Calcolo delle probabilità e statistica
Breve guida per insegnanti

Febbraio 2008

Prof. Alberto Gandolfi
Dipartimento di Matematica U. Dini
Università di Firenze
gandolfi@math.unifi.it

Indice

Capitolo 1. INTRODUZIONE 5

SPAZI DI PROBABILITÀ E VARIABILI ALEATORIE 8

Capitolo 2. PROBABILITÀ UNIFORMI 9
- 2.1. Spazi di probabilità uniformi 9
- 2.2. Calcolo combinatorio 10

Capitolo 3. PROBABILITÀ SU INSIEMI FINITI 12
- 3.1. Spazi di probabilità su insiemi finiti 12
- 3.2. Indipendenza 13
- 3.3. Utilizzo della probabilità 16
- 3.4. Probabilità condizionate 17
- 3.5. Formula di Bayes 19
- 3.6. Alcuni calcoli di probabilità in genetica 20

Capitolo 4. VARIABILI ALEATORIE FINITE 23
- 4.1. Variabili aleatorie 23
- 4.2. Valore atteso 25
- 4.3. Indipendenza tra variabili aleatorie 30
- 4.4. Deviazioni dalla media 35

Capitolo 5. PROBABILITÀ NEL CONTINUO 37
- 5.1. Variabili aleatorie continue 37
- 5.2. Funzioni di distribuzione 39
- 5.3. Trasformazioni di densità 42
- 5.4. Simulazione di variabili aleatorie continue 44

Capitolo 6. DISEGUAGLIANZE E LIMITI DI SOMME DI VARIABILI ALEATORIE 45
- 6.1. Diseguaglianze 45
- 6.2. Limiti di somme di variabili aleatorie 46

Capitolo 7. VARIABILI ALEATORIE CONGIUNTE 49
- 7.1. Variabili aleatorie congiunte 49
- 7.2. Vettori aleatori continui 51
- 7.3. Correlazione 54

STATISTICA 57

Capitolo 8. STATISTICA DESCRITTIVA 58

8.1. L'istogramma 58
8.2. Elementi di Statistica Descrittiva 59

Capitolo 9. STATISTICA INFERENZIALE 60
9.1. Stima di parametri: il metodo dei momenti 60
9.2. Test di ipotesi: lo z-test 61

CAPITOLO 1

INTRODUZIONE

Il calcolo delle probabilità e la statistica costituiscono quella parte della matematica e della scienza più in generale che si occupa dei fenomeni casuali; questo fatto crea talvolta qualche difficoltà nel comprenderne scopi e metodi. Per tentare di ovviare a queste incomprensioni, noi iniziamo qui presentando alcuni problemi, tratti da situazioni concrete del mondo reale, che, opportunamente formalizzati e predisposti per un approccio scientifico alla loro trattazione, faranno da guida per lo sviluppo di quasi tutta la teoria che presenteremo. Il lettore è invitato a tentare di fornire qualche risposta, sia inizialmente che durante lo sviluppo della materia, nonché a riferirsi a questi problemi quando stia perdendo di vista il senso di quanto viene discusso. Tutto ciò renderà la teoria molto semplice e naturale.

Esempio 1. *Supponiamo di partecipare ad un gioco in cui viene lanciata, da un addetto, una moneta* 1000 *volte, e supponiamo che noi si vinca 1$ se esce testa e se ne perda 1 se esce croce. Vogliamo valutare la nostra reazione ai vari possibili valori di* k*, il numero di teste sui* 1000 *lanci. Quale reazione sarebbe ragionevole esprimere se* k *fosse* 510*? E* 492*? Oppure* 459*,* 423*,* 397*,* 354*,* 299*,* 212*,* 154*,* 22*?*

Esempio 2. *Su una scatola di componenti elettronici contenente 100 pezzi è indicato che ognuno ha tempo medio di funzionamento di 1000 ore. Come nell'esempio precedente, proviamo ad immaginare quale sarebbe una reazione ragionevole se il tempo totale di funzionamento dell'intera scatola fosse 99800, oppure 95400 oppure 90200 oppure 84000 oppure 72000 ore.*

Esempio 3. *Supponiamo che in una popolazione vi siano solo due possibili forme di un certo carattere, per semplificare diciamo capelli biondi o castani, determinati dalla trasmissione genetica che ci è ora nota, con biondo carattere recessivo. Se nella popolazione i biondi sono il* 30%*, su* 100 *figli di genitori entrambi castani, quanti ce ne aspettiamo di biondi?*

Problemi come questi si riferiscono ad esperimenti di cui non si riesce a determinare con certezza l'esito. Ci sono di solito varie o anche

molte possibili alternative e le informazioni in nostro possesso non ci permettono di prevedere quale di queste si verificherà .

Non ci interessa discutere qui se questo sia solo un problema di mancanza di informazioni (come ad esempio nei problemi che si riferiscono alla meccanica classica) o se l'indeterminatezza sia intimamente connessa con la natura dell'esperimento (come si assume, per esempio, in meccanica quantistica o in una interpretazione soggettiva della materia qui esposta). Ci interessa soltanto che tale indeterminatezza rende per noi l'esperimento ad esito incerto. Chiameremo tali esperimenti casuali.

Poiché però spesso a noi interessa tentare qualche previsione dell'esito dell'esperimento descriveremo qui un metodo che è stato sviluppato a partire dal XVII secolo (e presubilmente già nel mondo greco) per prepararci agli esiti più frequenti. L'idea principale è quella di introdurre una nuova quantità, a cui si dà comunemente il nome di probabilità, che appunto serva come misura di quanto spesso o quanto facilmente si realizza un certo esito di un esperimento casuale.

Naturalmente non importa che tale quantità esista realmente, così come non è molto importante sapere se la lunghezza di un tavolo esiste davvero, quest'ultima cosa essendo difficilmente accertabile anche per il fatto che alcuni atomi si attaccano e si staccano continuamente dalle estremità e che queste non sono affatto liscie quando osservate al microscopio. Quello che importa però , sia per le lunghezze che per le probabilità , è che queste quantità siano misurabili con relativa facilità in vari casi interessanti e che si sia poi in grado, sulla base di tali misurazioni, di dire qualcosa di utile per qualche problema di un certo rilievo.

In considerazione di queste finalità pratiche non deve sorprendere che, essendovi vari diversi ambiti di applicazione, vi siano definizioni di probabilità non uguali tra loro e che vi siano discussioni anche accese sui meriti delle varie definizioni; discussioni che spesso sfociano nella sterile diatriba di quale sia quella più 'vera'. La discussione filosofica è poi ulteriormente complicata dall'ovvia presenza di avvenimenti con esito non determinato a cui non è semplice associare una valutazione di probabilità, e dalla meno ovvia esistenza di altri il cui risultato è ben determinato, ma di complessità tale che per analizzarli conviene inventarsi una probabilità.

Noi presenteremo qui alcune di queste definizioni e ne discuteremo lo sviluppo della teoria elementare ed alcune applicazioni di rilievo. Lo scopo di questa presentazione è duplice. Da una parte una serie di risultati nell'ambito del calcolo delle probabilità forniscono una migliore interpretazione di esperimenti, osservazioni, dati, insomma del mondo circostante, inclusi in questo presentazioni ed interpretazioni dei dati (come ad esempio il risultato di sondaggi di opinione). Dall'altra il

calcolo delle probabilità offre la possibilità di approfondire uno sviluppo matematico rigoroso, con tutti i benefici intellettuali che un tale esercizio mentale fornisce. Quindi la presentazione sara' inizialmente incentrata sullo sviluppo rigoroso di risultati partendo da alcune assunzioni di base, e solo successivamente, quando la trattazione diventa matematicamente troppo complessa, verranno citati alcuni risultati senza dimostrazione.

