siano U e V due variabili congiuntamente distribuite. Se U e V sono indipendenti allora sono anche non correlate.

Dimostrazione (caso discreto).

$$\begin{aligned}
 \mathbb{E}(UV) &= \sum_i \sum_j u_i v_j \mathbb{P}(U = u_i, V = v_j) \\
 (\text{per HP } U \text{ e } V \text{ indipendenti}) &= \sum_i \sum_j u_i v_j \mathbb{P}(U = u_i) \mathbb{P}(V = v_j) \\
 &= \sum_i u_i \mathbb{P}(U = u_i) \sum_j v_j \mathbb{P}(V = v_j) \\
 &= \mathbb{E}(U) \cdot \mathbb{E}(V)
 \end{aligned}$$

□

Dimostrazione (caso continuo).

$$\begin{aligned}
 \mathbb{E}(UV) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} uv f_{(U,V)}(u, v) du dv \\
 (f_{(X,Y)} \text{ si fattorizza per indipendenza HP}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} uv f_U(u) f_V(v) du dv \\
 &= \int_{-\infty}^{+\infty} u f_U(u) du \cdot \int_{-\infty}^{+\infty} v f_V(v) dv \\
 &= \mathbb{E}(U) \cdot \mathbb{E}(V)
 \end{aligned}$$

□

Esempio 5.6.1

Si determini la varianza del numero di "teste" nel lancio di 10 monete oneste.

$$S_{10} = X_{1,1} + X_{1,2} + \dots + X_{1,10}$$

Siccome ogni lancio è indipendente dall'altro possiamo dire sfruttare il teorema per calcolare facilmente la varianza:

$$\begin{aligned}
 \mathbb{D}^2(S_{10}) &= \mathbb{D}^2\left(\sum_{i=1}^{10} X_{1,i}\right) = \sum_{i=1}^{10} \mathbb{D}^2(X_{1,i}) \\
 &= \mathbb{D}^2(X_1) \sum_{i=1}^{10} 1 \\
 &= 10 \cdot p \cdot (1 - p) \\
 (\text{monete oneste}) \quad &= 10 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{2}
 \end{aligned}$$

Esempio 5.6.2

Si calcoli la varianza della somma dei punteggi nel lancio di 10 dadi onesti.

$$S_{10} = Z_{1,1} + \dots + Z_{1,10}$$

L'indipendenza implica che le variabili non sono correlate, dunque:

$$\mathbb{D}^2\left(\sum_{i=1}^{10} Z_{1,i}\right) = \sum_{i=1}^{10} \mathbb{D}^2(Z_{1,i}) = \sum_{i=1}^{10} \mathbb{D}^2(Z_1) = \mathbb{D}^2(Z_1) \sum_{i=1}^{10} 1 = 10 \cdot \frac{35}{2} = \frac{175}{2}$$

Esempio 5.6.3 (Esempio chiarificatore significato covarianza)

Sia $(\Omega, \mathcal{F}, \mathbb{P})$ una terna di probabilità e $A, B \in \mathcal{F}$ due eventi. Consideriamo poi questi numeri aleatori:

$$1_A = \begin{cases} 1, & \text{se } \omega \in A \\ 0, & \text{altrimenti} \end{cases} \quad 1_B = \begin{cases} 1, & \text{se } \omega \in B \\ 0, & \text{altrimenti} \end{cases} \quad 1_A 1_B = \begin{cases} 1, & \text{se } \omega \in A \cap B \\ 0, & \text{altrimenti} \end{cases}$$

Le leggi di questi numeri aleatori sono:

$$1_A \cdot \begin{pmatrix} 1 & 0 \\ \mathbb{P}(A) & 1 - \mathbb{P}(A) \end{pmatrix} \quad 1_B \cdot \begin{pmatrix} 1 & 0 \\ \mathbb{P}(B) & 1 - \mathbb{P}(B) \end{pmatrix} \quad 1_A 1_B \cdot \begin{pmatrix} 1 & 0 \\ \mathbb{P}(A \cap B) & 1 - \mathbb{P}(A \cap B) \end{pmatrix}$$

Calcoliamo adesso la covarianza di 1_A e 1_B :

$$\begin{aligned} \text{Cov}(1_A, 1_B) &= \mathbb{E}(1_A 1_B) - \mathbb{E}(1_A)\mathbb{E}(1_B) \\ &= \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(1_A 1_B = 1) - \mathbb{P}(1_A = 1)\mathbb{P}(1_B = 1) \\ &= \mathbb{P}(1_A = 1, 1_B = 1) - \mathbb{P}(1_A = 1)\mathbb{P}(1_B = 1) \end{aligned}$$

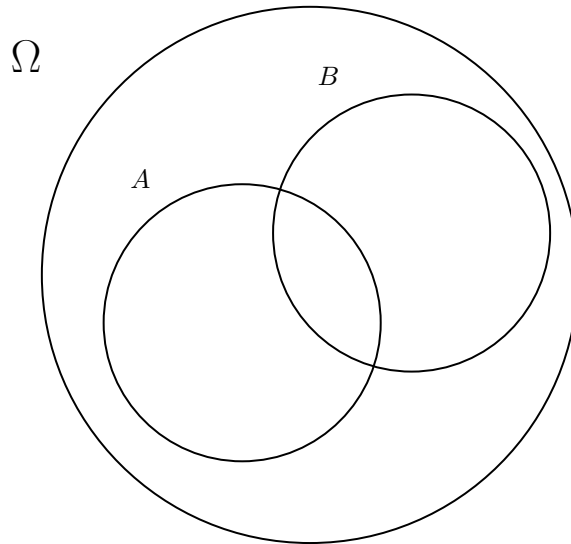
Da questo risultato possiamo dire che:

$$\begin{aligned} \text{Cov}(1_A, 1_B) > 0 &\iff \mathbb{P}(1_A = 1, 1_B = 1) > \mathbb{P}(1_A = 1)\mathbb{P}(1_B = 1) \\ &\iff \frac{\mathbb{P}(1_A = 1, 1_B = 1)}{\mathbb{P}(1_B = 1)} > \mathbb{P}(1_A = 1) \\ &\iff \mathbb{P}(1_A = 1 | 1_B = 1) > \mathbb{P}(1_A = 1) \end{aligned}$$

Quando la covarianza è positiva, la probabilità $\mathbb{P}(1_A = 1 | 1_B = 1)$ (cioè che 1_A sia uguale a 1 sapendo che 1_B è uguale a 1) è aumentata rispetto alla probabilità secca $\mathbb{P}(1_A = 1)$ (cioè la probabilità che $1_A = 1$ senza sapere nulla) e quando questo accade la covarianza è positiva (è quindi un se e solo se).

Se la covarianza è uguale a 0, le due probabilità considerate sono uguali.

La covarianza di due variabili statistiche o variabili aleatorie è un valore numerico che fornisce una misura di quanto le due varino assieme, ovvero della loro dipendenza. Un'interpretazione potrebbe essere che quando c'è un'intersezione consistente tra gli eventi A e B allora la covarianza assume un valore maggiore di 0. Quando non c'è intersezione, solo uno tra 1_A e 1_B può assumere il valore 1 (le variabili in questo caso sono indipendenti e hanno covarianza uguale a 0).



5.7 Disuguaglianza di Markov

Proposizione 5.7.1 (Disuguaglianza di Markov)

Se X è una variabile aleatoria che non è mai negativa (si dice che X è *quasi certamente* positivo, cioè $\mathbb{P}(X > 0) = 1$) allora per ogni $a > 0$ si ha che:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Dimostrazione (caso continuo con densità f)

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{+\infty} x f_X(x) dx \\ &= \int_0^a x f_X(x) dx + \int_a^{+\infty} x f_X(x) dx \\ (\text{entrambe le quantità sono maggiori di } 0) &\geq \int_a^{+\infty} x f_X(x) dx \\ (\text{perché } x \geq a \text{ nella regione di integrazione}) &\geq \int_a^{+\infty} a \cdot f_X(x) dx = a \int_a^{+\infty} f_X(x) dx = a \cdot \mathbb{P}(X \geq a) \end{aligned}$$

e quindi:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

□

5.7.1 Quand'è che un evento è quasi certamente 1 o quasi certamente impossibile?

Mettiamoci nell'esperimento del lancio di una moneta un numero indefinito di volte. La terna di probabilità di questo esperimento è definita come:

$$(\{C, T\}^\infty, \mathcal{F} = \sigma((T_n)_{n \in \mathbb{N}}), \mathbb{P}(T_n) = p \ n \in \mathbb{N})$$

Vediamo un esempio di un evento quasi certamente impossibile. Consideriamo la successione costituita da tutte teste. Qual'è la probabilità di questo evento? E' necessaria un'operazione di passaggio a limite:

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}(T_1 \cap T_2 \cap \dots \cap T_n \cap \Omega) \\ (\text{la successione è decrescente Teorema 12}) &= \mathbb{P}\left(\lim_{n \rightarrow \infty} T_1 \cap T_2 \cap \dots \cap T_n \cap \Omega\right) \\ (n \rightarrow \infty \text{ quindi è una successione di tutte teste}) &= \mathbb{P}(T_1 \cap T_2 \cap \dots \cap T_n \cap \dots) \\ (\text{eventi indipendenti - il limite fa zero perché } p \in (0, 1)) &= \lim_{n \rightarrow \infty} p^n = 0 \end{aligned}$$

Abbiamo trovato un evento diverso dall'evento impossibile ma che ha probabilità uguale a 0. Diciamo che questo evento è **quasi certamente impossibile** perché questa successione di eventi appartiene allo spazio campione $\{T, C\}^\infty$ e quindi non è del tutto impossibile. Questa situazione si presenta quando la cardinalità dello spazio campione è maggiore di quella del numerabile.

Questo comporta che la probabilità di ogni successione è nulla. Tale risultato è un bene perché se non fosse così troveremmo una crepa nel sistema assiomatico (una contraddizione).

Supponiamo per assurdo che ogni successione abbia probabilità $\epsilon > 0$. Siccome la cardinalità delle sequenze è maggiore del numerabile, troviamo che la somma delle probabilità di tutte queste sequenze (infinite, maggiore del numerabile) si troverà ad avere una probabilità maggiore di 1 (contraddizione va contro gli assiomi della probabilità).

5.8 Disuguaglianza di Chebyshev

Proposizione 5.8.1 (Disuguaglianza di Chebyshev)

Se X è una variabile aleatoria con media finita μ_X e varianza σ^2 , allora per ogni $\epsilon > 0$ si ha che:

$$\mathbb{P}(|X - \mu_X| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Dimostrazione.

Gli eventi $\{|X - \mu_X| \geq \epsilon\}$ e $\{(X - \mu)^2 \geq \epsilon^2\}$ coincidono e sono quindi equiprobabili. Visto che $(X - \mu_X)^2$ è una variabile aleatoria non negativa, possiamo applicarle la disuguaglianza di Markov con $a = \epsilon^2$:

$$\begin{aligned} \mathbb{P}(|X - \mu_X| \geq \epsilon) &= \mathbb{P}[(X - \mu_X)^2 \geq \epsilon^2] \\ (\text{Disuguaglianza di Markov}) &\leq \frac{\mathbb{E}[(X - \mu_X)^2]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2} \end{aligned}$$

□

Esempio 5.8.1 (Disuguaglianza di Chebyshev applicata al numero esponenziale)

Sia $X \sim \text{Esp}(\alpha)$ con media $\mathbb{E}(X) = \frac{1}{\alpha}$ e varianza $\mathbb{D}^2(X) = \frac{1}{\alpha^2}$. Determiniamo genericamente questa probabilità applicando la disuguaglianza di Chebyshev:

$$\mathbb{P}(|X - \frac{1}{\alpha}| > \epsilon) \leq \frac{1/\alpha^2}{\epsilon^2}$$

Se poniamo $\epsilon = \frac{1}{\alpha}$ otteniamo che questa probabilità non sarà mai più grande di 1 (limite superiore):

$$\mathbb{P}(|X - \frac{1}{\alpha}| > \frac{1}{\alpha}) \leq \frac{1/\alpha^2}{1/\alpha^2} = 1$$

Se invece poniamo $\epsilon = \frac{2}{\alpha}$ otteniamo che questa probabilità non sarà mai più grande di $\frac{1}{4}$:

$$\mathbb{P}(|X - \frac{1}{\alpha}| > \frac{2}{\alpha}) \leq \frac{1/\alpha^2}{4/\alpha^4} = \frac{1}{4}$$

Esempio 5.8.2

Il numero di pezzi prodotti da una fabbrica durante una settimana è una variabile aleatoria di media 50.

1. Cosa si può dire sulla probabilità che la produzione superi occasionalmente i 75 pezzi?
2. Se si suppone nota anche la varianza, pari a 25, cosa si può dire sulla probabilità che la produzione sia compresa tra i 40 e i 60 pezzi?

