

Appunti di Probabilità

Stefano Bonaccorsi
Dipartimento di Matematica,
Università degli Studi di Trento
<http://www.science.unitn.it/~probab>

Anno Accademico 2004/05

Contents

Introduzione	1
1 La teoria assiomatica del calcolo delle probabilità	3
1.1 Concetti chiave	3
1.1.1 Evento	3
1.1.2 Spazio di probabilità	3
1.1.3 Probabilità condizionata	4
1.1.4 Indipendenza	7
1.2 Calcolo combinatorio	8
1.2.1 Configurazioni di particelle	9
2 Variabili aleatorie o casuali	11
2.1 Variabili aleatorie reali	11
2.1.1 Funzione di ripartizione	11
2.1.2 Variabili aleatorie discrete	12
2.1.3 Variabili aleatorie continue	14
2.2 Vettori aleatori	16
2.2.1 Vettori aleatori continui	17
2.2.2 Indipendenza	18
2.3 Funzioni di variabili aleatorie	19
2.3.1 Cambiamento di variabile	19
2.3.2 Distribuzione della somma di due variabili aleatorie	19
2.4 Media	21
2.4.1 Momenti	24
2.4.2 Varianza	25
2.5 Funzione generatrice dei momenti	26
2.6 Alcune leggi legate alla statistica	28
2.6.1 Leggi χ^2	28
2.6.2 Leggi t di Student	30
3 Ripetizione di esperimenti indipendenti	31
3.1 Schema di Bernoulli	31
3.1.1 Formule approssimate	32
3.1.2 Probabilità di un valore compreso entro limiti assegnati	35
3.2 Teorema limite centrale	36
3.3 Legge dei grandi numeri	37

4	Processi di Markov	43
4.1	Introduzione	43
4.1.1	Classificazione dei processi stocastici	43
4.1.2	Introduzione ai processi di Markov	44
4.2	Le catene di Markov	44
4.2.1	La relazione di Chapman-Kolmogorov	45
4.2.2	Determinazione dello stato del sistema	46
4.2.3	Classificazione degli stati in una catena di Markov	46
4.3	Distribuzioni limite	47
4.3.1	Catene di Markov regolari	48
4.3.2	Catene periodiche	49
4.4	Catene a stati numerabili	49
4.5	Passeggiate casuali con barriere	50
4.5.1	Identità di Wald	51
4.6	Esercizi conclusivi	55
5	Processo di Poisson	59
5.1	Processi di conteggio	59
5.2	Processo di Poisson	60
5.3	Tempi di attesa e tempi di arrivo	60
5.3.1	Distribuzione condizionata	61
5.4	Ulteriori proprietà	62
5.4.1	Decomposizione di un processo di Poisson	62
5.4.2	Processi con divisione in classi dipendente dal tempo	62
5.4.3	Costruzione di un processo di Poisson spaziale	63
5.4.4	Comportamento asintotico	64
5.4.5	Teoremi limite	64
5.5	Alcune applicazioni alla teoria delle code	65
5.6	Esercizi conclusivi	67
6	Statistica	71
6.1	Statistica descrittiva	71
6.1.1	Indici	73
6.2	Statistica inferenziale	75
6.2.1	Stima per intervalli	77
6.2.2	Stima per intervalli della media di una popolazione qualsiasi, per grandi campioni	78
6.2.3	Stima per intervalli di una proporzione (frequenza)	79
6.2.4	Stima per intervalli della varianza di una popolazione normale	79
6.2.5	Stima della differenza tra le medie di due popolazioni normali	80
6.3	Esercizi di ricapitolazione	80

Introduzione

Il calcolo delle probabilità è quella disciplina matematica che affronta lo studio di tutti gli *esperimenti (casuali)*⁽¹⁾ il cui esito non è, in un certo istante⁽²⁾, prevedibile con certezza, ma su cui siamo interessati a dare un giudizio.

Si parla di probabilità sempre in relazione ad un evento; in altre parole si parla sempre di probabilità di una affermazione. Parleremo di affermazioni (in senso lato, una negazione sarà considerata una particolare affermazione) e di eventi come sinonimi. Intenderemo la probabilità di un evento un numero tra 0 ed 1, attribuendo probabilità 0 agli eventi impossibili e probabilità 1 agli eventi certi.

La definizione di *probabilità* è abbastanza controversa e diverse teorie sono state sviluppate in proposito; le riportiamo schematicamente nella prossima tabella. Qualunque impostazione si decida di accettare in proposito, vi è comunque un accordo generale sulla matematica del calcolo delle probabilità, che viene presentata nel prossimo capitolo.

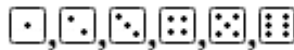
Teoria classica	Teoria frequentista	Teoria soggettiva
Formalizzata da Laplace (1812), è già implicita in Galileo, Pascal e Bernoulli, tra gli altri. Probabilità di un evento è il rapporto tra il numero di casi favorevoli e il numero dei casi totali, supposto che essi siano ugualmente possibili	Si basa sulla <i>legge empirica del caso</i> : in una serie di prove ripetute, tutte nelle stesse condizioni, ciascuno degli eventi possibili si presenta con una frequenza che, al crescere del numero delle prove, approssima la sua probabilità. Probabilità di un evento è il limite cui tende la frequenza al divergere del numero delle prove ripetute	Probabilità di un evento è la misura del grado di fiducia che un individuo coerente attribuisce, secondo le sue informazioni e opinioni, al verificarsi dell'evento. Si consideri anche la seguente interpretazione, dovuta a De Finetti: probabilità di un evento è il prezzo equo che un individuo è disposto a pagare per ottenere una vincita unitaria al verificarsi dell'evento. Con prezzo <i>equo</i> si intende il prezzo per cui l'individuo è disposto sia a fare la scommessa, che a riceverla

Nelle prossime pagine introdurremo la *teoria assiomatica* del calcolo delle probabilità. Vogliamo però dare subito un esempio di quelli che sono gli spazi di probabilità fondamentali, ovvero quelli più importanti sia dal punto di vista teorico che applicativo.

¹A seconda dei casi useremo, invece dell'aggettivo casuale, il termine aleatorio (dal latino *alea*: dado) o stocastico (dal greco *στοχος*: bersaglio, mira, congettura)

²indipendentemente dal fatto che l'esperimento sia stato già completato o meno, e neanche intendendo che debba essere necessariamente effettuato

Spazio campionario finito. Il lancio di un dado



è un fenomeno casuale, a cui associamo lo spazio campionario $\Omega = \{1, 2, \dots, 6\}$. In maniera analoga, spazi campionari finiti corrispondono al lancio di una moneta, all'estrazione di una carta da un mazzo o di una pallina da un'urna (estrazione del lotto).

Solitamente, non avremo ragioni per giudicare in maniera diversa i possibili esiti dell'esperimento. Se Ω possiede n elementi, porremo $\mathbb{P}(x) = 1/n$, dove $x \in \Omega$; questa scelta ci determina in maniera univoca la misura di probabilità.

Spazio di Poisson. Abbiamo ora a che fare con un esperimento i cui possibili esiti sono tutti i numeri interi maggiori o uguali a zero. Consideriamo ad esempio il numero di chiamate in coda ad un centralino ad un dato istante: non abbiamo ragione per ritenere che possa essere fissato un numero massimo di possibili chiamate, ma non sarà possibile ritenere tutti gli esiti equiprobabili. La variabile aleatoria connessa con questo esempio è detta avere distribuzione di Poisson.

Spazio di Bernoulli. Due giocatori prendono parte ad un gioco; lanciando a turno una moneta, il vincitore è colui che per primo ottiene lo stesso risultato del lancio precedente. I possibili esiti del gioco sono descritti dal seguente schema:

$$tt, ctt, tctt, ctctt, \dots \\ cc, tcc, ctcc, tctcc, \dots$$

Inoltre, possiamo pensare che le facce della moneta continuino ad alternarsi, seguendo uno dei due schemi

$$tctctct \dots, ctctctc \dots$$

Lo spazio campionario Ω che corrisponde al nostro "esperimento" è definito dall'insieme di tutte le successioni $\{x_i\}$ di $\{c, t\}$, con la convenzione che al lancio i è uscito come risultato x_i , ed è infinito. Consideriamo, per ogni j fissato, la proiezione X_j che ad ogni elemento $\{x_j\}$ di Ω associa x_j . L'evento $\{X_j = t\}$ rappresenta allora tutte gli elementi di Ω in cui esce testa al lancio j . Vogliamo che in \mathcal{A} siano contenuti, in primo luogo, gli eventi del tipo $\{X_j = t\}$. Allora anche gli eventi del tipo $\{X_j = c\}$ appartengono ad \mathcal{A} , così come gli eventi del tipo

$$\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}.$$

Cercheremo una misura di probabilità che, per ogni n fissato, ad ogni evento di questo tipo associa il valore $(\frac{1}{2})^n$. Fortunatamente, una tale misura esiste, ma la dimostrazione della sua esistenza non è per nulla banale!

Spazio uniforme. Consideriamo la scelta di un numero reale a caso nell'intervallo $[\alpha, \beta]$. Indichiamo con U il numero scelto. Lo spazio campione Ω coinciderà con l'intervallo $[\alpha, \beta]$, mentre vorremo che siano in \mathcal{A} tutti gli eventi del tipo $\{U \leq t\}$ per ogni $t \in [\alpha, \beta]$. Inoltre, se supponiamo che il numero sia scelto con probabilità uniforme all'interno dell'intervallo, la probabilità che il numero scelto appartenga ad un dato intervallo $[x, y]$ dipenderà solo dalla sua lunghezza $y - x$, e quindi

$$\mathbb{P}(U \in [x, y]) = \frac{y - x}{\beta - \alpha}, \quad \alpha \leq x \leq y \leq \beta.$$

Chapter 1

La teoria assiomatica del calcolo delle probabilità

1.1 Concetti chiave

1.1.1 Evento

Evento è un possibile risultato di un esperimento concettuale, con ciò intendendo che non ci interessa che l'esperimento sia stato realizzato o meno, o che non sia neanche realizzabile in pratica. La definizione degli eventi caratterizza l'esperimento, e viceversa non si può dare una definizione di evento senza riferirsi a un esperimento.

1.1.2 Spazio di probabilità

È una terna $(\Omega, \mathcal{A}, \mathbb{P})$ che formalizza tutte le nostre conoscenze sull'esperimento. Allora:

Ω : spazio campionario, contiene tutti i possibili esiti dell'esperimento;

\mathcal{A} : una collezione di sottoinsiemi di Ω (formalmente, una σ -algebra), che contiene tutti gli eventi a cui possiamo assegnare una probabilità. \mathcal{A} deve verificare le seguenti proprietà:

1. $\Omega \in \mathcal{A}$;
2. è stabile per passaggio al complementare $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;
3. è stabile per unione (e intersezione) infinita: se $A_i \in \mathcal{A}$, $i = 1, 2, \dots$, allora $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

\mathbb{P} : una misura di probabilità su (Ω, \mathcal{A}) , ossia una funzione che assegna ad ogni elemento $A \in \mathcal{A}$ un numero $\mathbb{P}(A)$. Si deve verificare che:

1. $\mathbb{P}(A) \geq 0$ per ogni $A \in \mathcal{A}$ e $\mathbb{P}(\Omega) = 1$;
2. se A_i , $i = 1, 2, \dots$ è una collezione di elementi di \mathcal{A} a due a due disgiunti, allora $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Vogliamo sottolineare come nella scelta dello spazio di probabilità, vi sia una notevole dose di arbitrarietà. Ciò, innanzitutto, perché non è necessariamente univoca la scelta dello spazio campionario. Nel caso del lancio di un dado, si potrà prendere $\Omega = \{1, 2, \dots, 6\}$, ma altrettanto legittima sarebbe scegliere $\Omega = \{0, 1, 2, \dots\}$.

Inoltre, una volta fissato l'insieme Ω , la scelta della famiglia \mathcal{A} è arbitraria e dipende *essenzialmente* dal criterio con cui decidiamo che alcuni possibili esiti sono interessanti ed altri no, ossia dal criterio con cui scegliamo gli eventi da porre in essa.

Infine, ancora meno scontata è la scelta della misura di probabilità \mathbb{P} , che dipende dal nostro giudizio sul possibile esito dell'esperimento.

La scelta del “modello matematico” $(\Omega, \mathcal{A}, \mathbb{P})$ è – in ogni caso – un'operazione *pre-matematica*. Chiedersi se una certa scelta sia “giusta o sbagliata” non ha dunque senso: o perlomeno non ha lo stesso senso che chiedersi se sono giusti o sbagliati determinati conti eseguiti nell'ambito di un particolare modello scelto.⁽¹⁾

Teoremi di base

- la probabilità dell'evento impossibile è nulla: $\mathbb{P}(\emptyset) = 0$.
- Un evento ed il suo complemento riempiono lo spazio (noto anche come *teorema delle probabilità totali*):

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

In particolare,

$$\mathbb{P}(A) + \mathbb{P}(A^c) = 1.$$

- La probabilità dell'unione di due eventi qualsiasi è data dalla formula

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

che possiamo esprimere dicendo che la probabilità dell'evento intersezione va contata una sola volta.

- Se $B \subseteq A$ allora $\mathbb{P}(B) \leq \mathbb{P}(A)$: quando l'evento B è contenuto in A il verificarsi del primo implica il secondo.
- Osserviamo che $(A \cup B)^c = A^c \cap B^c$: si ottiene allora $\mathbb{P}(\cup_n A_n) = 1 - \mathbb{P}(\cap_n A_n^c)$.

1.1.3 Probabilità condizionata

Per quanto abbiamo visto finora, la teoria delle probabilità sembra essere una parte dell'analisi. Il punto che la differenzia è il concetto di probabilità condizionata, che svilupperemo in questa sezione. Vogliamo rispondere alla seguente domanda. Supponiamo di avere assegnato una misura di probabilità su uno spazio campionario Ω , e di venire a conoscenza che un evento E è occorso. Come dobbiamo allora modificare le probabilità dei restanti eventi?

Ad un evento F possiamo assegnare due probabilità: una probabilità a priori $\mathbb{P}(F)$, che F possiede in Ω e una nuova probabilità, valutata appunto nell'ipotesi che E sia avvenuto, che chiameremo “probabilità condizionata di F dato E ”, e che indicheremo con $\mathbb{P}(F | E)$.

Esempio 1.1 Consideriamo l'esperimento di lanciare un dado. Sia X il risultato. Fissiamo l'evento F “esce il numero 6”, e sia E l'evento “esce un numero maggiore di 4”. A priori, tutti gli esiti sono equiprobabili, quindi assegneremo $\mathbb{P}(F) = 1/6$. Supponiamo ora che, dopo aver lanciato il dado, qualcuno ci informi che è accaduto E . Questo lascia due soli possibili risultati: 5 e 6. Senza ulteriori informazioni, dovremo ancora pensare che gli esiti siano equiprobabili, quindi la nuova probabilità che assegniamo a F sarà $1/2$, ossia $\mathbb{P}(F | E) = 1/2$.

△

¹Giorgio Letta, PROBABILITÀ ELEMENTARE, Zanichelli Editore, 1993

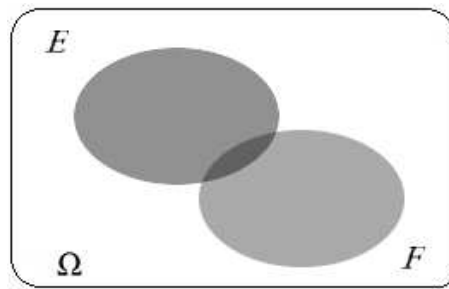


Figure 1.1: Diagramma di Venn

La probabilità condizionata dell'evento F , sapendo avvenuto l'evento E , si indica con la notazione $\mathbb{P}(F | E)$ e verifica

$$\mathbb{P}(F | E) \cdot \mathbb{P}(E) = \mathbb{P}(E \cap F). \quad (1.1)$$

Nel caso $\mathbb{P}(E) > 0$, $\mathbb{P}(F | E)$ è univocamente definita dalla relazione (1.1):

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

Più in generale, se A_1, A_2, \dots, A_n sono eventi, la probabilità della loro intersezione è data (in termini di probabilità condizionate) dalla seguente formula (*formula della moltiplicazione*)

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Di questa formula si può dare una dimostrazione per induzione.

Esempio 1.2 Ritorniamo al caso considerato nell'esempio 1.1; vogliamo giustificare la definizione (1.1). Supponiamo di aver effettuato numerose prove di estrazione, e indichiamo con $N_n(E)$ il numero di volte in cui si verifica l'evento E in n ripetizioni. Se considero solo gli $N_n(E)$ tentativi in cui è accaduto E , l'evento F è avvenuto $N_n(E \cap F)$ volte; la frequenza relativa allora verifica

$$\frac{N_n(E \cap F)}{N_n(E)} = \frac{N_n(E \cap F)/n}{N_n(E)/n} \xrightarrow{n \rightarrow \infty} \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)},$$

il primo termine è una misura della probabilità dell'evento F sapendo che E è accaduto, come volevamo mostrare.

Calcoliamo infine il valore di $\mathbb{P}(F | E)$ con l'aiuto della definizione (1.1):

$$\mathbb{P}(F | E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{1/6}{2/6} = \frac{1}{2}.$$

△

Il problema delle due monete.

Lanciamo due monete uguali ed equilibrate. (a) Qual è la probabilità che escano due teste, sapendo che la prima moneta mostra testa? (b) Qual è la probabilità che escano due teste, sapendo che almeno una moneta mostra testa?

Per quanto possa sembrare paradossale, le due domande non sono equivalenti e conducono a risposte diverse. Indichiamo con A e B gli eventi: la prima (risp. la seconda) moneta mostra testa. (a) Osserviamo che l'uscita di due teste è equivalente all'evento $A \cap B$; allora $\mathbb{P}(A \cap B | A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{1/4}{1/2} = \frac{1}{2}$. In questo caso dobbiamo calcolare $\mathbb{P}(A \cap B | A \cup B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A \cup B)} = \frac{1/4}{3/4} = \frac{1}{3}$.

Il paradosso dei compleanni.

In un'urna $\Omega = \{1, 2, \dots, s\}$ vengono scelti a caso n elementi. Ogni elemento è riposto nell'urna una volta estratto. Calcolare la probabilità dell'evento $A =$ "gli elementi estratti sono tutti diversi tra loro". Poniamo $A_k =$ "il k -esimo elemento estratto è diverso dai precedenti". Ovviamente

$$\mathbb{P}(A_1) = \frac{s}{s} = 1; \quad \mathbb{P}(A_2 | A_1) = \frac{s-1}{s}; \quad \mathbb{P}(A_3 | A_1 \cap A_2) = \frac{s-2}{s};$$

$$\mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}) = \frac{s-n+1}{s}.$$

Quindi, per la formula della moltiplicazione:

$$P(A) = \frac{s(s-1)(s-2)\dots(s-n+1)}{s^n} \cdot (2)$$

Questo esempio visto ci mostra come la legge della moltiplicazione permetta di risolvere problemi evitando le insidie e le difficoltà del calcolo combinatorio.

La formula precedente permette di risolvere il classico problema dei compleanni quando $s = 365$. Infatti

$$1 - \frac{(365)_n}{365^n}$$

è la probabilità che tra n persone scelte a caso ve ne siano almeno due che festeggiano il compleanno lo stesso giorno. Semplificando, faremo l'ipotesi che ogni persona abbia probabilità $1/365$ di nascere in uno qualsiasi dei giorni dell'anno (ed escluderemo gli anni bisestili). La seguente tabella fornisce alcuni valori numerici:

n	10	22	23	40	50	60	70
\mathbb{P}	.117	.476	.507	.891	.970	.994	.999

Contrariamente a quanto si potrebbe pensare, basta un insieme di 23 persone perché la probabilità che due di esse festeggino il compleanno lo stesso giorno (a parte l'anno, naturalmente) sia maggiore del 50%; con 70 persone, la probabilità che non vi siano sovrapposizioni è inferiore a un millesimo.

Formula di Bayes

Si tratta di una formula di frequente utilizzo nelle applicazioni, che segue facilmente dalla definizione di probabilità condizionata. Supponiamo che $\{H_i\}_{i=1}^N$ sia una *partizione* di Ω : gli H_i sono eventi a due a due disgiunti, di probabilità positiva, la cui unione è tutto Ω . Sia B un evento di probabilità positiva, e supponiamo di conoscere $\mathbb{P}(B | H_i)$ per ogni i . Quanto vale $\mathbb{P}(H_i | B)$? si ottiene:

$$\mathbb{P}(H_i | B) = \frac{\mathbb{P}(H_i)\mathbb{P}(B | H_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(H_i)\mathbb{P}(B | H_i)}{\sum_{j=1}^N \mathbb{P}(H_j)\mathbb{P}(B | H_j)}.$$

Questa formula ha una interpretazione suggestiva in termini di cause e effetto. Indichiamo con H_i le possibili "cause" dell'evento B . Allora ci interessa sapere qual è la probabilità che l'evento H_i sia stato la "causa" di B sapendo che B è occorso.

²Si noti che il numeratore $s(s-1)(s-2)\dots(s-n+1)$, abbreviato spesso con la scrittura $(s)_n$, chiamato fattoriale decrescente, è anche il numero di parole lunghe n su un alfabeto di s lettere, con lettere tutte distinte, oppure è il numero di disposizioni (semplici) di s oggetti presi n volte.

Esempio 1.3 Si supponga che nel rispondere ad una domanda, che ha s risposte, uno studente conosca la risposta esatta con probabilità p e non la conosca con probabilità $1 - p$. Se lo studente conosce la risposta sia 1 la probabilità che risponda esattamente, se invece non la sa, sia $1/s$ la probabilità che indovini. Determinare la probabilità che lo studente conosca la risposta nell'ipotesi che abbia risposto esattamente.

Ad esempio, uno studente liceale alla fine dell'anno scolastico, prima degli scrutini, ha un voto medio uguale a 5 e rischia di essere rimandato in quella materia. Il professore gli fa una domanda con solo due risposte per aumentargli il voto, e quindi promuoverlo, nel caso che risponda bene. Se lo studente risponde esattamente, quale dovrebbe essere ora il voto?

Il problema si risolve rapidamente con la legge di Bayes. Consideriamo gli eventi H = “lo studente conosce la risposta” ; A = “lo studente risponde esattamente” così che $\mathbb{P}(H) = p$, $\mathbb{P}(A | H) = 1$, $\mathbb{P}(A | H^c) = 1/s$; ne segue $\mathbb{P}(H | A) = \frac{ps}{1 - p + ps}$.

Nel caso dello studente liceale, è $s = 2$ e $p = 0.5$. Infatti uno studente con voto 5 ha probabilità $5/10$ di rispondere esattamente: questa è una misura della sua bravura. Se risponde esattamente alla domanda del docente il suo voto è ora $\mathbb{P}(H | A) = 2/3 \simeq 0,666$, cioè quasi 7.

△

1.1.4 Indipendenza

Due eventi A e B si dicono indipendenti se e solo se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. L'indipendenza equivale a dire che la conoscenza di uno dei due eventi non cambia la probabilità dell'altro, quindi si ha $\mathbb{P}(A | B) = \mathbb{P}(A)$, $\mathbb{P}(B | A) = \mathbb{P}(B)$.

Questa definizione si estende a più eventi come segue: gli eventi A_1, A_2, \dots, A_N sono indipendenti se per ogni scelta di indici i_1, \dots, i_k , tutti distinti e compresi tra 1 e N , si ha

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k}).$$

Esempio 1.4 Supponiamo di avere una moneta che dia testa con probabilità p , e croce con probabilità $q = 1 - p$. Supponiamo anche che questa moneta sia lanciata due volte. È ragionevole assegnare all'uscita di (t, t) la probabilità p^2 , all'uscita di (t, c) la probabilità pq , e così via. Sia E l'evento che esca testa al primo lancio e F l'evento è uscito croce al secondo lancio. Vogliamo verificare che con le probabilità assegnate, i due eventi sono indipendenti, come ci possiamo aspettare.

Abbiamo $\mathbb{P}(E) = p^2 + pq = p$, $\mathbb{P}(F) = pq + q^2 = q$. Infine, $\mathbb{P}(E \cap F) = pq$, quindi $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.

△

Esempio 1.5 Non è sempre così evidente a priori se due eventi sono indipendenti. Consideriamo una moneta perfettamente bilanciata, che viene lanciata due volte. Consideriamo gli eventi A = “al primo lancio esce testa” e B = “i due lanci sono uguali”. Allora

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}\{(t, t)\}}{\mathbb{P}\{(t, t), (c, c)\}} = \frac{1/4}{1/2} = \frac{1}{2} = \mathbb{P}(B).$$

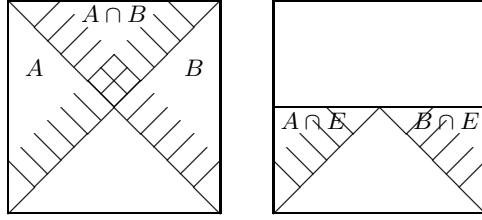
Quindi A e B sono indipendenti, anche se il risultato non era così ovvio.

△

Esempio 1.6 Per concludere, diamo un esempio di eventi che non sono indipendenti. Nell'esperimento precedente, siano I l'evento “al primo lancio esce testa” e J = “escono due teste”. Allora $\mathbb{P}(I) = \frac{1}{2}$ e $\mathbb{P}(J) = \frac{1}{4}$. L'evento $I \cap J$ è l'evento “escono due teste” ed ha probabilità $\frac{1}{4}$. Allora I e J non sono indipendenti dato che $\mathbb{P}(I)\mathbb{P}(J) = \frac{1}{8} \neq \mathbb{P}(I \cap J)$.

△

Esempio 1.7 La proprietà di indipendenza non si conserva cambiando la misura di probabilità. Costruiamo un esempio in cui A e B sono indipendenti rispetto a \mathbb{P} , E è un evento, e $\mathbb{P}(A | E)\mathbb{P}(B | E) \neq \mathbb{P}(A \cap B | E)$.



Nel quadrato $[0, 1] \times [0, 1]$, sia A l'evento $\{X < Y\}$, sia B l'evento $\{X + Y \geq 1\}$: sono indipendenti, dato che $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$, $\mathbb{P}(A \cap B) = \frac{1}{4}$. Consideriamo l'evento $E = \{Y \leq \frac{1}{2}\}$. Allora $\mathbb{P}(A | E) = \mathbb{P}(B | E) = \frac{1}{4}$, ma $\mathbb{P}(A \cap B | E) = 0$.

△

1.2 Calcolo combinatorio

Nello studio dei più semplici giochi, estrazioni o problemi di ordinamento, saremo portati a trattare spazi campionari finiti, in cui attribuiamo la stessa probabilità a tutti gli esiti elementari. Per calcolare la probabilità di un evento A basterà allora dividere il numero di punti contenuti in A (*casi favorevoli*) per il numero totale di punti di Ω (*casi totali*). Questo si può ottenere più facilmente seguendo poche regole che vogliamo ora introdurre.

Definizioni

Si dice **disposizione** di n oggetti di classe k ($k \leq n$) ogni ordinamento (sequenza ordinata) di k oggetti scelti tra n senza ripetizioni.

Indichiamo con D_k^n il numero di tali disposizioni:

$$D_k^n = (n)_k = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

Due disposizioni possono differire o per gli oggetti contenuti o per l'ordine degli oggetti stessi.

Se $n = k$ abbiamo le **permutazioni** di n oggetti:

$$P_n = D_n^n = n!^{(3)}$$

³Questo numero è detto *n* fattoriale, ed è uguale a $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$. L'espressione $0!$ è definita essere 1 per rendere più semplici alcune formule. Diamo i primi valori di questa funzione:

n	$n!$	n	$n!$
0	1	6	720
1	1	7	5040
2	2	8	40320
3	6	9	362880
4	24	10	3628800
5	120		

Si osservi come questa funzione cresce molto rapidamente.

Possiamo anche scegliere k oggetti tra gli n con ripetizione; ogni scelta viene effettuata sull'intera popolazione, cosicché ogni elemento può essere estratto più volte. Ognuno delle k scelte può essere effettuata in n modi, così esistono n^k disposizioni con ripetizione.

Sia E un insieme di n elementi. Si dice **combinazione** di classe k tra n oggetti ogni sottoinsieme di E costituito da k elementi. Indichiamo con C_k^n il numero di tali sottoinsiemi:

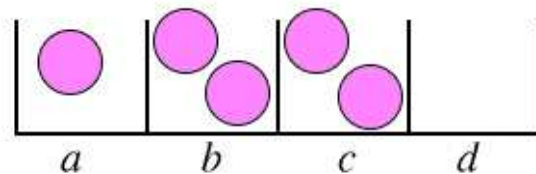
$$C_k^n = \binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{1}{k!} D_k^n.$$

1.2.1 Configurazioni di particelle

Statistica di Bose-Einstein.