La caratteristica principale della probabilità è che, al contrario della lunghezza che si misura attraverso la comparazione fisica con un oggetto campione, essa si misura attraverso la riflessione astratta, usando logica e calcolo matematico. Queste note espongono vari metodi di calcolo di probabilità e varie applicazioni dei risultati.

Essendo un prodotto di calcoli logici, la probabilità verrà misurata tramite numeri puri; spesso nel linguaggio comune essa viene riportata in frazione di 100, ossia in percentuale: così si parla di 30% o 2%. E' però più comodo matematicamente esprimere la probabilità in frazione di 1, indicando quindi $0,3 = 30\%$ oppure $0,02 = 2\%$, principalmente perchè la moltiplicazione di probabilità riesce così più comoda: il 2% del 30% è lo $0,6\%$, direttamente ottenibile da $0,02 \cdot 0,3 = 0,006$.

Si noti che anche le frequenze sono espresse tramite percentuali (come fatto nell'esempio 3), ma che si tratta per lo più di una coincidenza (così come, per esempio, sia pesi che lunghezze si esprimono tramite numeri decimali, in quel caso la dimensionalità ci salva dalla confusione, ma le percentuali sono appunto numeri puri).

Avendo tuttavia deciso di misurare le probabilità con valori in $[0, 1]$ si vede che per due casi estremi è possibile determinare subito il valore della probabilità. Agli eventi logicamente impossibili da realizzare è naturale assegnare probabilità 0, mentre a quelli certi va assegnata probabilità 1. A tutte le altre situazioni andrà assegnata una probabilità nell'intervallo $[0, 1]$ chiuso (ciò signfica che potranno esserci altre situazioni che avranno probabilità 0, che saranno logicamente possibili ma senza la possibilità di realizzarsi). Ora cominciamo a vedere vari metodi per tale assegnazione.

Esercizio 1. *Calcolare il* 3% *del* 25%.

Esercizio 2. *Calcolare il* 90% *del* 20% *dell'*80%.

Esercizio 3. *Calcolare l'*80% *del* 120%.

Esercizio 4. *Calcolare, se possibile, la probabilità che una asserzione falsa venga correttamente ritenuta vera.*

Esercizio 5. *Calcolare, se possibile, la probabilità di un evento la cui probabilità sia uguale ad* 1 *meno la probabilità stessa.*

SPAZI DI PROBABILITÀ E VARIABILI ALEATORIE

CAPITOLO 2

PROBABILITÀ UNIFORMI

2.1. Spazi di probabilità uniformi

Iniziamo ora a definire questa nuova quantità, la probabilità, in modo che serva per alcune situazioni semplici, come ad esempio:

Esempio 4. *Nel lancio di una moneta calcolare la probabilità che venga testa.*

Esempio 5. *Nel lancio di un dado calcolare la probabilità che il dado mostri la faccia 3.*

In questi esempi dobbiamo naturalmente fare un piccolo sforzo di astrazione. Assumiamo che il risultato di un lancio sia necessariamente una faccia (non una moneta verticale o la sparizione del dado), e che prima del lancio vi sia stato un adeguato mescolamento. Cosa questo sia non è ben determinato, ma a noi interessa l'esito di tale mescolamento, e cioè che, per quanto ne sappiamo, ognuna delle facce si comporta in modo equivalente a tutte le altre. Ossia, se dobbiamo assegnare una probabilità ad una, dobbiamo assegnare la stessa probabilità alle altre. Avendo già deciso che la probabilità che qualcosa avvenga è 1 ne consegue che per queste situazioni è adeguata la prossima definizione. E' chiaro che il discorso esposto finora è euristico, ossia non rigoroso ma fatto cercando di interpretare la realtà esterna, mentre da ora in poi si inizia a fare matematica partendo da una definizione precisa e sviluppandone le conseguenze. Per decidere a quali situazioni si potrà applicare si ritorna a fare discorsi euristici: a tutte quelle situazioni in cui vi siano un numero finito di alternative equivalenti dal punto di vista probabilistico.

Visto che si parla di un numero finito di alternative conviene considerare un insieme finito ed adottare quindi la terminologia delle teoria degli insiemi.

Definizione 1. *(Spazi di probabilità uniformi). Sia S un insieme finito. I suoi sottinsiemi $A \subseteq S$ sono detti eventi, e la probabilità uniforme su S è una funzione P definita su ogni evento A da*

$$P(A) = \frac{|A|}{|S|},$$

avendo indicato con $|A|$ il numero di elementi dell'insieme A.

Si dice talvolta che queste probabilità sono definite come rapporto tra il numero di *casi favorevoli* ed il numero di *casi possibili*.

Inoltre, gli elementi di S sono anche detti eventi elementari.

Dalla definizione discendono alcune proprietà elementari:

LEMMA 1. *(i)* $P(\emptyset) = 0, P(S) = 1$
(ii) per ogni $A \subseteq S$, $0 \leq P(A) \leq 1$;
(iii) se $A \cap B = \emptyset$ *allora* $P(A \cup B) = P(A) + P(B)$
(iv) per ogni $s \in S$, $P(s) = \frac{1}{|S|}$;
(v) per ogni $A \subseteq B \subseteq S$, $P(A) \leq P(B)$.

In questo modo si calcolano le probabilità di eventi relativamente ad un singolo lancio di una moneta, di un dado, all'estrazione di carte o di numeri nel lotto ecc., talvolta, usando (iii), a partire dalle probabilità di eventi più semplici.

ESEMPIO 6. *Calcoliamo la probabilità di uscita di un numero (diciamo il* 23*) su una certa ruota nel lotto (in cui si estraggono senza reinserimento* 5 *numeri da* 90*). Se* A *è l'evento che esce il* 23 *ed* A_i *è l'evento che il* 23 *esce all'i-sima estrazione, per* $i = 1, \ldots, 5$, *si ha che gli* A_i *sono disgiunti e* $P(A_i) = 1/90$ *(attenzione, qui si intende che il* 23 *esce alla i-sima prova senza sapere nulla delle altre, torneremo su questo calcolo), per cui*

$$P(A) = P(\cup_{i=1}^{5} A_i) = \sum_{i=1}^{5} P(A_i) = 5/90$$

2.2. Calcolo combinatorio

La necessità di stabilire delle cardinalità di vari insiemi ha portato allo sviluppo del *calcolo combinatorio*, le cui formule principali sono le seguenti:

Il numero di campioni ordinati ossia delle k-ple ordinate con ripetizione da n elementi, dette disposizioni con ripetizione, è dato da

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

Il numero dei campioni senza ripetizione ordinati ossia delle k-ple ordinate senza ripetizione da n elementi, dette disposizioni senza ripetizione, è dato da

$$D_{n,k} = n(n-1) \cdot \cdots \cdot (n-k+1) = (n)_k.$$

Il numero delle sottopopolazioni, ossia delle k-ple non ordinate senza ripetizione da n elementi, dette combinazioni senza ripetizione, è dato dal coefficiente binomiale

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

Il numero delle sottopopolazioni con ripetizione, ossia delle k-ple non ordinate con ripetizione da n elementi, dette combinazioni con ripetizione, è dato da

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

Il numero delle permutazioni di n elementi è dato da

$$P_n = D_{n,n} = n!.$$

Il numero di ripartizioni in sottopopolazioni di $k_1, k_2, \ldots, k_r$ elementi di un insieme di n elementi, con $k_1 + k_2 + \cdots + k_r = n$, è dato dal coefficiente multinomiale

$$C_{n,(k_1,k_2,\ldots,k_r)} = \frac{n!}{k_1!k_2!\ldots k_r!}.$$

CAPITOLO 3

PROBABILITÀ SU INSIEMI FINITI

3.1. Spazi di probabilità su insiemi finiti

Nel caso delle probabilità uniformi ci trovavamo di fronte ad alternative tutte equivalenti. Tuttavia ci sono eventi (per esempio il colpire un bersaglio) ad esito non determinato per cui sarebbe utile assegnare una probabilità ma per i quali non c'è un ragionevole conteggio da fare. Per questi potremmo provare ad assegnare un qualche valore di probabilità sulla base di altre riflessioni (ad esempio un rapporto di aree, l'esperienza accumulata ecc.).