Denotiamo con X la variabile aleatoria che indica il numero di pezzi prodotti in una settimana. Risposta 1: per la disuguaglianza di Markov,

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}(X)}{75} = \frac{50}{75} = \frac{2}{3}$$

Risposta 2: applicando la disuguaglianza di Chebyshev,

$$\mathbb{P}(40 \leq X \leq 60) = \mathbb{P}(|X - 50| \geq 10) \leq \frac{25}{10^2} = \frac{1}{4}$$

5.9 La legge debole dei grandi numeri

Teorema 14 (La legge debole dei grandi numeri)

Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. (*independenti e identicamente distribuite*), tutte con media e varianza finita ed uguale (chiamiamo la media $\mathbb{E}(X) = \mu$ e la varianza $\mathbb{D}^2(X) = \sigma^2$). Allora, se $n \in \mathbb{N}$ e $S_n = \sum_{i=1}^n X_i$ è la variabile aleatoria ottenuta sommando tutte le X_i :

$$\frac{S_n}{n} \rightarrow \mu \quad (\text{converge in probabilità})$$

$\frac{S_n}{n}$ rappresenta la media aritmetica e questa converge alla media comune μ .
Dimostrazione.

$$\begin{aligned} n \in \mathbb{N} \quad \mathbb{E}(S_n) &= \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i) = n \cdot \mu \\ \text{quindi} \quad \mathbb{E}\left(\frac{S_n}{n}\right) &= \frac{1}{n} \mathbb{E}(S_n) = \frac{n\mu}{n} = \mu \end{aligned}$$

Vediamo la varianza:

$$\begin{aligned}\mathbb{D}^2(S_n) &= \mathbb{D}^2\left(\sum_{i=1}^n X_i\right) \\ (\text{indipendenza - proprietà varianza}) &= \sum_{i=1}^n \mathbb{D}^2(X_i) \\ (\text{tutte le variabili hanno la stessa legge}) &= n \cdot \sigma^2\end{aligned}$$

$$\textbf{quindi} \quad \mathbb{D}^2\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \mathbb{D}^2(S_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

□

Applichiamo la disuguaglianza di Chebycev a $\frac{S_n}{n}$:

$$n \in \mathbb{N} \quad \epsilon > 0, \quad \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\mathbb{D}^2(S_n/n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Siccome vale per ogni $n \in \mathbb{N}$, è possibile calcolare il limite:

$$\lim_{n \rightarrow \infty} \left(\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \right) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2/\epsilon^2}{n} = 0$$

Quindi il limite è minore o uguale a 0. Quindi:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0 \iff \frac{S_n}{n} \rightarrow \mu$$

5.9.1 Teorema di Bernoulli

Questa è la situazione della legge debole nel caso del modello del lancio della moneta un numero indefinito di volte. Questa situazione è un caso particolare della legge debole.

Sia $X_n \sim B(1, p)$ con $n \in \mathbb{N}$. Per ogni $i = 1, \dots, n$ abbiamo che tutte le X_i sono indipendenti ed identicamente distribuite. Sappiamo che $X_n = 1$ con probabilità p e $X_n = 0$ con probabilità $1 - p = q$. La media è $\mathbb{E}(X_n) = p$ e la varianza è $\mathbb{D}^2(X_n) = pq$.

In questo caso, la media aritmetica $\frac{S_n}{n}$ rappresenta la **frequenza relativa delle teste nei primi lanci** e questa quantità converge in probabilità alla media p :

$$\frac{S_n}{n} \rightarrow p \quad (\text{converge in probabilità})$$

Statistica descrittiva

6 Statistica descrittiva

6.1 Definizione di statistica descrittiva

Con il termine statistica descrittiva si intende un insieme di tecniche e strumenti finalizzati ad assolvere uno dei principali compiti assegnati alla statistica:

*descrivere, rappresentare e sintetizzare in maniera opportuna un insieme di dati relativamente ad una o più **caratteristiche** di una popolazione di interesse*

Per **popolazione** (o *collettivo statistico*) si intende la totalità dei casi, ovvero dei *membri* (o *unità statistiche*) sui quali è possibile *rilevare* uno o più *caratteri* che rivestono particolare importanza per il fenomeno che si sta studiando.

Esempio 6.1.1

Durante il semestre viene proposto agli studenti dei corsi di Laurea offerti in Ateneo, il questionario sulla valutazione della struttura didattica nella sua complessità.

Qui la *popolazione* è costituita dagli studenti con o senza obbligo di frequenza ciascuno dei quali è un *membro*. Per lo studio in questione ci si serve di un questionario avente un certo numero di domande raggruppate per sezioni. In questo contesto ciascuna domanda del questionario corrisponde ad un *carattere*. Per la maggior parte delle domande, lo studente può rispondere scegliendo una tra le quattro possibili risposte tra loro alternative:

Decisamente no Più no che sì Più sì che no Decisamente sì

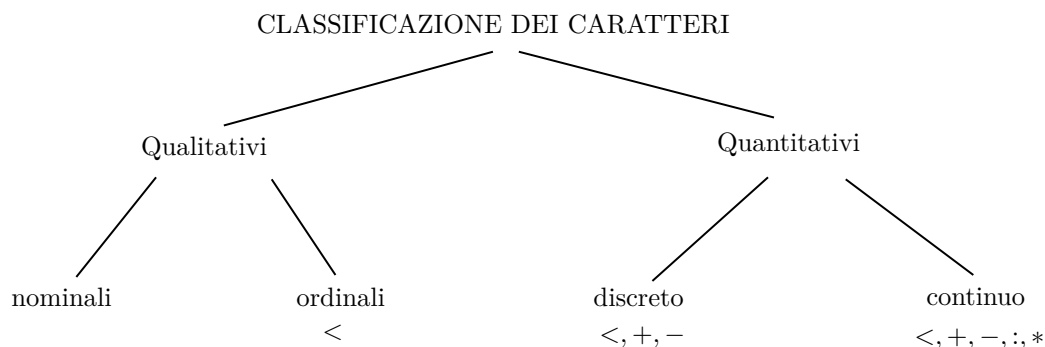
Esempio 6.1.2

In un allevamento di bufale da latte si vuole mettere in relazione la produzione giornaliera di latte con la grandezza delle mammelle e con lo stato di salute delle stesse.

Qui la *popolazione* è costituita da tutte le bufale dell'allevamento ciascuna delle quali è un *membro*. Per lo studio in questione i *caratteri* da rilevare su ciascun membro sono: la produzione di latte, una misura lineare delle mammelle, lo stato di salute delle mammelle.

6.2 Tipo e modalità dei caratteri

Le diverse espressioni con le quali si manifesta un carattere si chiamano **modalità**. Una modalità di un carattere è uno dei suoi possibili valori. Per caratteri quantitativi si usa spesso anche il termine valore. Le proprietà (logiche, aritmetiche o di altra natura) delle modalità permettono di classificare il carattere:



Nell'Esempio 6.1.1 le modalità, per ciascun carattere, sono rappresentate dalle quattro risposte alternative tra loro. Ciascuno dei caratteri, ovvero una domanda proposta nel questionario, è di tipo **qualitativo ordinale**. Infatti, si osservi che esse sono delle etichette e che in relazione al gradimento espresso dallo studente risulta:

Decisamente no < Più no che sì < Più sì che no < Decisamente sì

Invece, nell'Esempio 6.1.2, i caratteri "produzione giornaliera di latte" e "grandezza delle mammelle" sono **quantitativi continui** in quanto per la loro determinazione sono necessari uno strumento di misurazione di una capacità (volume) e uno strumento per la misurazione di una lunghezza e pertanto i dati ottenuti sono numeri decimali appartenenti ad un conveniente intervallo. Nello stesso esempio, il carattere "stato di salute delle mammelle" è di tipo qualitativo ordinale e le modalità sono i diversi valori della scala prescelta (ad esempio, mammelle sane, infiammazione lieve, infiammazione moderata, infiammazione grave). Il fatto stesso che si parla

di scala comporta la presenza di un ordinamento tra le etichette.

Per un esempio di carattere **qualitativo nominale** si pensi al *gruppo sanguigno* che, prescindendo dal fattore Rh, si manifesta con le modalità: A, B, AB e 0.

Per un esempio di carattere **quantitativo discreto** si pensi al numero delle persone presenti nello stato di famiglia dei residenti nel comune di Napoli alla data del più recente censimento ISTAT.

6.3 Distribuzioni di frequenza

Si può senz'altro pensare ad una popolazione come ad un insieme. La cardinalità della popolazione *rilevata* è detta **taglia**; essa, di solito, si designa con la lettera N . Un carattere viene designato con una lettera latina maiuscola mentre i valori rilevati vengono rappresentati con la stessa lettera ma in minuscolo. Se, allora, Y rappresenta il carattere sotto studio non continuo, la sequenza:

$$\underline{y} = (y_1, y_2, \dots, y_N)$$

rappresenta l'intera rilevazione dei dati. Denotiamo ora con x_1, x_2, \dots, x_k (con $k \leq N$) le modalità del carattere Y . Per $j = (1, 2, \dots, k)$, il numero n_j che rappresenta quante volte è presente la modalità x_j in \underline{y} è detto **frequenza assoluta** (o semplicemente *frequenza*) della modalità x_j . In aggiunta:

$$j = (1, 2, \dots, k), \quad f_j = \frac{n_j}{N}$$

è detta **frequenza relativa** della modalità x_j . La frequenza relativa è più informativa della frequenza assoluta in quanto tiene conto anche della taglia. E' del tutto ovvio che la somma delle k frequenze assolute vale N mentre la somma delle k frequenze relative vale 1.

Infine,

$$j = (1, 2, \dots, k), \quad F_j = f_1 + f_2 + \dots + f_j = F_{j-1} + f_j$$

è detta **frequenza assoluta cumulata** di x_j . E' del tutto ovvio che $F_k = 1$.

La rappresentazione tabellare di quanto appena esposto è detta *distribuzione di frequenza* della rilevazione dati $\underline{y} = (y_1, y_2, \dots, y_N)$ (la tabella è ordinata):

modalità del carattere	frequenza assoluta	frequenza relativa	frequenza relativa cumulata
x_1	n_1	$f_1 = \frac{n_1}{N}$	$F_1 = f_1$
x_2	n_2	$f_2 = \frac{n_2}{N}$	$F_2 = F_1 + f_2$
\dots	\dots	\dots	\dots
\dots	\dots	\dots	\dots
x_{k-1}	n_{k-1}	$f_{k-1} = \frac{n_{k-1}}{N}$	$F_{k-1} = F_{k-2} + f_{k-1}$
x_k	n_k	$f_k = \frac{n_k}{N}$	1
	N	1	

Se vogliamo trovare delle analogie con ciò che abbiamo studiato nei capitoli precedenti, possiamo dire che un carattere rappresenta un numero aleatorio. Le modalità del carattere rappresenta il suo spettro e le frequenze relative rappresentano le probabilità di ogni modalità. Siccome:

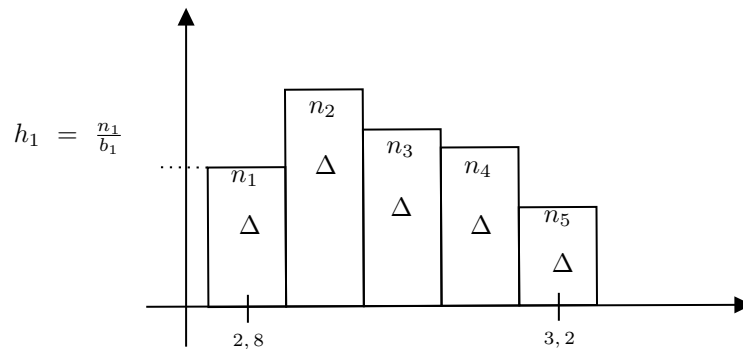
$$x_j \in S_X, \quad F_X(x_j) = \mathbb{P}(X \leq x_j) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + \dots + \mathbb{P}(X = x_j)$$

è chiaro che la somma di tutte le frequenze relative cumulate è uguale a 1.

6.4 Rappresentare un carattere quantitativo continuo - Istogramma

Per un carattere quantitativo continuo è necessario dapprima procedere alla suddivisione in *classi di modalità* dell'intervallo nel quale si manifesta il carattere stesso. Qui bisogna fare attenzione a rendere le classi contigue ma senza ingenerare dubbi di collocazione di un dato nelle classi stesse. Allo scopo, se i dati (y_1, y_2, \dots, y_N) sono riportati con s cifre decimali, è sufficiente tenere conto dell'operazione di arrotondamento e rappresentare gli estremi delle classi con $s + 1$ cifre decimali. Ad esempio, supponiamo che una rilevazione di un peso è effettuata con bilancia digitale precisa all'ettogrammo. Sia 3,2 kg il peso minore rilevato: 3,2 rappresenta tutte le misurazioni comprese nell'intervallo $[3, 15; 3, 25]$. Allo stesso modo sia 4,4 kg il peso maggiore rilevato: 4,4 rappresenta tutte le misurazioni comprese nell'intervallo $[4, 35; 4, 45]$. Quindi se è vero che l'intervallo nel quale si osservano i dati è $[3, 2; 4, 4]$ è a maggior ragione vero che senza l'operazione di arrotondamento esso sarebbe stato $[3, 15; 4, 45]$. Allora, è quest'ultimo intervallo che deve essere suddiviso nel numero desiderato di

classi e queste devono avere come estremi dei numeri aventi due cifre decimali. Dopo di ciò, per un carattere quantitativo, la prima colonna contiene le classi di modalità così individuate. Raffiguriamo ciò che abbiamo detto con un **istogramma**:



dove la quantità Δ è definita in questo esempio come:

$$\Delta = \frac{M + 0,05 - (m - 0,05)}{5} = \frac{M - m + 0,1}{5} = 0,1$$

L'altezza dei rettangolini è la *frequenza assoluta* oppure la *frequenza relativa* fratto la base del rettangolo. La frequenza per ogni rettangolo è proprio l'area.

6.5 Moda e quartili

Definizione 6.5.1 (Definizione di moda)

Si consideri un carattere (di qualsiasi tipo). La modalità corrispondente alla frequenza (assoluta o relativa) più grande viene detta *moda* (M_0) della rilevazione dati.

□

Definizione 6.5.2 (Definizione di mediana o secondo quartile)

Si consideri un carattere (non qualitativo nominale). La modalità corrispondente alla più piccola frequenza relativa cumulata maggiore o uguale a 0,5 viene detta *mediana* (M_1) oppure *secondo quartile* (Q_2) della rilevazione dati. La mediana suddivide la rilevazione dati ordinata:

$$(y_1, y_2, \dots, y_N) \quad \text{con} \quad y_1 \leq y_2 \leq \dots \leq y_N$$

in due parti: i dati minori della mediana sono nello stesso numero dei dati maggiori della mediana.