Consideriamo un sistema di n particelle indistinguibili; dividiamo queste particelle in s regioni, in modo che ognuna delle particelle cada in una e una sola regione. Lo stato del sistema è descritto dal numero delle particelle in ognuna delle regioni, quindi è definito univocamente dal vettore (n_1, \dots, n_s) , dove n_k è il numero di particelle nella regione k , $k = 1, \dots, s$.

Definiamo su questo sistema la *statistica di Bose-Einstein*, ossia definiamo gli eventi elementari come tutte le configurazioni distinte di particelle, ad ognuna delle quali assegnamo la stessa probabilità.



Vogliamo calcolare quante sono le configurazioni distinte.

Pensiamo alle scatole come agli intervalli tra $(s+1)$ sbarre, e usiamo un asterisco $*$ per le palline. Il simbolo $|*|**|*|*$ identifica la configurazione precedente, in cui una pallina è nella scatola a , due nella scatola b , due nella c e nessuna nella scatola d . La distribuzione delle palline nelle scatole è fissata nel momento in cui abbiamo fissato i posti in cui troviamo gli asterischi. Ma vi sono $n+s-1$ posizioni possibili per gli asterischi, quindi il numero di possibili configurazioni è

$$C_n^{n+s-1} = \binom{n+s-1}{n}.$$

Statistica di Fermi-Dirac.

Definiamo su un sistema di n particelle e s regioni la *statistica di Fermi-Dirac*, in cui si assume che non più di una particella può occupare una regione e tutte le configurazioni possibili sono equiprobabili.

Particelle atomiche, come ad esempio gli elettroni, sono soggette alla statistica di Fermi-Dirac; i fotoni, al contrario, sono soggetti alla statistica di Bose-Einstein.

L'espressione $n!$ entrerà in molti conti, e avremo bisogno di stimare il suo valore quando n è grande. Chiaramente, non è possibile, praticamente, dare il risultato esatto in questo caso. Useremo invece la *formula di Stirling*, che permette di approssimare $n!$ quando n è grande con l'espressione

$$n! \simeq \sqrt{2\pi n} n^n e^{-n}. \quad (1.2)$$

Esempio 1.8 Vogliamo trovare il numero di configurazioni (equiprobabili) di n particelle divise in s regioni, soggette alla statistica di Fermi-Dirac. Lo stato del sistema è identificato dalle regioni occupate; le configurazioni distinte sono

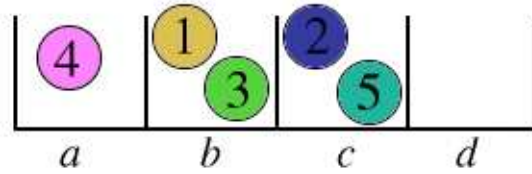
$$C_n^s = \binom{s}{n}.$$

△

Statistica di Maxwell-Boltzmann.

Nel caso di particelle distinguibili, si ricade nel caso classico; si parla allora di *statistica di Maxwell-Boltzmann*. Si gettano a caso n biglie in s scatole (disposizione) o si formano parole lunghe n con un alfabeto di s lettere.

Ad esempio, alla disposizione di 5 biglie in 4 scatole a, b, c, d : la prima e la terza nella scatola b , la seconda e la quinta in c , la quarta in a ,



corrisponde la parola $bcbac$. In tal caso lo spazio campionario Ω ha s^n elementi.

Chapter 2

Variabili aleatorie o casuali

2.1 Variabili aleatorie reali

Al termine di un esperimento, si è spesso condotti a considerare delle quantità scalari che sono funzioni dell'esito dell'esperimento stesso, piuttosto che gli eventi stessi definiti dall'esperimento. Si usa parlare allora di variabili casuali quando i possibili risultati dell'esperimento sono esprimibili mediante numeri.

Potremo dire coincisamente: variabile aleatoria o casuale è una funzione sullo spazio di probabilità ai cui valori si associa la probabilità del loro manifestarsi, è tale che due valori distinti vanno considerati come eventi escludentesi a vicenda, e l'insieme dei valori considerati esaurisce tutte le possibilità.

Definizione

L'applicazione $X : \Omega \rightarrow \mathbb{R}$ è una variabile casuale sullo spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ se gli eventi $(X \leq x)$:

$$(X \leq x) = \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}(]-\infty, x])$$

sono contenuti in \mathcal{A} per ogni $x \in \mathbb{R}$.

Le variabili aleatorie sono definite, in pratica, assegnando la loro funzione di ripartizione $F(x) = \mathbb{P}(X \leq x)$. Due particolari tipi di variabili aleatorie hanno il massimo interesse nelle applicazioni:

- (a) la funzione di ripartizione $F(x)$ è una funzione a scalini: la variabile aleatoria associata è detta variabile aleatoria discreta;
- (b) la funzione di ripartizione $F(x)$ è una primitiva di una funzione $f(x) \geq 0$. In tal caso si parla di variabile aleatoria continua. La funzione $f(x)$ prende il nome di densità di probabilità.

2.1.1 Funzione di ripartizione

La funzione di ripartizione di una variabile aleatoria X viene definita come la probabilità che si manifesti un qualunque valore della variabile aleatoria non superiore ad un valore prefissato x :

$$F_X(x) = F(x) = \mathbb{P}(X \leq x).$$

La funzione di ripartizione di una generica variabile casuale X ha le seguenti proprietà:

- (a) $\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1;$

- (b) F è una **funzione monotona non decrescente**: $F(x) \leq F(y)$ se $x < y$. Infatti, vale il corrispondente ordinamento tra gli insiemi: $(X \leq x) \subset (X \leq y)$ se $x < y$.
- (c) per ogni punto $x \in \mathbb{R}$ sono definiti i limiti

$$\lim_{\epsilon \searrow 0} F(x - \epsilon) = F(x^-), \quad \lim_{\epsilon \searrow 0} F(x + \epsilon) = F(x^+).$$

La funzione di ripartizione è continua a destra:

$$F(x^+) = F(x)$$

per ogni $x \in \mathbb{R}$; si ha inoltre $F(x^-) \leq F(x)$. I punti in cui vale la disuguaglianza stretta: $F(x^-) < F(x)$ sono al più una infinità numerabile e in tali punti si ha

$$\mathbb{P}(X = x) = F(x) - F(x^-),$$

che si legge: il valore del salto della funzione di ripartizione in un punto di discontinuità $x \in \mathbb{R}$ corrisponde alla probabilità che la variabile aleatoria assuma il valore x ;

- (d) se esistono due punti $a, b \in \mathbb{R}$ per cui $F(a) = 0$, $F(b) = 1$, allora la variabile aleatoria assume con certezza valori nell'intervallo $[a, b]$, e viceversa.

Densità di probabilità

Una funzione reale $f(x)$ è una densità di probabilità se:

- (a) **assume valori non negativi**: $f(x) \geq 0$;
- (b) **ha “massa” totale 1**: $\int_{\mathbb{R}} f(x) dx = 1$.

Si parlerà di densità di probabilità in relazione alle variabili aleatorie continue. In particolare, data la funzione di ripartizione $F(x)$ di una variabile di tale genere, la funzione $f(x) = \frac{dF}{dx}(x)$ è la sua⁽¹⁾ funzione densità di probabilità.

Data una funzione densità, è possibile risalire alla funzione di ripartizione tramite un integrale:

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

In altre parole, $F(x)$ è la primitiva di $f(x)$ determinata dalle condizioni $F(-\infty) = 0$, $F(+\infty) = 1$.

Osserviamo ora che per una variabile aleatoria continua X , con densità $f(x)$, per ogni punto $x \in \mathbb{R}$ si ha $\mathbb{P}(X = x) = 0$, quindi $f(x)$ *non* rappresenta certamente questa probabilità. La funzione $f(x)$ *non* rappresenta la probabilità di alcun evento riguardante la variabile aleatoria X . Solo quando viene integrata, ad esempio su un intervallo $I = [x, x + dx]$, essa rappresenta una probabilità.

2.1.2 Variabili aleatorie discrete

Indicheremo con variabile aleatoria discreta una variabile aleatoria la cui funzione di ripartizione ha la seguente proprietà:

la funzione di ripartizione $F(x)$ è una funzione a scalini: esiste un insieme di punti di discontinuità $\{x_n\}$ in cui la differenza tra i limiti destro e sinistro è positiva, mentre tra due punti di discontinuità la funzione assume valore costante.

¹con ciò intendendo che è univocamente determinata dalla funzione di ripartizione, e, viceversa

Ne risulta che l'insieme dei possibili valori $\{x_n\}$ che la variabile assume con probabilità positiva ha cardinalità al più infinitamente numerabile. Parleremo allora di “distribuzione” della variabile aleatoria X intendendo l'insieme delle coppie $\{(x_n, p_n)\}$, dove $p_n = \mathbb{P}(X = x_n)$.

Ogni insieme finito o infinito numerabile di numeri positivi p_i , compresi tra 0 e 1 e soddisfacenti la condizione

$$\sum_i p_i = 1 \quad (2.1)$$

può essere utilizzato per definire una variabile aleatoria discreta. Basterà, a tal fine, associare ognuno di questi valori a un particolare valore della variabile.

Esempio 2.1 Consideriamo il lancio di una moneta avente probabilità p di ottenere testa. Diremo variabile aleatoria di Bernoulli la variabile che assume valore 1 nel caso esca testa, e valore 0 nel caso esca croce, ossia

$$\mathbb{P}(X = 0) = q = 1 - p, \quad \mathbb{P}(X = 1) = p.$$

△

Distribuzione binomiale

Consideriamo una variabile casuale X che può assumere tutti i valori interi k compresi tra 0 e n , e le corrispondenti probabilità siano date dalla formula

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

dove $0 < p < 1$, $q = 1 - p$.

Si osserva che le singole probabilità p_k sono date dai termini dello sviluppo del binomio

$$(p + q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}.$$

Osservando che $(p + q)^n = 1^n = 1$, si ricava immediatamente che la condizione (2.1) è soddisfatta.

In questo caso, la funzione di ripartizione è esprimibile solo tramite la somma parziale

$$F(x) = \sum_{k=0}^j \binom{n}{k} p^k q^{n-k}, \quad j \leq x < j + 1, \quad 0 \leq x \leq n,$$

e $F(x) = 0$ per $x < 0$, $F(x) = 1$ per $x \geq n$.

Esempio 2.2 Siano I_1, I_2, \dots, I_n variabili aleatorie di Bernoulli, che descrivono il risultato del lancio di una moneta: $\mathbb{P}(I_1 = 1) = p$ è la probabilità che esca testa, $\mathbb{P}(I_1 = 0) = q = 1 - p$ è la probabilità che esca croce. Allora $S_n = I_1 + I_2 + \dots + I_n$ il numero di teste (=successi) sugli n lanci della moneta. Si dimostra che per ogni $k = 0, 1, 2, \dots, n$, la probabilità $\mathbb{P}(S_n = k)$ di avere k teste su n lanci, vale

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (\text{Bernoulli, 1713})$$

Infatti, $p^k (1 - p)^{n-k}$ è la probabilità di avere un numero di teste uguale a k , ed $n - k$ croci, in un ordine specificato. Ma vi sono $\binom{n}{k}$ modi di scegliere tra gli n lanci quei k in cui fare uscire le teste. Da qui il prodotto.

△

Distribuzione geometrica

Nota. È utile ricordare la somma della serie geometrica: per ogni $-1 < \alpha < 1$, vale

$$\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}.$$

La variabile X assuma i valori interi $0, 1, 2, \dots$; le probabilità corrispondenti siano date da $p_n = p(1-p)^n$, dove p è un parametro che si può supporre noto e tale che $0 < p < 1$. La variabile aleatoria X è quindi definita da

$$\mathbb{P}(X = n) = p(1-p)^n, \quad n = 0, 1, 2, \dots$$

e viene detta distribuzione geometrica di parametro p .

La funzione di ripartizione è

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - (1-p)^{n+1}, & n \leq x < n+1. \end{cases}$$

Distribuzione di Poisson

Sia λ un numero positivo. La distribuzione di Poisson di parametro λ è definita da

$$\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n = 0, 1, 2, \dots$$

La proprietà (2.1) si verifica facilmente attraverso lo sviluppo in serie di Taylor della funzione esponenziale:

$$e^\lambda = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}.$$

In molti esperimenti, il contare il **numero di ripetizioni di un fenomeno** può essere modellato da una variabile di Poisson. Ad esempio, il numero di particelle radioattive che decadono in un intervallo di tempo fissato, il numero di chiamate in arrivo ad un centralino, il numero di errori di stampa in una pagina di libro, sono tutti fenomeni che sappiamo (dall'esperienza) poter essere modellati con questa distribuzione.

Esempio 2.3 *In molti casi, la distribuzione di Poisson può essere usata come approssimazione per una distribuzione binomiale di parametri n e p , con n molto grande e p piccolo. Si parla allora di legge degli eventi rari; un senso preciso a questo verrà dato nella Sezione 3.1.1.*

△

2.1.3 Variabili aleatorie continue

In molte situazioni, è naturale considerare le variabili aleatorie (ad esempio, nella misura di quantità fisiche, come la massa, la distanza o il tempo) “continue” piuttosto che discrete. Intuitivamente, possiamo definire una variabile aleatoria continua X su uno spazio Ω come una funzione $X(\omega)$ tale che

$$\mathbb{P}(X = x) = 0, \quad x \in \mathbb{R},$$

ossia, **le che X assuma ogni specifico valore con probabilità nulla.** Questa proprietà si manifesta quando la funzione di ripartizione è continua in ogni punto, ma non necessariamente ammette densità.

In effetti, per ragioni di semplicità matematica, richiederemo qualcosa di più.

Definizione.

X è una variabile aleatoria continua se e solo se la sua funzione di ripartizione è assolutamente continua⁽²⁾.

In molti casi, le proprietà delle variabili aleatorie continue sono definite in termini della funzione densità di probabilità.

Distribuzione esponenziale

La densità di probabilità di una variabile avente distribuzione esponenziale è data da

$$f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$$

La corrispondente funzione di ripartizione è

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

Le distribuzioni esponenziali appaiono naturalmente nello studio dei tempi di decadimento di fenomeni radioattivi ed in altri modelli che riguardano tempi d'attesa, come il tempo di attesa prima della rottura di un'apparecchiatura, il tempo trascorso in una coda, ecc.

Un'importante proprietà delle distribuzioni esponenziali è che, se X è una tale variabile,

$$\mathbb{P}(X > a + b) = \mathbb{P}(X > a)\mathbb{P}(X > b), \quad a \geq 0, b \geq 0.$$

Una forma più suggestiva, ma equivalente, della relazione precedente si scrive

$$\mathbb{P}(X > a + b \mid X > a) = \mathbb{P}(X > b), \quad a \geq 0, b \geq 0. \quad (2.2)$$

Pensiamo a X come il tempo necessario perché un'apparecchiatura si rompa; allora, condizionato al fatto di non avere avuto rotture fino al tempo a , la probabilità di non avere rotture per altre b unità di tempo è pari alla probabilità di non avere rotture nelle prime b unità di tempo. Questo implica che l'invecchiamento del pezzo non modifica (non aumenta né diminuisce) la probabilità di rotture in un certo intervallo.

Infine, sia X è una variabile aleatoria, non negativa, per cui è verificata la (2.2): allora:

$$\text{o } \mathbb{P}(X = 0) = 1,$$

oppure X ha distribuzione esponenziale.

Distribuzione gaussiana o normale

La distribuzione Gaussiana possiede molte interessanti proprietà, così, spesso, variabili aleatorie di cui non si conosce la distribuzione sono supposte essere Gaussiane, soprattutto in fisica e astronomia. Sebbene questa possa essere un'ipotesi errata, è sovente una buona approssimazione, grazie ad un risultato sorprendente noto come *teorema limite centrale*. Questo teorema afferma che la distribuzione della somma di un gran numero di variabili aleatorie indipendenti, che possiedono media e varianza finita, tende alla distribuzione Gaussiana. Molti esperimenti "classici", come la distribuzione dei voti di un esame, le altezze di una classe di popolazione, ecc., seguono a grandi linee una distribuzione Gaussiana, con pochi elementi estremi e molti elementi vicino alla media.

²questo ci consente di affermare che $F(x)$ è derivabile per quasi ogni x e che la sua derivata $f(x)$ è la densità di probabilità di X : $\int_{-\infty}^x f(t) dt = F(x)$

La distribuzione Gaussiana standard, che indichiamo con il simbolo $\mathcal{N}(0, 1)$, è definita tramite la funzione densità di probabilità

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right).$$

Questa densità è chiaramente una funzione simmetrica rispetto all'origine.

Non si può esprimere una primitiva di $e^{-x^2/2}$ in termini finiti; è tuttavia possibile verificare che $\varphi(x)$ ha "massa" totale 1. I valori della funzione di ripartizione $\Phi(x)$ sono riportati in apposite tavole. In particolare, dato che φ è simmetrica, è necessariamente $\Phi(0) = 1/2$ ed inoltre

$$\Phi(x) = 1 - \Phi(-x), \quad x \in \mathbb{R}.$$

Sia ora X una variabile aleatoria con distribuzione Gaussiana standard e poniamo $Y = m + \sigma X$, $\sigma > 0$. Si può allora verificare che Y ha densità di probabilità

$$\varphi_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - m)^2\right).$$

e la distribuzione Gaussiana corrispondente si indica con $\mathcal{N}(m, \sigma^2)$. Dato che non abbiamo ancora parlato di media e varianza, per il momento pensiamo a m e σ come due parametri che indicizzano la classe delle distribuzioni Gaussiane.

Vedremo più avanti altre proprietà delle distribuzioni Gaussiane. Per il momento, ci limitiamo a osservare che la somma di variabili Gaussiane indipendenti è ancora una variabile Gaussiana: Se X_1, X_2 sono variabili Gaussiane indipendenti, di distribuzione $\mathcal{N}(m_i, \sigma_i^2)$ per $i = 1, 2$, allora $Y = X_1 + X_2$ ha distribuzione $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

2.2 Vettori aleatori

Nell'introdurre il concetto di variabile aleatoria, si è osservato che al termine di un esperimento è possibile essere interessati a più aspetti dello stesso, il che conduce a definire più funzioni $X_1(\omega), X_2(\omega)$, ecc., sullo stesso spazio $(\Omega, \mathcal{A}, \mathbb{P})$. Se si considerano *simultaneamente* i valori assunti dalle funzioni X_1, X_2 , ecc., il risultato sarà un *vettore aleatorio* $(X_1(\omega), X_2(\omega), \dots)$.

Come nel caso scalare, lo studio dei vettori aleatori viene condotto attraverso l'analisi della funzione di ripartizione e (nel caso continuo) della funzione densità.

Consideriamo un vettore aleatorio X avente k componenti, ossia $X = (X_1, \dots, X_k)$. Se esiste un insieme (finito o numerabile) di punti $\{x_j, j \in \mathbb{N}, x_j = (x_j^1, \dots, x_j^k)\}$ tale che $f(x_j^1, \dots, x_j^k) := \mathbb{P}(X = x_j) > 0$, $\sum_{j=0}^{\infty} f(x_j^1, \dots, x_j^k) = 1$, diremo che X è un vettore aleatorio discreto; l'insieme $\{(x_j, p_j)\}$ è detto la distribuzione di X .

Esiste una terminologia tradizionale collegata ai vettori aleatori. Così, *parleremo di distribuzione congiunta per indicare la distribuzione del vettore $X = (X_1, \dots, X_k)$ e di distribuzione marginale della variabile aleatoria X_j per indicare la distribuzione della componente X_j .*

Consideriamo il vettore aleatorio discreto (X, Y) , con distribuzione $\{(z_k, p_k), z_k = (x_k, y_k) \in \mathbb{R}^2\}$. Per determinare la distribuzione della variabile aleatoria X , basterà eseguire la somma delle probabilità degli eventi (incompatibili)

$$p_X(x) = \mathbb{P}(X = x) = \sum_{\{(x_k, y_k) : x_k = x\}} \mathbb{P}((X, Y) = (x_k, y_k)) = \sum_{\{(x_k, y_k) : x_k = x\}} p_k.$$

Similmente,

$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{\{(x_k, y_k) : y_k = y\}} \mathbb{P}((X, Y) = (x_k, y_k)) = \sum_{\{(x_k, y_k) : y_k = y\}} p_k.$$

In altre parole, conoscendo la distribuzione congiunta del vettore (X, Y) è possibile ricavare le distribuzioni delle componenti X e Y .

Consideriamo l'esperimento di lanciare una moneta e girare un dado. Intuitivamente, sappiamo che qualunque sia il risultato del lancio della moneta, questo non influenzerà l'esito del lancio del dado, e viceversa. Vogliamo ora tradurre in termini matematici questa intuizione. Sia X la variabile che assume valore 0 o 1 a seconda che ne lancio della moneta esca testa oppure croce, e sia Y il numero uscito dal lancio del dado; allora il vettore aleatorio (X, Y) rappresenta l'esito del nostro esperimento. Se assumiamo che gli eventi $(X = i)$ e $(Y = j)$, $i = 0, 1$, $j = 1, \dots, 6$ sono indipendenti, deve essere

$$\mathbb{P}((X, Y) = (i, j)) = \mathbb{P}(X = i)\mathbb{P}(Y = j), \quad \forall (i, j).$$

In altre parole, indicando con $\{(z_k, p_k), z_k = (x_k, y_k) \in \mathbb{R}^2\}$ la distribuzione del vettore aleatorio discreto $\vec{Z} = (X, Y)$, diremo che le variabili aleatorie X e Y sono indipendenti se e solo se

$$\mathbb{P}(\vec{Z} = \vec{z}_k) = \mathbb{P}(X = x_k)\mathbb{P}(Y = y_k).$$

2.2.1 Vettori aleatori continui

Per ragioni di semplicità, ci limiteremo al caso di due variabili casuali X e Y . Poiché sappiamo che $(X \leq x)$ e $(Y \leq y)$ sono eventi, tale è anche la loro intersezione $(X \leq x, Y \leq y)$.

Indichiamo con $F = F_{(X,Y)}$ la funzione di ripartizione congiunta di X e Y definita da

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

La funzione di ripartizione può essere utilizzata per calcolare la probabilità che la coppia (X, Y) appartenga ad un rettangolo del piano. Dato il rettangolo

$$R = \{(x, y) : a < x \leq b, c < y \leq d\},$$

dove $a \leq b$, $c \leq d$, allora

$$\mathbb{P}((X, Y) \in R) = F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

Diremo che X e Y hanno *densità congiunta* se esiste una funzione f integrabile, $f \geq 0$, tale che

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

Naturalmente **esistono vettori aleatori che non sono né discreti né dotati di densità. Una delle svariate circostanze in cui ciò si manifesta è quella in cui, ad esempio, X è discreta e Y è invece continua.**

Una volta dato il vettore aleatorio bi-dimensionale (X, Y) è possibile da esso definire le due variabili aleatorie X e Y ; abbiamo visto come, nel caso discreto, dalla funzione di ripartizione del vettore sia possibile, con operazioni elementari, ottenere le funzioni di ripartizione delle due variabili aleatorie considerate singolarmente.

Nel caso di vettore aleatorio continuo, l'operazione necessaria per ottenere le funzioni di densità marginali, corrispondente alla sommatoria del caso discreto, sarà data da un integrale

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx.$$

Le funzioni di ripartizione marginali di X e Y si potranno ottenere direttamente dalla funzione di ripartizione F del vettore (X, Y) con la formula

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y), \quad F_Y(y) = \lim_{x \rightarrow \infty} F(x, y).$$

Mentre dalla distribuzione bidimensionale è sempre possibile ricavare le marginali, il viceversa non vale in generale: la conoscenza delle funzioni di probabilità o delle densità marginali di un vettore aleatorio (X, Y) non è, in generale, sufficiente per la conoscenza della distribuzione del vettore stesso.

2.2.2 Indipendenza

Definizione.

Le variabili aleatorie X_1, \dots, X_m si dicono *indipendenti* se e solo se per ogni valore di $a_1 \leq b_1, \dots, a_m \leq b_m$ si ha

$$\mathbb{P}(a_1 < X_1 \leq b_1, \dots, a_m < X_m \leq b_m) = \mathbb{P}(a_1 < X_1 \leq b_1) \cdots \mathbb{P}(a_m < X_m \leq b_m).$$

La precedente definizione, nel caso di due variabili aleatorie X e Y , si riduce a chiedere che sia

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \mathbb{P}(a < X \leq b) \mathbb{P}(c < Y \leq d) \quad (2.3)$$

per ogni scelta di $a \leq b, c \leq d$.

Abbiamo già visto come l'indipendenza di due variabili aleatorie discrete X e Y si possa caratterizzare in termini delle distribuzioni: posto \vec{Z} il vettore (X, Y) , con distribuzione $\{\vec{z}_k, p_k\}$, si ha

$$p_k = \mathbb{P}(\vec{Z} = \vec{z}_k) = \mathbb{P}(X = x_k) \mathbb{P}(Y = y_k).$$

È facile vedere che questa condizione è equivalente alla definizione di indipendenza.

Una condizione equivalente alla (2.3) è la seguente: indicando con F la funzione di ripartizione congiunta e con F_X, F_Y le funzioni di ripartizione delle variabili X e Y rispettivamente, allora X e Y sono indipendenti se e solo se per ogni coppia (x, y) di punti del piano si ha

$$F(x, y) = F_X(x) F_Y(y).$$

Supponiamo che X e Y abbiano densità congiunta f e marginali f_X, f_Y ; allora la relazione (2.3) diviene

$$\int_a^b \int_c^d f(x, y) dy dx = \int_c^d f_Y(y) dy \int_a^b f_X(x) dx.$$

Questa eguaglianza è verificata se e solo se

$$f(x, y) = f_X(x) f_Y(y) \quad (2.4)$$

per ogni $(x, y) \in \mathbb{R}^2$.

In particolare, per determinare l'indipendenza di due variabili aleatorie X e Y , basta conoscere la loro densità congiunta $f(x, y)$: a partire da f si possono calcolare le densità marginali f_X e f_Y , e quindi verificare se vale la (2.4).

Esempio. Distribuzione uniforme nel disco unitario

Sia $Z = (X, Y)$ una variabile aleatoria con densità

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1, \\ 0, & \text{altrimenti.} \end{cases}$$

Allora, per ogni sottoinsieme A (sufficientemente regolare) del disco,

$$\mathbb{P}((X, Y) \in A) = \int \int_A f(x, y) dy dx = \frac{\text{area di } A}{\pi},$$

che coincide con la nostra intuizione di uniformità. La densità marginale f_X è data da

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{2}{\pi} \sqrt{1 - x^2}$$

per $x \in (-1, 1)$ e $f_X(x) = 0$ altrimenti. La densità marginale $f_Y(y)$ è data dalla stessa formula, sostituendo x con y .

Per $x \in (-1, 1)$, $y \in (-1, 1)$ si ha $f_X(x)f_Y(y) = \frac{4}{\pi^2} \sqrt{(1-x^2)(1-y^2)}$ che non coincide con la densità congiunta $f(x, y)$. Ne segue che X e Y sono variabili aleatorie dipendenti. Questo è in accordo con la nostra intuizione di indipendenza, visto che sappiamo che quando X è vicino ad 1, ad esempio, allora deve essere necessariamente Y vicino a 0: la conoscenza di X influenza il nostro giudizio di probabilità su Y .

2.3 Funzioni di variabili aleatorie

2.3.1 Cambiamento di variabile

Consideriamo una funzione *continua* $g : \mathbb{R} \rightarrow \mathbb{R}$. Se X è una variabile aleatoria su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, allora $Y = g(X)$ è ancora una variabile aleatoria sullo stesso spazio.

Così, se X è una variabile aleatoria, tale sono anche $|X|$, X^2 , $aX + b$ (dove $a, b \in \mathbb{R}$), e^X , perché le funzioni $|x|$, x^2 , $ax + b$, e^x sono funzioni continue.

Il problema che si intende risolvere è allora il seguente: conoscendo la distribuzione di X , è possibile risalire a quella di $Y = g(X)$? Per risolvere il problema, si osservi che se A è l'evento $A = \{x \in \mathbb{R} : g(x) \leq z\}$, allora $(X \in A)$ e $(Y \leq z)$ sono equivalenti, quindi

$$F_Y(z) = \mathbb{P}(Y \leq z) = \mathbb{P}(X \in A).$$

Questa formula ci permette il calcolo della funzione di ripartizione di Y in termini della distribuzione di X .