Sorge quindi naturale di considerare delle probabilità che non vengano da conteggi di insiemi, ma siano semplicemente dei valori in $[0, 1]$ soddisfacenti a certe regole.

Rimaniamo però su un insieme finito e richiediamo per la probabilità che soddisfi alcune delle proprietà che abbiamo verificato essere vere nel caso uniforme. In particolare, si vede che le proprietà principali da cui derivano tutte le altre sono le (i)-(iii) del lemma 1. Assumiamole quindi come ipotesi:

Definizione 2. *Dato un insieme finito S si dice probabilità (finita) su S ogni funzione P definita sui sottinsiemi di S tale che:*
(1) $P(S) = 1$
(2) per ogni $A \subseteq S$, $P(A) \in [0, 1]$*;*
(3) se $A \cap B = \emptyset$ *allora* $P(A \cup B) = P(A) + P(B)$.

Esercizio 6. *Un metodo per costruire probabilità finite è quello di partire dall'insieme finito S e da una funzione non negativa f su S. Si ottiene una probabilità ponendo:*
(a) per ogni $s \in S$, $P(S) = P(\{s\}) = \frac{f(s)}{\sum_{s \in S} f(s)}$
(b) per ogni evento A, $P(A) = \sum_{s \in A} P(s)$.
Verificare che P è una probabilità.

A volte è più agevole dedurre probabilità di certi eventi da altri per i quali la probabilità si deriva più facilmente: ora vedremo cosa fare nel caso che l'evento che ci interessa possa essere visto come unione di altri.

LEMMA 2. *Per ogni* $A, B \subseteq S$ *si ha*
(1) se $A \subseteq B$ *allora* $P(B \backslash A) = P(B) - P(A \cap B)$
(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(3) se $A_i \cap A_j = \emptyset$ *allora* $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

DIMOSTRAZIONE. (1) $P(B) = P((A \cap B) \cup B \backslash A) = P(A \cap B) + P(B \backslash A)$

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

(3) $P(\cup_{i=1}^n A_i) = P(\cup_{i=1}^{n-1} A_i \cup A_n) = P(\cup_{i=1}^{n-1} A_i) + P(A_n) = \cdots = \sum_{i=1}^n P(A_i)$

□

Esiste poi una formula più generale per la probabilità dell'unione di n eventi anche non disgiunti, che non riportiamo.

In taluni casi è più semplice calcolare la probabilità della negazione di un evento A, ossia del suo complemento insiemistico $A^c = S \backslash A$. Si ha

COROLLARIO 1. *(I) Per ogni evento* $A \subseteq S$ *si ha* $P(A^c) = 1 - P(A)$.

3.2. Indipendenza

La teoria della probabilità esposta finora è stata composta da elementi di teoria degli insiemi ed alcune proprietà algebriche. C'è però un concetto intuitivo la cui traduzione nell'ambito della teoria le conferisce uno sviluppo automono. Si tratta dell'indipendenza che noi percepiamo tra vari eventi, ad esempio tra i risultati di lanci successivi di una moneta o di un dado (a patto che siano stati opportunamente mescolati tra un lancio e l'altro).

Per capire come inserire tale concetto all'interno della teoria consideriamo un esempio semplice: in due lanci successivi di una moneta il conteggio ci dice che la probabilità di due teste è 1/4, che risulta quindi uguale a 1/2 moltiplicato per 1/2. In altre situazioni che riteniamo indipendenti si verifica la stessa proprietà per cui è naturale porre la definizione seguente. Come al solito, queste erano riflessioni euristiche e da qui comincia la teoria.

DEFINIZIONE 3. *(i) due eventi* $A, B \subseteq S$ *si dicono indipendenti se*

$$P(A \cap B) = P(A)P(B); \tag{3.1}$$

(ii) n *eventi* $A_i \subseteq S$, $i = 1, \ldots, n$ *si dicono (collettivamente) indipendenti se per ogni sottofamiglia* $J \subseteq \{1, \ldots, n\}$ *di indici,*

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

ESEMPIO 7. *Nel lancio di un dado, se* A *è l'evento che esce un pari,* B *l'evento che esce un numero minore o uguale a* 2 *e* c *l'evento che*

esce un numero minore o uguale a 3*, allora* $P(A) = 1/2$*,* $P(B) = 1/3$*,* $P(C) = 1/2$ *e* $P(A \cap B) = P(A \cap C) = 1/6$ *per cui* A *e* B *sono indipendenti e* A *e* C *non lo sono.*

Spesso però si fa il percorso inverso: si assume che certi eventi siano indipendenti e, senza verificarlo all'interno dello spazio di probabilità , si utilizza la formula (3.1), ossia $P(A \cap B) = P(A)P(B)$, per calcolare $P(A \cap B)$ dagli altri due valori (questo metodo può suggerire qualche problema logico, nel senso che si usa l'indipendenza senza averla provata: ogni dubbio è rimosso nella teoria più completa dal teorema di esistenza di eventi indipendenti, che qui omettiamo).

Esempio 8. *Due lanci ripetuti di dado sono indipendenti, quindi se* A *indica l'uscita di due* 3 *e* A_i*,* $i = 1, 2$*, indica l'uscita del* 3 *all'i-simo dado, allora* $A = A_1 \cap A_2$ *e* $P(A) = P(A_1)P(A_2) = \frac{1}{36}$*.*

Questo caso è ancora facilmente verificabile, ma con più dadi questo diventa un po' più complesso.

Esempio 9. *Anche* n *lanci ripetuti di dado sono collettivamente indipendenti, quindi se* A *indica l'uscita del* 3 *in tutti i dadi e* A_i*,* $i = 1, \ldots, n$*, indica l'uscita del* 3 *all'i-simo dado, allora* $P(A) = \prod_{i=1}^{n} P(A_i) = \frac{1}{6^n}$*.*

L'indipendenza si può dedurre per i complementi.

Lemma 3. *Se* A, B *sono indipendenti allora anche* A^c *e* B *sono indipendenti, e di conseguenza anche la coppia* A*,* B^c *e la coppia* A^c*,* B^c*.*

Dimostrazione.

$$\begin{aligned} P(A^c \cap B) &= P(B \backslash (A \cap B)) \\ &= P(B) - P(A \cap B) \\ &= P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^c) \end{aligned}$$

□

Da qui non è difficile dedurre che anche per n eventi collettivamente indipendenti tutte le combinazioni di eventi o complementi sono collettivamente indipendenti.