□

Definizione 6.5.3 (Definizione di terzo quartile)

Si consideri un carattere (non qualitativo nominale). La modalità corrispondente alla più piccola frequenza relativa cumulata maggiore o uguale a 0,75 viene detta *terzo quartile* (Q_3) della rilevazione dati. Il terzo quartile corrisponde anche alla mediana dei dati maggiori della mediana.

□

Definizione 6.5.4 (Definizione quartile di ordine zero)

Si consideri un carattere (non qualitativo nominale). Il dato più piccolo y_1 è il *quartile di ordine zero* Q_0 della rilevazione dati.

□

Definizione 6.5.5 (Definizione di quarto quartile)

Si consideri un carattere (non qualitativo nominale). Il dato più grande y_N è il *quarto quartile* (Q_4) della rilevazione dati.

6.6 Rappresentazioni grafiche

Le distribuzioni di frequenza possono essere rappresentate in forma grafica con scelta eseguita opportunamente rispetto al tipo di carattere. Per i caratteri qualitativi e quantitativi discreti, oltre al *diagramma circolare*, la rappresentazione grafica più usata è quella del *diagramma a barre verticali*: ogni barra verticale è centrata attorno ad una modalità ed ha altezza pari alla frequenza assoluta (oppure relativa) e la larghezza di ogni barra verticale non conta a niente.

Per i soli *caratteri quantitativi continui* è opportuno utilizzare la rappresentazione grafica detta *istogramma*.

La differenza qualitativa tra un istogramma ed un diagramma a barre verticali è che nell'istogramma le barre verticali **devono essere contigue** (in effetti sono dei rettangoli e, quindi, dotati di base e altezza). Ma la differenza sostanziale è che nell'istogramma la frequenza della classe di modalità rappresenta l'area del rettangolo (e non la sua altezza come nel diagramma a barre).

Pertanto, per qualsiasi $j \in \{1, 2, \dots, k\}$, se b_j rappresenta l'ampiezza della j -ma classe di modalità, l'altezza h_j del relativo rettangolo è ottenuto mediante la formula inversa per l'area e quindi:

$$j = (1, 2, \dots, k), \quad h_j = \frac{n_j}{b_j}.$$

6.7 Indici di posizione

Si vuole ora considerare il problema di **sintetizzare** la rilevazione dati

$$\underline{y} = (y_1, y_2, \dots, y_N)$$

con un unico valore di *tendenza centrale*, ossia un valore che fornisca un'indicazione di massima sulla locazione di \underline{y} . Ciò è utile non solo per una più immediata comprensione dei risultati dell'indagine ma anche per istituire un confronto del fenomeno studiato con altri fenomeni dello stesso tipo.

Per i caratteri quantitativi è possibile fare ricorso ai due indici media e mediana. Nel caso di una taglia N dispari, un modo pratico per ottenere la mediana senza costruire la distribuzione di frequenza è quello di ordinare i dati dal più piccolo al più grande e poi, ricorsivamente, depennare il minimo e il massimo fino a quando resta un unico elemento che è la mediana. Se invece N è pari, alla fine di tutti i depennamenti restano due elementi. In tal caso, bisogna separare due casi:

1. i due elementi restanti sono uguali (il loro valore comune coincide con la mediana)
2. i due elementi restanti sono diversi

Nel primo caso, con un *carattere quantitativo*, la mediana è la semisomma dei due elementi restanti. Nel secondo caso con un *carattere qualitativo ordinale* la mediana è indeterminata.

Per i *caratteri quantitativi* ci sono almeno altri due approcci teorici in grado di far ottenere indici di tendenza centrale:

- le *medie analitiche*
- i *centri*

6.7.1 Medie analitiche

Sia $\underline{y} = (y_1, y_2, \dots, y_N)$ una rilevazione dati su un *carattere quantitativo*. Per ottenere una **media analitica** bisogna dapprima specificare un criterio C (o *funzione di circostanza*) rispetto al quale si vuole ottenere la valutazione di tendenza centrale. Dopo di ciò bisogna determinare un numero reale y per il quale, indicata con $y^* = (y, y, \dots, y)$ una rilevazione dati (fittizia) di taglia N aventi tutti gli elementi uguali a y , la valutazione della funzione di circostanza C su \underline{y} deve coincidere con la valutazione di C su y^* . In simboli:

$$C(y_1, y_2, \dots, y_N) = C(y, y, \dots, y).$$

Esempio 6.7.1 (Media aritmetica)

Si scelga come funzione di circostanza C la somma dei dati:

$$C(y_1, y_2, \dots, y_N) = y_1 + y_2 + \dots + y_N.$$

Dopo di ciò,

$$\begin{aligned} C(y_1, y_2, \dots, y_N) &= C(y, y, \dots, y) \\ \iff y_1 + y_2 + \dots + y_N &= y + y + \dots + y \quad (n \text{ volte}) \\ \iff y_1 + y_2 + \dots + y_N &= N \cdot y. \end{aligned}$$

In definitiva, la soluzione dell'equazione di circostanza è la *media aritmetica* dei dati:

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_N}{N} = \frac{1}{N} \sum_{i=1}^N y_i.$$

□

Esempio 6.7.2 (Media geometrica)

Si scelga come funzione di circostanza C il prodotto dei dati (che devono essere rilevati da un carattere positivo):

$$C(y_1, y_2, \dots, y_N) = y_1 \cdot y_2 \cdots y_N$$

Dopo di ciò,

$$\begin{aligned} C(y_1, y_2, \dots, y_N) = C(y, y, \dots, y) &\iff y_1 \cdot y_2 \cdots y_N = y \cdot y \cdots y \quad (n \text{ volte}) \\ &\iff y_1 \cdot y_2 \cdots y_N = y^N. \end{aligned}$$

In definitiva, la soluzione dell'equazione di circostanza è la *media geometrica* dei dati:

$$M_g = \sqrt[N]{y_1 \cdot y_2 \cdots y_N} = \sqrt[N]{\prod_{i=1}^N y_i}$$

□

Esempio 6.7.3 (Media armonica)

Note del professore

6.7.2 Centri

Sia $\underline{y} = (y_1, y_2, \dots, y_N)$ una rilevazione dati su un *carattere quantitativo* Y e sia:

$$\forall x \in \mathbb{R}, \quad d(x, \underline{y}) \geq 0$$

una funzione che si ritiene adatta a rappresentare la *distanza* di un generico valore reale x da tutti gli elementi della rilevazione dati \underline{y} . Si definisce *centro* di una rilevazione dati \underline{y} , e lo si indica con $\xi(\underline{y})$, il punto di minimo assoluto dalla funzione $d(x, \underline{y})$; in simboli:

$$\xi(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d(x, \underline{y})$$

In particolare, sono molto spesso considerate le seguenti funzioni di distanza *di tipo potenze*:

$$\forall x \in \mathbb{R}, \quad d_0(x, \underline{y}) = \frac{1}{N} \sum_{i=1}^N |x - y_i|^0$$

e

$$\forall r \in \mathbb{N}, \quad \forall x \in \mathbb{R}, \quad d_r(x, \underline{y}) = \sqrt[r]{\frac{1}{N} \sum_{i=1}^N |x - y_i|^r}.$$

La funzione $d_r(x, \underline{y})$ è detta anche *distanza di ordine r* e il suo punto di minimo assoluto è detto *centro di ordine r* . Quindi il centro di ordine r di una rilevazione dati \underline{y} è il numero reale $\xi_r(\underline{y})$ che rende minima la distanza di ordine r . In simboli:

$$\xi_r(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_r(x, \underline{y}).$$

Teorema 15 (Il centro di ordine 0 di \underline{y} coincide con la moda di \underline{y})

Il centro di ordine 0 della rilevazione dati \underline{y} , ovvero $\xi_0(\underline{y})$, coincide con la moda della rilevazione dati.

Dimostrazione.

Per definizione,

$$d_0(x, \underline{y}) = \frac{1}{N} \sum_{i=1}^N |x - y_i|^0$$

rappresenta la distanza di ordine 0 di x dall'intera rilevazione dati \underline{y} , mentre $|x - y_i|^0$ rappresenta la distanza tra x e il generico dato y_i .

Se ne ricava che quando x coincide con y_i la distanza è nulla ovvero l'addendo i -mo non porta contributo alla distanza complessiva. Pertanto, si ha:

$$d_0(x, \underline{y}) = \begin{cases} 1, & \text{se } x \notin \{y_1, y_2, \dots, y_n\} \\ 1 - \frac{n_x}{N} < 1, & \text{se } x \in \{y_1, y_2, \dots, y_n\}. \end{cases}$$

n_x rappresenta il numero delle volte che si presenta il dato x . Allora, il minimo si trova tra gli elementi di \underline{y} e precisamente è quel dato al quale compete la frequenza maggiore che per definizione è la moda della rilevazione dati:

$$\xi_0(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^N |x - y_i|^0 = M_0.$$

□

Teorema 16 (Il centro di ordine 1 di \underline{y} coincide con la mediana)

Il centro di ordine 1 della rilevazione dati \underline{y} , ovvero $\xi_1(\underline{y})$, coincide con la mediana:

$$\xi_1(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_1(x, \underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^N |x - y_i| = M_1 \equiv Q_2.$$

Senza dimostrazione.

□

Teorema 17 (Il centro di ordine 2 della rilevazione dati \underline{y} coincide con la media aritmetica)

Il centro di ordine 2 della rilevazione dati \underline{y} , cioè $\xi_2(\underline{y})$, coincide con la media aritmetica.

Dimostrazione.

Per definizione,

$$d_2(x, \underline{y}) = \sqrt[2]{\frac{1}{N} \sum_{i=1}^N |x - y_i|^2} = \sqrt[2]{\frac{1}{N} \sum_{i=1}^N (x - y_i)^2}$$

D'altra parte, dal momento che la funzione radice quadrata è strettamente crescente nel suo dominio $[0, +\infty[$, il minimo della distanza di ordine 2 viene raggiunto in corrispondenza del minimo del radicando, ovvero della funzione:

$$x \in \mathbb{R}, \quad f(x) = \frac{1}{N} \sum_{i=1}^N (x - y_i)^2.$$

Pertanto,

$$\xi_2(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_2(x, \underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} f(x)$$

Procediamo con la ricerca dei punti stazionari di $f(x)$:

$$x \in \mathbb{R}, \quad f'(x) = \frac{2}{N} \sum_{i=1}^N (x - y_i) \implies f'(x) = 0 \iff \sum_{i=1}^N (x - y_i) = 0 \iff x = \frac{1}{N} \sum_{i=1}^N y_i \equiv \bar{y}.$$

Siccome:

$$f''(x) = \frac{2}{N} \sum_{i=1}^N 1 = 2 \implies f''(\bar{y}) = 2 > 0$$

questo implica che \bar{y} è un punto di minimo relativo per $f(x)$.

Troviamo adesso il minimo assoluto di $f(x)$:

$$\bar{y} = \operatorname{argmin}_{x \in \mathbb{R}} f(x)$$

in quanto:

$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow -\infty} f(x) = +\infty$$

e la derivata prima di $f(x)$ è definita in \mathbb{R} , e $f(x)$ ammette un unico punto di minimo relativo.

In definitiva,

$$\xi_2(\underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} d_2(x, \underline{y}) = \operatorname{argmin}_{x \in \mathbb{R}} f(x) = \bar{y}$$

□

Definizione 6.7.1 (Centro di ordine infinito)

Il centro di ordine infinito di una rilevazione dati \underline{y} si indica con il simbolo $\xi_\infty(\underline{y})$ ed è definito dalla posizione:

$$\xi_\infty(\underline{y}) := \lim_{r \rightarrow +\infty} \xi_r(\underline{y})$$

□

Definizione 6.7.2 (Centro di ordine m)

$$\xi_m = \operatorname{argmin}_{n \in \mathbb{N}} \left(\frac{1}{N} \sum_{i=1}^N |x - y_i|^m \right)^{\frac{1}{m}}$$

Definizione 6.7.3 (Valore centrale)

Il valore centrale di una rilevazione dati \underline{y} è la semisomma tra il dato minimo e il dato massimo:

$$\frac{y_{(1)} + y_{(N)}}{2} = \frac{Q_0 + Q_4}{2}$$

□

Teorema 18 (Il centro di ordine infinito è uguale al valore centrale)

Il centro di ordine infinito della rilevazione dati \underline{y} coincide con il valore centrale:

$$\xi_\infty(\underline{y}) = \frac{y_{(1)} + y_{(N)}}{2} = \lim_{r \rightarrow +\infty} \xi_r = \frac{y_{(1)} + y_{(N)}}{2}$$

6.8 Indici di dispersione

Un *indice di dispersione* (o indicatore di dispersione o indice di variabilità o indice di variazione) serve per descrivere sinteticamente la misura con la quale una rilevazione dati di un *carattere quantitativo* è distante da una sua tendenza centrale. La dispersione esprime la bontà o la inadeguatezza di un indice di tendenza centrale quale descrittore di una distribuzione di frequenza.

Per i caratteri qualitativi si usano gli *indici di diversità* dei quali quello maggiormente usato è l'*indice di ricchezza* che opera un semplice conteggio del numero delle modalità presenti nella rilevazione dati.