Esempio 2.4 Nel caso X sia una variabile aleatoria discreta, con distribuzione $\{(x_k, p_k)\}$, allora $Y = g(X)$ sarà necessariamente discreta, in quanto i suoi valori saranno dati da $g(x_1), g(x_2), \dots$, anche se non necessariamente distinti. Le rispettive probabilità si ottengono dalla formula

$$p_Y(y) = \sum_{\{x : g(x)=y\}} p_X(x),$$

per ogni possibile valore y .

△

Esempio 2.5 Sia X una variabile aleatoria con distribuzione uniforme in $[0, \mu]$, dove $\mu > 0$ è fissato, e sia $g(x) = -\lambda \log(x/\mu)$, per un qualche numero λ reale positivo. Definiamo $Y = g(X) = -\lambda \log(X/\mu)$; allora $(Y \leq t)$ corrisponde a $X \geq \mu e^{-t/\lambda}$, da cui

$$F_Y(t) = \mathbb{P}(X \geq \mu e^{-t/\lambda}) = 1 - e^{-t/\lambda}.$$

Si ottiene in particolare che Y ha distribuzione esponenziale di parametro $\frac{1}{\lambda}$.

△

2.3.2 Distribuzione della somma di due variabili aleatorie

Siano X e Y variabili aleatorie con densità congiunta f . In molti casi, ci interessa costruire una variabile aleatoria Z definita in termini di X e Y (diciamo $Z = \phi(X, Y)$, dove ϕ è una funzione continua) di cui

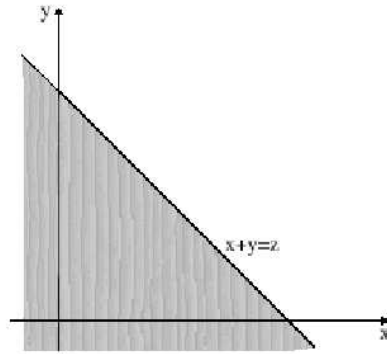


Figure 2.1: Distribuzione della somma di due variabili aleatorie

vogliamo calcolare la distribuzione. Per ogni $z \in \mathbb{R}$, l'evento $(Z \leq z)$ è equivalente all'evento $((X, Y) \in A_z)$, dove A_z è l'insieme dei punti del piano definito da

$$A_z = \{(x, y) : \phi(x, y) \leq z\}.$$

Nel caso $Z = X + Y$, l'insieme $A_z = \{(x, y) : x + y \leq z\}$ è il semipiano a sinistra della retta $x + y = z$, frontiera inclusa, si veda la figura precedente.

Supponiamo per semplicità che X e Y siano indipendenti; allora l'espressione

$$\int \int_{A_z} f(x, y) dy dx = \int_0^\infty \int_0^{z-x} f_X(x) f_Y(y) dy dx$$

rappresenta la probabilità di $(Z \leq z)$. Per ottenere la densità della variabile Z è necessario prendere la derivata dell'espressione precedente; si ottiene infine

$$f_Z(z) = \int_{\mathbb{R}} f_Y(z - x) \cdot f_X(x) dx. \quad (2.5)$$

Esempio 2.6 Nel caso X e Y possano assumere solo valori positivi, l'espressione (2.5) si semplifica:

$$f_Z(z) = \int_0^z f_Y(z - x) \cdot f_X(x) dx.$$

Inoltre, l'espressione (2.5) può essere adattata al caso di variabili aleatorie discrete X e Y , a valori interi, per cui $Z = X + Y$ assume valori interi anch'essa e vale

$$\mathbb{P}(Z = z) = \sum_{x \in \mathbb{Z}} \mathbb{P}(Y = z - x) \mathbb{P}(X = x), \quad z \in \mathbb{Z}.$$

△

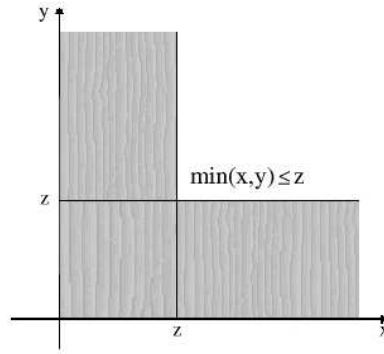


Figure 2.2: Distribuzione del minimo di due variabili aleatorie

Consideriamo in dettaglio anche la distribuzione del minimo di due variabili aleatorie. Siano X e Y variabili aleatorie positive; sia $Z = \min(X, Y)$: allora $A_z = \{(x, y) : x \geq 0, y \geq 0, \min(x, y) \leq z\}$ è la regione di piano evidenziata nella figura seguente.

Per ottenere $F_z(z) = \mathbb{P}(Z \leq z)$ conviene rifarsi al seguente calcolo:

$$\mathbb{P}(Z \leq z) = 1 - \mathbb{P}(Z > z) = 1 - \mathbb{P}(\min(X, Y) > z) = 1 - \mathbb{P}(X > z, Y > z) = 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z)$$

da cui

$$F_Z(z) = 1 - (1 - F_X(z))(1 - F_Y(z)), \quad z \geq 0. \quad (2.6)$$

Esempio 2.7 Siano X e Y distribuzioni esponenziali, di parametro λ e μ , rispettivamente. Sia $Z = \min(X, Y)$. Possiamo allora calcolare la distribuzione di Z tramite la (2.6):

$$F_Z(z) = 1 - e^{-(\lambda+\mu)z}, \quad z \geq 0.$$

Riconosciamo dunque che Z ha distribuzione esponenziale di parametro $\lambda + \mu$.

△

2.4 Media

Sia X una variabile aleatoria discreta, con distribuzione $\{(x_k, p_k), k = 0, 1, \dots\}$; diremo che X è *integrabile*, ovvero che ha *media finita*, se

$$\sum_{k=0}^{\infty} |x_k| p_k < \infty \quad (2.7)$$

ed in questo caso³, la media di X , che indichiamo con $\mathbb{E}(X)$, sarà uguale a

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} x_k p_k = \sum_{k=0}^{\infty} x_k \mathbb{P}(X = x_k). \quad (2.8)$$

³la condizione (2.7) serve a garantirsi che il risultato della serie in (2.8) sia indipendente dall'ordinamento dei termini. Se la condizione (2.7) non fosse verificata, la somma della serie in (2.8) potrebbe mutare secondo l'ordinamento prescelto. Quando la condizione (2.7) è verificata diremo che la serie in (2.8) è assolutamente convergente

I termini che compaiono in (2.8) non sono altro che i valori assunti da X moltiplicati per le loro rispettive probabilità.

Cambiando le serie in integrali, si ottiene la definizione di *media* per una variabile aleatoria continua X con densità di probabilità $f(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx,$$

a condizione che l'integrale converga assolutamente:

$$\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty.$$

Esempio 2.8 Sia X una variabile aleatoria con distribuzione

$$x_k = k, \quad p_k = \frac{1}{k(1+k)}, \quad k = 1, 2, \dots$$

X non ha media finita, dato che la condizione (2.7) non è verificata:

$$\sum_{k=1}^{\infty} |x_k| p_k = \sum_{k=1}^{\infty} \frac{1}{1+k} = +\infty.$$

△

Esempio 2.9 Sia X una variabile aleatoria con funzione densità di probabilità

$$f_X(x) = \frac{1}{2} e^{-|x|}, \quad -\infty < x < \infty.$$

Poiché

$$\begin{aligned} \int_{-\infty}^{\infty} |x| f_X(x) dx &= \frac{1}{2} \int_{-\infty}^0 (-x) e^x dx + \frac{1}{2} \int_0^{\infty} x e^{-x} dx \\ &= \int_0^{\infty} x e^{-x} dx = 1 < \infty, \end{aligned}$$

la speranza matematica di X esiste finita ed è data da

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{2} \int_{-\infty}^0 x e^x dx + \frac{1}{2} \int_0^{\infty} x e^{-x} dx = 0.$$

△

Sia ora $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione continua, $\vec{X} = (X_1, \dots, X_n)$ un vettore aleatorio (discreto) di lunghezza n e poniamo $Z = \phi(\vec{X})$. È facile convincersi che Z è una variabile aleatoria discreta, il cui valore di aspettazione è dato da

$$\mathbb{E}[Z] = \sum_{\vec{x}} \phi(\vec{x}) \mathbb{P}(\vec{X} = \vec{x}), \quad (2.9)$$

supposto che la serie a secondo membro verifichi

$$\sum_{\vec{x}} |\phi(\vec{x})| \mathbb{P}(\vec{X} = \vec{x}) < \infty.$$

Esempio 2.10 Sia X una variabile aleatoria con distribuzione di probabilità geometrica

$$x_k = k, \quad p_k = \frac{2}{3} \left(\frac{1}{3}\right)^{k-1}, \quad k = 1, 2, \dots$$

Poniamo $Y = g(X) = (-1)^X$. La distribuzione di Y è data da

$$\mathbb{P}(Y = -1) = \frac{3}{4}, \quad \mathbb{P}(Y = 1) = \frac{1}{4},$$

pertanto

$$\mathbb{E}(Y) = (-1)\frac{3}{4} + (1)\frac{1}{4} = -\frac{1}{2}.$$

Ma, senza utilizzare la distribuzione di Y , la (2.9) fornisce

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}[(-1)^X] = \sum_{k=1}^{\infty} (-1)^k \frac{2}{3} \left(\frac{1}{3}\right)^{k-1} = -\frac{2}{3} \sum_{k=0}^{\infty} \left(-\frac{1}{3}\right)^k \\ &= -\frac{2}{3} \frac{1}{1 - (-\frac{1}{3})} = -\frac{1}{2}. \end{aligned}$$

△

Proprietà della speranza matematica

Le seguenti proprietà sono di facile verifica:

1. siano X e Y variabili aleatorie integrabili, a e b numeri reali; allora

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

2. siano X e Y variabili aleatorie integrabili indipendenti; allora

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]. \quad (2.10)$$

3. se X è una variabile aleatoria discreta, che assume valori interi positivi, potremo scrivere

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

Se X è una variabile aleatoria continua, con densità $f(x)$, $X \geq 0$, risulta

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > x) dx.$$

Esempio 2.11 Sia (X, Y) un vettore aleatorio avente la seguente densità di probabilità:

$$f_{(X,Y)} = \begin{cases} 2 & 0 \leq x \leq 1, \ 0 \leq y \leq 1, \ x + y \leq 1, \\ 0 & \text{altrove.} \end{cases}$$

Determiniamo la media di $Z = X + Y$. Possiamo allora scrivere

$$\begin{aligned}\mathbb{E}[Z] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_{(X,Y)}(x,y) \, dy \, dx = \int_0^1 \int_0^{1-x} 2(x+y) \, dy \, dx \\ &= \int_0^1 2x(1-x) + (1-x)^2 \, dx = \int_0^1 1 - x^2 \, dx = \frac{2}{3}.\end{aligned}$$

Altrimenti, determiniamo le densità marginali di X e Y :

$$f_X(x) = \int_0^{1-x} 2 \, dy = 2(1-x), \quad 0 \leq x \leq 1,$$

e anche

$$f_Y(y) = \int_0^{1-y} 2 \, dx = 2(1-y), \quad 0 \leq y \leq 1.$$

Allora

$$\mathbb{E}[X] = \int_0^1 2x(1-x) \, dx = \frac{1}{3} = \mathbb{E}[Y],$$

quindi $\mathbb{E}[X + Y] = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$.

△

2.4.1 Momenti

Definiamo il **momento di ordine k della variabile aleatoria X** (centrato rispetto allo 0)

$$\mu_k = \mathbb{E}[X^k]$$

dove, scelto $\phi(x) = x^k$, $\mathbb{E}[\phi(X)]$ è il valore di aspettazione della variabile $\phi(X)$. Il momento di ordine 1: $\mu_1 = \mathbb{E}[X]$ **coincide con la media della variabile X** .

Se il momento viene invece calcolato intorno ad un punto a , si ottiene

$$\mu_k(a) = \mathbb{E}[(X - a)^k].$$

I momenti vengono spesso calcolati intorno alla media. Questi, detti *momenti centrali*, sono indicati con μ'_k e sono definiti da

$$\mu'_k = \mathbb{E}[(X - \mu)^k],$$

ed in particolare $\mu'_1 = 0$. Il **secondo momento intorno alla media è uguale alla varianza**

$$\mu'_2 = \sigma^2$$

mentre $\sigma = \sqrt{\mu'_2}$ è chiamata deviazione standard.

Si osservino, comunque, le seguenti proprietà

1. se X ha momento di ordine k finito, possiede anche tutti i momenti di ordine $\leq k$ finiti;
2. se X e Y hanno momento di ordine k finito, allora anche $X + Y$ ha momento di ordine k finito.
3. se X ha momento di ordine k finito, allora ha anche momento centrato di ordine k finito.

2.4.2 Varianza

Chiameremo *varianza* della variabile aleatoria X il momento centrato del secondo ordine:

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

Un diverso modo per calcolare la varianza di una variabile aleatoria è dato dalla formula

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

La radice quadrata aritmetica di σ^2 , indicata con σ , è detta scarto quadratico medio o deviazione standard

$$\sigma = \sqrt{\text{Var}[X]} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}.$$

Si osservi che lo scarto quadratico medio è espresso nella stessa unità di misura di X , mentre la varianza si esprime con il quadrato dell'unità di misura.

La varianza di una variabile aleatoria X è una quantità non negativa: $\text{Var}[X] \geq 0$ e vale l'uguaglianza se e solo se X assume valore costante con probabilità 1.

La varianza è una misura della dispersione intorno alla media dei valori assunti da una variabile aleatoria; più una variabile è concentrata intorno alla media, minore è la sua varianza. Per stimare la dispersione di una variabile aleatoria intorno alla sua media, è sovente utile il seguente classico risultato

Diseguaglianza di Chebyshev: per ogni $\varepsilon > 0$ si ha

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2}. \quad (2.11)$$

Si osservi infatti la seguente formula, di immediata deduzione a partire da (2.11):

$$\mathbb{P}(|X - \mathbb{E}[X]| \leq \lambda \sigma_X) \geq 1 - \frac{1}{\lambda^2}.$$

In particolare, per ogni variabile aleatoria che ammette momento secondo finito,

$$\mathbb{P}(|X - \mathbb{E}[X]| \leq 2\sigma_X) \geq 0,75$$

sicché la probabilità che X assuma valori nell'intervallo $(m - 2\sigma, m + 2\sigma)$ è almeno uguale a 0,75. Con $\lambda = 3$, si ottiene che la probabilità di assumere valori nell'intervallo $(m - 3\sigma, m + 3\sigma)$ è almeno uguale a $8/9 \simeq 0,89$.

1. Se $Y = aX + b$, dove X è una variabile aleatoria con varianza finita e $a \neq 0$, b sono numeri reali, allora

$$\text{Var}[Y] = a^2 \text{Var}[X].$$

2. La varianza della somma di due variabili aleatorie indipendenti X e Y è pari alla somma delle varianze:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y], \quad \text{se } X \text{ e } Y \text{ sono indipendenti.}$$

Indichiamo con $\text{Cov}(X, Y)$ la *covarianza* di X e Y :

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

Date due variabili aleatorie X e Y , si ottiene

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}(X, Y).$$

In generale, per una somma di variabili aleatorie, la varianza è espressa dalla formula

$$\text{Var}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i^2 \text{Var}[X_i] + 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}(X_i, X_j).$$

La covarianza di due variabili X e Y fornisce una misura di quanto esse siano correlate, e la quantità

$$\rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sigma_X \sigma_Y}$$

dove σ_X, σ_Y sono le deviazioni standard di X e Y , è detta *coefficiente di correlazione* di X e Y .

Diremo che X e Y sono *non correlate* se $\rho_{X,Y} = 0$, ossia se $\text{Cov}(X,Y) = 0$. Variabili indipendenti sono, in particolare, non correlate; il viceversa in generale è falso.

Giova infine ricordare che un valore di ρ prossimo a zero non significa che tra le due variabili aleatorie manchi una qualche relazione, anche funzionale, ma solo che è assente (o debole) una relazione di tipo lineare.

2.5 Funzione generatrice dei momenti

La funzione generatrice dei momenti (f.g.m.) è una trasformazione integrale che associa ad una distribuzione di probabilità un'altra funzione (quando esiste) con la quale è possibile operare per risolvere in modo più agevole alcuni problemi.

Definizione. Si dice funzione generatrice dei momenti (f.g.m.) di una v.a. X la funzione del parametro reale t definita da

$$M_X(t) = \mathbb{E}[e^{tX}] = \begin{cases} \sum_{k=1}^{\infty} e^{tx_k} p_k, & X \text{ v.a. discreta} \\ \int_{\mathbb{R}} e^{tx} f_X(x) dx, & X \text{ continua,} \end{cases}$$

se la serie (o l'integrale) esiste finito in un intorno completo dell'origine, ossia se esiste $\epsilon > 0$ e $M_X(t)$ esiste finito per ogni $-\epsilon < t < \epsilon$.

Naturalmente si ha sempre $M_X(0) = 1$ essendo $M_X(0) = \mathbb{E}[e^{0X}] = \mathbb{E}[1]$. Per l'esistenza della f.g.m. si richiede però che l'integrale o la serie esista finito in un intorno completo dell'origine, e questo non è sempre il caso.

Esempio 1. Le seguenti distribuzioni non ammettono f.g.m.: **(a)** X v.a. continua, con densità $f_X(x) = \frac{1}{x^2}$ per $x \geq 1$ e 0 altrove. Si dimostra che per $t \leq 0$

$$\int_1^{\infty} e^{tx} \frac{1}{x^2} dx \leq 1,$$

ma per $t > 0$

$$\int_1^{\infty} e^{tx} \frac{1}{x^2} dx = +\infty,$$

quindi non esiste alcun intorno dell'origine in cui l'integrale esiste finito.

(b) X v.a. discreta, con distribuzione $\{(k, p_k = \frac{1}{k(k+1)}), k = 1, 2, \dots\}$. Anche in questo caso, la serie vale $+\infty$ per ogni $t > 0$ e quindi non esiste la f.g.m.

Esempio 2. Sia X una v.a. con distribuzione esponenziale di parametro $\lambda > 0$. Allora la f.g.m. esiste, ed è ben definita per ogni $t < \lambda$:

$$M_X(t) = \begin{cases} \frac{\lambda}{\lambda - t}, & t < \lambda \\ +\infty, & t \geq \lambda. \end{cases}$$

Il seguente risultato mostra che, se una v.a. possiede f.g.m., la conoscenza di questa mi garantisce la conoscenza della distribuzione. A parole, significa che se due v.a. hanno la stessa f.g.m., allora devono essere equidistribuite.

Proposizione 1. Se X e Y sono v.a. con f.g.m. che esiste finita, allora

$$M_X(\cdot) = M_Y(\cdot) \iff F_X(\cdot) = F_Y(\cdot).$$

A questo proposito, rimandiamo alla tabella 1, in cui sono riportate le f.g.m. di alcune importanti distribuzioni. Un altro importante risultato, che spiega anche il nome adottato per questa funzione, è contenuto nella seguente

Proposizione 2. Se X ammette f.g.m. $M_X(t)$ allora essa è sviluppabile in serie di Taylor intorno all'origine

$$M_X(t) = \sum_{n=0}^{\infty} \frac{d^n M_X}{dt^n} \Big|_{t=0} \frac{t^n}{n!}, \quad -\epsilon < t < \epsilon$$

ed inoltre vale

$$\frac{d^n M_X}{dt^n} \Big|_{t=0} = \mathbb{E}[X^n], \quad n = 0, 1, 2, \dots$$

Se dunque una v.a. X ammette f.g.m., allora ha anche tutti i momenti finiti, e questi si possono calcolare a partire da quella.

Esempio 3. Se X ha distribuzione esponenziale di parametro λ , allora la f.g.m. esiste per $t < \lambda$ e vale $M_X(t) = \frac{\lambda}{\lambda - t}$; sviluppando questa funzione in serie si ottiene

$$M_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - (t/\lambda)} = \sum_{k=0}^{\infty} \left(\frac{t}{\lambda}\right)^k = \sum_{k=0}^{\infty} \frac{k!}{\lambda^k} \frac{t^k}{k!}$$

quindi $\mathbb{E}[X^k] = \frac{k!}{\lambda^k}$.

Nome della distribuzione	distribuzione o densità	f.g.m.
Binomiale	$\binom{N}{x} p^x (1-p)^{N-x}$ $N \geq 1, p \in (0, 1), x = 0, 1, \dots, N$	$(1 - p + pe^t)^N$ $t \in \mathbb{R}$
Poisson	$e^{-\lambda} \frac{\lambda^x}{x!}$ $\lambda > 0, x \in \mathbb{N}$	$e^{\lambda(e^t - 1)}$ $t \in \mathbb{R}$
Geometrica	pq^{x-1} $p + q = 1, x = 1, 2, \dots$	$\frac{pe^t}{1 - qe^t}$ $t < -\log(q)$
Uniforme continua	$\frac{1}{b-a}$ $a < x < b$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$ $t \in \mathbb{R}$
Gaussiana	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$, $x \in \mathbb{R}$	$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$ $t \in \mathbb{R}$
esponenziale	$\lambda e^{-\lambda t}, t > 0$ $t > 0$	$\frac{\lambda}{\lambda - t}, t < \lambda$ $t < \lambda$

Table 2.1: Alcune distribuzioni notevoli.

Passiamo a considerare come si comporta la f.g.m. nel caso di trasformazioni di v.a..

Proposizione 3. Se X ammette f.g.m. $M_X(t)$ e se $Y = \mu + \sigma X$, allora

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{(\mu + \sigma X)t}] = e^{\mu t} M_X(\sigma t).$$

Proposizione 4. Se X e Y sono v.a. indipendenti, che ammettono f.g.m. $M_X(t)$ e $M_Y(t)$, allora la v.a. $X + Y$ ammette f.g.m. e vale $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Esempio 4.

(a) Se X ha distribuzione gaussiana standard, allora $M_X(t) = e^{t^2/2}$; prendendo le derivate di questa funzione si ottiene

$$\frac{d}{dt}M_X(t) = te^{t^2/2}, \quad \frac{d^2}{dt^2}M_X(t) = (1 + t^2)e^{t^2/2}$$

e calcolando il valore in $t = 0$ otteniamo i primi momenti:

$$\mathbb{E}[X] = 0, \quad V[X] = \mathbb{E}[X^2] = 1.$$

(b) Se $Y = \mu + \sigma X$, dove $X \sim \mathcal{N}(0, 1)$, allora $M_Y(t) = e^{\mu t} e^{\frac{1}{2}(\sigma t)^2}$, da cui riconosciamo un Y una legge gaussiana $\mathcal{N}(\mu, \sigma^2)$. Si ottiene allora

$$\mathbb{E}[Y] = \mu + \sigma \mathbb{E}[X] = \mu, \quad V[Y] = \sigma^2 V[X] = \sigma^2.$$

Quindi i parametri μ e σ^2 individuano, rispettivamente, media e varianza della distribuzione gaussiana $\mathcal{N}(\mu, \sigma^2)$.

(c) Se X e Y sono v.a. indipendenti con distribuzione gaussiana $\mathcal{N}(\mu, \sigma^2)$ e $\mathcal{N}(\nu, \tau^2)$, allora

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2} e^{t\nu + \frac{1}{2}\tau^2 t^2} = \exp\left(t(\mu + \nu) + \frac{1}{2}t^2(\sigma^2 + \tau^2)\right)$$

da cui riconosciamo una legge gaussiana $\mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$. In altre parole, la somma di gaussiane indipendenti è ancora gaussiana.

2.6 Alcune leggi legate alla statistica

2.6.1 Leggi χ^2

Consideriamo una successione di v.a. X_1, \dots, X_n indipendenti ed equidistribuite, di legge gaussiana standard $\mathcal{N}(0, 1)$; posto $Y = \sum_{i=1}^n X_i^2$, Y si dice avere legge χ^2 a n gradi di libertà, e si indica con $\chi^2(n)$. Sebbene le leggi χ^2 abbiano densità che può essere calcolata, per i nostri scopi è sufficiente considerare i valori che troviamo sulle tavole.

- Possiamo calcolare i primi momenti, che risultano

$$\mathbb{E}[Y] = \sum_{i=1}^n \mathbb{E}[X_i^2] = n, \quad \text{Var}[Y] = 2n.$$

- Se n è grande, diciamo $n > 30$, risulta che la legge $\chi^2(n)$ è approssimativamente normale $\mathcal{N}(n, 2n)$.
- Se $Y \sim \chi^2(n)$ e $Z \sim \chi^2(m)$ sono v.a. indipendenti, allora $Y + Z \sim \chi^2(n + m)$.

Esempio 2.12 (Quantili delle leggi χ^2) I quantili della legge χ^2 a n gradi di libertà sono indicati con $x_\alpha(n)$ e sono forniti sulle tavole. Ad esempio, si trova

$n \setminus \alpha$	0.01	0.025	0.05	0.95	0.975	0.99
10	2.5582	3.2470	3.9403	18.307	20.483	23.209
20	8.2604	9.5908	10.8508	31.410	34.170	37.566

da cui si ottiene che (se $Y \sim \chi^2(20)$) $\mathbb{P}(Y < 9.59) = 0.025$, e $\mathbb{P}(Y > 24.17) = 1 - 0.975 = 0.025$.

È immediato osservare che $Y \geq 0$ non ha distribuzione simmetrica rispetto alla media. Per ottenere un intervallo bilatero, ossia $\mathbb{P}(a < Y < b) = \alpha$, si cerca un intervallo a code uguali, cioè $\mathbb{P}(Y < a) = \mathbb{P}(Y > b) = \frac{1-\alpha}{2}$. Così, per $Y \sim \chi^2(10)$, si ha $\mathbb{P}(Y \in (3.247, 20.483)) = 0.95$.

△

Esempio 2.13 Consideriamo una legge χ^2 con $n = 50$ gradi di libertà; allora per calcolare un intervallo unilatero $\mathbb{P}(Y \leq x_\alpha) = \alpha$ si ottiene

$$\mathbb{P}\left(\frac{Y - n}{\sqrt{2n}} \leq \frac{x_\alpha - n}{\sqrt{2n}}\right) = \alpha \quad \Longleftrightarrow \quad \frac{x_\alpha - n}{\sqrt{2n}} = \phi_\alpha$$

(ϕ_α è il quantile di ordine α per la legge normale), da cui

$$x_\alpha = n + \sqrt{2n}\phi_\alpha.$$

△

La seguente proposizione riassume i punti principali che ci interessano.

Lemma 2.1 Siano X_1, \dots, X_n v.a. indipendenti ed equidistribuiti con legge $\mathcal{N}(\mu, \sigma^2)$.

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

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

(c) la varianza campionaria $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$ è indipendente da \bar{X} e si ha $\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1)$
(quest'ultima affermazione è equivalente a (b)).

Esempio 2.14 Una fabbrica di lastre di vetro sa che lo spessore delle sue lastre segue una legge normale, con deviazione standard $\sigma = 1.32 \times 10^{-4}$. Da ogni partita di lastre viene estratto un campione di $n = 25$ pezzi, di cui si misura lo spessore; qual è la probabilità che la varianza campionaria superi il valore critico di $\tilde{\sigma}^2 = 2 \times 10^{-4}$?

In base alla proprietà (c) del lemma, $\frac{(n-1)}{\sigma^2} S^2$ ha distribuzione $\chi^2(n-1)$, quindi la probabilità di superare il valore critico $\tilde{\sigma}^2$ è

$$\begin{aligned} \mathbb{P}(S^2 > \tilde{\sigma}^2) &= \mathbb{P}\left(Y > \frac{(n-1)\tilde{\sigma}^2}{\sigma^2}\right) \quad \text{con } Y \sim \chi^2(24) \\ &= \mathbb{P}(Y > 36,364) \simeq 0.9 \quad (\text{dalle tavole}). \end{aligned}$$

△

2.6.2 Leggi t di Student

Siano $X \sim \mathcal{N}(0, 1)$ e $Y \sim \chi^2(n)$ due v.a. indipendenti; diremo legge t di Student a n gradi di libertà la legge della v.a. $T = \frac{X}{\sqrt{Y/n}}$. Anche in questo caso, è possibile calcolare esplicitamente la densità, ma a noi interesserà soprattutto conoscere i quantili di questa legge, che si possono ricavare dalle apposite tavole.

Uno studio della legge $t(n)$ mostra che essa è simmetrica e inoltre converge, per $n \rightarrow \infty$, ad una legge Gaussiana standard. Le tavole mostrano i quantili per diversi valori di n , ma non tutti; per $n > 120$ si utilizzano i quantili della legge normale, mentre se (ad esempio) $n = 87$, e ci accontentiamo di una minore precisione (ad esempio, limitata a due cifre decimali), possiamo approssimare $t_\alpha(87)$ con i valori tabulati per l'intero n più vicino (in questo caso, 90).