Esempio 10. *Se* A *indica l'uscita di almeno una testa in* 3 *lanci di una moneta ed* A_i *l'uscita di testa all'i-simo lancio, allora gli* A_i *sono collettivamente indipendenti e* $A = \cup_{i=1}^{3} A_i$*; ma gli eventi non sono disgiunti e quindi il calcolo si complica. Tuttavia,* $A^c = \cap_{i=1}^{3} A_i^c$ *e dall'indipendenza collettiva si ha* $P(A^c) = \prod_{i=1}^{3} P(A_i^c) = (\frac{1}{2})^3$*, così che* $P(A) = 1 - 1/8 = 7/8$*.*

Facciamo una pausa per riassumere le regole che abbiamo visto per calcolare probabilità. Per calcolare la probabilità di un evento A si può provare a:

(1) contare gli elementi di A e di S;
(2) provare a passare a A^c;
(3) vedere A come unione, ossia $A = \cup_{i=1}^n B_i$ per certi B_i, e poi
(3_1) se B_i sono disgiunti si ha $P(A) = \sum_{i=1}^n P(B_i)$,
(3_2) se B_i non sono disgiunti ma ci sono solo due eventi (ossia $n = 2$) si ha $P(A) = P(B_1) + P(B_2) - P(B_1 \cap B_2)$, se invece $n > 2$ c'è una formula che non abbiamo descritto qui;
(4) vedere A come intersezione, ossia $A = \cap_{i=1}^n B_i$ per certi B_i, e poi
(4_1) se B_i sono collettivamente indipendenti si ha $P(A) = \prod_{i=1}^n P(B_i)$,
(4_2) se B_i non sono collettivamente indipendenti non abbiamo ancora una formula adeguata.

Esercizio 7. *In un esempio che dette origine a molti di questi calcoli, nel 1654 un incallito giocatore, Chevalier De Méré, scrisse a Pascal chiedendogli se l'uscita di almeno un* 6 *in* 4 *lanci di dado avesse la stessa probabilità dell'uscita di almeno un doppio* 6 *in* 24 *lanci di due dadi. Una storia racconta che avesse perso una grossa somma assumendo che in entrambi i casi si trattasse di 'due probabilità su tre' di vincere. Pascal rispose al De Méré, poi scrisse a sua volta a Fermat e questo fu l'inizio della teoria che stiamo presentando. Calcolare le probabilità di questi eventi.*

Con i metodi precedenti si può anche scrivere in forma esplicita la probabilità che escano esattamente k teste in n lanci di una moneta. Scriviamo ora l'espressione in forma ancora più generale.

Esempio 11. *Supponiamo di effettuare n esperimenti indipendenti tali che in ciascuno la probabilità di successo sia un certo valore $p \in [0, 1]$, ad esempio potrebbe essere $p = 1/6$ se per successo intendessimo l'uscita del* 3 *in un dado e così via. Se $p(k, n, p)$ indica la probabilità di esattamente k successi su n prove indipendenti ognuna con probabilità di successo p ed A_i indica il successo alla i-sima prova si ha:*

$$\begin{aligned} p(k,n,p) &= P(\cup_{J\subseteq\{1,\dots,n\},|J|=k}(\cap_{j\in J}A_j \cap \cap_{j\in J^c}A_j^c)) \\ &= \sum_{J\subseteq\{1,\dots,n\},|J|=k} P(\cap_{j\in J}A_j \cap \cap_{j\in J^c}A_j^c) \\ &= \sum_{J\subseteq\{1,dots,n\},|J|=k} \prod_{j\in J} P(A_j) \prod_{j\in J^c} P(A_j^c) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Ci si riferisce a queste probabilità chiamandole distribuzione binomiale o di Bernoulli di parametri n e p od usando il simbolo $B(n, p)$.

Nel caso delle monete siamo interessati alla distribuzione di Bernoulli $B(n, 1/2)$. L'espressione ora ottenuta permette di fare un passo formale verso la soluzione del problema 1 posto all'inizio, ma ancora calcolare queste probabilità nel caso di 1000 monete non sarebbe

agevole e non sapremmo nemmeno bene con cosa confrontare i valori così ottenuti.

Prima di analizzare meglio questo problema generalizziamo il concetto di probabilità.

3.3. Utilizzo della probabilità

Prima di passare alla generalizzazione facciamo una pausa di riflessione: abbiamo visto alcuni ragionamenti che conducono al calcolo di probabilità e forse è il momento di tentare una prima connessione tra i valori di tali probabilità e la reltà. Non abbiamo ancora abbastanza strumenti per una discussione approfondita, ma possiamo fare alcune osservazioni.

1) E' ragionevole aspettarsi che un evento che ha probabilità 1% si verifichi circa 1 volta su 100 esperimenti indipendenti. Dopo tutto la probabilità che si realizzi esattamente una volta, data dalla distribuzione di Bernoulli $B(100, \frac{1}{100})$, è

$$p(1, 100, \frac{1}{100}) = 100\frac{1}{100^1}(\frac{99}{100})^{100} \equiv \frac{1}{e} \equiv 37\%$$

che quindi abbastanza grande per poter procedere.

Esercizio 8. *Se non vi sembra abbastanza grande stimate la probabilità che un evento di probabilità* 1% *si verifichi* $0, 1$ *o* 2 *volte su* 100 *prove indipendenti.*

Analogamente, eventi di probabilità $1/m$ si verificheranno circa 1 volta ogni m prove indipendenti (con alta probabilità).

Alla luce di questo è ragionevole affermare che se individuiamo un evento A *a priori*, ossia prima che sia effettuato l'esperimento che ne verificherà il realizzarsi o meno, e se la probabilità di A è circa $1/m$ con m grande ($m = 50, 100, 1000$), allora sarà sorprendente (tanto più quanto minore è la probabilità) vedere A realizzato già alla prima di queste prove.

2) D'altra parte se abbiamo 100 eventi disgiunti di probabilità ognuno 1% segue dal Lemma 2 che la probabilità dell'unione è 100%, ossia uno di questi accade di sicuro. Questo accade per esempio nella tombola: ogni numero ha probabilità 1/90, ma in ogni estrazione viene estratto un numero.

Questo si può leggere così : è ben diversa la situazione *a posteriori*: dopo che un esperimento è stato realizzato possiamo selezionare eventi che avevano probabilità (a priori!) piccolissima e che si sono realizzati.

Esercizio 9. *Calcolare la probabilità che 10 lanci successivi di un dado (quindi indipendenti) risultino in una successione data di facce, per esempio* (1234565432).

I calcoli a priori vengono definiti **probabilità** ed in questo ambito siamo riusciti a dare un senso ai nostri calcoli, mentre quelli a posteriori

(in cui rientrano i calcoli relativi all'esempio 1 sulle monete) vengono definiti **statistica** e per questo la nostra analisi è ancora insufficiente.

OSSERVAZIONE 1. *Le riflessioni di questo paragrafo sembrano coincidere con una frase di Cicerone, che dice che gettando degli schizzi di colore a caso sul muro sarà molto facile osservare dei tratti che assomiglino ad una faccia, ma sarà assai difficile che questa possa essere quella della Venere di Milo.*

3.4. Probabilità condizionate

In alcune situazioni si vuol calcolare la probabilità di un evento quando si sa che un altro si è già realizzato. Un buon sistema per trattare questa situazione è restringere l'ambito dei possibili risultati all'evento che si sa essere realizzato rendendone 1 la probabilità. Questo risulta naturale se l'evento già realizzato aveva probabilità positiva (ché altrimenti ci si immette in una considerazione forse non priva di senso, ma il cui senso richiede certamente un dettagliato chiarimento):

DEFINIZIONE 4. *Dato uno spazio di probabilità (S, P) e due eventi $A, B \subseteq S$, con $P(B) \neq 0$, si dice probabilità condizionata di A dato B il valore*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{3.2}$$

ESEMPIO 12. *Nel lancio di un dado, se $A = \{3\}$ e $B = \{dispari\}$ (nello stesso lancio), $P(A|B) = \frac{1/6}{1/2} = \frac{1}{3}$ è la probabilità che sia uscito il 3 sapendo che è uscito un dispari.*

Come già accaduto per l'indipendenza, talvolta è conveniente 'dedurre' il valore della probabilità condizionata dal problema ed usare (3.2) per ricavare uno degli altri termini. Anche in questo caso utilizziamo il valore della probabilità condizionata che ci viene dalla situazione concreta omettendone la verifica.