Sia $\underline{y} = (y_1, y_2, \dots, y_N)$ una rilevazione dati su un carattere quantitativo. Sono indici di dispersione i seguenti:

- La differenza tra il dato più grande da quello più piccolo (*campo o intervallo di variazione*):

$$\Gamma = y_{(N)} - y_{(1)} = Q_4 - Q_0$$

- La differenza tra il terzo e il primo quartile (*differenza interquartilica*):

$$\gamma = Q_3 - Q_1$$

- La media del valore assoluto delle differenze dei dati dalla loro mediana Q_2 (*scarto mediano assoluto*):

$$S_{Q_2} = \frac{1}{N} \sum_{i=1}^N |y_i - Q_2|$$

- La media del valore assoluto delle differenze dei dati dalla loro media aritmetica \bar{y} (*scarto medio assoluto*):

$$S_{\bar{y}} = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}|$$

- La media del quadrato delle differenze dei dati dalla loro media aritmetica M_2 (*varianza*):

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

La varianza è associata alla media aritmetica \bar{y} .

- La radice quadrata della varianza (*scarto tipo o deviazione standard*):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}$$

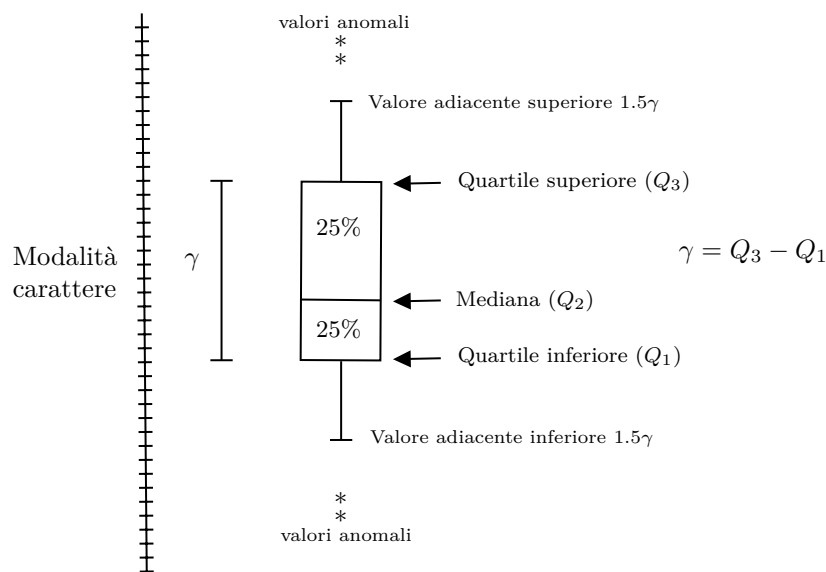
La deviazione standard è associata alla media aritmetica \bar{y} .

6.9 Tabella indici di sintesi e indici di dispersione

SINTESI	$\xi_1 = Me$	$\xi_2 = \bar{y}$	$\xi_\infty = \frac{y_{(1)} + y_{(N)}}{2}$
DISPERSIONE	$\frac{1}{N} \sum_{i=1}^N y_i - Me $	$\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}$	$\Gamma = Q_4 - Q_0$ $= y_{(N)} - y_{(1)}$

6.10 Diagramma scatola con baffi

Un metodo grafico per rappresentare una distribuzione di frequenze che mette in risalto anche la dispersione intorno alla mediana è il *grafico della scatola con baffi* (*Box Plot*):



La linea interna alla scatola rappresenta la *mediana* della distribuzione (cioè Q_2). Le linee estreme della scatola rappresentano il primo ed il terzo quartile (rispettivamente Q_1 e Q_3).

La *distanza interquartilica* γ , è una misura della **dispersione** della distribuzione. Il 50% dei dati si trovano

comprese tra questi due valori. Se l'intervallo interquartilico è piccolo, tale metà delle osservazioni si trova fortemente concentrata intorno alla mediana; all'aumentare della distanza interquartilica aumenta la dispersione del 50% dei dati centrali intorno alla mediana.

Le distanze tra ciascun quartile e la mediana forniscono informazioni relativamente alla **forma** della distribuzione. Se una distanza è diversa dall'altra allora la distribuzione è asimmetrica.

Le linee che si allungano dai bordi della scatola (*baffi*) individuano gli intervalli in cui sono posizionati i valori rispettivamente minori di Q_1 e maggiori di Q_3 ; i punti estremi dei “baffi” evidenziano i *valori adiacenti*. Il *valore adiacente inferiore* (VAI) è il valore più piccolo tra i dati che risulta maggiore o uguale a $Q_1 + 1,5\gamma$. Il *valore adiacente superiore* (VAS), invece, è il valore più grande tra i dati che risulta minore o uguale a $Q_3 + 1,5\gamma$.

I valori esterni ai valori adiacenti (chiamati in genere *valori fuori limite* oppure *valori anomali*), vengono segnalati individualmente nel box-plot per meglio evidenziarne la presenza e la posizione. Questi valori infatti costituiscono una “anomalia” rispetto alla maggior parte dei valori osservati e pertanto è necessario identificarli per poterne analizzare le caratteristiche e le eventuali cause che li hanno determinati. Essi forniscono informazioni ulteriori sulla dispersione e sulla forma della distribuzione. Quando il valore adiacente superiore coincide con il dato più grande e il valore adiacente inferiore coincide con il dato più piccolo, allora non comparirà alcun valore anomalo.

Statistica

7 Statistica

7.1 Distribuzione delle statistiche campionarie

Definizione 7.1.1 (Campione casuale semplice (c.c.s.) di taglia n)

Un campione casuale semplice (abbreviato c.c.s.) di taglia n è una n -upla di numeri aleatori indipendenti e identicamente distribuiti:

$$\underline{X} := (X_1, X_2, \dots, X_n) \text{ indipendenti e identicamente distribuiti}$$

Tutte le componenti del campione è *un'osservazione*. Il campione genera la rilevazione dati. □

Se \underline{X} è un campione casuale semplice di taglia n possiamo dire che vale questa uguaglianza:

$$F_{\underline{X}}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n) = F_{X_1}(x_1) \cdot F_{X_1}(x_2) \cdot \dots \cdot F_{X_1}(x_n) = \prod_{i=1}^n F_X(x_i)$$

poiché tutte le X_i (con $i = 1, \dots, n$) sono indipendenti tra loro e identicamente distribuiti (hanno tutti la stessa legge di distribuzione).

Definizione 7.1.2 (Genitrice del campione casuale semplice)

Si dice genitrice del campione casuale semplice \underline{X} , un *qualsiasi* numero aleatorio che ha la distribuzione uguale a quella comune a X_1, X_2, \dots, X_n . □

Definizione 7.1.3 (Definizione di statistica)

Si dice statistica ogni *funzione misurabile* di un campione casuale semplice. □

Se \underline{X} è un campione casuale semplice, allora:

$$T = g(\underline{X}) = \sum_{i=1}^N X_i \text{ è una statistica}$$

$$\prod_i = g_i(\underline{X}) = X_i \text{ è una statistica}$$

$$X \sim G(p) \sum_{i=1}^n X_i - p \text{ è una statistica}$$

$$X \sim \mathcal{N}(3, \sigma^2) \frac{X_1 - 3}{\sigma} \text{ NON è una statistica perché } \sigma^2 \text{ potrebbe non essere nota}$$

$$X \sim \mathcal{N}(3, 9) \frac{X_1 - 3}{3} = \frac{\prod_1(\underline{X}) - 3}{3} \text{ è una statistica}$$

Definizione 7.1.4 (Media campionaria)

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice estratto da X (genitrice del c.c.s.). La sua media campionaria è definita come:

$$\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \text{ è una statistica e si chiama media campionaria}$$

□

Teorema 19 (La media della media campionaria è uguale alla media della genitrice)

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice estratto da X (genitrice del c.c.s.). Se la genitrice X ammette la media $\mu := \mathbb{E}(X)$ (notiamo che $\mu = \mu_1 = \mu_2 = \dots = \mu_n$) allora possiamo affermare che:

$$\mathbb{E}(\overline{X}) = \mu$$

Dimostrazione.

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

□

Teorema 20 (La varianza della media campionaria è uguale alla varianza della genitrice fratto la taglia)

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice di taglia n estratto dalla genitrice X . Se la genitrice ammette $\mu' := \mathbb{E}(X^2)$ finita (e quindi ammette anche la varianza $\mathbb{D}^2(X)$) allora possiamo affermare che:

$$\mathbb{D}^2(\bar{X}) = \frac{\mathbb{D}^2(X)}{n}$$

Dimostrazione.

$$\mathbb{D}^2(\bar{X}) = \mathbb{D}^2\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathbb{D}^2\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}^2(X_i) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}^2(X) = \frac{n \cdot \mathbb{D}^2(X)}{n^2} = \frac{\mathbb{D}^2(X)}{n}$$

Questo risultato ci porta alla conclusione che più la taglia è alta più si ha una varianza bassa. □

7.2 Il teorema del limite centrale

In questa sezione affrontiamo uno dei risultati più notevoli della teoria della probabilità, il *teorema del limite centrale*⁵. In termini semplicistici, esso afferma che la somma di un numero elevato di variabili aleatorie indipendenti, tende ad avere distribuzione approssimativamente normale. L'importanza è duplice: da un lato siamo in grado di ottenere stime approssimative delle probabilità che riguardano la somma di variabili aleatorie indipendenti, dall'altro abbiamo giustificato il fatto notevole che la distribuzione empirica delle frequenze di un gran numero di popolazioni naturali esibisca forme a campana (in realtà, gaussiane).

L'enunciato, presentato nella sua forma più semplice, è il seguente:

Teorema 21 (Teorema del limite centrale)

Siano X_1, X_2, \dots, X_n delle variabili aleatorie i.i.d. (indipendenti e identicamente distribuite), tutte con media μ e varianza σ^2 . Allora se n è grande, la somma:

$$X_1 + X_2 + \dots + X_n$$

è approssimativamente normale con media $n \cdot \mu$ e varianza $n \cdot \sigma^2$

Si può anche normalizzare la somma precedente in modo da ottenere una distribuzione approssimativamente normale *standard*. Si ha infatti che il numero aleatorio così ottenuto converge in distribuzione alla variabile aleatoria normale standardizzata:

$$Y_n := \frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sqrt{\sigma^2 \cdot n}} \sim \mathcal{N}(0, 1)$$

dove con il simbolo \sim si intende "è approssimativamente distribuito come". Ciò significa che per n abbastanza grande e x qualsiasi vale l'approssimazione:

$$F_{Y_n}(x) = \mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < x\right) \approx \phi(x)$$

dove $\phi(x)$ è la funzione di distribuzione di una variabile aleatoria normale standard.

Esempio 7.2.1

Una compagnia di assicurazione ha 25.000 polizze auto attive. Il risarcimento dovuto annualmente per ogni singolo assicurato è una variabile aleatoria con media 320 e deviazione standard 540. Quanto vale approssimativamente la probabilità che in un determinato anno le richieste di indennizzi superino 8.3 milioni?

Sia X la richiesta annuale complessiva di indennizzi. Numeriamo gli assicurati, e sia X_i il risarcimento dovuto all'assicurato i -esimo, per $i = 1, 2, \dots, n$, con $n = 25.000$. E' chiaro che $X = \sum_{i=1}^n X_i$, e segue dal teorema del limite centrale che X ha approssimativamente distribuzione normale con media $320 \times 25.000 = 8 \times 10^6$ e deviazione standard $540\sqrt{25000} \approx 8.54 \times 10^4$. Perciò, se Z denota una variabile aleatoria con distribuzione $\mathcal{N}(0, 1)$:

$$\mathbb{P}(X > 8.3 \times 10^6) = \mathbb{P}\left(\frac{X - 8 \times 10^6}{8.54 \times 10^4} > \frac{8.3 \times 10^6 - 8 \times 10^6}{8.54 \times 10^4}\right) \approx \mathbb{P}\left(Z > \frac{0.3 \times 10^6}{8.54 \times 10^4}\right) \approx \mathbb{P}(Z > 3.51) \approx 0$$

⁵Spesso lo si trova abbreviato negli acronimi TLC o CLT, dove il secondo deriva ovviamente dall'espressione inglese corrispondente, *central limit theorem*.

7.2.1 Distribuzione approssimata della media campionaria tramite il teorema del limite centrale

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice di taglia n ottenuto dalla genitrice X con media μ e varianza σ^2 . Per il teorema del limite centrale, per n grande possiamo approssimare la distribuzione di:

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0, 1)$$

Vediamo come il teorema del limite centrale ci permette di approssimare la distribuzione della media campionaria:

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^n X_i - n\mu}{\frac{\sqrt{n\sigma^2}}{n}} = \frac{\sum_{i=1}^n X_i}{n} - \frac{n\mu}{n} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0, 1)$$

Dunque possiamo concludere che, per $n \gg 1$ (molto grande) e per ogni $x \in \mathbb{R}$:

$$F_{\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}}(x) \approx \phi(x)$$

Esempio 7.2.2

Una popolazione formata da operai maschi, presenta dei pesi corporei (in libbre) di media 167 e deviazione standard 27.

1. Se si seleziona un campione di 36 elementi, quanto vale circa la probabilità che la media campionaria dei loro pesi stia tra 163 e 171?