Lemma 2.2 Siano X_1, \dots, X_n v.a. indipendenti ed equidistribuite, con legge $\mathcal{N}(\mu, \sigma^2)$. Allora, date \bar{X} la media campionaria e S^2 la varianza campionaria

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

e posto $T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}$, T ha distribuzione $t(n-1)$ di Student con $n-1$ gradi di libertà.

Dimostrazione.

Osserviamo che

$$T = \frac{\frac{\sqrt{n}}{\sigma} \bar{X} - \mu}{\sqrt{\frac{n-1}{n-1} S^2 / \sigma}}$$

$\xi := \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu)$ ha distribuzione normale standard; sotto radice si ha la v.a. $\eta := \sqrt{(n-1)S^2/\sigma^2}$ che ha legge $\chi^2(n-1)$ ed inoltre è indipendente da ξ ; allora $T = \frac{\xi}{\sqrt{\eta/(n-1)}}$ ha, per definizione, legge $t(n-1)$.

Chapter 3

Ripetizione di esperimenti indipendenti

Consideriamo un esperimento (come il lancio di un dado) che abbia un numero finito o numerabile di possibili esiti. Possiamo allora esprimere il risultato di questo esperimento con il valore di una variabile aleatoria X . Supponiamo che l'esperimento sia ripetuto n volte. Il risultato globale corrisponde ad osservare le variabili aleatorie X_1, X_2, \dots, X_n , dove X_i è il risultato dell'esperimento i -esimo. Se questo esperimento viene ripetuto nelle stesse condizioni, è naturale richiedere che le distribuzioni di probabilità delle variabili aleatorie siano identiche. Se inoltre supponiamo che il ripetersi dell'esperimento non condizioni l'esito di una nuova prova, è naturale richiedere che le variabili aleatorie X_i siano tra loro indipendenti.

3.1 Schema di Bernoulli

Il più semplice esempio è quello di un esperimento in cui i possibili esiti sono solo due: successo o perdita, ad esempio il lancio di una moneta, oppure l'estrazione di un asso da un mazzo di carte ben mischiato, ecc. Indichiamo con X la distribuzione di un tale esperimento, $p = \mathbb{P}(X = 1)$ la probabilità di vittoria, $q = 1 - p$ la probabilità di sconfitta. Una variabile aleatoria con tale distribuzione viene detta variabile *Bernoulliana*. La ripetizione di n tentativi indipendenti conduce alla successione X_1, X_2, \dots, X_n di variabili aleatorie indipendenti ed equi-distribuite. Parleremo allora di schema di Bernoulli di parametro p .

Il risultato di n esperimenti ripetuti di Bernoulli è espresso dal vettore aleatorio $\vec{X} = (X_1, \dots, X_n)$: l'informazione contenuta nel vettore è completa, e ci descrive in quali esperimenti si è avuto un successo e in quali una perdita. Spesso, tuttavia, ciò che ci interessa sono alcuni elementi dell'esperimento. Riportiamo alcune variabili aleatorie riconducibili allo schema di Bernoulli.

- Consideriamo il numero totale di successi S_n ottenuti in n tentativi. Si ha $S_n = X_1 + \dots + X_n$ e S_n ha distribuzione binomiale di parametri n e p .
- Sia G_1 il numero di sconfitte prima di ottenere il primo successo ($G_1 = 0$ se si ottiene successo al primo tentativo). Allora G ha distribuzione geometrica di parametro p . La stessa distribuzione compete alla variabile G_k che conta il numero di sconfitte tra la $k - 1$ -esima vittoria e la k -esima. Si dimostra che le G_k sono tra loro indipendenti ed equidistribuite.
- Consideriamo ora il numero di tentativi $T(n)$ necessari per ottenere n successi. Calcoliamo la distribuzione di $T(n)$.

1. il tipo di sequenza che realizza l'evento $T(n) = k$ è dato da

$$T(n) = k \quad \text{se e solo se} \quad \vec{X} = (\underbrace{0, \dots, 0, 1, 0, \dots, 0, \dots, 0}_{k-1 \text{ termini con } n-1 \text{ "1"}}, 1, *, \dots)$$

dove con $*$ indichiamo un qualunque valore 0 o 1.

2. La probabilità di un singolo evento di questo tipo è $p^n(1-p)^{k-n}$.
3. Il numero di sequenze di questo tipo è pari al numero di sottoinsiemi di $n-1$ elementi tra $k-1$ elementi, ovvero a C_{n-1}^{k-1} , ossia $\binom{k-1}{n-1}$.
4. Di conseguenza, la probabilità di $T(n) = k$ è data dalla formula

$$\mathbb{P}(T(n) = k) = \binom{k-1}{n-1} p^n q^{k-n}$$

per ogni $k \geq n$, e zero altrimenti.

- Le variabili $T(n)$ e G_k introdotte in precedenza sono legate dalla relazione

$$T(n) = G_1 + \dots + G_n.$$

Ricordando che la media di una variabile aleatoria con distribuzione geometrica è q/p , si ottiene

$$\mathbb{E}[T(n)] = n/p.$$

Esempio 3.1 Consideriamo uno schema di Bernulli X_1, X_2, \dots , in cui sia p la probabilità di avere un successo. Definiamo le variabili aleatorie T, U rispettivamente come il primo ed il secondo istante in cui si osserva un successo, ovvero:

$$T \equiv \inf\{n \geq 1 : X_n = 1\} \text{ e } U \equiv \inf\{n \geq T + 1 : X_n = 1\}$$

Calcolare:

- (a) la distribuzione congiunta di T e U e le loro distribuzioni marginali;
- (b) definita la nuova variabile aleatoria $V \equiv U - T$, calcolare la distribuzione congiunta di T e V e dire se le due variabili sono indipendenti;
- (c) determinare la distribuzione marginale di V e riconoscerne la legge.

△

3.1.1 Formule approssimate

Consideriamo una variabile aleatoria binomiale S_n , con parametro n grande. Ricordiamo che S_n determina la probabilità che in n tentativi si ottengano k successi, $k \leq n$, e che

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n.$$

Il calcolo della precedente probabilità diventa rapidamente oneroso al crescere di n . È sembrato quindi utile ricercare delle formule approssimate, capaci di rendere agevole tale calcolo ed accettabile l'errore derivante dall'approssimazione.

La formula di Poisson

Fissata la costante positiva λ , si ponga per ogni $n > \lambda$

$$p = p_n = \frac{\lambda}{n}. \quad (3.1)$$

Si consideri la successione delle variabili aleatorie $S(n)$ con distribuzione binomiale di parametro n e p_n (quindi, al variare di n , $S(n)$ non è il risultato di uno stesso schema di Bernoulli!)

$$P_k := \mathbb{P}(S(n) = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (3.2)$$

Quando n tende a infinito, restando valida la posizione (3.1), la successione di leggi di probabilità definite in (3.2) converge verso una legge di Poisson di parametro λ

$$Q_k = e^{-\lambda} \frac{\lambda^k}{k!}$$

nel senso che

$$\lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

A titolo esemplificativo, riportiamo i valori di P_k dedotti dalla formula (3.1) con $n = 100$, $p = 0.05$ e $n = 500$, $p = 0.01$, rispettivamente, ed i valori approssimati Q_k ottenuti tramite la formula (3.2), con $\lambda = np = 5$:

k	n=100, p=0.05	n=500, p=0.01	$\lambda=5$
0	0.0059205	0.0065705	0.0067379
5	0.180018	0.176351	0.175467
10	0.0167159	0.0178585	0.0181328
15	0.0000988	0.0001442	0.0001572
20	8.442×10^{-8}	2.143×10^{-7}	2.641×10^{-7}
25	1.543×10^{-11}	8.818×10^{-11}	1.295×10^{-10}

Formula di Gauss

Si riprenda la formula

$$P_k = \binom{n}{k} p^k q^{n-k} \quad (3.3)$$

che fornisce la probabilità di ottenere k successi su n prove nello schema di Bernoulli. Consideriamo la corrispondente variabile standardizzata

$$S_n^* = \frac{S_n - np}{\sqrt{npq}};$$

osserviamo che s_n^* assume valori $x_k^* = \frac{k-np}{\sqrt{npq}}$ per $k = 0, 1, \dots, n$.

Proposizione.

Per n sufficientemente grande, per ogni $k = 0, 1, \dots, n$, vale l'approssimazione

$$\mathbb{P}(S_n = k) = \mathbb{P}(S_n^* = x_k^*) \simeq \frac{1}{\sqrt{2\pi npq}} \exp(-(x_k^*)^2/2) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right). \quad (3.4)$$

Si consideri la formula di Stirling per l'approssimazione del fattoriale $n!$, data da

$$n! \sim \sqrt{2\pi n} n^n e^{-n}, \quad n \rightarrow \infty,$$

dove $a_n \sim b_n$ significa che

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1.$$

Sostituendo in (3.3) si ottiene

$$\begin{aligned} P_k &= \frac{n!}{k!(n-k)!} p^k q^{n-k} \simeq \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi(n-k)} (n-k)^{n-k} e^{-(n-k)}} p^k q^{n-k} \\ &\simeq \frac{1}{\sqrt{2\pi n} (k/n)^{k+1/2} (1-k/n)^{n-k+1/2}} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} \frac{1}{(k/np)^{k+1/2} [(n-k)/nq]^{n-k+1/2}}. \end{aligned}$$

Consideriamo il secondo termine dell'eguaglianza precedente:

$$A = \frac{1}{(k/np)^{k+1/2} [(n-k)/nq]^{n-k+1/2}}.$$

Prendendo il logaritmo di ambo i membri, si ricava

$$\log(A) = -(k + \frac{1}{2}) \log(k/np) - (n - k + \frac{1}{2}) \log[(n - k)/nq].$$

Ricordando che $n = np + nq$, si ha

$$\log(A) = -(np + (k - np) + \frac{1}{2}) \log(1 + \frac{k - np}{np}) - (nq - (k - np) + \frac{1}{2}) \log(1 - \frac{k - np}{nq}). \quad (3.5)$$

Facciamo tendere n all'infinito con la posizione

$$x = \frac{k - np}{\sqrt{npq}} \quad \text{fissato.} \quad (3.6)$$

Possiamo allora scegliere n abbastanza grande perché sia

$$\left| \frac{k - np}{np} \right| < 1, \quad \left| \frac{k - np}{nq} \right| < 1;$$

sviluppiamo allora il logaritmo in serie di Taylor intorno a 1:

$$\log(1 + t) = t - \frac{1}{2}t^2 + R_2(t);$$

sostituendo in (3.5)

$$\begin{aligned} \log(A) = & - \left[np \frac{k - np}{np} + (k - np) \frac{k - np}{np} + \frac{1}{2} \frac{k - np}{np} - nq \frac{k - np}{nq} + (k - np) \frac{k - np}{nq} - \frac{1}{2} \frac{k - np}{nq} \right. \\ & \left. - \frac{1}{2} np \left(\frac{k - np}{np} \right)^2 - \frac{1}{2} nq \left(\frac{k - np}{nq} \right)^2 + \delta \right] \end{aligned}$$

dove δ contiene tutti i termini del prodotto che non abbiamo esplicitato. Semplificando l'espressione precedente e ricordando la posizione (3.6) si ottiene

$$\log(A) = -\frac{x^2}{2} + \delta$$

ed inoltre si può dimostrare che $\delta \rightarrow 0$ per $n \rightarrow \infty$.

Il risultato appena stabilito è noto come *teorema di De Moivre - Laplace* e la formula (3.4) è la *formula di Gauss*. La conclusione va interpretata come segue: data una variabile binomiale, la probabilità che

si manifesti un particolare valore k di questa è tanto più prossima a quella deducibile dalla formula di Gauss (3.4) tanto più n cresce, restando p fissato.

Osservazione.

Il risultato di De Moivre - Laplace appena stabilito rappresenta una approssimazione accettabile della formula di Bernoulli per valori non grandi di n solo se $p = q$; se invece p e q sono molto diversi tra loro, l'approssimazione diviene accettabile solo per n molto grande; in tal caso, è conveniente utilizzare la formula di Poisson.

3.1.2 Probabilità di un valore compreso entro limiti assegnati

Nel paragrafo precedente ci siamo occupati del calcolo della probabilità che compete all'evento $(S_n = k)$, in uno schema di Bernoulli di parametro p . È però interessante, nelle applicazioni, anche il problema di calcolare la probabilità che compete all'evento $(a < S_n \leq b)$, fissati a e b .

Da un punto di vista teorico, non vi sono difficoltà; si tratta infatti di sommare le probabilità di eventi incompatibili, per cui la probabilità cercata è

$$\mathbb{P}(S_n = a + 1) + \cdots + \mathbb{P}(S_n = b);$$

il vero problema consiste nell'escogitare un procedimento comodo per il calcolo di questa espressione.

Per semplicità di esposizione, consideriamo il caso di intervalli centrati nella media; sia

$$\mathbb{P}(-\lambda\sigma \leq S_n - np \leq \lambda\sigma) = \mathbb{P}(|S_n - np| \leq \lambda\sigma)$$

la probabilità che, nello schema di Bernoulli, si manifesti uno scarto dalla media superiore a $\lambda\sigma$, dove $\sigma = \sqrt{npq}$ è lo scarto quadratico medio. Vogliamo mostrare che al crescere di n si ha

$$\mathbb{P}(-\lambda\sigma \leq S_n - np \leq \lambda\sigma) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-x^2/2} dx. \quad (3.7)$$

Consideriamo la variabile standardizzata

$$S_n^* = \frac{S_n - np}{\sqrt{npq}};$$

abbiamo visto come S_n^* abbia distribuzione che soddisfa la relazione

$$\mathbb{P}(S_n^* = x^*) \simeq \frac{1}{\sqrt{2\pi npq}} \exp[-(x^*)^2/2],$$

che possiamo anche esprimere nella forma

$$\sqrt{npq} \mathbb{P}(S_n^* = x^*) \simeq \frac{1}{\sqrt{2\pi}} \exp[-(x^*)^2/2]. \quad (3.8)$$

Per calcolare la probabilità che si manifesti un qualunque valore di S_n^* compreso tra $-\lambda$ e λ , indichiamo con k_1 e k_2 i valori di k che corrispondono al minimo e al massimo tra i valori che x^* assume all'interno di tale intervallo:

$$\sum_{k=k_1}^{k_2} \mathbb{P}(S_n^* = x_k^*).$$

Notiamo poi che $x_k^* = \frac{(k - np)}{\sqrt{npq}}$, da cui è anche $(x_{k+1}^* - x_k^*) = \frac{1}{\sqrt{npq}}$, ossia

$$\sum_{k=k_1}^{k_2} \mathbb{P}(S_n^* = x_k^*) = \sum_{k=k_1}^{k_2} (x_{k+1}^* - x_k^*) \sqrt{npq} \mathbb{P}(S_n^* = x_k^*) \simeq \sum_{k=k_1}^{k_2} (x_{k+1}^* - x_k^*) \frac{1}{\sqrt{2\pi}} \exp[-(x_k^*)^2/2].$$

Se consideriamo la funzione

$$\varphi_n(x) = \frac{1}{\sqrt{2\pi}} \exp[-x_k^2/2], \quad x_k \leq x < x_{k+1},$$

l'espressione precedente si può esprimere come integrale

$$\int_{x_{k_1}}^{x_{k_2+1}} \frac{1}{\sqrt{2\pi}} \varphi_n(x) dx.$$

Al tendere di n all'infinito, si avrà che φ_n converge alla funzione $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp[-x^2/2]$; dato che x^* varia di $1/\sqrt{npq}$ quando k varia di 1, si ottiene anche che $x_{k_1}^* \rightarrow -\lambda$, $x_{k_2+1}^* \rightarrow \lambda$, quindi la probabilità cercata converge all'integrale

$$\int_{-\lambda}^{\lambda} \frac{1}{\sqrt{2\pi}} \exp[-x^2/2] dx.$$

3.2 Teorema limite centrale

L'approssimazione di De Moivre - Laplace è solo un caso particolare di un risultato generale. Per introdurre questo, si inizia con l'osservare che S_n rappresenta la somma di variabili aleatorie indipendenti, tutte con la stessa distribuzione di Bernoulli, aventi media e varianza finita. Allora la formula di De Moivre - Laplace (3.7) può essere espressa dicendo che la variabile aleatoria

$$\frac{\sum_{k=1}^n X_k - n\mathbb{E}(X)}{\sqrt{n\text{Var}[X]}}$$

è approssimativamente distribuita secondo la legge Gaussiana $\mathcal{N}(0, 1)$.

Ci possiamo allora aspettare che una somma di variabili aleatorie indipendenti, sotto precise condizioni, sia distribuita secondo una legge Gaussiana. È questo il contenuto del teorema limite centrale.

Dobbiamo preliminarmente introdurre cosa intendiamo con convergenza in distribuzione.

Definizione

Si dice che la successione $(X_n)_{n \geq 0}$ converge in distribuzione alla variabile aleatoria X se la successione delle funzioni di ripartizione F_n converge verso la funzione di ripartizione F di X in ogni punto di continuità di quest'ultima:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{per } x \text{ punto di continuità di } F.$$

Si osservi che se X_n, X sono variabili aleatorie discrete, che assumono gli stessi valori $\{y_1, y_2, \dots\}$ allora la convergenza in distribuzione di X_n verso X si verifica se e solo se

$$\mathbb{P}(X_n = y_i) \rightarrow \mathbb{P}(X = y_i) \quad \text{per ogni } i, \text{ quando } n \rightarrow \infty.$$

Possiamo ora enunciare in maniera precisa quanto osservato nei precedenti paragrafi.

1. La successione di variabili aleatorie $(X_n)_{n \geq \lambda}$ aventi distribuzione binomiale $B(n, \lambda/n)$ convergono in distribuzione ad una legge di Poisson di parametro λ .
2. Il teorema di De Moivre - Laplace afferma che la successione di variabili aleatorie S_n^* converge, in distribuzione, ad una variabile aleatoria con legge Gaussiana $\mathcal{N}(0, 1)$.

Possiamo ora enunciare il teorema centrale.

Teorema centrale

Sia $(X_n)_{n \geq 1}$ una successione di variabili aleatorie indipendenti, equidistribuite, aventi media μ e varianza $\sigma^2 > 0$ finite. Allora la variabile

$$Z_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma^2}}$$

converge in distribuzione ad una variabile aleatoria con legge Gaussiana standard $\mathcal{N}(0, 1)$.

La generalità di questo teorema è notevole. Le variabili aleatorie X_n possono essere discrete, oppure continue o nessuno dei due tipi. Inoltre, la conclusione resta vera anche se non esistono i momenti di ordine superiore a 2. Un'altra sorprendente affermazione del teorema è che la distribuzione limite è indipendente dalla legge di X_n (supposto che le ipotesi del teorema siano soddisfatte).

Si può dimostrare che il teorema centrale rimane vero (cioè che S_n^* converge in distribuzione ad una legge Gaussiana standard) se S_n^* è una variabile aleatoria che si può scrivere come somma di un numero crescente di variabili aleatorie indipendenti, tutte “piccole” (in un senso da precisare), anche senza l'ipotesi che siano equidistribuite. Per questo motivo, ad esempio, si assume che un errore di misurazione segua una legge normale: in assenza di errori sistematici, è plausibile pensare che la discrepanza tra il valore vero e quello misurato sia la risultante di tanti piccoli errori che si sono sovrapposti.

3.3 Legge dei grandi numeri

Media campionaria

Introdurremo il concetto di convergenza in probabilità col classico teorema di Bernoulli.

Consideriamo uno schema di Bernoulli di parametro p e sia $\{X_k\}$ la successione di variabili aleatorie indipendenti, aventi tutte la stessa distribuzione, che rappresentano l'esito dell'esperimento k -esimo. Per ogni numero naturale $n \geq 1$ chiamiamo *media campionaria* di X_1, \dots, X_n la nuova variabile aleatoria

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Ovviamente, le variabili aleatorie X_k hanno la stessa media $\mu = p$ e la stessa varianza $\sigma^2 = p(1 - p)$. Notiamo che

1. anche il valore atteso della media campionaria \bar{X}_n è uguale a μ ;
2. la varianza di \bar{X}_n è uguale a $\frac{\sigma^2}{n}$.

Si noti che la varianza della media campionaria diventa sempre più piccola – più precisamente, tende a zero – al tendere di n all'infinito. Intuitivamente ciò significa che quanto più è ampio il campione (cioè quanto più è grande N) tanto più i valori della media campionaria sono concentrati intorno al valore atteso (che è μ).

La disuguaglianza di Chebyshev fornisce allora, per ogni $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Poiché l'ultimo termine, per ε fissato ed n tendente all'infinito converge a zero, per ogni $\eta > 0$ si può determinare un valore n sufficientemente grande tale che

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \eta.$$

Da ciò discende che per n sufficientemente grande, si avrà

$$\mathbb{P}(|\bar{X}_n - \mu| \leq \varepsilon) \geq 1 - \eta.$$

In altre parole, si potrà determinare n tanto grande da far sì che la probabilità della disuguaglianza $|\bar{X}_n - \mu| \leq \varepsilon$ sia prossima ad 1 quanto si vuole.

Questo fatto è noto come *teorema di Bernoulli* ed è il punto di partenza di numerosi altri teoremi in probabilità e statistica.

Definizione

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in probabilità alla variabile aleatoria X se

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0 \quad \text{per ogni } \varepsilon > 0.$$

Legge dei grandi numeri

Consideriamo una successione $\{X_n\}$ di variabili aleatorie e sia S_n la successione delle somme parziali

$$S_n = (X_1 + \cdots + X_n).$$

Si dice che X_n obbedisce alla legge (debole) dei grandi numeri rispetto alla successione $\{B_n\}$, $B_n > 0$, $B_n \rightarrow \infty$, se esiste una successione $\{A_n\}$ tale che

$$\frac{S_n - A_n}{B_n} \xrightarrow{P} 0.$$

1. Se $\{X_n\}$ è una successione di variabili aleatorie a due a due non correlate, con $\mathbb{E}[X_i] = \mu_i$, $\text{Var}[X_i] = \sigma_i^2$ e se

$$B_n := \sum_{i=1}^n \sigma_i^2 \rightarrow \infty \quad \text{quando } n \rightarrow \infty,$$

allora $\{X_n\}$ obbedisce alla legge dei grandi numeri rispetto a B_n essendo $A_n = \sum_{i=1}^n \mu_i$:

$$\frac{S_n - \sum_{i=1}^n \mu_i}{\sum_{i=1}^n \sigma_i^2} = \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sum_{i=1}^n \sigma_i^2} \xrightarrow{P} 0.$$

2. Un caso particolare è quello in cui le variabili X_n sono indipendenti ed equidistribuite. In tal caso, essendo $\mathbb{E}[X_i] = \mu$, $\text{Var}[X_i] = \sigma^2 > 0$, si ha

$$\frac{S_n - n\mu}{n\sigma^2} \xrightarrow{P} 0.$$

Possiamo allora esprimere in maniera rigorosa il teorema di Bernoulli. Siano $\{X_n\}$ variabili aleatorie indipendenti e con distribuzione di Bernoulli $B(p)$; allora, posto $\bar{X}_n = \frac{1}{n}S_n$,

$$\frac{\bar{X}_n - \mu}{\sigma^2} \xrightarrow{P} 0$$

ossia, anche

$$\bar{X}_n \xrightarrow{P} \mu.$$

smallskip

Si noti che la varianza della media campionaria diventa sempre più piccola – più precisamente, tende a zero – al tendere di n all'infinito. Intuitivamente ciò significa che quanto più è ampio il campione (cioè quanto più è grande N) tanto più i valori della media campionaria sono concentrati intorno al valore atteso (che è μ).

La disuguaglianza di Chebyshev fornisce allora, per ogni $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Poiché l'ultimo termine, per ε fissato ed n tendente all'infinito converge a zero, per ogni $\eta > 0$ si può determinare un valore n sufficientemente grande tale che

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \eta.$$

Da ciò discende che per n sufficientemente grande, si avrà

$$\mathbb{P}(|\bar{X}_n - \mu| \leq \varepsilon) \geq 1 - \eta.$$

In altre parole, si potrà determinare n tanto grande da far sì che la probabilità della disuguaglianza $|\bar{X}_n - \mu| \leq \varepsilon$ sia prossima ad 1 quanto si vuole.

Questo fatto è noto come *teorema di Bernoulli* ed è il punto di partenza di numerosi altri teoremi in probabilità e statistica.

Definizione

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in probabilità alla variabile aleatoria X se

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0 \quad \text{per ogni } \varepsilon > 0.$$

Legge dei grandi numeri

Consideriamo una successione $\{X_n\}$ di variabili aleatorie e sia S_n la successione delle somme parziali

$$S_n = (X_1 + \cdots + X_n).$$

Si dice che X_n obbedisce alla legge (debole) dei grandi numeri rispetto alla successione $\{B_n\}$, $B_n > 0$, $B_n \rightarrow \infty$, se esiste una successione $\{A_n\}$ tale che

$$\frac{S_n - A_n}{B_n} \xrightarrow{P} 0.$$

1. Se $\{X_n\}$ è una successione di variabili aleatorie a due a due non correlate, con $\mathbb{E}[X_i] = \mu_i$, $\text{Var}[X_i] = \sigma_i^2$ e se

$$B_n := \sum_{i=1}^n \sigma_i^2 \rightarrow \infty \quad \text{quando } n \rightarrow \infty,$$

allora $\{X_n\}$ obbedisce alla legge dei grandi numeri rispetto a B_n essendo $A_n = \sum_{i=1}^n \mu_i$:

$$\frac{S_n - \sum_{i=1}^n \mu_i}{\sum_{i=1}^n \sigma_i^2} = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sum_{i=1}^n \sigma_i^2} \xrightarrow{P} 0.$$

2. Un caso particolare è quello in cui le variabili X_n sono indipendenti ed equidistribuite. In tal caso, essendo $\mathbb{E}[X_i] = \mu$, $\text{Var}[X_i] = \sigma^2 > 0$, si ha

$$\frac{S_n - n\mu}{n\sigma^2} \xrightarrow{P} 0.$$

Possiamo allora esprimere in maniera rigorosa il teorema di Bernoulli. Siano $\{X_n\}$ variabili aleatorie indipendenti e con distribuzione di Bernoulli $B(p)$; allora, posto $\bar{X}_n = \frac{1}{n}S_n$,

$$\frac{\bar{X}_n - \mu}{\sigma^2} \xrightarrow{P} 0$$

ossia, anche

$$\bar{X}_n \xrightarrow{P} \mu.$$

Esempio 3.2 Una roulette americana ha 38 caselle, 18 rosse, 18 nere e due verdi, indicate con 0 e 00. Un giocatore decide di puntare 1\$ sull'uscita del rosso: se indichiamo con X l'esito della giocata, si ottiene $X = 1$ con probabilità $p = \frac{18}{38} = \frac{9}{19}$ e si ottiene $X = -1$ con probabilità $q = 1 - p = \frac{10}{19}$. Il giocatore è interessato al guadagno totale $S_n = X_1 + \dots + X_n$ dopo n partite:

1. ogni partita ha guadagno medio $\mu = -\frac{1}{19} \simeq 0.053$ e varianza $\sigma^2 = 1 - \frac{1}{19^2} \simeq 1$;
2. il guadagno medio, dopo n giocate, è dato da $\mathbb{E}[S_n] = n\mu = -\frac{n}{19}$;
3. la legge dei grandi numeri afferma che

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} + \frac{1}{19} \right| > \varepsilon \right) = 0 \quad \text{per ogni } \varepsilon > 0$$

cioè il guadagno medio per giocata (con n grande) è prossimo a $-\frac{1}{19}$ con probabilità elevata;

4. il teorema centrale afferma che

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq y \right) = \Phi(y)$$

dove $\Phi(x)$ è la funzione di ripartizione di una variabile Y gaussiana standard, cioè, per n grande, si ha approssimativamente

$$S_n \simeq n\mu + \sigma\sqrt{n}Y.$$

Nel nostro caso, se per esempio scegliamo $n = 100$,

$$S_{100} \simeq -5.3 + 10Y,$$

cosicché la probabilità di un guadagno positivo è

$$\mathbb{P}(S_{100} > 0) \simeq \mathbb{P}(-5.3 + 10Y > 0) = \Phi(0.53) \simeq 0.2981$$

Dopo 100 giocate, il giocatore sta vincendo con una probabilità di circa il 30% (e naturalmente è in perdita con probabilità di circa il 70%).