ESEMPIO 13. *Estraendo senza reinserimento da un'urna contenente 3 palline bianche e 5 rosse, che indicheremo schematicamente come $|3B \quad 2R|$, ed indicando con B_i e R_i gli eventi che escano una pallina bianca o rossa, rispettivamente, alla i-sima estrazione si ha che $P(B_1) = 3/8$, $P(R_2|B_1) = 5/7$ da cui $P(B_1 \cap R_2) = 18/56$.*

Dalla conoscenza delle probabilità condizionate a certi eventi è poi possibile risalire alle probabilità non condizionate degli eventi stessi. A tale scopo diciamo che una famiglia di insiemi $B_1, \ldots, B_n \subseteq S$ è una partizione di S se

(a) $B_i \cap B_j = \emptyset$ se i è diverso da j;

(b) $\cup_{i=1}^{n} B_i = S$.

Vale ora il seguente risultato, talvolta detto teorema delle probabilità totali o delle probabilità composte.

TEOREMA 1. *Se B_i, $i = 1, \ldots, n$, costituiscono una partizione di S con $P(B_i)$ diverso da 0 per ogni i, allora per ogni evento A si ha*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{3.3}$$

DIMOSTRAZIONE. Dalla definizione di probabilità condizionata si ha

$$\begin{aligned} \sum_{i=1}^{n} P(A|B_i)P(B_i) &= \sum_{i=1}^{n} \frac{P(A \cap B_i)}{P(B_i)} P(B_i) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= P(\cup_{i=1}^{n} A \cap B_i) = P(A) \end{aligned} \tag{3.4}$$

in cui le due ultime uguaglianze seguono, la prima dal fatto che i B_i sono disgiunti e quindi così sono gli $A \cap B_i$, e la seconda dal fatto che $\cup_{i=1}^{n} B_i = S$.

□

ESEMPIO 14. *Nella situazione e con le notazioni dell'ultimo esempio relativo alle estrazioni da $|3B \quad 2R|$ si ha:*

$$P(R_2) = P(R_2|B_1)P(B_1) + P(R_2|R_1)P(R_1) = \frac{5}{7} \cdot \frac{3}{8} + \frac{4}{7} \cdot \frac{5}{8} = \frac{5}{8}.$$

Si noti che $P(R_2) = P(R_1)$, come già in precedenza argomentato.

Questo teorema è molto utile, ma a volte bisogna saper trovare la partizione giusta a cui condizionare:

ESEMPIO 15. *In un gioco ci sono tre porte e dietro ad una sola vi è un premio. Un concorrente sceglie una porta senza aprirla, poi dal conduttore gli viene aperta una porta senza premio ed offerta la possibilità di cambiare la sua scelta (necessariamente nell'unica porta finora non menzionata). E' conveniente cambiare o la cosa è irrilevante in quanto comunque aveva scelto una porta a caso? O avremmo preferito che venisse aperta una porta senza premio prima della nostra scelta?*

Diciamo di giudicare sulla base della probabilità di trovare il premio, che vogliamo massima, e fissiamo due strategie: cambiare o non cambiare (al momento della scelta). Per calcolare le probabilità di trovare il premio la cosa migliore sembra essere scegliere la seguente partizione dopo aver scelto la prima porta: se a è la porta scelta, siano A_a l'evento che il premio è in a ed A_a^c il suo complemento (che necessariamente formano una partizione dell'insieme dei possibili esiti, quali che essi siano). Ora, detto V l'evento che trovo il premio, si ha: $P(V) = P(V|A_a)P(A_a) + P(V|A_a^c)P(A_a^c)$. Naturalmente, visto che la posizione del premio è casuale (o quantomeno ignota al concorrente) $P(A_a) = 1/3$. Nella strategia di cambiare, $P(V|A_a) = 0$ e $P(V|A_a^c) = 1$

e viceversa nella strategia di non cambiare, per cui quando si cambia $P(V) = 2/3$ *e non cambiando* $P(V) = 1/3$. *Quindi conviene cambiare.*

Se poi fosse stata aperta una porta prima della scelta, avremmo avuto probabilità $1/2$ *di vincere, quindi sempre meglio fare la scelta, vedere la porta vuota e cambiare.*

Combinando il teorema delle probabilità totali con la definizione di probabilità condizionata si possono calcolare probabilità condizionate non direttamente evidenti.

Esercizio 10. *Calcolare la probabilità che il secondo numero estratto nella tombola (quindi senza reinserimento) sia il 30 sapendo che il primo estratto è pari.*

3.5. Formula di Bayes

La formula presentata in questa sezione permette di tornare indietro, ossia calcolare la probabilità che avevano certi eventi già realizzati (ma del cui esito non abbiamo naturalmente informazione diretta).

Esempio 16. *Supponiamo di avere due urne,* $|3B \quad 2R|$ *e* $|2B \quad 5R|$, *e di usare il seguente procedimento di estrazione: scelgo un'urna a caso e poi da questa estraggo una pallina a caso. Se la pallina è rossa, qual è la probabilità che l'urna estratta sia la prima?*

In questa situazione conosciamo le probabilità iniziali di scelta dell'urna e poi le probabilità condizionate di prendere una certa pallina scelta l'urna, ma appunto il problema chiede di tornare indietro.

Teorema 2. *Data una partizione Se* B_i, $i = 1, \ldots, n$ *ed un evento* A *sottinsieme di un insieme* S, *con* $P(A) \neq 0$, *si ha*

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \tag{3.5}$$

Dimostrazione. Dalla definizione di probabilità condizionata e dalla formula delle probabilità totali si ha

$$\frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} = \frac{P(A \cap B_1)}{P(A)} = P(B_1|A).$$

□

Esempio 17. *Nell'esempio precedente siano* B_1 *l'evento che viene scelta l'urna* $|3B \quad 2R|$ *e* B_2 *l'evento che viene scelta l'altra urna;* B_1 *e* B_2 *formano una partizione. Se* A *è l'evento che è stata estratta una pallina rossa si risponde al problema determinando* $P(A|B_1)$. *E dalla formula di Bayes:*

$$P(A|B_1) = \frac{\frac{2}{5} \cdot \frac{1}{2}}{\frac{2}{5} \cdot \frac{1}{2} + \frac{5}{7} \cdot \frac{1}{2}} = \frac{14}{39}.$$

3.6. Alcuni calcoli di probabilità in genetica

Torniamo ora al secondo problema posto all'inizio di queste lezioni. La questione deve essere formulata in termini scientifici. La teoria rivelatasi più adeguata è quella genetica (che qui esponiamo in forma molto semplificata) in cui si assume:

1. ogni individuo ha, relativamente al carattere in oggetto, due alleli;
2. tali alleli possono essere o b per il carattere biondo o C per il carattere castano;
3. b è un allele recessivo, C dominante per cui il fenotipo (ossia la forma espressa del carattere da parte dell'individuo portatore di due alleli è sempre castano salvo quando i due alleli sono entrambi b);
4. (legge di Hardy-Weinberg) se p_b è la frequenza di alleli b in una popolazione (ossia il rapporto tra il numero di loci con allele b rispetto al numero totali di loci, quest'ultimo essendo due volte il numero di individui della popolazione) allora la frequenza di alleli C sarà $p_C = 1-p_b$ e la frequenza di individui con alleli bb sarà $p_{bb} = p_b \cdot p_b$ e di quelli con alleli CC sarà $p_{CC} = p_C \cdot p_C$.