Sia Z una variabile aleatoria normale standard. Dal teorema del limite centrale segue che la media campionaria è approssimativamente normale con media 167 e deviazione standard $27/\sqrt{36} = 4.5$. Quindi:

$$\begin{aligned}\mathbb{P}(163 < \bar{X} < 171) &= \mathbb{P}\left(\frac{163 - 167}{4.5} < \frac{\bar{X} - 167}{4.5} < \frac{171 - 167}{4.5}\right) \\ &= \mathbb{P}(-0.8889 < Z < 0.8889) \\ &= 2 \cdot \mathbb{P}(Z < 0.8889) - 1 \approx 0.63\end{aligned}$$

2. E se si selezionano 144 operai?

Con una ampiezza del campione di 144, \bar{X} sarà approssimativamente normale di media 167 e deviazione standard $27/\sqrt{144} = 2.25$. Quindi:

$$\begin{aligned}\mathbb{P}(163 < \bar{X} < 171) &= \mathbb{P}\left(\frac{163 - 167}{2.25} < \frac{\bar{X} - 167}{2.25} < \frac{171 - 167}{2.25}\right) \\ &\approx \mathbb{P}(-1.7778 < Z < 1.7778) \\ &= 2 \cdot \mathbb{P}(Z < 1.7778) - 1 \approx 0.92\end{aligned}$$

7.2.2 Quando un campione è abbastanza numeroso?

Il teorema del limite centrale lascia aperta la questione di quanto grande debba essere la numerosità del campione n , affinché l'approssimazione normale sia valida. In effetti la risposta dipende dalla distribuzione da cui vengono campionati i dati. Ad esempio, se la distribuzione della popolazione è normale, allora \bar{X} sarà a sua volta normale indipendentemente dall'ampiezza del campione (questo perché la distribuzione normale è riproducibile cioè la somma di variabili aleatorie normali e indipendenti ha essa stessa distribuzione normale). Una buona regola empirica è che si può essere confidenti nella validità dell'approssimazione se n è almeno 30. Questo vuole dire che, per quanto "poco gaussiana" sia la distribuzione considerata, la media campionaria di un gruppo di dati di numerosità 30 risulta comunque approssimativamente normale. Si tenga presente comunque che in molti casi è possibile che questo accada anche per n molto più piccolo, e in effetti spesso $n = 5$ è sufficiente ad ottenere approssimazioni non troppo sbagliate.

7.3 Funzione generatrice dei momenti della media campionaria

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice ottenuto dalla genitrice X dotata di funzione generatrice dei momenti. La media campionaria è:

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

Determiniamo la sua funzione generatrice dei momenti:

$$\phi_{\bar{X}}(t) = \phi_{\frac{1}{n} \sum_{i=1}^n X_i}(t) = \phi_{\sum_{i=1}^n X_i}\left(\frac{t}{n}\right) = \phi_X^n\left(\frac{t}{n}\right)$$

7.3.1 Caso in cui la genitrice è gaussiana

Se la genitrice del campione casuale semplice ha distribuzione normale $X \sim \mathcal{N}(\mu, \sigma^2)$ allora otteniamo il seguente risultato:

$$\begin{aligned} \phi_{\bar{X}}(t) &= \phi_X^n\left(\frac{t}{n}\right) \\ \text{siccome } X &\sim \mathcal{N}(\mu, \sigma^2) \quad = \left[e^{\frac{t}{n}\mu} \cdot e^{\frac{\sigma^2}{n} \cdot \frac{t^2}{2n}} \right]^n = \left(e^{\frac{t}{n}\mu} \right)^n \cdot \left(e^{\frac{\sigma^2}{n} \cdot \frac{t^2}{2n}} \right)^n = e^{t\mu} \cdot e^{\frac{\sigma^2}{n} \cdot \frac{t^2}{2}} \end{aligned}$$

Dunque:

$$\phi_{\bar{X}}(t) = e^{t\mu} \cdot e^{\frac{\sigma^2}{n} \cdot \frac{t^2}{2}}$$

E possiamo affermare che:

La genitrice X è gaussiana $\implies \bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ la media campionaria è gaussiana

7.4 Varianza campionaria

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice proveniente dalla genitrice X con media μ e varianza σ^2 . Sia \bar{X} la sua media campionaria. Introduciamo una nuova *statistica*.

Definizione 7.4.1 (Varianza campionaria)

La statistica S^2 , definita da:

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

si dice *varianza campionaria*. La sua radice quadrata, $S = \sqrt{S^2}$ prende invece il nome di *deviazione standard campionaria*.

Proposizione 7.4.1

Sia dato un insieme di dati x_1, x_2, \dots, x_n e sia \bar{x} la sua media campionaria, allora:

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

Dimostrazione.

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \\ (\text{per definizione di } \bar{x} \text{ media campionaria}) &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{aligned}$$

□

Dalla Proposizione 7.4.1 possiamo scrivere S^2 come:

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

(sostituiamo la sommatoria con la media campionaria del secondo ordine) $= \frac{n}{n-1} (\bar{X}^{(2)} - \bar{X}^2)$

dove la media campionaria del secondo ordine è:

$$\bar{X}^{(2)} := \frac{1}{n} \sum_{i=1}^n X_i^2$$

Quindi, S^2 si può scrivere come:

$$S^2 = \frac{n}{n-1} (\bar{X}^{(2)} - \bar{X}^2)$$

7.5 Stimatori

Definizione 7.5.1 (Stimatore)

Si dice *stimatore* per una funzione del parametro della genitrice, $\omega(\theta)$, una statistica che è finalizzata per la stima di $\omega(\theta)$ (è usata per stimare un qualcosa di ignoto).

7.5.1 Introduzione ai stimatori con un esempio

Esempio 7.5.1

La media campionaria \bar{X} è stimatore per la media μ (in generale θ) di una genitrice X con media ignota. Lo stimatore dà un'approssimazione della media della genitrice X . Ricordiamo che:

$$\mu \in \mathbb{R}, \mathbb{E}(\bar{X}) = \mu$$

La media di \bar{X} è uguale a μ , per ogni valore di μ : quando succede questo si dice che \bar{X} è uno stimatore *corretto* per μ . La correttezza dello stimatore è importante perché è possibile trovarsi in una situazione in cui la media di \bar{X} è uguale a μ solamente in una certa parte dell'insieme (detto *regione parametrica*) dove risiede il parametro θ . In questo esempio, la regione parametrica è tutto \mathbb{R} . La correttezza dice che in media lo stimatore coincide con la quantità ignota per ogni valore della regione parametrica.

7.5.2 Correttezza dello stimatore S^2 per la varianza σ^2

Vogliamo vedere se lo stimatore S^2 è uno stimatore corretto per la varianza σ^2 . La regione parametrica di σ^2 è $]0, +\infty[$. Partiamo con:

$$\sigma^2 > 0 \quad \mathbb{E}(X_i^2) = \mathbb{E}(X^2) = \sigma^2 + \mu^2$$

$$\mathbb{E}(\bar{X}^2) = \mathbb{D}^2(\bar{X}) + \mathbb{E}^2(\bar{X}) = \frac{\sigma^2}{n} + \mu^2$$

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] &= \mathbb{E} [n(\bar{X}^{(2)} - \bar{X}^2)] = \mathbb{E} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right] = \mathbb{E} \left(\sum_{i=1}^n X_i^2 \right) - n\mathbb{E}(\bar{X}^2) \\ &= \sum_{i=1}^n \mathbb{E}(X_i^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \\ &= n(\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \\ &= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = \sigma^2(n-1) \end{aligned}$$

Allora:

$$\sigma^2 > 0 \quad \mathbb{E}(S^2) = \frac{1}{n-1} \mathbb{E} \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] = \frac{\sigma^2(n-1)}{(n-1)} = \sigma^2$$

Scopriamo quindi che S^2 è uno stimatore corretto per la varianza σ^2 . Se S^2 fosse stato diviso per n e non per $n - 1$, la proprietà di correttezza non sarebbe stata soddisfatta. Per questo motivo si divide per $n - 1$. Infatti, consideriamo la seguente statistica:

$$S_\alpha^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Se avessimo fatto gli stessi calcoli di prima, avremmo ottenuto:

$$\sigma^2 > 0, \quad \mathbb{E}(S_\alpha^2) = \sigma^2 \cdot \frac{n-1}{n}$$

La proprietà di correttezza non è soddisfatta. Possiamo però dire che è *asintoticamente corretto* perché se proviamo a fare il limite per $n \rightarrow +\infty$ otteniamo:

$$\lim_{n \rightarrow +\infty} \mathbb{E}(S_\alpha^2) = \lim_{n \rightarrow +\infty} \sigma^2 \frac{n-1}{n} \iff \mathbb{E}(S_\alpha^2) = \sigma^2$$

7.5.3 Definizione di correttezza

Definizione 7.5.2 (Correttezza di uno stimatore)

Uno stimatore $T = g(\underline{X})$ si dice *corretto* per la stima di $\omega(\theta)$ se e solo se:

$$\theta \in \Theta, \quad \mathbb{E}(T) = \mathbb{E}_T(\theta) = \omega(\theta)$$

dove Θ indica la regione parametrica di θ .

7.5.4 Ordinamento tra gli stimatori corretti

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice dalla genitrice X di media μ incognita e di varianza $\sigma^2 < \infty$. Consideriamo il seguente stimatore:

$$T = \frac{X_1 + X_n}{2}$$

Scopriamo che tale stimatore è corretto per la media μ , infatti:

$$\mu \in \mathbb{R} \quad \mathbb{E}_T(\mu) = \frac{1}{2} \mathbb{E}(X_1) + \mathbb{E}(X_n) = \frac{1}{2}(\mu + \mu) = \mu$$

Ma questo stimatore non è un granché perché considera solo due campioni (X_1 e X_2). Lo stimatore \bar{X} è decisamente migliore di T . Per confrontare due stimatori corretti e verificare chi è più affidabile, viene usata la varianza:

$$\begin{aligned} \mathbb{D}^2(\bar{X}) &= \frac{\sigma^2}{n} \\ \mathbb{D}^2(T) &= \frac{1}{4} [\mathbb{D}^2(X_1) + \mathbb{D}^2(X_2)] = \frac{1}{4} 2\sigma^2 = \frac{\sigma^2}{2} \end{aligned}$$

Da come si può facilmente intuire, è decisamente meglio lo stimatore \bar{X} . Formalmente, si dice che \bar{X} è strettamente preferibile di T . Per confrontare gli stimatori usiamo dunque la varianza.

7.5.5 Rischio quadratico medio

Il rischio quadratico medio è una funzione che dipende dal parametro θ . Esso è definito come:

$$\theta \in \Theta, \quad R_T(\theta) = \mathbb{E}\{[T - \omega(\theta)]^2\}$$

L'espressione:

$$[T - \omega(\theta)]^2$$

viene chiamato *errore quadratico medio*. Calcolando la media dell'errore quadratico medio, otteniamo un numero (cioè il rischio quadratico medio) che dipende dal parametro θ .

Il rischio quadratico medio viene utilizzato al posto della varianza perché permette di confrontare stimatori anche non corretti, anche detti stimatori distorti (riesce ad essere più generale).

Possiamo esprimere il rischio quadratico medio anche in questa forma (poniamo $\mathbb{E}(T) = \mu_T$):

$$\begin{aligned}\theta \in \Theta, \quad R_T(\theta) &= \mathbb{E}\{[T - \omega(\theta)]^2\} \\ (\text{aggiungo e sottraggo } \mu_T) &= \mathbb{E}[(T - \mu_T)^2] + [\mu_T - \omega(\theta)]^2 + 2(\mu_T - \omega(\theta))\mathbb{E}[(T - \mu_T)] \\ (\text{la media dello scarto è zero } \mu_T - \mu_T = 0) &= \mathbb{E}[(T - \mu_T)^2] + [\mu_T - \omega(\theta)]^2 \\ (\text{il primo addendo è la varianza}) &= \mathbb{D}_T^2(\theta) + [\mu_T - \omega(\theta)]^2\end{aligned}$$

Otteniamo quindi che:

$$\theta \in \Theta, \quad R_T(\theta) = \mathbb{E}\{[T - \omega(\theta)]^2\} = \mathbb{D}_T^2(\theta) + [\mu_T - \omega(\theta)]^2$$

La radice quadrata dell'addendo che viene sommato alla varianza si chiama *distorsione*. Esso è zero se lo stimatore è corretto.

$$d(\theta) = \mu_T - \omega(\theta)$$

Quando lo stimatore è corretto, la distorsione è uguale a zero e il rischio quadratico medio è uguale alla varianza.

7.5.6 Ordinamento tra due stimatori generici (anche non corretti)

Siano T ed S due stimatori per $\omega(\theta)$ dotati di varianza finita.

Definizione 7.5.3 (Relazione di preferibilità)

Si dice che S è *preferibile* a T (la relazione di preferibilità si denota con la lettera greca $\underline{\alpha}$ sottolineata), se vale questa proprietà:

$$S \underline{\alpha} T \iff \theta \in \Theta, \quad R_S(\theta) \leq R_T(\theta)$$

□

Definizione 7.5.4 (Relazione di stretta preferibilità)

Si dice che S è *strettamente preferibile* a T (la relazione di preferibilità si denota con la lettera greca α), se vale questa proprietà:

$$S \alpha T \iff S \underline{\alpha} T \text{ ed esiste un } \theta_0 \in \Theta \text{ tale che } R_S(\theta_0) < R_T(\theta_0)$$

Uno stimatore è strettamente preferibile ad un altro stimatore se è preferibile all'altro ed esiste almeno un $\theta_0 \in \Theta$ nella regione parametrica tale che $R_S(\theta_0) < R_T(\theta_0)$

Definizione 7.5.5 (Stimatore ammissibile)

Uno stimatore S è *ammissibile* per la stima di $\omega(\theta)$ quando non esiste un altro stimatore a lui strettamente preferibile.

La relazione di preferibilità (stretta o non) NON è una relazione totale: esistono stimatori non confrontabili.

Stima parametrica

8 Stima parametrica

8.1 Introduzione

Dobbiamo rispondere a questa domanda: *chi li fornisce gli stimatori?* Consideriamo un campione $\underline{X} = (X_1, X_2, \dots, X_n)$ casuale semplice estratto dalla genitrice X . Tale genitrice X avrà una sua F_X , conosciamo il tipo di legge ma non la conosciamo precisamente perché non si conosce il parametro θ . Allora denotiamo la funzione di distribuzione della genitrice X in questa maniera:

$$F(x; \theta)$$

E quindi a questo punto sorge il problema di come ottenere tali stimatori che svelano il parametro θ (in maniera approssimativa).