\triangle

Chapter 4

Processi di Markov

4.1 Introduzione

Consideriamo l'osservazione di un fenomeno che evolve nel tempo. Molto spesso si palesa l'impossibilità di una previsione deterministica sull'evoluzione del sistema; in questi casi la sola altra maniera di procedere consiste nell'associare una variabile casuale ad ogni istante futuro di osservazione, e nel dedurre dalle caratteristiche di questa le previsioni che ci interessano. Introduciamo quindi una famiglia di variabili aleatorie

$$\{X_t, t \in T\}, \quad T \text{ è un sottoinsieme di } \mathbb{R}_+,$$

a valori in un sottoinsieme E di \mathbb{R} , che viene chiamato *spazio delle fasi*.

È facile immaginare vari sistemi in cui l'insieme dei tempi che descrive le osservazioni future è un intervallo (finito o infinito) oppure un insieme discreto. Inoltre, possiamo immaginare fenomeni i cui esiti sono in numero finito o infinito numerabile, così come fenomeni caratterizzabili mediante i numeri appartenenti ad un intervallo (limitato o infinito).

4.1.1 Classificazione dei processi stocastici

In base a quanto è stato detto, possiamo introdurre una prima classificazione dei processi stocastici. Abbiamo:

1. processi stocastici discreti nel tempo e nello spazio (delle fasi);
2. processi stocastici continui nel tempo e discreti nello spazio;
3. processi stocastici discreti nel tempo e continui nello spazio;
4. processi stocastici continui nel tempo e nello spazio.

Si consideri ora una successione di istanti di osservazione e si supponga di conoscere la funzione di ripartizione delle singole variabili casuali associate ad essi. La probabilità che in un dato istante si manifesti un certo valore della corrispondente variabile può dipendere dai valori che si sono manifestati nelle precedenti osservazioni o può esserne indipendente. Questa duplice eventualità ci conduce a considerare una seconda classificazione dei processi stocastici:

1. processi stocastici costituiti da una successione di variabili aleatorie indipendenti;
2. processi stocastici in cui la generica variabile casuale dipende dalla evoluzione del sistema, cioè dalla particolare successione di valori già ottenuti.

4.1.2 Introduzione ai processi di Markov

Sia dato un processo stocastico del secondo tipo, e consideriamo una successione di osservazioni; l'insieme dei valori rilevati costituisce ciò che chiamiamo una “realizzazione del processo”:

$$\{(t_1, x_{t_1}; t_2, x_{t_2}; \dots; t_n, x_{t_n})\}.$$

Sia t_n l'istante in cui è stata effettuata l'ultima osservazione e quindi t_{n+1} l'istante in cui dovrà essere effettuata la prossima osservazione. Vogliamo considerare il problema di prevedere quale sarà lo stato del sistema al tempo t_{n+1} : ossia vogliamo assegnare la distribuzione di probabilità della variabile $X_{t_{n+1}}$ che descrive il sistema al tempo t_{n+1} . Se il sistema ha un insieme discreto di stati, il problema sarà risolto assegnando le probabilità che al tempo t_{n+1} il sistema si trovi nello stato x_j , $j = 1, 2, 3, \dots$

Poiché si tratta delle probabilità di passaggio da uno stato x_{t_n} NOTO ad un certo stato x , le indicheremo con il nome di *probabilità di transizione*. Se queste probabilità fossero indipendenti dalla realizzazione del processo sino a t_n incluso, saremmo di fronte ad un processo a variabili indipendenti.

Supponiamo invece che le probabilità di transizione siano condizionate dalla storia del sistema; se il processo ha avuto inizio in t_0 , si potrà scrivere

$$\mathbb{P}(X_{t_{n+1}} = x \mid X_{t_0} = y_0; X_{t_1} = y_1; \dots; X_{t_n} = y_n) \quad (4.1)$$

per indicare la probabilità che passi allo stato x , sapendo che in t_0 si è trovato in x_0 , in t_1 si è trovato in y_1 , ecc.

Un caso notevole di processi stocastici di tale tipo è quello che si presenta quando la dipendenza dipende solo dallo stato del sistema all'istante attuale; il passato non ha alcuna influenza. I processi stocastici caratterizzati in tale modo si dicono *markoviani*; per essi la (4.1) si semplifica in

$$\mathbb{P}(X_{t_{n+1}} = x \mid X_{t_n} = y_n).$$

4.2 Le catene di Markov

La nostra esposizione si limita a considerare processi di Markov discreti nel tempo e nello spazio. Indicheremo con E lo spazio degli eventi; in prima lettura, supporremo $E = \{1, \dots, M\}$, ma gran parte della teoria rimane valida per catene con un numero infinito numerabile di stati, come verrà mostrato nella sezione 4.4. Indicheremo con $\{t_n\}$ la successione degli istanti di osservazione, ordinata per tempi crescenti ($t_{n+1} > t_n$).

Poiché la successione degli istanti di osservazione è a priori nota, l'evoluzione del sistema sarà completamente determinata prendendo in considerazione il numero d'ordine delle transizioni. Risulta quindi

$$\mathbb{P}(X_{t_{n+1}} = x_j \mid X_{t_n} = x_i) = p_{ij}(n)$$

dove la coppia di indici “ ij ” indica che si considera la probabilità della transizione $x_i \rightarrow x_j$, mentre l'indice “ n ” tra parentesi sta ad indicare che si tratta della n -esima transizione.

Se l'insieme degli stati E ha cardinalità M , possiamo considerare la seguente disposizione matriciale:

$$P = \begin{pmatrix} p_{1,1}(n) & p_{1,2}(n) & \dots & p_{1,M}(n) \\ p_{2,1}(n) & p_{2,2}(n) & \dots & p_{2,M}(n) \\ \dots & \dots & \dots & \dots \\ p_{M,1}(n) & p_{M,2}(n) & \dots & p_{M,M}(n) \end{pmatrix}$$

La matrice di transizione P gode delle proprietà seguenti.

- Il generico elemento è $p_{i,j}(n)$, di cui abbiamo già chiarito il significato. In particolare, tutti gli elementi della matrice sono numeri reali, non negativi e non maggiori di 1.

- Gli elementi appartenenti alla i -esima riga sono le probabilità di transizione dallo stato i ad un qualunque stato j (compreso i stesso).
- Per il principio delle probabilità totali, la somma degli elementi di ogni riga è 1:

$$\sum_{j=1}^M p_{ij}(n) = 1.$$

Questo numero è infatti la probabilità della transizione da i verso uno qualunque dei possibili stati del sistema.

- La matrice è quadrata, perché le righe corrispondono agli stati iniziali e le colonne agli stati finali della transizione, che evidentemente sono uguali.

Definiamo *matrice stocastica* una matrice quadrata, i cui elementi sono numeri reali compresi nell'intervallo $[0, 1]$, in cui la somma degli elementi di ogni riga è uno.

Il caso più elementare e più studiato di catena di Markov è quello caratterizzato da matrici stocastiche i cui elementi non dipendono dal numero d'ordine della transizione considerata. Questa condizione si esprime anche dicendo che la catena è *omogenea* nel tempo. In tal caso potremo semplificare la notazione eliminando l'indice “(n)” dagli elementi di P .

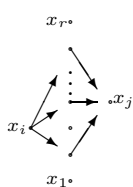
Abbiamo visto come ad ogni catena di Markov omogenea nel tempo, con insieme finito di stati, sia associata la matrice di transizione che risulta essere una matrice stocastica P ; viceversa, si può dimostrare che data una matrice stocastica P è possibile costruire un processo di Markov la cui matrice di transizione sia P .

4.2.1 La relazione di Chapman-Kolmogorov

Consideriamo una catena di Markov omogenea, avente un insieme finito di stati; ad essa è associata una matrice di transizione, che racchiude tutte le informazioni sulla evoluzione del sistema.

Poniamo $p_{ij}^{(m)}$ la probabilità di transizione dallo stato x_i allo stato x_j in m passi; notiamo che se $m = 1$, $p_{ij}^{(1)} = p_{ij}$; inoltre, la probabilità che ciò avvenga in due transizioni è data da

$$p_{ij}^{(2)} = \sum_{k=1}^r p_{ik} p_{kj}. \quad (4.2)$$



Infatti l'evento che stiamo considerando è del tipo $x_i \rightarrow x_k \rightarrow x_j$, dove k può assumere un qualunque valore tra 1 e r . Fissato k , per il principio della probabilità composta, $p_{ik} p_{kj}$ è la probabilità della transizione attraverso il particolare stato intermedio k . Poiché questo può essere qualunque, e al variare di k gli eventi sono mutualmente esclusivi, basterà applicare il principio delle probabilità totali per ottenere la (4.2).

La matrice $P^{(2)}$ il cui termine generico è $p_{ij}^{(2)}$ si ottiene come prodotto di P con se stessa:

$$P^{(2)} = \left(p_{ij}^{(2)} \right) = P \cdot P = P^2.$$

L'osservazione è facilmente generalizzabile: se si suppone nota la matrice $P^{(2)}$, il calcolo della probabilità che il sistema passi dallo stato x_i allo stato x_j in tre transizioni si effettua applicando il ragionamento già indicato, e si arriva a scrivere

$$p_{ij}^{(3)} = \sum_{k=1}^r p_{ik}^{(2)} p_{kj}$$

e quindi la matrice $P^{(3)}$ i cui elementi sono $p_{ij}^{(3)}$ si ottiene come prodotto tra il quadrato di P e P , cioè

$$P^{(3)} = P^2 \cdot P = P^3.$$

In generale si ha

$$p_{ij}^{(n)} = \sum_{k=1}^r p_{ik}^{(n-1)} p_{kj}, \quad P^{(n)} = \begin{pmatrix} p_{ij}^{(n)} \end{pmatrix} = P^{(n-1)} \cdot P = P^{n-1} \cdot P = P^n$$

e

$$p_{ij}^{(n+m)} = \sum_{k=1}^r p_{ik}^{(n)} p_{kj}^{(m)}. \quad (4.3)$$

L'ultima relazione prende il nome di equazione di Chapman - Kolmogorov ed è caratteristica dei processi di Markov.

4.2.2 Determinazione dello stato del sistema

Consideriamo una catena di Markov a valori in $E = \{x_1, \dots, x_r\}$; poiché le variabili aleatorie assumono valori in E , la legge di X_n è individuata dai numeri π_1, π_2, \dots dove, per $k = 1, \dots, r$:

$$\pi_k = \pi_k(n) = \mathbb{P}(X_n = x_k).$$

Posto $\pi(n) = (\pi_1(n), \dots, \pi_r(n))$, $\pi(n)$ è un vettore riga di dimensione r . Supponiamo che X_0 abbia legge $\pi(0)$; calcoliamo la legge di X_n :

$$\pi_k(n) = \sum_{j=1}^r \mathbb{P}(X_n = x_k \mid X_0 = x_j) \mathbb{P}(X_0 = x_j) = \sum_{j=1}^r p_{jk}^{(n)} \pi_j(0)$$

cioè i vettori $\pi(0)$ e $\pi(n)$ sono legati dalla relazione

$$\pi(n) = \pi(0) P^n.$$

Un caso particolare di distribuzione iniziale si ha quando X_0 assume il valore x_j con probabilità 1. In tal caso diremo che la catena di Markov *parte* dallo stato x_j e la distribuzione di $\pi(n)$ coincide con la riga j -esima di P^n .

4.2.3 Classificazione degli stati in una catena di Markov

Abbiamo stabilito, al termine del paragrafo 4.2.1, che per determinare la variabile casuale da associare alla n -esima osservazione basta calcolare la potenza n -esima della matrice di transizione. Ci si può chiedere ora quale sia il comportamento delle successive variabili casuali quando il numero di transizioni tende a infinito. La risoluzione di questo problema conduce a classificare gli stati del sistema in *stati transitori* e *stati finali*. Ci limiteremo a considerare il caso di catene di Markov omogenee con insieme degli stati finito.

Il criterio che conduce a tale classificazione nasce dalla seguente osservazione. Lo spazio degli stati può suddividersi (ma non necessariamente) in due sottoinsiemi C_1 e C_2 , nel primo dei quali compaiono elementi che comunicano tra loro¹ ma per cui è nulla la probabilità di transizione verso gli elementi

¹ diremo che x_i comunica con x_j se esiste $n \geq 1$ per cui $p_{ij}^{(n)} > 0$. In altre parole, esiste una successione di stati y_1, \dots, y_n tra loro distinti, per cui la probabilità della traiettoria $x_i \rightarrow y_1 \rightarrow \dots \rightarrow y_n \rightarrow x_j$ è non nulla

del secondo. Nel caso in cui tutti gli stati del sistema comunichino tra loro, diremo che la matrice di transizione è *irriducibile*.

Se invece tale suddivisione è realizzabile, parleremo di matrice di transizione *scindibile*. Sarà allora possibile, previo un riordino opportuno degli stati, scrivere la matrice nella forma

$$\begin{matrix} & C_1 & C_2 \\ \begin{matrix} C_1 \\ C_2 \end{matrix} & \begin{pmatrix} P & 0 \\ R & Q \end{pmatrix} \end{matrix}$$

0 sta ad indicare una matrice nulla ($m \times n$) dove m è il numero degli stati in C_1 ed n quello degli stati in C_2 . Gli elementi della classe C_1 sono *stati finali* del sistema, in quanto che quando il sistema raggiunge la classe C_1 non potrà in seguito occupare stati dell'altra classe.

L'insieme C_2 può a sua volta essere scindibile in due sottoinsiemi di stati, C'_2 e C''_2 , tali che da un qualunque stato del primo è impossibile passare ad uno stato del secondo (e agli stati di C_1). Un nuovo ordinamento degli stati conduce a scrivere la matrice di transizione nella forma

$$\begin{matrix} & C_1 & C'_2 & C''_2 \\ \begin{matrix} C_1 \\ C'_2 \\ C''_2 \end{matrix} & \begin{pmatrix} P & 0 & 0 \\ 0 & P' & 0 \\ R' & R'' & Q' \end{pmatrix} \end{matrix}$$

Procedendo in modo analogo è possibile porre in evidenza un insieme di “gruppi finali”; gli stati che appartengono ad una classe finale sono detti *ricorrenti*, mentre gli elementi che non appartengono a nessuna di tali classi verranno detti *stati transitori* perché per essi è maggiore di zero la probabilità di passare da uno di questi stati ad uno stato finale, e nulla la probabilità di tornare.

Diremo inoltre che un elemento è *assorbente* se forma da solo una classe finale.

4.3 Distribuzioni limite

Si consideri una catena di Markov omogenea su uno spazio degli stati finito. Siamo interessati a stabilire il comportamento asintotico della successione delle variabili casuali che definiscono la catena; questo ci conduce a studiare il problema dell'esistenza di *probabilità invarianti* o stazionarie per la catena.

Diremo che una distribuzione di probabilità su E

$$\pi = (\pi_1, \dots, \pi_M)$$

è invariante per la matrice P (e anche per la catena di Markov a P associata) se

$$\pi P = \pi.$$

Se la legge dello stato iniziale è una misura invariante π , allora tutti gli stati seguenti hanno la stessa legge.

Teorema 4.1 (Teorema di Markov-Kakutani) *Una matrice di transizione su uno spazio finito di stati ha sempre almeno una probabilità invariante.*

Osserviamo che la probabilità invariante non è necessariamente unica. Anzi, se π^0 e π^1 sono due probabilità invarianti, allora per ogni $\lambda \in (0, 1)$, la legge $\pi^\lambda = \lambda\pi^1 + (1 - \lambda)\pi^0$ è invariante per P . Vale tuttavia il seguente risultato:

se una catena è irriducibile, allora possiede al più una probabilità invariante.

4.3.1 Catene di Markov regolari

Abbiamo detto che se la distribuzione π dello stato iniziale X_0 di una catena di Markov è invariante per la matrice di transizione, allora la distribuzione di tutte le variabili X_n coincide con π . Ci possiamo ora chiedere cosa succede se consideriamo l'evoluzione di un sistema a partire da una generica distribuzione iniziale. In questo caso, come noto, si ottiene

$$\pi^{(n)} = \pi^{(0)} P^{(n)} = \pi^{(0)} P^n,$$

dove P^n è la potenza n -esima della matrice P . Rimane quindi individuato il problema della determinazione di un limite per P^n quando n cresce oltre ogni limite. Non daremo una risposta completa al problema, ma ci limiteremo ad enunciare il teorema di Markov, che risponde al problema nel caso di matrici *regolari*.

Diremo *regolare* una matrice di transizione che verifica:

per un qualche valore di n gli elementi della matrice P^n hanno tutti i valori (strettamente) positivi.

Osserviamo subito che se per un qualche n la matrice P^n ammette tutti gli elementi > 0 , allora ciò è vero anche per tutte le potenze seguenti. Infatti

$$p_{ij}^{(n+1)} = \sum_{k=1}^r p_{ik} p_{kj}^{(n)} \quad (4.4)$$

e, per ipotesi, almeno un addendo è maggiore di zero; quindi, tutti gli elementi $p_{ij}^{(n+1)}$ sono positivi.

Teorema 4.2 (Teorema di Markov) *Se la matrice P è regolare, allora*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = p_j,$$

ossia i limiti delle probabilità $p_{ij}(n)$ esistono e sono indipendenti dall'indice i . La distribuzione di probabilità $\pi = (p_1, \dots, p_r)$ è l'unica distribuzione invariante per la catena.

Conseguenza del teorema è la seguente proposizione: nelle ipotesi enunciate, la distribuzione del sistema, dopo un numero elevato di transizioni, tende a coincidere con la distribuzione stazionaria. In particolare, questa convergenza è verificata per ogni distribuzione iniziale, cioè, la distribuzione del sistema, dopo un numero elevato di transizioni, tende a essere indipendente dalla distribuzione iniziale. Un sistema in cui sia rilevabile questa indipendenza si dice *ergodico*.

Nello studio delle proprietà di una catena di Markov, è utile considerare il seguente criterio:

$$\begin{aligned} &\text{se tutti gli stati comunicano tra loro ed inoltre esiste } j \text{ tale che } p_{jj} > 0, \\ &\text{allora la catena è regolare.} \end{aligned} \quad (4.5)$$

La costruzione precedente mostra che una catena regolare ammette una distribuzione limite, ed inoltre questa è unica. Vogliamo indicare qui un procedimento di calcolo per tale distribuzione.

Poiché $p_{ij}^{(n)} = \sum_{k=1}^r p_{ik}^{(n-1)} p_{kj}$ e poiché per ipotesi $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = p_j$, $\lim_{n \rightarrow \infty} p_{ik}^{(n-1)} = p_k$, si ha, per n tendente a infinito:

$$p_j = \sum_{k=1}^r p_k p_{kj}, \quad j = 1, 2, \dots, r \quad (4.6)$$

Il sistema (4.6), associato alla condizione ovvia

$$\sum_{j=1}^r p_j = 1,$$

permette di determinare la distribuzione limite.

4.3.2 Catene periodiche

Per ogni punto $x \in E$, costruiamo l'insieme

$$D_x = \{n : p_{x,x}^{(n)} > 0\}$$

insieme dei tempi n in cui è positiva la probabilità di tornare in x sapendo di partire dallo stato x stesso. D_x è un insieme di numeri, di cui posso calcolare il massimo comun denominatore MCD, che indico con d_x (ricordiamo che il MCD d_x è il più grande intero che divide n per ogni $n \in D_x$).

Lemma 4.3 *Se P è una matrice irriducibile, allora $d_x = d$ è costante per ogni $x \in E$.*

Per una catena irriducibile, il valore d è detto il *periodo* della catena, e una catena irriducibile con periodo $d = 1$ è detta a-periodica.

Si dimostra che una catena irriducibile a-periodica è regolare, quindi ergodica; una catena periodica di periodo $d \geq 2$ non è regolare né ergodica (quindi, la distribuzione limite $\lim_{n \rightarrow \infty} \pi^{(n)} = \lim_{n \rightarrow \infty} \pi^{(0)} P^n$ dipende dalla distribuzione iniziale $\pi^{(0)}$).

4.4 Catene a stati numerabili

Per ogni stato $x \in E$, poniamo $T^x = \min\{n \mid X_n = x, X_0 = x\}$ il tempo di primo ritorno in x per una catena che parte da x . Diremo che uno stato è transitorio se

$$\rho_x = \mathbb{P}(T^x < +\infty) = \sum_{n \geq 1} p_{x,x}^{(n)} < 1,$$

e ricorrente se $\rho_x = 1$.

Osserviamo che se x è ricorrente allora $\sum_{n \geq 1} \mathbb{P}(T^x = n) = 1$, quindi T^x è una v.a. (in senso classico) mentre se x è transitorio, T^x è una v.a. in senso esteso, cioè ammettiamo che T^x possa assumere valore $+\infty$ con probabilità positiva.

Se x è transitorio, necessariamente $\mathbb{E}[T^x] = +\infty$; se x è ricorrente, allora T^x è una v.a., di cui possiamo calcolare la media. Se risulta $\mathbb{E}[T^x] < +\infty$, allora diremo che x è *ricorrente positivo*, mentre se $\mathbb{E}[T^x] = +\infty$, diremo che x è *ricorrente nullo*.

Da questa definizione segue che se x è ricorrente e $x \rightarrow y$, allora anche y è ricorrente; allora, una catena irriducibile ha tutti gli stati che sono ricorrenti oppure transitori, ed una catena finita ha sempre almeno uno stato ricorrente.

Possiamo anche calcolare il numero di volte che una catena raggiunge lo stato x ; per semplicità, supporremo che lo stesso x sia lo stato iniziale (parleremo quindi di ritorni); poniamo N_x il numero di ritorni a partire dallo stato x . Se x è ricorrente, allora

$$\mathbb{P}(N_x = m) = 0 \quad \forall m \in \mathbb{N},$$

cioè $\mathbb{P}(N_x = +\infty) = 1$; se invece x è transitorio,

$$\mathbb{P}(N_x = m) = \rho_x^m (1 - \rho_x) \quad \forall m \in \mathbb{N},$$

e quindi $\mathbb{P}(N_x < \infty) = 1$ e inoltre $\mathbb{E}[N_x] = \frac{\rho_x}{1 - \rho_x}$.

Si ha inoltre il seguente criterio per determinare se uno stato è ricorrente oppure transitorio, che generalizza la definizione: dato una coppia di stati $x, y \in E$, se la serie $\sum_{n \geq 1} p_{y,x}^{(n)}$ è convergente allora lo stato x è *transitorio*, se la serie è divergente allora lo stato x è *ricorrente*.

Terminiamo con la seguente distinzione tra gli stati ricorrenti per le catene a stati numerabili. Indichiamo con $m_x = \mathbb{E}[T^x]$ il tempo medio di ritorno in x . Se lo stato è transitorio, abbiamo visto che $\mathbb{E}[T^x] = \infty$; se lo stato è ricorrente, la media può esistere finita oppure essere infinita. Diremo che lo stato è ricorrente positivo se $\mathbb{E}[T^x] < \infty$, altrimenti diremo che lo stato è ricorrente nullo.

Vale la seguente caratterizzazione: per una catena irriducibile a stati ricorrenti, o tutti gli stati sono ricorrenti positivi, oppure sono ricorrenti nulli.

Il seguente teorema è molto importante (e vale anche nel caso di una catena finita). Con le stesse notazioni della sezione precedente, poniamo N_x^n il numero di ritorni a partire dallo stato x effettuati entro il passo n -esimo.

Teorema 4.4 *Per una catena irriducibile si ha*

$$\lim_{n \rightarrow \infty} \frac{1}{n} N_x^n = \frac{1}{m_x},$$

Inoltre, se $m_x < \infty$ e poniamo $\pi_x = \frac{1}{m_x}$, allora $\pi = (\pi_1, \pi_2, \dots)$ è l'unica distribuzione invariante della catena. Viceversa, se $m_x = +\infty$ per un qualche $x \in E$, allora non esiste alcuna distribuzione invariante.

4.5 Passeggiate casuali con barriere

Consideriamo un sistema di prove ripetute, che possiamo modellare tramite una successione di variabili aleatorie discrete $\{X_n\}$ indipendenti ed equidistribuite. Consideriamo la successione di variabili aleatorie

$$S_n = x + X_1 + \dots + X_n, \quad n \geq 1,$$

dove $x \in \mathbb{R}$ e scriveremo $S_0 = x$ per indicare la condizione iniziale del sistema. È facile convincersi che $\{S_n\}$ è una catena di Markov (dato lo stato attuale del sistema S_n , la posizione al tempo $n + 1$ dipende solo dalla distribuzione di X_{n+1} e non da quella delle variabili ai tempi precedenti).

Nel caso di una evoluzione non banale, in cui cioè $\mathbb{P}(X_1 = 0) < 1$, allora lo spazio degli stati E del sistema è infinitamente numerabile. Per evitare questo problema, fissiamo due valori $a < b$, con $a \leq x \leq b$, per cui il sistema, giunto in a (oppure in b), evolve in maniera banale (rimane costante in a , oppure in b , rispettivamente). In tal modo, possiamo scrivere

$$E = \{a, a + 1, \dots, b - 1, b\}$$

e la matrice di transizione associata al sistema è una matrice quadrata P .

Esempio: passeggiata casuale semplice

Consideriamo il caso in cui X ha distribuzione

$$X = \begin{cases} -1, & q \\ 0, & r \\ 1, & p. \end{cases}$$

ossia ad ogni passo il sistema può salire di una unità, rimanere costante, scendere di una unità. L'evoluzione non è banale se $r < 1$, come da ora in poi richiederemo. Se ad esempio le X fossero le variabili del lancio di una moneta equilibrata, sarebbe $p = q = \frac{1}{2}$, $r = 0$.

Siano $a < b$ le barriere scelte, $a \leq x \leq b$; la matrice di transizione del sistema ha la forma

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ q & r & p & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & q & r & p & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & q & r & p & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & q & r & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix}$$

L'analisi della matrice conduce a riconoscere in a e b due stati *assorbenti*: per ognuno di essi è positiva la probabilità di ingresso, e nulla la probabilità di uscita dallo stato (essi comunicano solo con se stessi). Tutti gli altri stati si classificano come stati transitori.

Ritorniamo al caso generale. Per una passeggiata casuale con barriere, è di primaria importanza capire se le barriere sono effettivamente raggiunte in un tempo finito. Indichiamo con T la variabile aleatoria definita da

$$T = \inf\{n \geq 1 : S_n \leq a \text{ oppure } S_n = b\}.$$

T è il primo istante in cui la catena di Markov S_n tocca una barriera; il valore S_T deve essere pertanto uguale ad a oppure b . È possibile dimostrare che per un sistema non banale (ossia, ripetiamo, con $\mathbb{P}(X = 0) < 1$) T è finita.

A questo punto, possiamo rispondere alle principali domande di una passeggiata casuale con barriere:

- se la condizione iniziale è x , qual è la probabilità di raggiungere b prima di a ?
- qual è il tempo medio $\mathbb{E}(T)$ necessario per raggiungere una delle barriere?

4.5.1 Identità di Wald

Se fossimo di fronte ad una passeggiata casuale senza barriere, potremo calcolare la posizione media $\mathbb{E}(S_n)$, per $n \geq 1$, utilizzando la linearità della funzione media e l'equidistribuzione delle variabili aleatorie X_i :

$$\mathbb{E}(S_n) = x + n\mathbb{E}(X).$$

La *prima identità di Wald* calcola il valore medio S_T di una passeggiata casuale con barriere *al tempo di arrivo su una barriera*:

$$\mathbb{E}(S_T) = x + \mathbb{E}(T)\mathbb{E}(X).$$

Quanto vale, per una passeggiata casuale senza barriere, la varianza di S_n ? Sfruttando l'indipendenza supposta delle variabili X_i , si ha

$$\text{Var}(S_n) = n\text{Var}(X).$$

Nel caso di v.a. con media $\mu = 0$ e varianza σ^2 finita, la *seconda identità di Wald* calcola la varianza $\text{Var}(S_T)$ di una passeggiata casuale con barriere *al tempo di arrivo su una barriera*:

$$\text{Var}(S_T) = \mathbb{E}(T)\text{Var}(X).$$

Passeggiata casuale semplice simmetrica

Diremo che la passeggiata casuale è simmetrica se $p = q$: si ha allora

$$\mathbb{E}(X) = 0, \quad \text{Var}(X) = p + q = 1 - r.$$

Nel caso di una passeggiata casuale simmetrica, le identità di Wald consentono di rispondere alle domande poste in precedenza.

Iniziamo dalla relazione $\mathbb{P}(S_T = a) + \mathbb{P}(S_T = b) = 1$; si ha

$$\begin{aligned} \mathbb{E}(S_T) &= a \cdot \mathbb{P}(S_T = a) + b\mathbb{P}(S_T = b) \\ &= a + (b - a)\mathbb{P}(S_T = b). \end{aligned}$$

Dalla prima identità di Wald si ottiene

$$x = \mathbb{E}(S_T) = a + (b - a)\mathbb{P}(S_T = b)$$

ossia $\mathbb{P}(S_T = b) = \frac{(x - a)}{(b - a)}$ ed anche $\mathbb{P}(S_T = a) = \frac{(b - x)}{(b - a)}$.