Sulla base di questi elementi, ed osservando che se un individuo è scelto a caso allora la probabilità di un certo evento coincide con la sua frequenza nella popolazione, è possibile determinare molte probabilità relativamente ai genotipi ed ai fenotipi di una popolazione in un dato momento.

Esempio 18. *La probabilità che un individuo sia eterozigote (ossia abbia due alleli diversi) è* $p_{bC} = 1 - p_{bb} - p_{CC} = 2p_b p_C$.

La probabilità che un individuo castano sia eterozigote è data dalla probabilità condizionata che l'individuo sia eterozigote dato che è castano: indicando con $\bar{C}$ *quest'ultimo evento, la probabilità è quindi*

$$P(bC|\bar{C}) = \frac{P(bC \cap \bar{C})}{P(\bar{C})} = \frac{2p_b p_C}{1-p_b^2} = \frac{2p_b(1-p_b)}{1-p_b^2} = \frac{2p_b}{1+p_b}.$$

Quest'ultima probabilità si può vedere come una applicazione, in un caso molto semplice, del teorema di Bayes in quanto è immediato desumere le probabilità di un certo fenotipo dato il genotipo (che sono peraltro 0 *o* 1*) e qui si richiede la probabilità condizionata inversa (esercizio).*

Tuttavia nel problema 2 dobbiamo anche prendere in considerazione la riproduzione in quanto siamo interessati a questioni di discendenza. Un'ipotesi ragionevole relativamente alla riproduzione è che

5. ogni figlio viene generato scegliendo due genitori scelti a caso (eventualmente lo stesso!) ed il suo genotipo è generato scegliendo a caso un allele da ogni genitore.

Questa ipotesi non è molto realistica per le popolazioni umane, ma sembra più ragionevole per animali inferiori e poi è molto semplice. Prima di procedere osserviamo che c'è però un problema di coerenza tra

quest'ultima ipotesi e le precedenti (questo problema fu portato verso il 1920 al matematico Hardy dal biologo Punnet e dette in seguito luogo alla legge che porta il nome di Hardy): se ad una certa generazione valgono le ipotesi 1-4 e la riproduzione segue l'ipotesi 5 si manterranno le condizioni 1-4 anche alla generazione successiva?

Per questa verifica si usa il teorema delle probabilità totali. Se indichiamo con $P_{\alpha\beta}$, $M_{\alpha\beta}$ e $F_{\alpha\beta}$, gli eventi che il padre, la madre o il figlio rispettivamente, hanno genotipo $\alpha\beta$ si ha:

$$\begin{aligned} P(F_{bb}) &= \sum_{\alpha,\beta,\alpha',\beta'=b,C} P(F_{bb}|P_{\alpha,\beta} \cap M_{\alpha',\beta'}) P(P_{\alpha,\beta} \cap M_{\alpha',\beta'}) \\ &= P(P_{b,b} \cap M_{b,b}) + \frac{1}{2} P(P_{b,C} \cap M_{b,b}) \\ &\qquad + \frac{1}{2} P(P_{b,b} \cap M_{b,C}) + \frac{1}{4} P(P_{b,C} \cap M_{b,C}) \\ &= p_b^4 + 2\frac{1}{2} 2p_b^3(1-p_b) + \frac{1}{4} 4p_b^2(1-p_b)^2 \\ &= p_b^2 = P(P_{bb}). \end{aligned} \tag{3.6}$$

La stessa cosa si può verificare per gli altri genotipi, per cui le condizioni 1-4 risultano in effetti stabili rispetto alla riproduzione modellizzata da 5.

Siamo ora in grado di fornire una risposta al problema 2. Se indichiamo con $F_{\bar{b}}$ ($F_{\bar{C}}$) l'evento che il figlio è biondo (o castano) ed analogamente denotiamo gli eventi che il padre o la madre hanno un certo fenotipo, una soluzione al problema viene data dal calcolo di

$$P(F_{\bar{b}}|P_{\bar{C}} \cap M_{\bar{C}}) = P(F_{bb}|P_{\bar{C}} \cap M_{\bar{C}}).$$

Dalla definizione di probabilità condizionate si ha:

$$\begin{aligned} P(F_{bb}|P_{\bar{C}} \cap M_{\bar{C}}) &= \frac{P(F_{bb} \cap P_{\bar{C}} \cap M_{\bar{C}})}{P(P_{\bar{C}} \cap M_{\bar{C}})} \\ &= \frac{\sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(F_{bb} \cap P_{\alpha,\beta} \cap M_{\alpha',\beta'})}{\sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(P_{\alpha,\beta} \cap M_{\alpha',\beta'})}. \end{aligned}$$

L'indipendenza nella scelta dei genitori implica che

$$\sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(P_{\alpha,\beta} \cap M_{\alpha',\beta'}) = (1-p_b^2)^2.$$

Per il calcolo del numeratore possiamo utilizzare di nuovo la definizione di probabilità condizionata, osservando che solo una delle probabilità

condizionate è diversa da 0:

$$\begin{aligned}
& \sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(F_{bb} \cap P_{\alpha,\beta} \cap M_{\alpha',\beta'}) \\
= & \sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(F_{bb}|P_{\alpha,\beta} \cap M_{\alpha',\beta'})P(P_{\alpha,\beta} \cap M_{\alpha',\beta'}) \\
= & \frac{1}{4}4(1-p_b)^2p_b^2.
\end{aligned}$$

Per cui si ha

$$P(F_{bb}|P_{\bar{C}} \cap M_{\bar{C}}) = \frac{(1-p_b)^2p_b^2}{(1-p_b^2)^2} = \frac{p_b^2}{(1+p_b)^2}.$$

Ora rimane solo da determinare p_b. In effetti, sappiamo che la frequanza di (fenotipi) biondi è il 30% e che la probabilità di un genotipo puro bb è p_b^2. Assumendo che questi due valori siano uguali si ha $p_b = \sqrt{0.3}$.

Per cui $P(F_{bb}|P_{\bar{C}} \cap M_{\bar{C}}) = \frac{0,3}{(1+\sqrt{0.3})^2} \simeq 0,125$.

Anche se era ovvio che la probabilità di un figlio biondo dovesse essere minore per due genitori castani rispetto a due qualunque genitori, quantificare tale riduzione è un risultato per nulla immediato.

CAPITOLO 4

VARIABILI ALEATORIE FINITE

4.1. Variabili aleatorie

In alcuni problemi si è portati a considerare funzioni definite su uno spazio di probabilità .