E' possibile stimare un parametro θ quando non lo si conosce soltanto a partire dai dati. Bisogna stimare θ tramite un'opportuna funzione del campione casuale e cioè tramite una statistica. La domanda è *chi fornisce tale statistica?* Un metodo per ottenere queste statistiche che verranno utilizzate come stimatori è il metodo della **massima verosimiglianza** (in inglese *likelihood* - parola intesa come verosimiglianza oppure plausibilità).

8.2 Stimatori di massima verosimiglianza

La massima verosimiglianza (in inglese *likelihood*) è un metodo per ottenere statistiche che verranno utilizzate come stimatori. In questo contesto, la parola *verosimiglianza* è una funzione e quello che andremo a fare è massimizzare questa funzione (trovare il massimo - *massima verosimiglianza*). E' il significato di questa funzione che ci dice che questo è un buon metodo per ottenere gli stimatori.

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un campione casuale semplice di taglia n estratto dalla genitrice X . Le possibili realizzazioni di \underline{X} sono costituite dalle n -uple $\underline{x} = (x_1, x_2, \dots, x_n)$.

8.2.1 Stimatori di massima verosimiglianza - caso continuo (variabile esponenziale)

Mettiamo nel caso assolutamente continuo. Sappiamo che per descrivere il campione casuale semplice \underline{X} abbiamo bisogno della funzione di densità di probabilità congiunta che prende come argomento una n -upla \underline{x} e il parametro θ che si fattorizza nel prodotto delle funzioni di densità di probabilità delle osservazioni del campione:

$$\begin{aligned} f_{\underline{X}}(\underline{x}; \theta) &\stackrel{\text{(indipendenti)}}{=} f_{X_1}(x_1, \theta) \cdot f_{X_2}(x_2, \theta) \cdot \dots \cdot f_{X_n}(x_n, \theta) \\ \text{(identicamente distribuite)} &= f_X(x_1, \theta) \cdot f_X(x_2, \theta) \cdot \dots \cdot f_X(x_n, \theta) \end{aligned}$$

Essendo $\underline{x} = (x_1, x_2, \dots, x_n)$ dati e quindi valori fissati, la funzione di densità di probabilità congiunta è adesso *in funzione del parametro θ* . In questa situazione, la funzione di probabilità di densità congiunta assume il ruolo di **funzione di verosimiglianza** (quando invece il parametro è noto e la n -upla no, è semplicemente la funzione di densità di probabilità congiunta normale):

$$\theta \in \Theta, \quad L(\theta; \underline{x}) = \prod_{i=1}^n f_X(x_i, \theta)$$

Dobbiamo massimizzare tale funzione. In particolare dobbiamo trovare il *punto di massimo assoluto* della funzione di verosimiglianza:

$$\hat{\theta}_{MV} := \underset{\theta \in \Theta}{\operatorname{argmax}} L(\theta; \underline{x})$$

Il punto di massimo assoluto viene chiamato **stima di massima verosimiglianza** e si denota con $\hat{\theta}_{MV}$. In generale, lo stimatore, denotato con $\hat{\theta}$, del parametro θ è:

$$\hat{\theta} := \underset{\theta \in \Theta}{\operatorname{argmax}} L(\theta; \underline{X})$$

Esempio 8.2.1 (Determinazione stimatore parametro α di una genitrice esponenziale)

Supponiamo che $X \sim \text{Esp}(\alpha)$. La funzione di densità di probabilità di una variabile esponenziale (che ha media $\mathbb{E}(X) = \frac{1}{\alpha}$) è:

$$\alpha > 0 \quad f_X(x; \alpha) = \alpha \cdot e^{-\alpha x} \quad x > 0$$

La funzione di verosimiglianza con $\underline{x} = (x_1, \dots, x_n)$ fissato è:

$$L(\alpha; \underline{x}) = \prod_{i=1}^n f_X(x_i, \alpha) = \prod_{i=1}^n \alpha \cdot e^{-\alpha \cdot x_i} = \alpha^n \cdot e^{-\alpha \cdot \sum_{i=1}^n x_i}$$

Poniamo $s = \sum_{i=1}^n x_i$, quindi la funzione diventa:

$$L(\alpha; \underline{x}) = \alpha^n \cdot e^{-\alpha \cdot s}$$

Adesso, avendo osservando l'ennupla \underline{x} , qual'è il valore di α più verosimile? E' quello che fornisce il massimo valore della funzione di massima verosimiglianza.

Facciamo la derivata della funzione L rispetto ad α :

$$\frac{dL(\alpha; \underline{x})}{d\alpha} = n \cdot \alpha^{n-1} \cdot e^{-\alpha \cdot s} - s \cdot \alpha^n \cdot e^{-\alpha \cdot s} = e^{-\alpha \cdot s} \cdot \alpha^{n-1} \cdot (n - s \cdot \alpha)$$

Capiamo quando la derivata prima di L è uguale a 0:

$$\frac{dL(\alpha; \underline{x})}{d\alpha} = 0 \iff n - s\alpha = 0 \iff \alpha = \frac{n}{s}$$

Questo è l'unico punto stazionario di tale funzione oppure si potrebbe trovare sulla frontiera il massimo. Per adesso pensiamo ai punti stazionari. Facciamo la derivata seconda:

$$\begin{aligned} \frac{d^2 L(\alpha; \underline{x})}{d^2 \alpha} &= -s \cdot e^{-\alpha \cdot s} \cdot \alpha^{n-1} \cdot (n - \alpha s) + e^{-\alpha \cdot s} \cdot (n-1) \cdot \alpha^{n-2} \cdot (n - \alpha s) - s \cdot e^{-\alpha \cdot s} \cdot \alpha^{n-1} \\ &= e^{-\alpha \cdot s} \cdot \alpha^{n-2} \cdot [-s \cdot \alpha \cdot (n - \alpha \cdot s) + (n-1)(n - \alpha \cdot s) - s \cdot \alpha] \end{aligned}$$

$$(\text{poniamo adesso } \alpha = n/s) \quad = -n \cdot \left(\frac{n}{s}\right)^{n-2} \cdot e^{-n} < 0$$

Quindi il punto $\alpha = \frac{n}{s}$ non solo è un punto stazionario ma è anche un punto di massimo relativo. Non ci resta che controllare sulla frontiera della funzione (ai limiti), quindi facciamo il seguente limite con $\alpha \rightarrow 0^+$ (α che tende a 0 dalla destra):

$$\lim_{\alpha \rightarrow 0^+} L(\alpha; \underline{x}) = \lim_{\alpha \rightarrow 0^+} \alpha^n \cdot e^{-\alpha \cdot s} = 0$$

Quindi la frontiera sinistra non si può paragonare con il punto di massimo relativo $\alpha = \frac{n}{s}$ perché in questo punto è positivo. Controlliamo la frontiera destra:

$$\lim_{\alpha \rightarrow +\infty} L(\alpha; \underline{x}) = 0$$

Scopriamo che anche qui il risultato è zero perché, nonostante il limite sia una forma indeterminata, $e^{-\alpha \cdot s}$ (che è 0) domina su α^n . Possiamo affermare che $\alpha = \frac{n}{s}$ non solo è un punto stazionario, non solo è un punto di massimo relativo ma è anche un punto di massimo assoluto. Possiamo affermare che la stima $\hat{\alpha}_{MV}$ è uguale a:

$$\hat{\alpha}_{MV} = \frac{n}{s} = \frac{n}{\sum_{i=1}^n x_i}$$

che, ricordiamo, non è una variabile aleatoria ma è un valore assunto da una statistica. Quindi lo stimatore di α è:

$$\hat{\alpha} := \frac{n}{\sum_{i=1}^n x_i} = (\bar{X})^{-1}$$

Lo stimatore di α è il *reciproco della media campionaria*.

□

Per rendere i calcoli più semplici è possibile affidarsi invece a questa funzione:

$$l(\theta; \underline{x}) = \ln L(\theta; \underline{x})$$

Massimizzare L o massimizzare l conduce allo stesso identico risultato perché \ln è una funzione crescente (non stiamo considerando le altezze ma stiamo considerando i valori sull'asse orizzontale, usare \ln non cambia il

risultato della funzione argmax). Quindi, conviene usare questa funzione per calcolare la stima di massima verosimiglianza:

$$\alpha > 0, \quad \hat{\theta}_{MS} = \operatorname{argmax}_{\theta \in \Theta} l(\theta; \underline{x})$$

Perché tale funzione è più facile da calcolare? Perché il logaritmo del prodotto è uguale alla somma dei logaritmi e questo ci permette di fare derivate di una somma.

Esempio 8.2.2 (Determinazione stimatore parametro α di una genitrice esponenziale usando $l(\alpha; \underline{x})$)

Calcoliamo $l(\alpha; \underline{x})$:

$$l(\alpha; \underline{x}) = \ln(\alpha^n \cdot e^{-\alpha \cdot s}) = \ln \alpha^n + \ln e^{-\alpha \cdot s} = n \cdot \ln \alpha - \alpha \cdot s$$

Calcoliamo la derivata prima:

$$\frac{dl(\alpha; \underline{x})}{d\alpha} = \frac{n}{\alpha} - s$$

Capiamo quando la derivata prima di l è uguale a 0:

$$\frac{dl(\alpha; \underline{x})}{d\alpha} = 0 \iff \frac{n}{\alpha} = s \iff \frac{\alpha}{n} = \frac{1}{s} \iff \alpha = \frac{n}{s} \quad (10)$$

La frontiera sinistra tende a $-\infty$ mentre la frontiera destra tende a 0. Ci serve solo stabilire che tale punto stazionario è un punto di massimo relativo:

$$\alpha > 0 \quad \frac{d^2 l(\alpha; \underline{x})}{d^2 \alpha} = -n \cdot \frac{1}{\alpha^2} < 0$$

quindi il punto $\alpha = \frac{n}{s}$ è un punto di massimo assoluto. Concludiamo che la stima di massima somiglianza è:

$$\hat{\alpha}_{MV} = \frac{n}{\sum_{i=1}^n x_i}$$

Quindi lo stimatore di α è:

$$\hat{\alpha}_{MV} = (\overline{X})^{-1}$$

□

Quindi, la funzione $l(\alpha; \underline{x})$ rende i calcoli più semplici. Questo perché, se:

$$L(\alpha, \underline{x}) = \prod_{i=1}^n \alpha \cdot e^{-\alpha \cdot x_i}$$

Allora, grazie alla proprietà dei logaritmi (il logaritmo di un prodotto è la somma dei logaritmi) il calcolo si riduce a:

$$l(\alpha; \underline{x}) = \sum_{i=1}^n \ln(\alpha \cdot e^{-\alpha \cdot x_i}) = \sum_{i=1}^n \ln \alpha + \sum_{i=1}^n (-\alpha \cdot x_i) = n \cdot \ln \alpha - \alpha \cdot \sum_{i=1}^n x_i = n \cdot \ln \alpha - \alpha \cdot s$$

Quindi, nel caso di una variabile $X \sim \operatorname{Esp}(\alpha)$, la funzione di verosimiglianza $l(\alpha; \underline{x})$ diventa:

$$l(\alpha; \underline{x}) = n \cdot \ln \alpha - \alpha \cdot s$$

8.2.2 Stimatori di massima verosimiglianza - caso discreto (variabile binomiale)

Supponiamo che $X \sim B(1, p)$ con $p \in (0, 1)$. Quindi $\mathbb{P}(X = 0) = q$ e $\mathbb{P}(X = 1) = p$. Nel caso generale, possiamo dire che:

$$\mathbb{P}(X = x) = p^{-x} \cdot q^{1-x}$$

La funzione di verosimiglianza è:

$$L(p; \underline{x}) = \prod_{i=1}^n \mathbb{P}(X = x_i) = \prod_{i=1}^n p^{x_i} \cdot q^{1-x_i}$$

Quando il parametro p non si conosce e si ha \underline{x} fissato, allora tale funzione non assume il ruolo di funzione di probabilità congiunta ma di funzione di verosimiglianza del parametro p incognito e \underline{x} fissato.