Passiamo ora a calcolare il tempo medio $\mathbb{E}(T)$. Per questo, è necessario utilizzare la seconda identità di Wald. Iniziamo con il calcolare $\mathbb{E}(S_T^2) = a^2 + (b^2 - a^2)\mathbb{P}(S_T = b)$, da cui

$$\text{Var}(S_T) = \mathbb{E}(S_T^2) - (\mathbb{E}(S_T))^2 = a^2 + (b^2 - a^2)\frac{(x - a)}{(b - a)} - x^2;$$

questa espressione si semplifica e si ottiene $\text{Var}(S_T) = (b - x)(x - a)$. Dalla seconda identità di Wald si ha anche

$$\text{Var}(S_T) = \mathbb{E}(T)\text{Var}(X),$$

quindi

$$\mathbb{E}(T) = \frac{1}{(b - x)(x - a)}(1 - r).$$

Passeggiata casuale non simmetrica

Consideriamo ora il caso $p \neq q$. Non abbiamo a disposizione la seconda disuguaglianza di Wald; possiamo però affrontare il problema direttamente. Per semplificare i conti, supporremo $a = 0$.

Indichiamo con $f(x) = \mathbb{P}(S_T = b \mid S_0 = x)$: un attimo di riflessione ci porta alla seguente relazione:

$$f(x) = pf(x + 1) + qf(x - 1) + rf(x),$$

che è una equazione alle differenze, con dati al bordo $f(0) = 0$, $f(b) = 1$. La soluzione è data dalla formula

$$f(x) = \mathbb{P}(S_T = b \mid S_0 = x) = \frac{1 - (q/p)^x}{1 - (q/p)^b}.$$

Possiamo allora calcolare il valor medio di S_T^x :

$$\mathbb{E}[S_T^x] = bf(x),$$

e infine dalla prima identità di Wald si ottiene che il numero medio di giocate per terminare la partita è

$$\mathbb{E}[T] = \frac{1}{p - q}(bf(x) - x).$$

Catene con un solo stato assorbente

Consideriamo il seguente esempio. Lancio ripetutamente una moneta, e voglio studiare l'uscita della sequenza CCTT. Scegliamo come spazio degli stati $E = \{x_0 = T, x_1 = C, x_2 = CC, x_3 = CCT, x_4 = CCTT\}$, supponendo che lo stato x_4 sia assorbente. Voglio determinare la velocità di ingresso nello stato assorbente. A tal fine, risponderò alla domanda: qual è la probabilità di giungere in x_4 entro il tempo K ? Sia p la probabilità di uscita di testa; posso allora scrivere la matrice di transizione

$$\begin{pmatrix} p & q & & & \\ p & & q & & \\ & & q & p & \\ & q & & & p \\ & & & & 1 \end{pmatrix}$$

(gli altri elementi sono 0).

Indichiamo con τ^i il primo tempo in cui si raggiunge lo stato assorbente a partire dallo stato i , e sia $g_i(n) = \mathbb{P}(\tau^i = n)$ la sua distribuzione. Si ottiene

$$g_i(n) = \sum p_{x_i, y_1} p_{y_1, y_2} \cdots p_{y_{n-1}, x_4}$$

dove sommo su tutte le traiettorie $x_i \rightarrow y_1 \rightarrow y_2 \rightarrow \cdots \rightarrow y_{n-1} \rightarrow x_4$ in cui $y_i \neq x_4$, ed anche

$$g_i(n) = \sum_{j=0}^3 p_{x_i, x_j} g_j(n-1).$$

Siamo interessati alla probabilità di raggiungere x_4 nei primi (\leq) K tentativi; posto $\phi_i(n) = \sum_{t=1}^n g_i(t) = \mathbb{P}(\tau^i \leq n)$, sommando nella formula precedente si ottiene la relazione

$$\phi_i(n+1) = p_{i, x_4} + \sum_{j=0}^3 p_{x_i, x_j} \phi_j(n).$$

Le seguenti righe sono state ottenute con OpenOffice

	A	B	C	D	E
1	tempo/probabilità	ϕ_0	ϕ_1	ϕ_2	ϕ_3
2	1	0	0	0	p
3	2	$= p * B2 + q * C2$	$= p * B2 + q * D2$	$= q * D2 + p * E2$	$= p + q * C2$
	12	0.533	0.615	0.766	0.808
	21	0.797	0.833	0.899	0.917
	51	0.987	0.990	0.994	0.995

dove gli ultimi valori sono stati calcolati con $p = q = \frac{1}{2}$.

Passeggiata casuale a stati numerabili

Consideriamo il seguente processo (nascita e morte): se ad un tempo n ho x individui, al passo $n+1$ ne avrò $x-1$, x oppure $x+1$ con probabilità, rispettivamente, q , r , p , dove $p > 0$, $q > 0$, $p + q + r = 1$, ma se al tempo n ho 0 individui, allora ho una probabilità $p_0 > 0$ di passare in 1 (immigrazione). Allora la catena è irriducibile (tutti gli stati comunicano), quindi tutti gli stati sono o ricorrenti o transitori; vogliamo vedere quale alternativa prevale.

Indichiamo con τ_0^i la probabilità di arrivare in 0 a partire dallo stato i : vale la formula

$$\rho_0 = p_{0,0} + p_{0,1}\rho_0^1,$$

quindi 0 è ricorrente se e solo se $\rho_0^1 = 1$, ossia se e solo se $\mathbb{P}(\tau_0^1 < +\infty) = 1$. Consideriamo uno stato $M \in E$: allora

$$\mathbb{P}(\tau_0^1 < \infty) \geq \mathbb{P}(\tau_0^1 < \tau_M^1)$$

(a sx, ho tutte le traiettorie che arrivano in 0, a dx, tutte le traiettorie che arrivano in 0 senza toccare il livello M). Il lato destro, d'altra parte, è la probabilità di bancarotta in un gioco con barriere 0 e M , partendo dal livello 1, quindi è pari a

$$\mathbb{P}(\tau_0^1 < \tau_M^1) = 1 - f(1) = 1 - \frac{1 - (q/p)}{1 - (q/p)^M} = \frac{(q/p) - (q/p)^M}{1 - (q/p)^M}$$

se $p \neq q$, e $\mathbb{P}(\tau_0^1 < \tau_M^1) = \frac{M-x}{M}$ se $p = q$. Studiamo il limite per $M \rightarrow \infty$ di queste relazioni: se $(q/p) < 1$ allora

$$\lim_{M \rightarrow \infty} \frac{(q/p) - (q/p)^M}{1 - (q/p)^M} = \frac{q}{p} < 1;$$

se $(q/p) \geq 1$ allora

$$\lim_{M \rightarrow \infty} \mathbb{P}(\tau_0^1 < \tau_M^1) = 1.$$

Quindi, se $q/p \geq 1$ si ottiene che $\mathbb{P}(\tau_0^1 < \infty) = 1$ e quindi tutti gli stati sono ricorrenti, se $q/p < 1$ allora tutti gli stati sono transitori. In questo caso, in particolare, fissato comunque un insieme di stati $\{0, 1, 2, \dots, N\}$, dopo un tempo finito la catena di Markov lascerà questo insieme per non tornarci più, quindi $\lim_{n \rightarrow \infty} X_n = +\infty$.

In questo esempio, possiamo anche calcolare la distribuzione invariante (quando esiste). Iniziamo con il fissare $q \geq p$, altrimenti la catena è transitoria e non può esistere la misura invariante. Indichiamo con $\pi = (\pi_0, \pi_1, \dots)$ la distribuzione invariante. Allora vale

$$\begin{cases} \pi_0 = \pi_0 r_0 + \pi_1 q, \\ \pi_1 = \pi_0 p_0 + \pi_1 r + \pi_2 q, \\ \pi_2 = \pi_1 p + \pi_2 r + \pi_3 q, \\ \dots \\ \pi_n = \pi_{n-1} p + \pi_n r + \pi_{n+1} q \end{cases}$$

che posso anche scrivere ($r = 1 - p - q$, $r_0 = 1 - p_0$)

$$\begin{cases} \pi_0 p_0 = \pi_1 q, \pi_1 q - \pi_0 p_0 = \pi_2 q - \pi_1 p, \dots \\ \pi_n q - \pi_{n-1} p = \pi_{n+1} q - \pi_n r \end{cases}$$

e quindi, per ricorrenza,

$$\begin{cases} \pi_1 = \pi_0 \frac{p_0}{q}, \pi_2 = \pi_1 \frac{p}{q} = \pi_0 \frac{p_0 p}{q^2}, \dots \\ \pi_{n+1} = \pi_0 \frac{p_0 p^n}{q^{n+1}}. \end{cases}$$

Rimane da determinare π_0 ; dalla relazione $\sum_n \pi_n = 1$ si ottiene

$$1 = \pi_0 + \pi_0 \frac{p_0}{q} \sum_{n=0}^{\infty} (p/q)^n$$

e quindi, se $p < q$, si ottiene

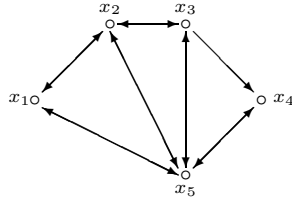
$$\pi_0 = \frac{q-p}{q-p+p_0}, \quad \pi_{n+1} = \frac{q-p}{q-p+p_0} \frac{p_0 p^n}{q^{n+1}}.$$

Concludiamo osservando che nel caso $p = q$ tutti gli stati sono ricorrenti nulli e non esiste distribuzione invariante.

4.6 Esercizi conclusivi

Grafi

Consideriamo un generico grafo G , e sia $E = \{x_1, \dots, x_r\}$ l'insieme dei suoi vertici.



Sia n_i il numero di vertici che si possono raggiungere da x_i ; definiamo la catena di Markov associata al grafo G attraverso la matrice di transizione

$$\begin{cases} p_{ij} = \frac{1}{n_i} & \text{se } x_j \text{ è raggiungibile da } x_i, \\ p_{ij} = 0 & \text{altrimenti.} \end{cases}$$

Nel caso della figura precedente, si ha

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 1 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{pmatrix}$$

Per determinare se la matrice è regolare, conviene osservare che il primo esponente per cui $p_{14}^{(n)} > 0$ è $n = 3$; ma si verifica facilmente che $p_{ij}^{(3)} > 0$ per ogni coppia di indici $i, j = 1, \dots, 5$, quindi la matrice è regolare e ammette una unica probabilità invariante.

Consideriamo ancora il grafo nella figura precedente, a cui aggiungiamo ora una freccia che collega $x_4 \rightarrow x_3$. Allora tutti i vertici sono collegati da una doppia freccia “ \leftrightarrow ”; in questo caso, esiste una unica misura invariante che è anche reversibile, ed è determinata da

$$\pi = (\pi_1, \dots, \pi_r), \quad \pi_j = \frac{n_j}{n_1 + \dots + n_r}.$$

Scriviamo la nuova matrice di transizione

$$P_1 = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{pmatrix}$$

Ancora con riferimento al grafo nella figura precedente, supponiamo di eliminare ora la freccia $x_4 \rightarrow x_3$ e di sostituirla con $x_4 \rightarrow x_4$. Allora lo stato x_4 è uno stato assorbente e forma l'unica classe finale del sistema; i rimanenti stati (x_1, x_2, x_3 e x_5) sono transitori.

Scriviamo la nuova matrice di transizione

$$P_2 = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{pmatrix}$$

Esercizio

Consideriamo il seguente sistema: sia $E = \{a, b, c, d\}$ lo spazio degli stati e sia P la matrice di transizione

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Sia $\pi(0) = (0 \ 0 \ 0 \ 1)$ lo stato iniziale del sistema.

- (a) Il sistema è irriducibile?
- (b) Il sistema è ergodico (la matrice P è regolare)?
- (c) Determinare la distribuzione invariante del sistema.

Soluzione.

Osserviamo che tutti gli stati del sistema comunicano: ad esempio, la traiettoria $a \rightarrow b \rightarrow c \rightarrow d \rightarrow c \rightarrow a$ ha probabilità $\frac{1}{64}$. Quindi tutti gli stati appartengono alla stessa classe, ossia il sistema è irriducibile.

Osserviamo inoltre che $p_{4,4} > 0$ quindi per il criterio enunciato in precedenza il sistema è regolare.

Vogliamo infine determinare la distribuzione invariante; a tal fine, è necessario risolvere il sistema

$$\begin{cases} \frac{1}{2}\pi_2 + \frac{1}{4}\pi_3 = \pi_1 \\ \frac{1}{2}\pi_1 + \frac{1}{4}\pi_3 = \pi_2 \\ \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{2}\pi_4 = \pi_3 \\ \frac{1}{2}\pi_3 + \frac{1}{2}\pi_4 = \pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases}$$

che ha soluzione $\pi = (1/6 \ 1/6 \ 1/3 \ 1/3)$.

Il sistema è ergodico, quindi anche partendo dallo stato $\pi(0)$ si ottiene $\lim_{n \rightarrow \infty} \pi(n) = \pi$. Si confronti con

$$\pi(10) = \pi(0)P^{10} = \left(\frac{341}{2048} \quad \frac{341}{2048} \quad \frac{341}{1024} \quad \frac{342}{1024} \right)$$

le cui componenti si discostano da quelle di π per un fattore 10^{-4} .

△

Esercizio.

Consideriamo il seguente sistema: sia $E = \{a, b, c, d\}$ lo spazio degli stati e sia P la matrice di transizione

$$P = \begin{pmatrix} 0 & p & 0 & q \\ p & 0 & q & 0 \\ 0 & p & 0 & q \\ p & 0 & q & 0 \end{pmatrix}.$$

Sia $\pi(0) = (0 \ 0 \ 0 \ 1)$ lo stato iniziale del sistema.

- (a) Descrivere il sistema.
- (b) Determinare la distribuzione invariante del sistema.
- (c) Determinare la distribuzione $\pi(10)$ del sistema a partire dallo stato $\pi(0)$.

Soluzione.

Osserviamo che tutti gli stati del sistema comunicano: ad esempio, la traiettoria $a \rightarrow b \rightarrow c \rightarrow d \rightarrow a$ ha probabilità $\frac{1}{64}$. Quindi tutti gli stati appartengono alla stessa classe, ossia il sistema è irriducibile.

Il sistema è periodico di periodo 2. Basta infatti osservare che a partire dallo stato 1, le possibilità di ritorno in 1 sono

$$p_{1,1}^{(2n)} = q, \quad p_{1,1}^{(2n+1)} = 0.$$

Vogliamo determinare la distribuzione invariante; a tal fine, è necessario risolvere il sistema

$$\begin{cases} q\pi_2 + q\pi_4 = \pi_1 \\ q\pi_1 + q\pi_3 = \pi_2 \\ p\pi_2 + p\pi_4 = \pi_3 \\ p\pi_1 + p\pi_3 = \pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases}$$

che ha soluzione $\pi = (\frac{q}{2} \quad \frac{q}{2} \quad \frac{p}{2} \quad \frac{p}{2})$.

Il sistema non è ergodico, tuttavia partendo dallo stato $\pi(0)$ si ottiene $\pi(2n) = (0 \quad p \quad 0 \quad p)$ mentre $\pi(2n+1) = (p \quad 0 \quad p \quad 0)$.

△

Esercizio.

Consideriamo due amici che giocano al rosso e nero su una roulette. Sia $A = 5\$$ il capitale del primo giocatore e $B = 10\$$ il capitale del secondo; se esce rosso, il primo giocatore riceve 1\$ dal secondo, se esce nero paga 1\$, se esce zero (verde) si ripete la scommessa.

Calcolare la probabilità che il giocatore con capitale minore vinca e il tempo atteso di gioco.

Soluzione.

Indichiamo con S_n^x il capitale del primo giocatore: si tratta di una passeggiata casuale simmetrica ($p = q = \frac{18}{37}$, $r = \frac{1}{37}$) con barriere $a = 0$ (il primo giocatore perde tutto) e $b = 15$ (il primo giocatore vince).

Dalla prima identità di Wald si ottiene $\mathbb{E}[S_T^x] = x$, ma è anche

$$x = \mathbb{E}[S_T^x] = a\mathbb{P}(S_T^x = a) + b\mathbb{P}(S_T^x = b) \quad \implies \quad \mathbb{P}(S_T^x = b) = \frac{x-a}{b-a} = \frac{1}{3}.$$

Dalla seconda identità di Wald si ottiene infine $\mathbb{E}[T] = \frac{\text{Var}[S_T^x]}{\text{Var}[X]} = \frac{37}{36}50 \simeq 51,4$.

△

Esercizio.

Modifichiamo l'esempio precedente: per aiutare il più povero, gli aumentiamo le probabilità di vittoria. Avevamo $A = 5\$$ il capitale del primo giocatore e $B = 10\$$ il capitale del secondo; se esce rosso o zero, il primo giocatore riceve 1\$ dal secondo, se esce nero paga 1\$. Vogliamo ancora calcolare la probabilità che il giocatore con capitale minore vinca e il tempo atteso di gioco.

Soluzione.

Indichiamo con S_n^x il capitale del primo giocatore: si tratta di una passeggiata casuale non-simmetrica ($p = \frac{18}{37}$, $q = \frac{17}{37}$) con barriere $a = 0$ (il primo giocatore perde tutto) e $b = 15$ (il primo giocatore vince), e capitale iniziale $x = 5$.

Utilizziamo la formula per calcolare la probabilità di vittoria

$$\mathbb{P}(S_T = b \mid S_0 = x) = \frac{1 - (17/18)^5}{1 - (17/18)^{15}} \simeq 0,43$$

quindi in media il capitale finale del primo giocatore è $\mathbb{E}[S_T] \simeq 6,5 > 5$ (in media, il primo giocatore vince 1,5\$ a partita). Dalla prima identità di Wald si ottiene la durata media di una partita: $\mathbb{E}[T] \simeq 54,6$.

△

Chapter 5

Processo di Poisson

In questo capitolo vogliamo considerare fenomeni che possono essere rappresentati come distribuzione di punti nello spazio, oppure di eventi nel tempo. Tra gli esempi tipici di tali fenomeni vi sono, ad esempio,

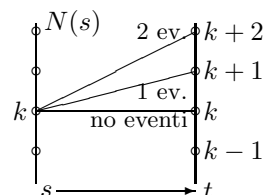
1. l'arrivo delle chiamate telefoniche in un centralino;
2. le richieste di servizio in un sistema di calcolo distribuito (ad esempio, l'accesso al disco);
3. la distribuzione delle colonie di batteri in una provetta di laboratorio; ecc.

Faremo sempre riferimento alla semiretta $\{t \geq 0\}$ come asse dei tempi. In questo caso, la realizzazione del fenomeno è costituita da un insieme di punti, al più infinitamente numerabile, t_1, t_2, \dots , dell'asse reale.

5.1 Processi di conteggio

Un processo $N(t)$, $t \geq 0$, è detto essere un *processo di conteggio* se “conta” il numero totale di eventi accaduti fino al tempo t . Si tratta di una speciale classe di processi stocastici, a *tempo continuo* e *stati* (possibili valori) *discreti*. Il numero di persone che sono entrate in un negozio, a partire dall'apertura e fino al tempo t , è un processo di conteggio, ma non lo è il numero di persone presenti nel negozio al tempo t . Per definizione, un processo di conteggio verifica:

- [i] $N(t) \geq 0$;
- [ii] $N(t)$ assume valori interi;
- [iii] se $s \leq t$ allora $N(s) \leq N(t)$;
- [iv] se $s < t$ allora $N(t) - N(s)$ conta il numero di eventi accaduti in $(s, t]$.



Definizione 5.1 Un processo di conteggio ammette incrementi indipendenti se per ogni $t \geq 0$, $h > 0$, le v.a. $N(t+h) - N(t)$ e $N(t)$ sono tra loro indipendenti.

Un processo di conteggio ammette incrementi stazionari se per ogni $h > 0$ fissato e per ogni $t, s \geq 0$ le v.a. $N(t+h) - N(t)$ e $N(s+h) - N(s)$ hanno la stessa distribuzione.

I processi di conteggio sono caratterizzati dalla successione dei tempi di arrivo dei vari eventi. In generale, per un processo di conteggio ad incrementi indipendenti e stazionari si verifica che i tempi di attesa tra due eventi consecutivi sono indipendenti ed equidistribuiti, con distribuzione arbitraria. Parleremo in tal caso di *processi di rinnovo* (come vedremo, un processo di Poisson è un processo di rinnovo, dove la distribuzione dei tempi di attesa è di tipo esponenziale).

5.2 Processo di Poisson

Un comune e semplice esempio di processi di conteggio è fornita dai processi di Poisson. La loro importanza è dovuta al gran numero di fenomeni fisici che possono essere descritti (almeno in prima approssimazione) da un tale processo.

Definizione 5.2 Il processo $N(t)$, $t \geq 0$, è detto essere un processo di Poisson con intensità (tasso, rate) λ se

$$[i] \ N(0) = 0;$$

[ii] il processo ha incrementi indipendenti;

[iii] per ogni $t \geq 0$, per ogni $h > 0$, la v.a. $N(t+h) - N(t)$ ha distribuzione di Poisson di parametro λh .

Osserviamo che, conseguenza della proprietà [iii], la media di $N(t)$ è pari a λt e che il processo ha incrementi stazionari. L'intensità λ ha le dimensioni dell'inverso del tempo, e rappresenta il numero di arrivi per unità di tempo.

Osserviamo che, per verificare che un processo stocastico è di Poisson, dovremo verificare le condizioni precedenti, e se le condizioni [i] e [ii] si possono determinare dalla conoscenza del processo, meno chiaro risulta come verificare la condizione [iii]. Daremo ora una definizione equivalente di processo di Poisson, che potrà essere utile negli esempi.

Definizione 5.3 Il processo $N(t)$, $t \geq 0$, è detto essere un processo di Poisson con intensità λ se

$$[a] \ N(0) = 0;$$

[b] il processo ha incrementi indipendenti e stazionari;

$$[c] \text{ per } h \text{ piccolo, } \mathbb{P}(N(h) = 1) = \lambda h + o(h);^{(1)}$$

$$[d] \text{ per } h \text{ piccolo, } \mathbb{P}(N(h) \geq 2) = o(h).$$

Teorema 5.1 Le due definizioni di processo di Poisson sono equivalenti.

Il fatto che la seconda definizione ci dia un processo di Poisson è una conseguenza del teorema di approssimazione di Poisson di una binomiale. Per vederlo, pensiamo di suddividere l'intervallo $[0, t]$ in $1/h$ intervalli di ampiezza th piccola. Per la proprietà [d], la probabilità di avere due eventi in un singolo intervallo $(t\frac{k}{h}, t\frac{k+1}{h}]$ tende a zero, quindi $N(t)$ tende ad essere uguale al numero di intervalli in cui avviene un evento. In ogni intervallo avviene un evento con probabilità circa pari a λh , quindi il numero totale di eventi avrà distribuzione binomiale di parametri $n = 1/h$ e $p = \lambda th$. Come noto, questa distribuzione converge, per $h \rightarrow 0$, ad una distribuzione di Poisson di parametro $np = \frac{1}{h} \lambda th = \lambda t$.

5.3 Tempi di attesa e tempi di arrivo

Indichiamo con T_n il tempo di attesa tra l'accadere dell'evento $(n-1)$ -esimo e dell'evento n -esimo. Definiamo inoltre $S_0 = 0$,

$$S_n = \sum_{k=1}^n T_k, \quad n \geq 1,$$

¹Una funzione $f(t)$ è detta $o(t)$ se verifica $\lim_{t \rightarrow 0} \frac{f(t)}{t} = 0$. Questo vuol dire che, per $t \rightarrow 0$, $f(t)$ va a zero più rapidamente di quanto lo faccia t . Ancora, possiamo dire che, per t piccolo, $f(t)$ deve essere piccolo rispetto a t .

i tempi di arrivo (tempi di rinnovo) del processo, ossia gli istanti in cui ha luogo il primo, secondo, ecc., evento.

Iniziamo con l'osservare che

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t},$$

quindi T_1 ha distribuzione esponenziale di parametro λ . In particolare, il tempo atteso per il primo evento è $\mathbb{E}[T_1] = \frac{1}{\lambda}$.

Se il primo evento è accaduto al tempo s , la probabilità che nessun nuovo evento accada tra s e $s+t$ è ancora pari a $e^{-\lambda t}$, per la stazionarietà degli incrementi di $N(t)$. Allora, anche T_2 (e in generale T_n) ha distribuzione esponenziale. Inoltre, T_2 è indipendente da T_1 per l'indipendenza degli incrementi di $N(t)$.

Qual è la distribuzione di S_n ? Osserviamo la relazione

$$N(t) \geq n \iff S_n \leq t,$$

e quindi

$$F_{S_n}(t) = \mathbb{P}(S_n \leq t) = \mathbb{P}(N(t) \geq n) = \sum_{j=n}^{\infty} e^{-\lambda t} \frac{1}{j!} (\lambda t)^j;$$

questa distribuzione si chiama *distribuzione Gamma* di parametri λ e n . La densità di probabilità si ottiene derivando l'espressione precedente:

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{1}{(n-1)!} (\lambda t)^{n-1}.$$

5.3.1 Distribuzione condizionata

Se un evento ha avuto luogo in $[0, t]$, ossia, se $N(t) = 1$, qual è la distribuzione di probabilità del suo accadere?

Si ricava da un calcolo diretto che questa distribuzione è uniforme in $[0, t]$:

$$\begin{aligned} \mathbb{P}(T_1 \leq s \mid N(t) = 1) &= \frac{\mathbb{P}(T_1 \leq s, N(t) = 1)}{\mathbb{P}(N(t) = 1)} = \frac{\mathbb{P}(N(s) = 1, N(t) - N(s) = 0)}{N(t) = 1} \\ &= \frac{\mathbb{P}(N(s) = 1) \mathbb{P}(N(t) - N(s) = 0)}{\mathbb{P}(N(t) = 1)} = \frac{e^{-\lambda s} (\lambda s) e^{-\lambda(t-s)}}{e^{-\lambda t} (\lambda t)} = \frac{s}{t}. \end{aligned}$$

Esempio. Un processo di Poisson con tempo aleatorio

Un ulteriore esempio di commistione tra processi di Poisson e tempi aleatori è dato dal seguente modello. Supponiamo che gli ingressi delle persone in una stazione seguano un processo di Poisson di parametro $\lambda = 20$ persone all'ora. Possiamo supporre che inizialmente la stazione sia vuota; il prossimo treno giungerà in un tempo aleatorio T , distribuito uniformemente nell'intervallo $[0, 1]$ (espresso in ore). Se tutti coloro che sono giunti in stazione prima del treno salgono, quante persone, in media, partiranno?

Soluzione.

Indichiamo con $N(t)$ le persone arrivate entro il tempo t . La distribuzione di $N(t)$ è di Poisson, di parametro λt . Per calcolare la distribuzione di $N(T)$ possiamo scrivere

$$\mathbb{P}(N(T) = x) = \int_0^1 \mathbb{P}(N(T) = x \mid T = s) f_T(s) ds = \int_0^1 \mathbb{P}(N(s) = x) ds;$$

per ottenere la media, sommiamo su tutti i possibili valori di x :

$$\mathbb{E}[N(T)] = \sum_{x=0}^{\infty} x \int_0^1 \mathbb{P}(N(s) = x) ds$$

scambiando l'ordine di somma e integrale otteniamo

$$\mathbb{E}[N(T)] = \int_0^1 \sum_{x=0}^{\infty} x \mathbb{P}(N(s) = x) ds = \int_0^1 \mathbb{E}[N(s)] ds = \int_0^1 \lambda s ds = \frac{1}{2} \lambda,$$

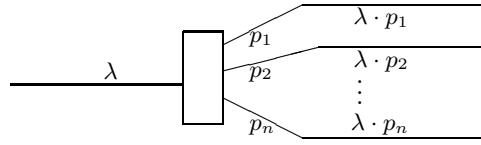
quindi, in media, partiranno 10 persone.

△

5.4 Ulteriori proprietà

5.4.1 Decomposizione di un processo di Poisson

Supponiamo di avere un processo di Poisson, di intensità λ , in cui ogni evento può essere classificato come evento di tipo $1, 2, \dots, k$ con probabilità rispettivamente p_1, \dots, p_k ($p_j \geq 0, p_1 + \dots + p_k = 1$). Costruiamo i processi stocastici $N_1(t), \dots, N_k(t)$ che contano il numero di eventi di classe $1, \dots, k$ rispettivamente, accaduti fino al tempo t . Risulta che essi sono ancora processi di Poisson, di intensità rispettivamente $p_1\lambda, \dots, p_k\lambda$, indipendenti.