Esempio 19. *Il numero di teste in* 100 *lanci di una moneta può essere visto come una funzione* $T_{100} : S \to \mathbb{R}$ *con* $S = \{0,1\}^{100}$ *che indica l'insieme dei possibili risultati del lancio di* 100 *monete, indicando testa con* 1 *e croce con* 0*, definita da*

$$T_{100}(a_1, \ldots, a_{100}) = \sum_{i=1}^{100} a_i;$$

ove su S *si considera la probabilità uniforme.*

Per distinguerla dalla teoria delle funzioni e per ricordare che stiamo parlando di fenomeni casuali, queste funzioni vengono chiamate in modo diverso:

Definizione 5. *Dato uno spazio di probabilità finito* (S, P) *ogni funzione* $X : S \to \mathbb{R}$ *è detta* **variabile aleatoria**.

Le variabili aleatorie costituiscono una generalizzazione del concetto di evento, infatti per ogni evento A in uno spazio di probabilità (S, P) la funzione indicatrice di $\mathbb{I}_A$ di A definita da

$$\mathbb{I}_A(s) = \begin{cases} 1 & \text{se } s \in A \\ 0 & \text{se } s \notin A \end{cases}$$

è una variabile aleatoria.

Spesso delle variabili aleatorie ci interessano i valori che esse possono assumere e le relative probabilità : denotiamo con $S_X = X(S) \subseteq \mathbb{R}$ il codominio di una variabile aleatoria X e per ogni $B \subseteq S_X$ con

$$\begin{aligned} P_X(B) &:= P(s \in S : X(s) \in B) \\ &= P(X^{-1}(B)) \\ &=: P(X \in B) \end{aligned}$$

la probabilità che X assuma valori in B. Conseguentemente, $P_X(x) = P_X(\{x\})$ indica la probabilità che X assuma il valore $x \in S_X$.

Lemma 4. *Per ogni variabile aleatoria finita* X*, la coppia* (S_X, P_X) *è uno spazio di probabilità finito detto* **distribuzione di** X.

Dimostrazione. Si ha:
(i) $P_X(S_X) = P(X^{-1}S_X) = P(S) = 1$
(ii) Per ogni $B \subseteq S_X$, $0 \leq P_X(B) \leq P(S) = 1$
(iii) Per ogni $B, C \subseteq S_X$, $B \cap C = \emptyset$,

$$\begin{aligned} P_X(B \cup C) &= P(X^{-1}(B \cup C)) \\ &= P(X^{-1}(B) \cup X^{-1}(C)) \\ &= P(X^{-1}(B)) + P(X^{-1}(C)) \\ &= P(B) + P(C) \end{aligned}$$

in cui la terza uguaglianza deriva, come sempre, dal fatto che gli eventi sono disgiunti e che quindi sono disgiunte le immagini inverse.

Infine, $|S_X| \leq |S| < \infty$. □

Osservazione 2. *Si noti che P_X risulta essere una probabilità proprio perché utilizziamo X^{-1} per trasportare P in P_X. Se $Z : T \to S$ fosse una fuzione a valori in uno spazio di probabilità (S, P) il trasporto nella direzione opposta dato da $P \circ Z$ non dà in generale uno spazio di probabilità perché le immagini di insiemi disgiunti non è detto che siano disgiunte.*

Se D indica una certa distribuzione e X è una variabile aleatoria si scrive $X \sim D$ (letto X ha distribuzione D) per indicare che X ha la distribuzione D. Ad esempio, se X è la la funzione identità su uno spazio di probabilità uniforme su $\{1, 2, \ldots, n\}$ indichiamo con $X \sim U(\{1, 2, \ldots, n\}) = U(n)$ il fatto la distribuzione di X è appunto lo spazio uniforme di partenza.

Poiché (S_X, P_X) è uno spazio di probabilità gran parte delle informazioni su X possono essere dedotte direttamente dalla sua distribuzione, tant'è che se due variabili aleatorie hanno la stessa distribuzione condividono gran parte delle proprietà . Per cui si pone:

Definizione 6. *Si dice che due variabili aleatorie X definita su (S_1, P_1) ed Y definita su (S_2, P_2) sono uguali in distribuzione, e si denota con $X =^d Y$, se $(S_X, P_X) = (S_Y, P_Y)$*

Da questo punto di vista sembra perfino che non sia necessario introdurre le variabili aleatorie, in quanto il loro studio si riconduce a quello di spazi di probabilità . Tuttavia, solo 'gran parte' delle proprietà dipendono dalla distributione, non tutte. In particolare, in qualunque modello di fenomeni reali si è interessati al valore assunto da una variabile aleatoria e non solo alla sua distribuzione; in questa direzione si osservi che variabili aleatorie uguali in distribuzione non sono necessariamente uguali, anzi possono essere sempre diverse:

Esempio 20. *Nel lancio di una moneta, la funzione indicatrice di testa $\mathbb{I}_T$ e la funzione indicatrice di croce $\mathbb{I}_C$ soddisfano $\mathbb{I}_T =^d \mathbb{I}_C$ ma $\mathbb{I}_T(s) \neq \mathbb{I}_C(s)$ per ogni $s \in S$; in altre parole, i guadagni di chi punta su testa sono opposti a quelli di chi punta su croce*

Come accaduto nell'ultimo esempio, capita di considerare due o più variabili aleatorie $X_1, X_2. \dots, X_n$ definite su uno stesso spazio di probabilità . Anche ogni combinazione $\phi(X_1, X_2, \dots, X_n)$ di queste determinata da una funzione $\phi : B \to \mathbb{R}$, con $B \subseteq \mathbb{R}^n$ tale che $S_{X_1} \times S_{X_2} \times \cdots \times S_{X_n} \subseteq B$, è una variabile aleatoria definita da

$$\phi(X_1, X_2, \dots, X_n)(s) = \phi(X_1(s), X_2(s), \dots, X_n(s)).$$

Così $X_1 + X_2$, $4X_1$ e X_1^2 sono variabili aleatorie sullo stesso spazio di X_1 e X_2.

La determinazione della distribuzione di

$$Y = \phi(X_1(s), X_2(s), \dots, X_n(s))$$

può essere laboriosa, come si intravede dall'esempio seguente:

Esempio 21. *Se X_1 e X_2 sono i risultati del lancio di due dadi e $X_1 + X_2$ è la loro somma, con una certa pazienza si puo' ricavare la distribuzione di $X_1 + X_2$ dalle probabilità dei risultati dei lanci dei due dadi. In particolare, $S_{X_1+X_2} = \{2, 3, \dots, 12\}$ e*

$$P_{X_1+X_2}(s) = \begin{cases} \frac{s-1}{36} & se\ s \leq 6 \\ \frac{12-s+1}{36} & se\ s \geq 7 \end{cases}$$

4.2. Valore atteso

Introduciamo ora una quantità , utilizzabile per l'analisi di variabili alatorie, che può essere giustificata in vari modi.

Esempio 22. *Storicamente fu descritta per la prima volta relativamente al gioco dei punti, in cui due giocatori A e B si disputano 24 gettoni a testa o croce: vince il primo che arriva a 2 vittorie. Naturalmente esiste un modo equo di dividersi i gettoni senza giocare: 12 ognuno. Se però A, che scommette su testa, ha vinto la prima partita, come ci si può dividere la posta senza giocare ulteriormente? Se si decidesse di fermarsi dopo un'ulteriore partita allora se in questa esce ancora testa A prende 24 gettoni, e se esce croce ci si puo' dividere 12 gettoni a testa; quindi dopo una vittoria di A è ragionevole dare i 12 ad A e dividersi gli altri 6 ciascuno, per cui A ha 18 gettoni. Si nota però che $18 = \frac{1}{2}24 + \frac{1}{2}12$. Se invece si decidesse di finire il gioco la probabilità di A di vincere tutta la posta dato che è uscita una testa sarebbe $P(T \cup CT) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ e si nota che di nuovo $18 = \frac{3}{4} \times 24$.*

Esempio 23. *Supponiamo di scommettere sul risultato di un dado, vincendo 3 se esce il 6 ed altrimenti perdendo 1. In ogni partita la nostra vincita sarà quindi una variabile aleatoria X tale che $X(6) = 3$ e $X(i) = -1$ per ogni $i = 1, \dots, 5$. Dopo 60 partite ci aspettiamo di aver vinto circa 10 partite e perse 50, con una vincita totale di -20. Si noti che*

$$-20 = 30 - 50 = 3 \times 10 - 1 \times 50 = 3\frac{60}{6} - 1\frac{60 \times 5}{6} = 60(3\frac{1}{6} - 1\frac{5}{6})$$

.