La domanda è *qual'è il valore di p che mi rende più verosimile l'ennupla fissata $\underline{x} = (x_1, \dots, x_n)$* ? Cioè, dobbiamo trovare il valore di p che assegna la massima probabilità possibile a ciò che abbiamo osservato. Passiamo alla funzione $l(p; \underline{x})$:

$$l(p; \underline{x}) = \sum_{i=1}^n \ln(p^{x_i} \cdot q^{1-x_i}) = \sum_{i=1}^n \ln p^{x_i} + \sum_{i=1}^n \ln q^{1-x_i} = \sum_{i=1}^n x_i \cdot \ln p + \sum_{i=1}^n (1-x_i) \ln q = s \cdot \ln p + (n-s) \ln(1-p)$$

Abbiamo ottenuto quindi che:

$$l(p; \underline{x}) = s \cdot \ln p + (n-s) \ln(1-p)$$

Dobbiamo trovare il massimo di tale funzione. Facciamo la derivata prima:

$$\frac{dl(p; \underline{x})}{dp} = \frac{s}{p} - \frac{n-s}{1-p}$$

Quand'è che tale derivata è uguale a 0?

$$\frac{dl(p; \underline{x})}{dp} = 0 \iff \frac{s}{p} = \frac{n-s}{1-p} \iff s \cdot (1-p) = (n-s) \cdot p \iff s - sp = np - sp \iff p = \frac{s}{n}$$

Questo è un massimo relativo. Verifichiamo se è massimo assoluto. Calcoliamo la derivata seconda:

$$\frac{d^2 l(p; \underline{x})}{dp^2} = -\frac{s}{p^2} - \frac{(n-s)}{(1-p)^2}$$

Ponendo $p = \frac{s}{n}$ scopriamo che:

$$\frac{d^2 l(p; \underline{x})}{dp^2} = -\frac{s}{p^2} - \frac{(n-s)}{(1-p)^2} < 0$$

Controlliamo le frontiere:

$$\lim_{p \rightarrow 0^+} \ln(p; \underline{x}) = -\infty$$

$$\lim_{p \rightarrow 1^-} \ln(p; \underline{x}) = -\infty$$

Quindi, abbiamo ottenuto:

$$\hat{p}_{MV} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\hat{p}_{MV} = \bar{X} \quad (\text{lo stimatore è la media campionaria})$$

□

Anche nel caso discreto, lo stimatore di un parametro θ è sempre e comunque dato dal massimo della funzione di verosimiglianza:

$$\hat{\theta}_{MV} = \underset{\theta \in \Theta}{\operatorname{argmax}} L(\theta; \underline{x}) = \underset{\theta \in \Theta}{\operatorname{argmax}} l(\theta; \underline{x})$$

8.3 Stimatori di massima verosimiglianza - variabile gaussiana

Sia $X \sim \mathcal{N}(\mu, \sigma^2)$ la genitrice con legge normale. Possono capitare tre casi:

1. μ incognita
2. σ^2 incognita
3. μ e σ^2 incognite

In questo caso, θ non è più unidimensionale ma bidimensionale (due parametri). Quindi, la funzione di verosimiglianza in questo caso assume questa forma:

$$\begin{aligned} L(\mu; \sigma^2; \underline{x}) &:= \prod_{i=1}^n f_X(x_i; \mu; \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma^2} \right)^{\frac{1}{2}} \cdot e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

Usando $l(\mu; \sigma^2; \underline{x})$ otteniamo:

$$l(\mu; \sigma^2; \underline{x}) = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

8.3.1 Determinazione stimatori - caso μ incognita e σ^2 nota

Consideriamo il caso μ incognita e σ^2 nota. La funzione di verosimiglianza è:

$$l(\mu; \underline{x}) = n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Calcoliamo la derivata prima:

$$\frac{dl(\mu; \underline{x})}{d\mu} = \frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)$$

Equazione punti stazionari (solo la sommatoria perché $\sigma^2 > 0$):

$$\sum_{i=1}^n (x_i - \mu) = 0 \iff \sum_{i=1}^n x_i = \sum_{i=1}^n \mu \iff n\mu = \sum_{i=1}^n x_i \iff \mu = \bar{x}$$

L'unico punto stazionario è la media campionaria. Entrambe le frontiere ($\mu \rightarrow -\infty$ e $\mu \rightarrow +\infty$) tendono a $-\infty$. Calcoliamo la derivata seconda:

$$\frac{d^2 l(\mu; \underline{x})}{d^2 \mu} = -\frac{n}{\sigma^2} < 0$$

Quindi \bar{x} è il massimo punto assoluto che cercavamo. Quindi la stima è:

$$\hat{\mu}_{MV} = \bar{x} \quad \underline{x} = (x_1, \dots, x_n)$$

Lo stimatore è:

$$\widehat{M}_{MV} = \bar{X} \quad \underline{X} = (X_1, \dots, X_n)$$

cioè è la media campionaria.

8.3.2 Determinazione stimatori - caso μ noto e σ^2 incognita

Consideriamo il caso μ noto e σ^2 incognita. Poniamo $\beta = \sigma^2$ per evitare confusione con le derivate. La funzione di verosimiglianza è:

$$l(\beta; \underline{x}) = n \ln \sqrt{2\pi} - \frac{n}{2} \ln \beta - \frac{1}{2\beta} \sum_{i=1}^n (x_i - \mu)^2$$

Calcoliamo la derivata prima rispetto a β :

$$\frac{dl(\beta; \underline{x})}{d\beta} = -\frac{n}{2} \frac{1}{\beta} + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{\beta^2}$$

Equazione punti stazionari:

$$-\frac{n}{2} \frac{1}{\beta} + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{\beta^2} = 0 \iff \frac{n}{2} \frac{1}{\beta} = \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{\beta^2} \iff n = \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{\beta} \iff \beta = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Entrambe le frontiere ($\beta \rightarrow 0^+$ e $\mu \rightarrow +\infty$) tendono a $-\infty$. Calcoliamo la derivata seconda:

$$\frac{d^2 l(\beta, \underline{x})}{d^2 \beta} = \frac{n}{2} \frac{1}{\beta^2} - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \frac{2}{\beta^3} = s_{d,\mu}^2$$

Poniamo $\beta = s_{d,\mu}^2$, otteniamo:

$$\frac{d^2 l(s_{d,\mu}^2; \underline{x})}{d\beta^2} = \frac{(s_{d,\mu}^2)^2}{2} \left[n - 2 \sum_{i=1}^n (x_i - \mu)^2 \cdot \frac{n}{\sum_{i=1}^n (x_i - \mu)^2} \right] = \frac{(s_{d,\mu}^2)^2}{2} \cdot n < 0$$

Essendo l'unico punto stazionario, non solo è massimo relativo ma anche assoluto. Quindi la stima di σ^2 è:

$$\widehat{\sigma^2}_{MV} = s_{d,\mu}^2$$

Lo stimatore è:

$$\widehat{S^2}_{MV} = S_{d,\mu}^2$$

Lo stimatore è distorto (non corretto) ma non importa.

8.3.3 Determinazione stimatori - caso μ incognita e σ^2 incognita

Consideriamo il caso μ incognita e σ^2 incognita. Poniamo $\beta = \sigma^2$ per evitare confusione con le derivate. La funzione di verosimiglianza è:

$$l(\mu; \beta; \underline{x}) = n \ln \sqrt{2\pi} - \frac{n}{2} \ln \beta - \frac{1}{2\beta} \sum_{i=1}^n (x_i - \mu)^2$$

Ci sono due variabili incognite, bisogna quindi fare la derivata rispetto ad entrambe. Il sistema che dobbiamo considerare è quello che annulla entrambe le derivate:

$$\begin{aligned} \frac{\partial l(\mu, \beta, \underline{x})}{\partial \mu} &= \frac{2}{2\beta} \sum_{i=1}^n (x_i - \mu) \\ \frac{\partial l(\mu, \sigma^2, \underline{x})}{\partial \beta} &= -\frac{n}{2} \frac{1}{\beta} + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{\beta^2} \end{aligned}$$

Sistema per trovare i punti stazionari:

$$\left\{ \begin{array}{l} \frac{1}{\beta} \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{1}{2\beta} \left[-n + \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{\beta} \right] = 0 \end{array} \right\} \iff \left\{ \begin{array}{l} \mu = \bar{x} \\ n = \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{\beta} \end{array} \right\} \iff \left\{ \begin{array}{l} \mu = \bar{x} \\ \beta = s_d^2 \end{array} \right.$$

Il punto stazionario è la coppia (\bar{x}, s_d^2) . Senza dimostrazione, diciamo che questo è il punto di massimo assoluto. Quindi, la stima di massima verosimiglianza per μ e σ^2 è:

$$\left\{ \begin{array}{l} \widehat{\mu}_{MV} = \bar{x} \\ \widehat{\sigma^2} = s_d^2 \end{array} \right.$$

Gli stimatori sono:

$$\left\{ \begin{array}{l} \widehat{M}_{MV} = \bar{X} \\ \widehat{S^2}_{MV} = S_d^2 \end{array} \right.$$

8.4 Stimatori di massima verosimiglianza - variabile di poisson

Sia $X \sim \prod(\lambda)$ con $\lambda > 0$. La funzione di verosimiglianza è:

$$L(\lambda; \underline{x}) = \prod_{i=1}^n f_X(x_i; \lambda) = \prod_{i=1}^n e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \cdot \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \dots x_n!}$$

Poniamo:

$$s = \sum_{i=1}^n x_i \quad \text{e} \quad A = x_1! x_2! \dots x_n!$$

Otteniamo quindi:

$$L(\lambda; \underline{x}) := \frac{1}{A} e^{-n\lambda} \lambda^s$$

Utilizzando $l(\lambda; \underline{x})$ otteniamo:

$$l(\lambda; \underline{x}) = -\ln A - n\lambda + s \ln \lambda$$

Dobbiamo trovare:

$$\underset{\lambda > 0}{\operatorname{argmax}} [-\ln A - n\lambda + s \ln \lambda]$$

Calcoliamo la derivata prima:

$$\frac{dl(\lambda, \underline{x})}{d\lambda} = 0 \iff n = \frac{s}{\lambda} \iff \lambda = \frac{s}{n} \iff \lambda = \bar{x}$$

Otteniamo come unico punto stazionario \bar{x} . La frontiera sinistra ($\lambda \rightarrow 0^+$ e $\lambda \rightarrow +\infty$) tendono a $-\infty$. Calcoliamo la derivata seconda ponendo $\underline{x} = \bar{x}$:

$$\frac{d^2 l(\lambda; \bar{x})}{d\lambda^2} = -\frac{s}{\lambda^2} < 0$$

Otteniamo quindi come stima:

$$\hat{\lambda}_{MV} = \bar{x}$$

Lo stimatore è quindi:

$$\hat{\lambda}_{MV} = \bar{X}$$

8.5 Stimatori di massima verosimiglianza - differenza caso discreto e caso continuo

In generale, l'unica differenza tra caso continuo e discreto è che nel caso discreto la f è una funzione di probabilità mentre nel caso continuo la f è una funzione di densità di probabilità.

8.6 Metodo dei momenti

Conosciamo due tipi di momenti, quelli *teorici* e quelli *campionari*:

$$r \in \mathbb{N} \quad \mathbb{E}(X^r) \quad (\text{teorico - calcolo delle probabilità})$$

$$\bar{X}^{(r)} = r \in \mathbb{N} \quad \frac{1}{n} \sum_{i=1}^n X_i^r \quad (\text{campionario - è una statistica})$$

Il metodo dei momenti si basa su una semplice idea e cioè quella di sostituire i momenti teorici con quelli campionari (e quindi noti perché una statistica si può sempre calcolare).

Supponiamo che θ ha dimensione k (quantità parametri). Si sceglie un $p \in \mathbb{N}$: $p \geq k$ e scegliamo p indici:

$$J_1, J_2, \dots, J_p \in \mathbb{N}$$

$$\left\{ \begin{array}{l} \mu'_{J_1} = \mathbb{E}(X^{J_1}) = f_{J_1}(\theta_1, \dots, \theta_k) \\ \mu'_{J_2} = \mathbb{E}(X^{J_2}) = f_{J_2}(\theta_1, \dots, \theta_k) \\ \dots \\ \mu'_{J_p} = \mathbb{E}(X^{J_p}) = f_{J_p}(\theta_1, \dots, \theta_k) \end{array} \right.$$

Il momento teorico di ordine J_i è funzione dei parametri. In statistica non conosciamo i parametri. Quindi l'incognita è proprio la k-upla $\underline{\theta} = (\theta_1, \dots, \theta_k)$. Quando risolviamo il sistema otteniamo proprio:

$$\begin{cases} \theta_1 = g_1(\mu'_{J_1}, \dots, \mu'_{J_p}) \\ \theta_2 = g_2(\mu'_{J_1}, \dots, \mu'_{J_p}) \\ \dots \\ \theta_k = g_k(\mu'_{J_1}, \dots, \mu'_{J_p}) \end{cases}$$

Quindi ogni θ_i sarà una funzione dei momenti teorici. A questo punto sostituiamo i momenti teorici con i momenti campionari e otteniamo gli stimatori che stavamo cercando.

Esempio 8.6.1

$$X \sim B(m, p) \quad \mathbb{E}(X) = m \cdot p \implies p = \frac{\mathbb{E}(X)}{m}$$

Al posto di $\mathbb{E}(X)$ ci mettiamo la media campionaria

□

Quindi, sostituiamo i momenti teorici con i momenti campionari:

$$\begin{aligned} \hat{\theta}_{1,MM} &= g_1(\overline{X}^{(J_1)}, \overline{X}^{(J_2)}, \dots, \overline{X}^{(J_p)}) \\ \hat{\theta}_{2,MM} &= g_2(\overline{X}^{(J_1)}, \overline{X}^{(J_2)}, \dots, \overline{X}^{(J_p)}) \\ &\dots \end{aligned}$$

Esempio 8.6.2

Sia $X \sim U(0, b)$ con legge uniforme nell'intervallo $]0, b[$. La quantità b è incognita e la regione parametrica è $b > 0$. Vogliamo ottenere una stima del parametro b con il metodo dei momenti:

$$\mu'_1 = \mathbb{E}(X) = \frac{b}{2}$$

Quindi:

$$b = 2 \cdot \mu'_1$$

Sostituiamo $\mu'_1 = \overline{X}$ e abbiamo trovato lo stimatore:

$$\hat{B}_{MM} = 2 \cdot \overline{X}$$

□

Esempio 8.6.3

Sia $X \sim U(-b, b)$ con b incognita e regione parametrica $b > 0$. Se scegliamo $J_1 = 1$, non possiamo stimare b perché la media del primo ordine è 0:

$$\mu'_1 = 0$$

Allora poniamo $J_2 = 2$ quindi il media del secondo ordine è:

$$\mu'_2 = \mathbb{E}(X^2) = \int_{-b}^b \frac{x^2}{2b} dx = \frac{1}{2b} \cdot \frac{x^3}{3} \Big|_{-b}^b = \frac{1}{6b} [b^3 - (-b)^3] = \frac{2b^3}{6b} = \frac{b^2}{3}$$

Quindi:

$$\mu'_2 = \mathbb{E}(X^2) = \frac{b^2}{3}$$

e allora:

$$b^2 = 3\mu'_2 \implies b = \sqrt{3\mu'_2}$$

Otteniamo quindi come stimatore di b :

$$\hat{B}_{MM} = \sqrt{3 \cdot \overline{X}^{(2)}}$$

Esempio 8.6.4

Sia $X \sim \mathcal{N}(\mu, \sigma^2)$ con μ e σ^2 incognite. Dobbiamo scegliere due equazioni. Scegliamo $J_1 = 1$ e $J_2 = 2$:

$$\begin{cases} \mu'_1 = \mathbb{E}(X) = \mu = f_1(\mu, \sigma^2) \\ \mu'_2 = \mathbb{E}(X^2) = \sigma^2 + \mu^2 = f_2(\mu, \sigma^2) \end{cases}$$

Otteniamo quindi che:

$$\begin{cases} \mu = \mu'_1 \\ \sigma^2 = \mu'_2 - \mu^2 \end{cases} \iff \begin{cases} \mu = \mu'_1 \\ \sigma^2 = \mu'_2 - (\mu'_1)^2 \end{cases}$$

Quindi, usiamo i momenti campionari invece dei momenti teorici e otteniamo:

$$\begin{cases} \widehat{M}_{MM} = \overline{X} \\ \widehat{S}_{MM}^2 = \overline{X^{(2)}} - (\overline{X})^2 \end{cases}$$

Lo stimatore ottenuto con il metodo della massima verosimiglianza di μ coincide con questo ottenuto con questo metodo.