Per dimostrarlo, consideriamo il caso $k = 2$: basterà mostrare che la distribuzione congiunta di $N_1(t)$ e $N_2(t)$ è pari al prodotto di due distribuzioni di Poisson, di intensità rispettivamente $p_1\lambda$ e $p_2\lambda$.

La seguente osservazione è cruciale. Condizionato al fatto che $N(t) = h + k$, la distribuzione di $N_1(t)$ è binomiale, di parametri p_1 e $h + k$, rispettivamente. Inoltre, dato che esistono solo due classi, questo identifica anche la distribuzione di $N_2(t)$. Allora:

$$\begin{aligned} \mathbb{P}(N_1(t) = k, N_2(t) = h) &= \mathbb{P}(N_1(t) = k, N_2(t) = h, N(t) = h + k) \\ &= \mathbb{P}(N_1(t) = k, N_2(t) = h \mid N(t) = h + k) \mathbb{P}(N(t) = h + k) \\ &= \binom{h+k}{k} p_1^k p_2^h e^{-\lambda t} \frac{(\lambda t)^{h+k}}{(h+k)!} = e^{-\lambda p_1 t} \frac{(\lambda p_1 t)^k}{k!} e^{-\lambda p_2 t} \frac{(\lambda p_2 t)^h}{h!} \end{aligned}$$

5.4.2 Processi con divisione in classi dipendente dal tempo

Abbiamo visto che, partendo da un processo di Poisson di intensità λ e classificando ogni evento in una delle possibili classi $1, 2, \dots, k$ con probabilità rispettivamente p_1, \dots, p_k ($p_j \geq 0, p_1 + \dots + p_k = 1$), i processi stocastici $N_1(t), \dots, N_k(t)$ siano ancora processi di Poisson, di intensità rispettivamente $p_1\lambda, \dots, p_k\lambda$, indipendenti.

Ci occupiamo ora del caso in cui la probabilità che un evento sia classificato di classe i dipende dall'istante in cui si è verificato. Specificatamente, supponiamo che un evento che avviene al tempo t si classifichi, indipendentemente da quanto successo in precedenza, come evento di classe i con probabilità $p_i(t)$.

Teorema 5.2 *Se indichiamo con $N_1(t), \dots, N_k(t)$ gli eventi di tipo $1, 2, \dots, k$ accaduti entro il tempo t , allora $N_1(t), \dots, N_k(t)$ sono v.a. di Poisson, indipendenti, aventi media rispettivamente*

$$\mathbb{E}[N_i(t)] = \lambda \int_0^t p_i(s) ds.$$

Esempio. Un server con infinite linee: distribuzione delle chiamate attive

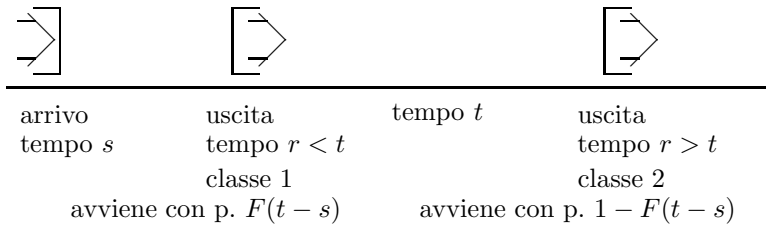
Supponiamo che le chiamate giungano ad un centralino secondo un processo di Poisson di intensità λ . Al momento dell'arrivo, ogni chiamata viene instradata lungo una delle infinite linee del centralino, e supponiamo che il tempo di servizio (il tempo in cui la linea rimane occupata) sia dato da una v.a. con f.d.r. fissata $F(x)$, indipendente da tutti gli altri eventi. Fissato un tempo $t > 0$, vogliamo determinare la distribuzione delle v.a. $X(t)$ e $Y(t)$ che contano, rispettivamente, le chiamate completate entro il tempo t e le chiamate in funzione al tempo t .

Decidiamo allora di chiamare evento di tipo 1 se una chiamata si chiude entro il tempo t , evento di tipo 2 se la chiamata è in atto al tempo t . Ora, se una chiamata giunge al tempo $s < t$, sarà classificata di tipo 1 se il suo tempo di servizio è minore di $t - s$, quindi con probabilità $F(t - s)$, altrimenti sarà di tipo 2. Per il teorema precedente, le distribuzioni di $X(t)$ e di $Y(t)$ sono indipendenti, di Poisson, di media

$$\mathbb{E}[X(t)] = \lambda \int_0^t F(t - s) ds = \lambda \int_0^t F(s) ds$$

e anche

$$\mathbb{E}[Y(t)] = \lambda \int_0^t (1 - F(t - s)) ds = \lambda \int_0^t (1 - F(s)) ds.$$



5.4.3 Costruzione di un processo di Poisson spaziale

Finora abbiamo considerato il caso in cui il parametro t rappresenta un tempo. Possiamo anche estendere questa costruzione a casi più generali.

Sia S una regione di spazio, di volume $|S|$, e consideriamo un insieme di regioni B_1, \dots, B_k , a due a due disgiunte, con unione S . È noto che, fissato un numero n di particelle che suppongo distribuite uniformemente in S , la distribuzione congiunta delle v.a. N_1, \dots, N_k (che contano le particelle nelle rispettive regioni) è data da: per ogni scelta di interi n_1, \dots, n_k , $n_j \geq 0$ e $\sum n_j = n$

$$\mathbb{P}(N_1 = n_1, \dots, N_k = n_k) = \frac{n!}{(n_1)! \dots (n_k)!} \left(\frac{|B_1|}{|S|} \right)^{n_1} \dots \left(\frac{|B_k|}{|S|} \right)^{n_k}.$$

Supponiamo ora che il numero di particelle non sia fissato a priori, ma si possa considerare come v.a. di tipo Poisson, di parametro $\lambda|S|$. Allora, il numero di particelle nella regione B_j è distribuito come v.a. di tipo Poisson, di parametro $\lambda|B_j|$; inoltre, le v.a. N_1, \dots, N_k sono tra loro indipendenti.

5.4.4 Comportamento asintotico

Possiamo domandarci se è possibile che un processo di Poisson (più in generale, un processo di rinnovo) assuma un valore infinito in un tempo finito. Questo non può succedere; per vederlo, osserviamo la relazione

$$N(t) = \max\{n \mid S_n \leq t\};$$

ora, dalla legge dei grandi numeri segue che

$$\frac{S_n}{n} \rightarrow 1/\lambda = \mathbb{E}[T], \quad \text{per } n \rightarrow \infty.$$

Questo implica che S_n deve essere dello stesso ordine di grandezza di n , quindi può essere limitato da t solo per un numero finito di n ; allora, anche $N(t)$ deve essere finito.

Tuttavia, anche se $N(t) < \infty$ per ogni t , è vero che (con probabilità 1)

$$N(\infty) = \lim_{t \rightarrow \infty} N(t) = \infty.$$

Infatti, l'unico modo per cui $N(\infty)$ sia finito è che un qualche tempo di attesa T_i assuma valore infinito, ma questo accade con probabilità zero.

Un altro interessante risultato di tutti i processi di rinnovo, che citiamo senza dimostrazione, è che

$$m(t) := \mathbb{E}[N(t)] < \infty \quad \text{per tutti i } t > 0.$$

Citiamo da S. Ross: “Qualche lettore potrebbe supporre che la finitezza di $m(t)$ segue direttamente dal fatto che, con probabilità 1, $N(t)$ è finito. Tuttavia, questo ragionamento è errato” (sapete dare un esempio di v.a. a valori interi positivi, con media infinita?)

Tra l'altro, si osservi che per un processo di Poisson $m(t) = \lambda t$ è una funzione *lineare* di t , e che, il processo di Poisson è l'unico processo di rinnovo per cui questo è vero.

5.4.5 Teoremi limite

Abbiamo già mostrato come $N(t)$ diverga per $t \rightarrow \infty$. Possiamo domandarci in quale modo ciò accada. Il primo risultato è una legge dei grandi numeri per i processi di Poisson.

Teorema 5.3 *Con probabilità 1,*

$$\frac{N(t)}{t} \rightarrow \lambda \quad \text{per } t \rightarrow \infty.$$

Prima di accennare la dimostrazione, consideriamo la v.a. $S_{N(t)}$. Ci possiamo domandare cosa rappresenta questa v.a.: ad esempio, se $N(t) = 3$, allora S_3 è il tempo di arrivo del terzo evento, e quindi $S_3 < t$, $S_4 > t$; $S_{N(t)}$ rappresenta il tempo di arrivo dell'ultimo evento accaduto prima del tempo t (e similmente $S_{N(t)+1}$ il tempo di arrivo del primo evento dopo t).

Dalla relazione

$$S_{N(t)} \leq t < S_{N(t)+1}$$

si ottiene

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)};$$

ma la legge dei grandi numeri⁽²⁾ implica che $S_{N(t)}/N(t) \rightarrow \frac{1}{\lambda}$ se $N(t) \rightarrow \infty$, cioè per $t \rightarrow \infty$, e analogamente per $\frac{S_{N(t)+1}}{N(t)}$; allora, per il teorema dei carabinieri (unicità del limite), anche $t/N(t) \rightarrow \frac{1}{\lambda}$. △

Abbiamo infine il seguente risultato, che citiamo senza dimostrazione.

²in una versione più forte di quella che abbiamo dimostrato nel corso

Teorema 5.4 Dato un processo di Poisson di parametro λ ,

$$\lim_{t \rightarrow \infty} \mathbb{P} \left(\frac{N(t) - \lambda t}{\sqrt{t\lambda}} < x \right) = \Phi(x) = \int_{-\infty}^x e^{-y^2/2} dy.$$

5.5 Alcune applicazioni alla teoria delle code

Esempio. Server singolo senza coda

Supponiamo che i potenziali clienti di una banca si presentino secondo un processo di Poisson di parametro λ , ma che entrino nella banca solamente se lo sportello è libero. Il tempo che un cliente rimane nella banca ha f.d.r. $F(t)$ fissata, indipendentemente dagli altri clienti. Possiamo allora domandarci:

- (a) qual è il tasso di ingresso dei clienti nella banca?
- (b) qual è la frazione dei potenziali clienti che entrano?

Per rispondere a queste domande, supponiamo che il primo cliente si presenti al tempo 0. Questo, in media, rimane nella banca un tempo μ_F (dove $\mu_F = \int_0^\infty (1 - F(x)) dx$); a questo punto, per le proprietà di rinnovo del processo di Poisson, il tempo medio per vedere comparire un altro potenziale cliente è $1/\lambda$, quindi il tempo medio tra due clienti consecutivi è $\mu = \mu_F + \frac{1}{\lambda}$. Allora, il tasso a cui i clienti entrano è

$$\frac{1}{\mu} = \frac{\lambda}{1 + \lambda\mu_F}.$$

D'altra parte, dato che il tasso di arrivo dei clienti potenziali è λ , risulta che la percentuale di coloro che entrano è

$$\frac{1/\mu}{\lambda} = \frac{1}{1 + \lambda\mu_F}.$$

△

Esempio. Server singolo con coda infinita

Consideriamo il caso di clienti che arrivano ad un server, se necessario si mettono in coda ad aspettare il proprio turno, quindi escono dal meccanismo. Supponiamo che il tempo di ingresso abbia distribuzione di tipo esponenziale di parametro λ , quindi che il processo di ingresso dei clienti sia un processo di Poisson. Supponiamo anche che il tempo necessario per essere servito abbia distribuzione esponenziale di parametro μ .

Le quantità che ci interessano sono date da

L , il numero medio di clienti nel sistema e

W , il tempo medio che il cliente spende nel sistema.

Queste quantità sono legate dall'*equazione di Little*

$$L = \lambda W.$$

Indichiamo con $X(t)$ il numero di clienti nel sistema al tempo t ($X(t)$ non è un processo di conteggio). Definiamo per $n \geq 0$

$$P_n = \lim_{t \rightarrow \infty} \mathbb{P}(X(t) = n)$$

la probabilità limite che vi siano esattamente n clienti nel sistema. Noi intenderemo queste quantità anche come le probabilità di verificare che, a regime, nel sistema vi siano esattamente n clienti, ed anche, in analogia con quanto visto per i processi di Markov, la frazione di tempo in cui il sistema ha esattamente n clienti.

Per analizzare il sistema, abbiamo a disposizione un principio generale che ci permette di determinare le probabilità P_n , ossia, che per ogni $n \geq 0$ il tasso in cui il sistema entra lo stato n è pari al tasso con cui il sistema lascia lo stato n ⁽³⁾. Determiniamo allora questi tassi. Consideriamo, per iniziare, il livello 0. Il sistema può lasciare il livello 0 solo se una nuova chiamata entra, e questo accade con un tasso pari a λP_0 . Viceversa, il sistema raggiunge lo stato 0 solo dallo stato 1, nel caso in cui la chiamata sia terminata, e questo accade con un tasso pari a μP_1 . Di conseguenza, dal principio di eguaglianza dei tassi, si ricava la prima equazione:

$$\lambda P_0 = \mu P_1.$$

Ora consideriamo lo stato 1. Il sistema può uscire da questo stato sia verso lo stato 0 (nel caso una chiamata sia terminata), sia verso lo stato 2 (nel caso arrivi una nuova chiamata). Dato che la frazione di tempo trascorsa dal sistema nello stato 1 è P_1 , il tasso di uscita da questo stato è $(\lambda + \mu)P_1$. D'altro canto, il sistema raggiunge lo stato 1 sia dallo stato 0 (nel caso entri una nuova chiamata) sia dallo stato 2 (nel caso una chiamata termini), quindi il tasso di ingresso è $\lambda P_0 + \mu P_2$. Questo ragionamento vale, in generale, per ogni stato $n \geq 1$, il che ci porta all'eguaglianza

$$(\lambda + \mu)P_n = \lambda P_{n-1} + \mu P_{n+1}, \quad \forall n \geq 1.$$

Le equazioni che abbiamo scritto, unite alla condizione che la somma dei P_n deve essere 1, conducono alla seguente soluzione⁽⁴⁾

$$P_0 = (1 - \frac{\lambda}{\mu}), \quad P_n = (1 - \frac{\lambda}{\mu}) \left(\frac{\lambda}{\mu} \right)^n, \quad n \geq 1.$$

Ora possiamo calcolare le quantità L e W introdotte in precedenza. Dalla relazione

$$L = \sum_{n=0}^{\infty} n P_n$$

si ottiene $L = \frac{\lambda}{\mu - \lambda}$, e dalla formula di Little $W = \frac{1}{\mu - \lambda}$.

Supponiamo che il numero di chiamate in ingresso segua un processo di Poisson di intensità $\lambda = 5$ chiamate ogni ora, e che il tempo medio per terminare una chiamata sia di $1/\mu = 8$ minuti. Allora (portando le grandezze alle stesse unità di misura) $\mu = 7,5$ chiamate ogni ora,

$$L = \text{media delle chiamate attive nel sistema} = 2,$$

$$W = \text{permanenza media nel sistema} = \frac{2}{5} = 24 \text{ minuti.}$$

Ora, supponiamo che l'intensità degli ingressi aumenti, passando da 5 a 6 chiamate ogni ora. Come cambiano L e W ? Si ottiene

$$L = 4, \quad W = 40 \text{ minuti.}$$

Vale a dire che un aumento del 20% degli ingressi si traduce in un raddoppio del numero medio di clienti nel sistema, ed in un aumento del 66% del tempo trascorso nel sistema.

³pensiamo ad un indice che è stato programmato per segnare lo stato del sistema, ossia il numero di clienti in coda. Allora, se sappiamo che in un certo periodo di tempo si è spostato fino al livello 1 k volte, sappiamo anche che il numero di volte che ha lasciato il livello 1 è pari a k – al massimo, si può discostare per una unità in più o in meno.

⁴questo ha senso se e solo se $\lambda < \mu$, altrimenti non esiste uno stato di regime, ma la coda tende a divergere

5.6 Esercizi conclusivi

Esercizio.

I clienti di un esercizio entrano al tasso di 1 al minuto; dalle nostre osservazioni, risulta che la probabilità che un cliente sia donna è doppia della probabilità che sia uomo.

- (a) Descrivere il processo che conta il numero di uomini entranti. Qual è la probabilità che nei primi 10 minuti entrino 8 uomini?
- (b) Sapendo che nei primi 10 minuti sono entrati 8 uomini, qual è la probabilità che siano entrate 6 donne? Qual è il numero medio di donne entrate in quei 10 minuti?
- (c) Qual è la probabilità che, tra i prossimi 5 clienti, vi siano almeno 2 uomini?

Esercizio.

Ogni pc dell'università è soggetto a rottura secondo una distribuzione esponenziale di media $1/\lambda = 30$ giorni. L'università ha un solo tecnico, che aggiusta ogni macchina in un tempo aleatorio, indipendentemente dalle altre, con distribuzione esponenziale di media $1/\mu = 10$ giorni. Se il tecnico è occupato, il pc viene rottamato.

Al tempo $t = 0$, nessuna macchina è rotta. Riconoscere il tipo di processo e rispondere alle seguenti domande:

- (a) Qual è il tempo medio tra due interventi del tecnico?
- (b) Qual è la frazione di pc rotti che viene aggiustata?

Esercizio.

Ogni pc dell'università è soggetto a rottura secondo una distribuzione esponenziale di media $1/\lambda = 30$ giorni. L'università ha un solo tecnico, che aggiusta ogni macchina in un tempo aleatorio, indipendentemente dalle altre, con distribuzione esponenziale di media $1/\mu = 10$ giorni. Se il tecnico è occupato, il pc viene messo in coda e riparato in un secondo momento.

Al tempo $t = 0$, nessuna macchina è rotta. Riconoscere il tipo di processo e rispondere alle seguenti domande:

- (a) Qual è il tempo medio necessario perché un pc venga aggiustato?
- (b) Qual è, in media, il numero di pc rotti che aspettano l'intervento del tecnico?

Esercizio.

Ogni pc dell'università è soggetto a rottura secondo una distribuzione esponenziale di media $1/\lambda = 30$ giorni. L'università ha numerosi tecnici (un numero così alto che possiamo supporre infinito), e ogni macchina si aggiusta in un tempo aleatorio, indipendentemente dalle altre, con distribuzione esponenziale di media $1/\mu = 10$ giorni.

Al tempo $t = 0$, nessuna macchina è rotta. Riconoscere il tipo di processo e rispondere alle seguenti domande:

- (a) Qual è la distribuzione del numero di pc che si sono rotti dopo un anno (tempo $t = 365$)?
- (b) Qual è la distribuzione del numero di pc che sono stati aggiustati dopo un anno? Quanti sono, in media, quelli che sono in riparazione?

Esercizio.

Dato un processo di Poisson di intensità $\lambda = 2$,

- (a) determinare la probabilità che nell'intervallo $[0, 2]$ si verifichino 2 arrivi e nell'intervallo $(5, 7]$ si verifichino 4 arrivi.
- (b) determinare la probabilità che nell'intervallo $[0, 2]$ si verifichino 0 arrivi e nell'intervallo $(1, 7]$ si verifichino 4 arrivi
- (c) determinare la probabilità che nell'intervallo $[0, 2]$ si verifichino 2 arrivi e nell'intervallo $(1, 7]$ si verifichino 4 arrivi

Esercizio.

Gli operai di una azienda sono divisi su tre turni, ed inizialmente il numero di operai in ciascun turno è determinato da una v.a. con distribuzione di Poisson di parametro λ_i , $i = 1, 2, 3$. Ogni giorno, un operaio cambia il proprio turno, indipendentemente dagli altri operai, secondo un processo di Markov di matrice di transizione P . Supponiamo che P sia regolare, con distribuzione invariante π .

- (a) Determinare il numero medio di operai su ogni turno al tempo $t = 0$.
- (b) Determinare la distribuzione del numero di operai su ogni turno, al tempo $t = 10$ gg., e la sua media.
- (c) Determinare la media del numero di operai su ogni turno, per un tempo t molto grande.

Soluzione.

Dato che la distribuzione di Poisson di parametro λ ha media λ , il numero medio di operai in ogni turno è λ_i , $i = 1, 2, 3$.

Consideriamo un operaio del primo turno. Dopo 1 giorno, si troverà nel turno i , $i = 1, 2, 3$, con probabilità $p_{1,i}$. Dopo due giorni, si troverà nel turno i con probabilità $p_{1,i}^{(2)}$, dove (in generale) $(p_{i,j}^{(n)})$ è la matrice di transizione a n passi P^n . Per calcolare il numero di operai che al primo giorno passa dal turno 1 al turno i , possiamo pensare a dividere in classi il processo di Poisson (operai del primo turno), secondo le probabilità $p_{1,i}$. Allora, il numero di operai che al primo giorno passa dal turno 1 al turno i è distribuito come una v.a. di Poisson di parametro $\lambda_1 p_{1,i}$. Al tempo $n = 10$, il numero di operai che è passato dal turno 1 al turno i è distribuito come una v.a. di Poisson di parametro $\lambda_1 p_{1,i}^{(10)}$. Si ha allora che il numero di operai che al tempo $n = 10$ sono nel turno i è pari alla somma di tre distribuzioni di Poisson indipendenti, quindi ha distribuzione di Poisson di parametro

$$\lambda_i^{(10)} = \lambda_1 p_{1,i}^{(10)} + \lambda_2 p_{2,i}^{(10)} + \lambda_3 p_{3,i}^{(10)},$$

e questo numero è anche la sua media.

Per n molto grande, dato che la matrice è regolare, si ha che $P^n \simeq \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \end{pmatrix}$ quindi per n grande il numero medio di operai al turno i tende ad essere costante pari a

$$\bar{\lambda}_i = \lambda_1 \pi_i + \lambda_2 \pi_i + \lambda_3 \pi_i = (\lambda_1 + \lambda_2 + \lambda_3) \pi_i.$$

△

Esercizio.

Nelle confezioni di merendine sono in regalo 10 personaggi di una raccolta, e la probabilità di trovare il tipo i è pari a p_i . Qual è il numero medio di confezioni da comprare per terminare la raccolta?

Soluzione.

Supponiamo di acquistare le confezioni secondo uno schema di Poisson di tasso $\lambda = 1$ confezioni al giorno. Allora $N_1(t), \dots, N_{10}(t)$ – numero di personaggi di tipo i raccolti entro il tempo t – sono processi di Poisson, di intensità $p_i\lambda$, $i = 1, \dots, 10$. Indico con T_1, \dots, T_{10} i tempi di primo arrivo di un evento di tipo i , e $T = \max\{T_1, \dots, T_{10}\}$. Per l'indipendenza dei processi N_i , anche le v.a. T_i sono indipendenti, e posso determinare la distribuzione di T :

$$\mathbb{P}(T \leq t) = \prod_{i=1}^{10} \mathbb{P}(T_i \leq t) = \prod_{i=1}^{10} (1 - e^{-\lambda p_i t})$$

da cui ricavo anche la media

$$\mathbb{E}[T] = \int_0^\infty \mathbb{P}(T > t) dt = \int_0^\infty \left(1 - \prod_{i=1}^{10} (1 - e^{-\lambda p_i t}) \right) dt.$$

Quale relazione esiste tra il tempo T e il numero di scatole $N(T)$ che compro? So che acquisto scatole in media al ritmo di una ogni giorno, ossia, se t fosse deterministico, $\mathbb{E}[N(t)] = t = \mathbb{E}[t]$. Voglio mostrare che questa relazione rimane vera per T : posto $f_T(t)$ la densità di T ,

$$\mathbb{E}[N(T)] = \int_0^\infty \mathbb{E}[N(T) \mid T = t] f_T(t) dt = \int_0^\infty t f_T(t) dt = \mathbb{E}[T]$$

dove l'ultima eguaglianza segue dalla definizione di media di una v.a.

△

Chapter 6

Statistica

La statistica riguarda lo studio di popolazioni di individui (dove entrambi i termini, popolazione e individui, devono intendersi in senso astratto): il carattere particolare che stiamo investigando, una popolazione presenta sempre una variabilità interna che è l'argomento dello studio statistico. Una parte della statistica, la statistica descrittiva, ha a che fare con la sintesi delle informazioni contenute nella descrizione di un campione estratto dalla popolazione; un'altra parte, la statistica inferenziale, vuole estrarre dalle informazioni che possiamo ricavare dal campione delle affermazioni che riguardano l'intera popolazione.

La *statistica descrittiva* è la parte che si occupa dell'analisi dei dati osservati, al fine di effettuare quella riduzione dei dati di cui si è detto prima. L'informazione rilevante può essere espressa mediante grafici, oppure indici numerici, che descrivano la distribuzione di una variabile sulla popolazione considerata.

Il problema a cui risponde la *statistica inferenziale* è quello di dire qualcosa sulla popolazione complessiva in esame, a partire dalle osservazioni fatte su un campione estratto dalla popolazione stessa. Le conclusioni cui si giunge non saranno certe, ma dovranno essere espresse nei termini del calcolo delle probabilità.

6.1 Statistica descrittiva

Abbiamo di fronte a noi il problema di analizzare una *variabile* (osservabile) che descrive una particolare proprietà di (un *campione*⁽¹⁾) estratto da) una popolazione.

Iniziamo con distinguere le osservazioni in tre tipi, a seconda del tipo di dati che osserviamo

tipi di variabili: $\left\{ \begin{array}{l} \text{numeriche} \\ \text{categoriche} \end{array} \right\} \left\{ \begin{array}{l} \text{discrete} \\ \text{continue} \end{array} \right\}$

I risultati dell'osservazione statistica producono un insieme di dati, che devono essere descritti:

insieme di dati: $\left\{ \begin{array}{l} \text{rappresentazione} \\ \text{sintesi} \end{array} \right\} \left\{ \begin{array}{l} \text{tabelle} \\ \text{grafici} \\ \text{indici di posizione} \\ \text{indici di dispersione} \end{array} \right\}$

¹Un campione è un sottoinsieme di una popolazione che è ottenuto tramite un qualche processo di scelta, magari prendendo a caso oppure attraverso un qualche insieme di regole, allo scopo di studiare le proprietà della popolazione globale. Perché il campione sia rappresentativo della popolazione, ciascun elemento deve avere la stessa probabilità di essere considerato.

Esempio 6.1 La tabella seguente contiene i dati relativi al numero di richieste di intervento che raggiungono un centralino, in 40 periodi consecutivi. Si tratta di un esempio di variabile numerica discreta.

0	2	1	4	3	1	2	2	3	8	0	2	1	3	3	1	3	2	2	5
5	4	4	2	4	3	3	1	1	2	4	4	2	3	3	3	3	3	5	2

La rappresentazione di questi dati avverrà sotto forma di *tabella di distribuzione di frequenza*. Nella prima colonna riportiamo le diverse classi (i valori osservati); nella seconda colonna il numero di osservazioni (frequenza assoluta, o semplicemente sequenza); nella terza colonna la frequenza relativa (questi valori sono gli analoghi della distribuzione di una v.a.); nella quarta colonna la frequenza cumulativa (in analogia con la funzione di ripartizione di una v.a.). La somma dei valori della seconda colonna è pari al numero di osservazioni; la somma degli elementi della terza colonna deve dare 1. In generale, non sempre saranno presenti tutte le colonne.

classi	frequenza	freq. rel.	freq. cum.
0	2	0,05	0,05
1	6	0,15	0,20
2	10	0,25	0,45
3	12	0,30	0,75
4	6	0,15	0,90
5	3	0,075	0,975
8	1	0,025	1,00
Σ	= 40	= 1,00	

Esempio 6.2 La costruzione di particolari sfere di metallo è soggetta a errori casuali, per cui il diametro della sfera prodotta varia; i risultati di un campione (espressi in cm.) sono riportati nella seguente tabella

2,08	1,72	1,90	2,11	1,82	2,04	2,04	1,82	2,04	2,07
1,79	1,86	1,80	1,91	1,84	1,86	1,80	1,85	2,08	2,03

In questo caso, la variabile osservata è di tipo *numerico continuo*, dato che a priori può assumere ogni valore (in un certo intervallo di variabilità, che in questo esempio è dato da $(1,70 \dots 2,20)$). Dividendo questo intervallo in classi di ampiezza opportuna⁽²⁾ possiamo compilare ancora la tabella di distribuzione di frequenza; ciascun punto di mezzo delle classi di intervallo da ora in poi rappresenterà tutti i valori del campione nell'ambito della classe.

Un secondo modo di rappresentare questi dati è attraverso un **diagramma steam and leaf**. Questo è il nome di un grafico in cui le osservazioni vengono tabulate, divise per classi, in modo crescente all'interno di ogni classe; il risultato fornisce, comunque, una rappresentazione visiva simile ad un istogramma (vedi esempio seguente).

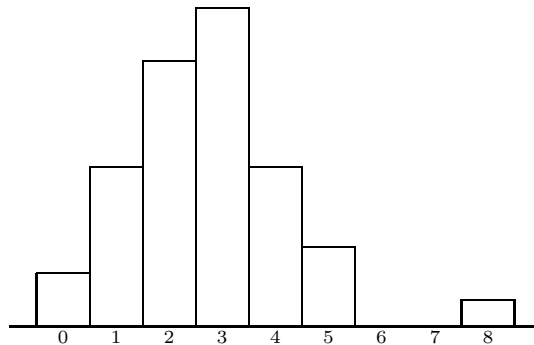
Costruiamo un diagramma *steam and leaf* in questo esempio, utilizzando due classi per ogni 0,1 cm.:

²Non esiste una regola per decidere qual è l'ampiezza "giusta"; si tratta di trovare un equilibrio caso per caso

ampiezze piccole (tante classi):	ampiezze grandi poche classi
⊕ maggiore dettaglio	⊕ maggiore chiarezza
⊖ maggiore peso, minore leggibilità	⊖ perdita di informazione

1,7.	2				
1,7 ⁺	9				
1,8.	0	0	2	2	4
1,8 ⁺	5	6	6		
1,9.	0	1			
1,9 ⁺					
2,0.	3	4	4	4	
2,0 ⁺	7	8	8		
2,1.	1				
2,1 ⁺					

Esempio 6.3 I dati dell'esempio 6.1 possono essere graficamente rappresentati sotto forma di istogramma

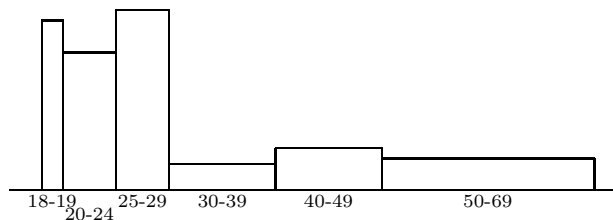


In un istogramma, l'area (base per altezza) di ogni rettangolo è proporzionale alla frequenza. Nel nostro caso, le basi sono uguali per tutti i rettangoli, in tal modo possiamo leggere le frequenze anche sull'asse delle ordinate.

Esempio 6.4 Nella seguente tabella sono riportati, già divisi per classi, le età in cui un campione ha conseguito la patente

Età	18-19	20-24	25-29	30-39	40-49	50-69
frequenza	8	13	17	5	8	12

I dati possono essere rappresentati sotto forma di istogramma; le basi dei rettangoli sono proporzionali alle ampiezze (diverse) delle classi, quindi non ha senso leggere le altezze dei rettangoli, ma solo le loro aree



6.1.1 Indici

Ci sono diversi concetti di “centro” di una distribuzione di frequenza; noi saremo interessati a due di questi: la *media campionaria* e la *mediana*.

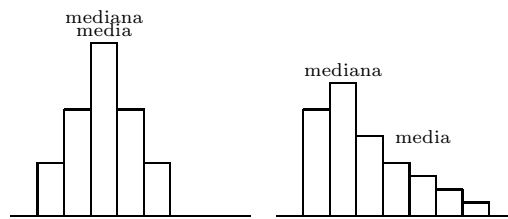
La *media aritmetica* è il più comune indice di posizione. Le osservazioni vengono sommate tra loro e quindi divise per n , così

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

dove x_i indica l' i -esima osservazione del campione.

La *mediana* è il valore al di sotto del quale cade la metà delle osservazioni. Se il numero di osservazioni è dispari, diciamo $2n+1$, allora esiste un solo valore centrale, x_{n+1} , e la mediana corrisponde a tale valore; se il numero di osservazioni è pari, diciamo $2n$, allora si sceglie la media aritmetica tra i due elementi centrali $\frac{1}{2}(x_n + x_{n+1})$.

Nella prossima figura confrontiamo le posizioni di media e mediana; in (a), la distribuzione è simmetrica, e media e mediana coincidono; in (b), la distribuzione presenta una coda a destra: la mediana si trova a sinistra della media.



Quartili

La mediana è anche detta il 50-esimo percentile, cioè il valore a sinistra del quale si trova il 50% delle osservazioni; analogamente, si definiscono il primo e il terzo quartile Q_1 e Q_3 come i valori a sinistra dei quali si trovano rispettivamente il 25% e il 75% dei valori osservati⁽³⁾.

A titolo di esempio, si considerino i seguenti valori

Osservazioni	6 44 47 49 15 43 8 41 7 39 43 41 42 45 46 36 54 14
Osservazioni ordinate	6 7 8 14 15 36 39 41 41 42 43 43 44 45 46 47 49 54
Quartile superiore	45
Quartile inferiore	15

Indici di dispersione

Benché la media possa essere la più importante caratteristica statistica del campione, è anche importante conoscere come si distribuiscono i dati attorno ad essa. Come nel caso degli indici di posizione, vi sono parecchie misure di variabilità. Iniziamo con la più semplice.

Il *range* di un campione è semplicemente la distanza tra il valore maggiore e il valore più piccolo del campione. L'inconveniente del range è che tutte le altre misure vengono ignorate e risulta sensibile ai valori estremi (outlier). Esempio: il range nella serie 65,73,89,56,73,52, 47 è $89 - 47 = 42$.

L'indice *IQR* (inter-quartile range) è la distanza tra il primo e il terzo quartile, e misura l'ampiezza dell'intervallo in cui si accumula il 50% centrale delle osservazioni.

³per determinare Q_1 si procede così: se $n/4$ è intero, allora $Q_1 = \frac{1}{2}(x_{\frac{n}{4}} + x_{\frac{n}{4}+1})$, mentre se $n/4$ non è intero, allora, se k è l'intero seguente, $Q_1 = x_k$

Passiamo poi a considerare la *deviazione standard*; questo indice tiene conto di tutte le osservazioni, e si calcola in analogia con la deviazione standard di una v.a. a partire dalla varianza

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

Esempio 6.5 Torniamo ora ad analizzare i dati dell'esempio 6.1. Riscriviamo la tabella

classi	f_i	$x_i f_i$	$x_i^2 f_i$
0	2	0	0
1	6	6	6
2	10	20	40
3	12	36	108
4	6	24	96
5	3	15	75
8	1	8	64
	$= 40$	$= 109$	$= 389$

La media campionaria è $\bar{x} = \frac{109}{40} = 2,725$; la deviazione standard è $\sigma = 1,516$.

6.2 Statistica inferenziale

Vogliamo ora occuparci dell'analisi sistematica dell'inferenza, cioè del passaggio da un gruppo parziale di informazioni (campione) a quello relativo alla popolazione dalla quale il campione proviene o può provenire.

La popolazione rappresenta l'insieme degli oggetti sui quali s'intende studiare una certa ipotesi. La popolazione viene studiata tramite le osservazioni effettuate su un campione, che viene identificato con una famiglia di v.a. indipendenti ed equidistribuite X_1, \dots, X_n le cui leggi dipendono da un parametro θ (che può essere un numero o un vettore). Per descrivere in maniera esatta la popolazione, saremo interessati a stimare non solo θ ma anche ogni sua funzione $\psi(\theta)$. Una regola che ci dice come calcolare una stima a partire dalle osservazioni contenute in un campione è detta *stimatore*.

Esempio 6.6 Consideriamo una moneta, di cui non conosciamo la parità; prenderemo come campione un certo numero N di lanci, i cui esiti sono espressi come v.a. bernoulliane di parametro p , che sarà il risultato del nostro studio.

Esempio 6.7 Consideriamo una popolazione di lampadine; possiamo supporre che la distribuzione del tempo di vita abbia legge esponenziale $\exp(\lambda)$; saremo interessati a stimare il parametro λ , ma anche $\psi(\lambda) = 1/\lambda$, che rappresenta il tempo medio di vita.

Consideriamo un campione X_1, \dots, X_n : una v.a. $T = t(X_1, \dots, X_n)$, funzione del campione (e quindi dipendente dal parametro θ), è una *statistica*. Diremo che T è uno *stimatore* di $\psi(\theta)$ se, date le osservazioni x_1, \dots, x_n , utilizzeremo il valore (calcolato) $t(x_1, \dots, x_n)$ al posto del valore incognito $\psi(\theta)$.

Consideriamo ad esempio il caso in cui $\psi(\theta) = \mathbb{E}[X]$; vogliamo trovare una statistica T che “stim” il valore vero di questa funzione. Pensiamo ad esempio

- $T_1 = t(X_1, \dots, X_n)$ dove $t(x_1, \dots, x_n) = x_1$,
- $T_2 = t(X_1, \dots, X_n)$ dove $t(x_1, \dots, x_n) = x_1 + \dots + x_n$,
- $T_3 = t(X_1, \dots, X_n)$ dove $t(x_1, \dots, x_n) = \frac{1}{n}(x_1 + \dots + x_n)$,

- $T_4 = t(X_1, \dots, X_n)$ dove $t(x_1, \dots, x_n) = \sqrt[n]{x_1 \cdot \dots \cdot x_n}$

Dato che ognuna delle statistiche precedenti è uno stimatore della media, esiste un criterio per preferirne uno o un'altro?

Diremo che uno stimatore T è corretto (non distorto) se per ogni valore del parametro θ si ha $\mathbb{E}[T] = \psi(\theta)$. Negli esempi precedenti, sia T_1 che T_3 sono stimatori corretti, mentre non lo sono T_2 e T_4 ; però, intuitivamente, T_3 è una stima migliore di T_1 , dato che tiene conto di tutte le osservazioni, mentre T_1 solo della prima. Diamo quindi un'ulteriore caratterizzazione. Diremo che uno stimatore T è *consistente* se, quando l'ampiezza n del campione tende a ∞ , $\text{Var}[T] \rightarrow 0$.

Esempio 6.8 Sia X_1, \dots, X_n un campione estratto da una popolazione, avente media μ e varianza σ^2 finite (ma incognite). Scegliremo allora come stimatore della media, la v.a. media campionaria, che viene solitamente indicata con \bar{X} :

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n).$$

Questo stimatore è corretto e consistente:

$$\mathbb{E}[\bar{X}] = \mu, \quad \text{Var}[\bar{X}] = \frac{1}{n}\sigma^2.$$

△

Esempio 6.9 Riprendiamo il caso dell'esempio 6.7; abbiamo visto che \bar{X} è un buon stimatore di $\mathbb{E}[X] = 1/\lambda$; cosa possiamo dire di λ ?

Consideriamo un campione X_1, \dots, X_n di v.a. con legge $\exp(\lambda)$; allora sappiamo che $Z = X_1 + \dots + X_n$ è l'istante di arrivo dell' n -esimo evento, di cui conosciamo la densità (si veda la sezione 5.3)

$$f_Z(t) = c_n t^{n-1} e^{-\lambda t}, \quad c_n = \frac{\lambda^n}{(n-1)!}.$$

Consideriamo come stimatore di λ l'inverso della media campionaria, $T = \frac{n}{X_1 + \dots + X_n} = \frac{n}{Z}$; è uno stimatore corretto? Calcoliamo

$$\begin{aligned} \mathbb{E}[T] &= n\mathbb{E}[1/Z] = n \int_0^\infty \frac{1}{t} f_Z(t) dt = nc_n \int_0^\infty t^{n-2} e^{-\lambda t} dt \\ &= \frac{nc_n}{c_{n-1}} \underbrace{\int_0^\infty c_{n-1} t^{n-2} e^{-\lambda t} dt}_{=1} = n \frac{\lambda^n}{(n-1)!} \frac{(n-2)!}{\lambda^{n-1}} = \frac{n}{n-1} \lambda. \end{aligned}$$

Ne risulta che uno stimatore corretto di λ è dato da $T_1 = \frac{n-1}{X_1 + \dots + X_n}$.

△

Leggi normali

Consideriamo un campione X_1, \dots, X_n di leggi normali $\mathcal{N}(\mu, \sigma^2)$. È questo un caso in cui il parametro $\theta = (\mu, \sigma^2)$ è bidimensionale, $\theta \in \Theta = \mathbb{R} \times \mathbb{R}_+$.

Abbiamo visto che **un buon stimatore per μ è dato dalla media campionaria** $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$. Ci possiamo chiedere come stimare la varianza σ^2 .

Nel caso la media μ sia nota, possiamo scrivere, in analogia con la definizione di varianza, la statistica

$$T_1 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2;$$

nel caso (più frequente in pratica) in cui anche μ è incognito, possiamo pensare di utilizzare la sua stima \bar{X} per definire la statistica

$$T_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Lemma 6.1 *La statistica T_1 è un buon stimatore per la varianza σ^2 . La statistica T_2 è uno stimatore distorto, in quanto risulta*

$$\mathbb{E}[T_2] = \frac{n-1}{n} \sigma^2.$$

Conseguenza del lemma è che un buon stimatore per la varianza σ^2 è la statistica campionaria

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

6.2.1 Stima per intervalli

Uno dei temi centrali della statistica inferenziale è lo studio degli intervalli di confidenza. In statistica viene solitamente utilizzata come riferimento la probabilità del 95% per definire un evento probabile e la probabilità del 5% per definirlo improbabile. Possiamo allora definire un intervallo (appunto del 95%) per il quale un evento ha una probabilità accettabile di verificarsi. In particolare è interessante definire tale intervallo relativamente ad una data statistica (media o varianza) campionaria: tale intervallo si chiama intervallo di confidenza al 95%.

Nell'intervallo sarà compreso il parametro reale della popolazione nel 95% dei casi, per una serie di intervalli calcolati su altrettanti campioni, di identica numerosità, estratti dalla popolazione. Non sapremo però mai quali di questi intervalli conterranno effettivamente il parametro.

Definizione 6.1 *Un intervallo di confidenza al livello α per la funzione del parametro $\psi(\theta)$ è un intervallo (aleatorio, perché dipende dalle osservazioni X_1, \dots, X_n) I_X tale che $\mathbb{P}(I_X \ni \psi(\theta)) = \alpha$.*

Esempio 6.10 *Studiamo l'altezza media degli studenti della classe di statistica, a partire da un campione di numerosità $n = 40$. Si può supporre, in prima approssimazione, che le altezze siano distribuite secondo una legge normale $\mathcal{N}(\mu, \sigma^2)$ e che la deviazione standard sia uguale a quella della popolazione adulta complessiva, che è nota essere $\sigma = 6.1 \text{ cm}$. La stima puntuale della media è $\hat{\mu} = \bar{X}$. Osserviamo che, se X_1, \dots, X_n sono Gaussiane $\mathcal{N}(\mu, \sigma^2)$, allora $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ e quindi*

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

Allora, fissato $\alpha \in (0, 1)$, possiamo affermare che

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

(dove ϕ_α è il quantile di ordine α della legge normale standard). Ad esempio, per $\alpha = 0.95$,

$$0.95 = \mathbb{P}(|\bar{X} - \mu| \leq \phi_{(1+\alpha)/2} \sigma/\sqrt{n}) = \mathbb{P}(|\bar{X} - \mu| \leq 1.89).$$

Questo significa che, prima di eseguire il campionamento, risulta pari a 0.95 la probabilità che $\mu \in I_X := (\bar{X} - 1.89, \bar{X} + 1.89)$; I_X è una v.a., perché lo è \bar{X} (che definisce I_X).

A posteriori, supponiamo di aver trovato il valore $\bar{x} = 178$; affermiamo allora che, con una confidenza del 95%, la media reale μ appartiene all'intervallo $(176.1, 179.9)$. Non possiamo più parlare di probabilità, perché ora sia μ che l'intervallo $(176.1, 179.9)$ sono fissati (anche se μ è incognita).

△

Stima per intervalli della media di una popolazione normale con varianza incognita

Consideriamo ora il caso che spesso si presenta in pratica, in cui la varianza σ^2 non è nota. Posso usare lo stimatore S^2 al posto di σ^2 ? La risposta è positiva, grazie al lemma 2.2; questo lemma implica anche che **dovremo usare non i quantili della legge normale, ma quelli della legge $t(n-1)$.**

Prima di eseguire il campionamento, si ottiene la seguente formula

$$\mathbb{P}\left(\left|\frac{\bar{X} - \mu}{\sqrt{S^2/n}}\right| < t_{\frac{1+\alpha}{2}}(n-1)\right) = \alpha$$

(dove $t_{\frac{1+\alpha}{2}}(n-1)$ è il quantile della legge $t(n-1)$ di Student con $n-1$ gradi di libertà). Possiamo allora scrivere il seguente intervallo di confidenza al livello $\alpha = 0.95$

$$(\bar{X} - t_{\frac{1+\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}}, \bar{X} + t_{\frac{1+\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}}).$$

Esempio 6.11 Riprendiamo l'esempio precedente, e supponiamo che sia stato osservato un valore $s = 6.1\text{cm}$ per la deviazione standard campionaria. Allora l'intervallo di confidenza al livello $\alpha = 0.95$ è

$$(\bar{x} - t_{0.975}(39)\sqrt{\frac{s^2}{40}}, \bar{x} + t_{0.975}(39)\sqrt{\frac{s^2}{40}})$$

e sostituendo i valori numerici (dalle tavole risulta $t_{0.975}(39) = 2.02$): si ottiene come intervallo di confidenza $(176, 180)$.

L'intervallo trovato in questo caso risulta più ampio di quello trovato in precedenza, come potevamo aspettarci, visto che ora possiamo usare meno informazione.

△

6.2.2 Stima per intervalli della media di una popolazione qualsiasi, per grandi campioni

Nel paragrafo precedente abbiamo ricavato un intervallo di confidenza per la media di una popolazione normale, con varianza incognita. L'ipotesi di normalità è stata usata per affermare che la v.a. $\frac{\bar{X} - \mu}{\sqrt{S^2/n}}$ ha distribuzione $t(n-1)$.

Come abbiamo già osservato nel corso, di conseguenza al teorema centrale del limite, **se la popolazione non è normale ma la sua densità non è troppo asimmetrica, e la numerosità del campione è abbastanza grande (diciamo $n \geq 30$), la v.a. \bar{X} è approssimativamente normale, e così la v.a. $\frac{\bar{X} - \mu}{\sqrt{S^2/n}}$ ha approssimativamente distribuzione $t(n-1)$.**

Di conseguenza,

$$\left(\bar{X} - t_{\frac{1+\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}}, \bar{X} + t_{\frac{1+\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}}\right)$$

è un intervallo di confidenza (approssimato) al livello α per μ .

Si ricordi infine che, se n è molto grande (diciamo $n > 120$), si può fare l'ulteriore approssimazione $t(n) \sim \mathcal{N}(0, 1)$ e utilizzare i quantili della legge normale standard per approssimare quelli della legge t di Student.

Se invece la varianza σ^2 si suppone nota, per campioni abbastanza grandi (diciamo $n \geq 30$), si può usare direttamente l'approssimazione normale

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

In questo caso l'intervallo di confidenza (approssimato) al livello α per μ è

$$\left(\bar{X} - \phi_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + \phi_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

6.2.3 Stima per intervalli di una proporzione (frequenza)

Ritorniamo al caso dell'esempio 6.6 in cui la popolazione è bernoulliana, per campioni numerosi ($n \geq 30$). Si vuole quindi stimare il parametro p , che rappresente la frequenza relativa (proporzione) con cui si presenta un carattere della popolazione.

Se X_1, \dots, X_n è il campione estratto da una popolazione bernoulliana $B(p)$, l'approssimazione normale della binomiale vale sotto l'ipotesi

$$np > 5, \quad n(1-p) > 5.$$

Poiché non conosciamo il vero valore di p , possiamo solo verificare che sia

$$n\bar{x} > 5, \quad n(1-\bar{x}) > 5$$

(i numeri $n\bar{x}$ e $n(1-\bar{x})$ rappresentano i casi favorevoli e i casi contrari contati nel campione). Se le condizioni precedenti non sono soddisfatte, il risultato trovato è privo di valore.

Se vale l'approssimazione normale risulta

$$\bar{X} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right).$$

Perciò, fissato $\alpha \in (0, 1)$, risulta

$$\mathbb{P}\left(\frac{|\bar{X} - p|}{\sqrt{p(1-p)/n}} < \phi_{\frac{1+\alpha}{2}}\right) \simeq \alpha.$$

Per calcolare l'intervallo di confidenza approssimato per p , si preferisce eseguire un'ulteriore approssimazione, stimando la varianza $p(1-p)/n$ mediante $\bar{X}(1-\bar{X})/n$. In definitiva, otteniamo come intervallo di confidenza approssimato per p al livello α

$$\left(\bar{X} - \phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + \phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \right).$$

6.2.4 Stima per intervalli della varianza di una popolazione normale

Supponiamo di estrarre un campione X_1, \dots, X_n da una popolazione che possiamo supporre avere legge normale $\mathcal{N}(\mu, \sigma^2)$, e di voler determinare l'intervallo di confidenza, al livello α , per σ^2 . Sfruttando il lemma 2.1, possiamo scrivere che

$$\alpha = \mathbb{P}\left(x_{\frac{1-\alpha}{2}}^2(n-1) < \frac{(n-1)}{\sigma^2} S^2 < x_{\frac{1+\alpha}{2}}^2(n-1)\right)$$

da cui ricaviamo che, l'intervallo di confidenza per σ^2 al livello α è

$$\left(\frac{(n-1)s^2}{x_{\frac{1+\alpha}{2}}^2(n-1)}, \frac{(n-1)s^2}{x_{\frac{1-\alpha}{2}}^2(n-1)} \right).$$

Si osservi che rispetto al valore s^2 che costituisce la stima puntuale di σ^2 , questo intervallo non è simmetrico, ma è stato costruito in modo da avere *code uguali*.

6.2.5 Stima della differenza tra le medie di due popolazioni normali

Siano (X_1, \dots, X_n) e (Y_1, \dots, Y_m) due campioni estratti da popolazioni di leggi normali, di media μ_X , μ_Y e varianza σ_X^2 , σ_Y^2 , rispettivamente.

Siccome \bar{X} e \bar{Y} sono gli stimatori di μ_X e μ_Y , sembra ragionevole (ed infatti può essere dimostrato) che $\bar{X} - \bar{Y}$ è un buon stimatore della differenza $\mu_X - \mu_Y$.

Per ottenere una stima per intervalli, occorre conoscere la distribuzione di $\bar{X} - \bar{Y}$. Dalle proprietà delle leggi normali si ottiene che

$$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}).$$

Se le varianze sono note, possiamo dedurre che con un livello di fiducia α , un intervallo di confidenza per la differenza $\mu_X - \mu_Y$ è

$$\left(\hat{x} - \hat{y} - \phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, \hat{x} - \hat{y} + \phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \right).$$

Nel caso le varianze σ_X^2 e σ_Y^2 non sono note, è abbastanza naturale tentare di sostituirle con le varianze campionarie, che di queste sono stimatori. Tuttavia, questo è possibile solo nel caso in cui si supponga che le varianze (incognite) siano uguali; denotiamo con σ^2 il loro valore (comune). Se poniamo

$$S^2 = \frac{n-1}{n+m-2} S_X^2 + \frac{m-1}{n+m-2} S_Y^2,$$

S^2 ha distribuzione χ^2 a $(n+m-2)$ gradi di libertà ed è uno stimatore di σ^2 ; otteniamo allora che

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{S^2(1/n + 1/m)}} \sim t(n+m-2).$$

Siamo quindi in grado di determinare gli intervalli di confidenza per $\mu_X - \mu_Y$. Ad esempio, si ottiene

$$\mathbb{P} \left(\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{S^2(1/n + 1/m)}} < t_\alpha(n+m-2) \right) = \alpha$$

da cui un intervallo di confidenza al livello α è

$$(\bar{x} - \bar{y} - \sqrt{S^2(1/n + 1/m)} t_\alpha(n+m-2), +\infty).$$

6.3 Esercizi di ricapitolazione

Esercizio.

Trovare gli indici di posizione per i dati dell'esempio 6.2, raggruppati nella tabella seguente

classe	x_i	f_i	$x_i f_i$
1,70-1,74	1,725	1	1,725
1,75-1,79	1,775	1	1,775
1,80-1,84	1,825	5	9,125
1,85-1,89	1,875	3	5,625
1,90-1,94	1,925	2	3,850
1,95-1,99	1,975	0	0
2,00-2,04	2,025	4	8,100
2,05-2,09	2,075	3	6,225
2,10-2,14	2,125	1	2,125
		= 20	= 38,550

Soluzione.

Possiamo calcolare la media dei dati raggruppati nella tabella; le osservazioni di ogni classe vengono approssimate da x_i (valore centrale di ogni classe): si ottiene

$$\bar{x} = \frac{1}{n} \sum_{i=1}^9 x_i f_i = \frac{1}{20} 38,550 = 1,928$$

Per ottenere i quartili, osserviamo che $20/4 = 5$ è intero, quindi Q_1 è la media aritmetica tra x_5 e x_6 ; questi valori sono nella stessa classe, quindi pongo Q_1 uguale al valore centrale di questa classe: $Q_1 = 1,825$. La mediana è pari alla media aritmetica tra x_{10} e x_{11} ,

$$M = Q_2 = \frac{1}{2}(1,825 + 1,875) = 1,85$$

Infine, si ottiene $Q_3 = 2,025$

△

Esercizio.

Su una linea soggetta a disturbi casuali, vengono trasmesse 9 copie di un segnale; al termine della linea vengono registrati i valori

5	8.5	12	15	7	9	7.5	6.5	10.5
---	-----	----	----	---	---	-----	-----	------

Determinare un intervallo di confidenza al livello $\alpha = 95\%$ per la media μ del segnale inviato.

Esercizio.

Il seguente campione è estratto da una popolazione che possiamo supporre di leggi normali, di parametri (μ, σ^2) , che dobbiamo stimare

1.35	1.79	2.11	2.49	3.16	3.22	3.33	3.57	4.17	4.58
------	------	------	------	------	------	------	------	------	------

Soluzione.

Stima di μ : usiamo come stimatore la media campionaria \bar{X} , quindi

$$\hat{\mu} = \frac{1}{10} \sum_{i=1}^{10} x_i = 2.977$$

Stima di σ^2 : usiamo come stimatore la **varianza campionaria S^2** , quindi

$$s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \hat{\mu})^2 = 1.07$$

Nel caso sapessimo il **valore vero della media, ad esempio $\mu = 3$** , allora avremo come stima

$$\hat{\sigma}^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \mu)^2 = 0.961$$

△

Esercizio.

Un sondaggio del 16 aprile 2004 riportò che il 48% della popolazione, con un margine di errore del 4%, era favorevole a mantenere le truppe italiane in Iraq. Cosa significa? Possiamo stimare quante persone furono intervistate?

Soluzione.

Al livello di fiducia $\alpha = 95\%$, l'affermazione principale è che l'intervallo di confidenza per la proporzione p di persone favorevoli è $(0.44, 0.52)$. Se vogliamo risalire ai dati, osserviamo che l'intervallo di confidenza viene costruito a partire dalla formula

$$p \in \left(\hat{p} - \phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right).$$

Con i dati forniti dal testo, risulta $\phi_{\frac{1+\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.04$, da cui $n \simeq 600$. Se fossero state intervistate 600 persone, 288 avrebbero espresso parere favorevole, 312 parere contrario.

△

Esercizio.

Abbiamo acquistato due impianti di produzione per sferette di acciaio, di cui ci hanno garantito che la prima è più precisa della seconda, nel senso che $\sigma_A^2 = 40$, $\sigma_B^2 = 100$. Abbiamo il sospetto che le sferette prodotte dalla prima siano più piccole di quelle della seconda; si effettua così un campionamento da entrambe, da cui si ottengono i seguenti dati

tipo A	34	35	36	36	37	38	41	44	44	44	51	52	53	54
tipo B	38	44	46	48	52	52	60	62	64	66	68	70		

Si determini un valore x per cui la differenza delle medie $\mu_B - \mu_A$ è superiore a x con confidenza $\alpha = 95\%$.

Soluzione.

Iniziamo con il calcolare le stime puntuali $\hat{\mu}_A = 42,8$ e $\hat{\mu}_B = 55,8$. La v.a. $\frac{\bar{X}_B - \bar{X}_A - (\mu_B - \mu_A)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}$ ha legge normale standard, da cui

$$\mathbb{P} \left(\frac{\bar{X}_B - \bar{X}_A - (\mu_B - \mu_A)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} < \phi_\alpha \right) = \alpha;$$

si ottiene un intervallo di confidenza unilatero al livello α

$$\left(\hat{\mu}_B - \hat{\mu}_A - \phi_\alpha \sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}, +\infty \right) \ni (\mu_B - \mu_A)$$

quindi, sostituendo i valori numerici, con confidenza 95% si ha $(\mu_B - \mu_A) > 7.47$.

△