Lo stimatore ottenuto con il metodo della massima verosimiglianza di σ^2 è:

$$\widehat{S}_{MV}^2 = S_d^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$$

8.7 Proprietà di consistenza degli stimatori

Ricordiamo che uno stimatore è una statistica $T = g(\underline{X})$, dipende dal campione e viene utilizzata per stimare i parametri. Il campione $\underline{X} = (X_1, X_2, \dots, X_n)$ è un campione casuale semplice di taglia n . Mettiamo in evidenza il numero della taglia denotando lo stimatore in questo modo:

$$T_n \in \mathbb{N}_n = g(\underline{X}_n)$$

quindi la media campionaria sarà:

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

Quando la taglia non è specificata, abbiamo una successione di stimatori, uno per ogni taglia fissata:

$$(T_n)_{n \in \mathbb{N}}$$

Tale successione può tendere ad un'altra variabile. La domanda è: *come tende? in che modo tende?*

Definizione 8.7.1 (Stimatore consistente)

$(T_n)_{n \in \mathbb{N}}$ è uno stimatore **consistente** (o coerente) per la stima del parametro θ se e solo se $T_n \rightarrow \theta$:

$$T_n \xrightarrow{\text{probabilità}} \theta$$

cioè, man mano che la taglia approssima sempre meglio il parametro θ . E' lo stesso tipo di convergenza della legge debole dei grandi numeri. □

Teorema 22 (Teorema consistenza di uno stimatore)

Condizione *sufficiente* per avere $(T_n)_{n \in \mathbb{N}}$ consistente è avere:

1. (T_n) è asintoticamente corretto
2. $\forall n \in \mathbb{N} \quad \exists \mathbb{D}^2(T_n) < +\infty$ che è infinitesima al crescere di n

Senza dimostrazione, gli enunciati si basano essenzialmente sul rischio dello stimatore T_n :

$$R_{T_n}(\theta) = \mathbb{D}_{T_n}^2(\theta) + d_{T_n}^2(\theta)$$

La prima condizione (1) dice che la distorsione tende a zero (è *asintoticamente corretto* - $d_{T_n}^2 \rightarrow 0$ per $n \rightarrow +\infty$) mentre la seconda condizione (2) dice che la varianza esiste e tende a 0 ($\mathbb{D}^2(T_n) \rightarrow 0$ per $n \rightarrow +\infty$). La dimostrazione si basa sulla disuguaglianza di Chebyshev. □

Su ipotesi non troppo restrittive (deve esistere almeno un momento), applicando tale teorema si dimostra che:

$$\overline{X}^{(r)} \text{ è consistente}$$

Quindi, quando, nel *metodo dei momenti* sostituiamo i momenti teorici con i momenti campionari, questi ultimi sono tutti consistenti (è *funzione di stimatori consistenti*).

Teorema 23 (Stimatore consistente e funzioni continue)

Se $(T_n)_{n \in \mathbb{N}}$ è consistente e g è una funzione continua, allora $[g(T_n)]_{n \in \mathbb{N}}$ è consistente per $g(\theta)$

□

Questo implica, che nel *metodo dei momenti*, quando risolviamo un sistema e otteniamo tutte funzioni continue, allora lo stimatore del metodo dei momenti è *consistente*.

Esempio 8.7.1

Consideriamo la genitrice $X \sim G(p)$ con $p \in (0, 1)$. Vogliamo stimare il parametro p con il metodo dei momenti. Scegliamo un indice $J_1 = 1$. La media di una variabile geometrica è:

$$\mu'_1 = \frac{1}{p}$$

Quindi:

$$p = \frac{1}{\mu'_1}$$

La funzione è la funzione reciproca ed è continua. Quindi, sostituendo il momento teorico con il momento campionario otteniamo:

$$\widehat{P}_{MM} = \frac{1}{\overline{X}}$$

La geometrica ammette il momento di ogni ordine perché ammette la funzione generatrice dei momenti definita nell'intorno di 0 (e quindi esistono tutti i momenti). Possiamo allora dire che la media campionaria è uno stimatore consistente. Questo implica che:

$$\widehat{P}_{MM} = \frac{1}{\overline{X}} \text{ è consistente per } p$$

Esempio 8.7.2

Sia $X \sim \Pi(\lambda)$ con $\lambda > 0$. Vogliamo stimare il parametro λ . Poniamo l'indice $J_1 = 1$. La media di una variabile di poisson è:

$$\mu'_1 = \lambda$$

Banalmente:

$$\lambda = \mu'_1$$

è quindi la funzione è la funzione d'identità ed è chiaramente continua. Sostituendo il momento teorico di ordine uno con il momento campionario di ordine uno, otteniamo:

$$\widehat{\Lambda}_{MM} = \overline{X}$$

Una variabile di poisson ammette la funzione generatrice dei momenti e quindi tutti i momenti esistono. Il momento campionario di ordine uno è uno stimatore consistente e questo implica che anche lo stimatore $\widehat{\Lambda}_{MM}$ è consistente essendo la funzione identica una funzione continua.

Consideriamo adesso la varianza $\mathbb{D}^2(X) = \mathbb{E}(X) = \lambda$. Sappiamo che $\mu'_1 = \lambda$. Quanto vale μ'_2 ?

$$\mathbb{D}^2(X) = \mu'_2 - (\mu'_1)^2 \implies \mu'_2 = \mathbb{D}^2(X) + (\mu'_1)^2 \implies \mu'_2 = \mu'_1 + (\mu'_1)^2 \implies \mu'_2 = \lambda + \lambda^2$$

Se provassimo a usare $J_1 = 2$ otterremmo:

$$\mu'_2 = \lambda + \lambda^2 \implies \lambda^2 + \lambda - \mu'_2 = 0 \implies \lambda = \frac{-1 + \sqrt{1 + 4\mu'_2}}{2}$$

Quindi, sostituendo il momento teorico di ordine due con il momento campionario di ordine due otteniamo:

$$\widehat{\Lambda}_{MM}^{[2]} = \frac{-1 + \sqrt{1 + 4 \cdot \overline{X}^{(2)}}}{2}$$

e anche questo stimatore è consistente perché la funzione è continua e il momento campionario di ordine due è uno stimatore consistente.

Abbiamo ottenuto due stimatori diversi semplicemente cambiando l'indice.

Esempio 8.7.3

Sia $X \sim \text{Esp}(\alpha)$ con $\alpha > 0$. La media di ordine r di una variabile esponenziale è:

$$\mu'_r = \mathbb{E}(X^r) = \frac{r!}{\alpha^r}$$

Se poniamo $J_1 = 1$ otteniamo:

$$\mu'_1 = \frac{1}{\alpha} \implies \alpha = \frac{1}{\mu'_1}$$

Tutti i momenti campionari di una variabile esponenziale sono consistenti perché la funzione generatrice dei momenti esiste, quindi il momento campionario di ordine uno è consistente. Inoltre, la funzione reciproco è continua, quindi lo stimatore (sostituendo il momento teorico di ordine uno con il momento campionario di ordine uno):

$$\hat{A}_{MM} = \frac{1}{\bar{X}} \quad (11)$$

è consistente.

Poniamo $J_1 = 2$, otteniamo:

$$\mu'_2 = \frac{2}{\alpha^2} \implies \alpha = \sqrt{\frac{2}{\mu'_2}} \quad (12)$$

La funzione ottenuta è continua. Il momento campionario di ordine due esiste ed è uno stimatore consistente. Quindi, lo stimatore:

$$\hat{A}_{MM}^{[2]} = \sqrt{\frac{2}{\bar{X}^{(2)}}}$$

Analizziamo il caso generale e cioè $J_1 \in \mathbb{N}$, otteniamo:

$$\mu'_{J_1} = \frac{J_1!}{\alpha^{J_1}} \implies \alpha^{J_1} = \frac{J_1!}{\mu'_{J_1}} \implies \alpha = \sqrt[J_1]{\frac{J_1!}{\mu'_{J_1}}}$$

Quindi, lo stimatore in generale di α sarà:

$$\hat{A}_{MM}^{[J_1]} = \sqrt[J_1]{\frac{J_1!}{\bar{X}^{(J_1)}}}$$

Abbiamo ottenuto una successione di stimatori.

Esempio 8.7.4

Sia $X \sim U(a, b)$ una genitrice con legge uniforme. Scegliamo $J_1 = 1$ e $J_2 = 2$, otteniamo:

$$\begin{cases} \mu'_1 = \mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2} \\ \mu'_2 = \mathbb{E}(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3} \end{cases}$$

Quindi:

$$\begin{aligned} \begin{cases} \mu'_1 = \frac{a+b}{2} \\ \mu'_2 = \frac{a^2+ab+b^2}{3} \end{cases} &\implies \begin{cases} a+b = 2\mu'_1 \\ a^2+ab+b^2 = 3\mu'_2 \end{cases} \implies \begin{cases} a+b = 2\mu'_1 \\ (a^2+2ab+b^2) - ab = 3\mu'_2 \end{cases} \implies \begin{cases} a+b = 2\mu'_1 \\ 4(\mu'_1)^2 - ab = 3\mu'_2 \end{cases} \\ &\implies \begin{cases} a+b = 2\mu'_1 \\ ab = 4(\mu'_1)^2 - 3\mu'_2 \end{cases} \end{aligned}$$

Chiamiamo $p = ab$ il prodotto e $s = a + b$ la somma. Dobbiamo risolvere questa equazione:

$$z^2 - sz + p = 0$$

questo perché, le due radici z_1 e z_2 corrispondono proprio ad a e b :

$$z_1 + z_2 = \frac{s}{1}, \quad z_1 \cdot z_2 = p$$

quindi:

$$z_{1,2} = \frac{s \pm \sqrt{s^2 - 4p}}{2}$$

Quindi:

$$a = \frac{s - \sqrt{s^2 - 4p}}{2}, \quad b = \frac{s + \sqrt{s^2 - 4p}}{2}$$

Sostituendo:

$$a = \frac{2\mu'_1 - \sqrt{4(\mu'_1)^2 - 16(\mu'_1)^2 + 12\mu'_2}}{2} \Rightarrow a = \frac{2\mu'_1 - \sqrt{12[\mu'_2 - (\mu'_1)^2]}}{2} \Rightarrow a = \frac{2\mu'_1 - 2\sqrt{3[\mu'_2 - (\mu'_1)^2]}}{2} \\ \Rightarrow a = \mu'_1 - \sqrt{3[\mu'_2 - (\mu'_1)^2]}$$

$$b = \mu'_1 + \sqrt{3[\mu'_2 - (\mu'_1)^2]}$$

Quindi:

$$\begin{cases} a = \mu'_1 - \sqrt{3[\mu'_2 - (\mu'_1)^2]} \\ b = \mu'_1 + \sqrt{3[\mu'_2 - (\mu'_1)^2]} \end{cases}$$

Quindi:

$$\hat{A}_{MM} = \bar{X} - \sqrt{3[\bar{X}^{(2)} - (\bar{X})^2]} \\ \hat{B}_{MM} = \bar{X} + \sqrt{3[\bar{X}^{(2)} - (\bar{X})^2]}$$

La genitrice X ammette la funzione generatrice dei momenti quindi tutti i momenti campionari sono stimatori consistenti. Le funzioni sono continue quindi gli stimatori ottenuti sono consistenti.

Esempio 8.7.5

Sia $X \sim U(0, b)$ una genitrice con legge uniforme nell'intervallo $[0, b]$. Vogliamo stimare il parametro $b > 0$ incognito con il *metodo della massima verosimiglianza*. Dobbiamo determinare:

$$\underline{x} \in S_{\underline{x}}, \quad \underset{b>0}{\operatorname{argmax}} L(b, \underline{x})$$

La funzione di densità di probabilità di una genetrici con legge uniforme nell'intervallo $[0, b]$ è:

$$f_X(x; b) = \begin{cases} \frac{1}{b}, & \text{se } x \in (0, b) \\ 0, & \text{altrimenti} \end{cases}$$

La problematica qui è che b compare anche nel supporto. Per trovare la funzione di verosimiglianza:

$$\prod_{i=1}^n f_X(x_i; b) = \frac{1}{b^n}$$

Lo stimatore di b è dato dall'osservazione più grande:

$$\hat{B}_{MV} = X_{(n)}$$

perché in questo modo tutti i dati sono più piccoli di questa stima e siamo sicuri che appartengono all'intervallo. Tale stima è la migliore perché ha la massima verosimiglianza ed è coerente con la rilevazione dei dati.

Con il metodo dei momenti abbiamo in precedenza ottenuto invece:

$$\hat{B}_{MM} = 2\bar{X}$$