In questi ed in molti altri esempi compare quindi la quantità somma dei valori vinti moltiplicati per la probabilità di vincerli e questa quantità può essere quindi un modo di valutare l'esito di una variabile aleatoria. Per cui si pone:

Definizione 7. *Data una variabile aleatoria X definita su uno spazio di probabilità (S, P) si dice* **valore atteso** *o* **speranza matematica** *o aspettazione o valor medio di X il valore*

$$E(X) = M(X) = \sum_{s \in S} X(s)P(s). \tag{4.7}$$

Il valore atteso dipende solo dalla distribuzione di X:

Lemma 5.

$$E(X) = \sum_{x \in S_X} xP_X(x)$$

Dimostrazione.

$$\begin{aligned} E(X) &= \sum_{s \in S} X(s)P(s) \\ &= \sum_{x \in S_X} \sum_{s \in S: X(s)=x} xP(s) \\ &= \sum_{x \in S_X} xP_X(x) \end{aligned}$$

□

Esempio 24. *Il valore atteso del risultato X del lancio di un dado è :*

$$E(X) = \sum_{i=1}^{6} i\frac{1}{6} = 3,5.$$

L'esempio mostra che non sempre il valore atteso è un valore che può essere assunto dalla variabile aleatoria e mostra quindi che l'idea che rappresenti il valore 'che ci aspettiamo' è soltanto approssimativa e va specificata nei sensi che abbiamo indicato in precedenza.

Esempio 25. *Il valore atteso di una variabile aleatoria $T_n \sim B(n, p)$ è :*

$$E(T_n) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np.$$

Alcune proprietà elementari del valore atteso sono:

Lemma 6. *Se $X_1, X_2, \ldots, X_n$ sono variabili aleatorie su uno spazio di probabilità (S, P) e $a_1, a_2, \ldots, a_n \in \mathbb{R}$ sono costanti si ha che:*
(i) $E(a) = a$

(ii) $E(a_1X_1 + a_2X_2) = E(X_1) + E(X_2)$ *e pertanto*
(iii) $E(\sum_{i=1}^n a_iX_i) = \sum_{i=1}^n a_iE(X_i)$

Dimostrazione. (i) $E(a) = \sum_{s\in S} aP(s) = a$.
(ii)

$$\begin{aligned} E(a_1X_1 + a_2X_2) &= \sum_{s\in S}(a_1X_1 + a_2X_2)(s)P(s) \\ &= \sum_{s\in S}(a_1X_1(s) + a_2X_2(s))P(s) \\ &= \sum_{s\in S} a_1X_1(s)P(s) + \sum_{s\in S} a_2X_2(s)P(s) \\ &= E(X_1) + E(X_2). \end{aligned}$$

(iii) segue per induzione su n □

Esprimere una variabile aleatoria come somma di altre variabili può quindi condurre ad una drastica semplificazione del calcolo del valore atteso, come si vede dall'esempio seguente.

Esempio 26. *Se* $T_{1000} \sim B(1000, 1/2)$ *è il numero di teste in* 1000 *lanci di una moneta allora* $T_{1000} = \sum_{i=1}^{1000} X_i$ *con* X_i *la funzione indicatrice di testa all'i-simo lancio. Si ha che* $E(X_i) = 1/2$ *per ogni* i *e*

$$E(T_n) = E(\sum_{i=1}^{1000} X_i) = \sum_{i=1}^{1000} E(X_i) = 500.$$

Pertanto il valore atteso riproduce in questo caso correttamente l'idea di quello che 'ci aspettiamo', così come era stato assunto nelle discussioni iniziali. Generalizzando si ottiene che se $T_n \sim B(n, p)$ *allora* $T_n = \sum_{i=1}^n X_i$ *con* $X_i \sim B(1, p)$. *Per cui* $E(X_i) = 1p - 0(1 - p) = p$ *e* $E(T_n) = np$.

In questo calcolo la proprietà di indipendenza della distribuzione di Bernoulli non è stata utilizzata ed infatti il Lemma vale in generale per tutte le variabili aleatorie.

Esempio 27. *Se* T_n *è il totale di* n *estrazioni dalla tombola senza reinserimento allora* $T_n = \sum_{i=1}^n X_i$, *con* X_i *i-simo numero estratto. poiché*

$$E(X_i) = \sum_{j=1}^{90} \frac{j}{90} = \frac{91}{2}$$

per ogni $i \leq 90$, *e* $E(X_i) = 0$ *per ogni* $i > 90$, *si ha*

$$E(T_n) = \begin{cases} n\frac{91}{2} & se\ n \leq 90 \\ 90\frac{91}{2} & se\ n \geq 90. \end{cases}$$

Si noti il caso $n = 90$, *in cui tutte le palline sono estratte e* $T_{90} \equiv 90\frac{91}{2}$ *con probabilità uno.*

Ci sono altri modi per indicare un 'valore medio' di una variabile aleatoria, utili in particolari contesti. Tra questi vi è il centro o i centri delle probabilità :

DEFINIZIONE 8. *Si dice* **mediana** *di una variabile aleatoria X definita su (S, P) ogni valore $m(X)$ tale che*

$$P(X \leq m(X)) \geq \frac{1}{2} \text{ e } P(X \geq m(X)) \geq \frac{1}{2}.$$

ESEMPIO 28. *Tutti i valori in* $[3, 4]$ *sono mediane del risultato del lancio di un dado.*

ESEMPIO 29. *La mediana del numero T_n di teste in* 100 *lanci di una moneta è* 50*, come si vede per simmetria, mentre per* 101 *lanci è un qualunque numero in* $[50, 51]$

Come si è visto dagli esempi, la mediana non è necessariamente unica e può coincidere o meno con il valore atteso.

DEFINIZIONE 9. *Si dice* **moda** *di una variabile aleatoria X ogni valore $mo(X)$ tale che per ogni $x \in S_X$*

$$P(X = mo(X)) \geq P(X = x).$$

E' moda di una variabile aleatoria $T_n \sim B(n, p)$ ogni valore in $[p(n+1)-1, p(n+1)] \cap \mathbb{N}$ mentre tutti i valori possibili sono moda per le variabili aleatorie uniformi. La moda di una variabile aleatoria è utile solo in alcune situazioni: ad esempio, se si è obbligati a scommettere su uno solo dei risultati possibili di un esperimento casuale la moda permette di massimizzare le probabilità di vittoria.

Quando si osservano dei dati $\{x_1, \ldots, x_n\}$ dopo aver condotto un esperimento casuale è possibile definire lo spazio di probabilità $(S_{\mathbb{O}}, P_{\mathbb{O}})$, che chiameremo empirico o campionario, in cui $S_{\mathbb{O}}$ è costituito dai valori osservati e $P_{\mathbb{O}}$ è determinato dalla frequenze, ossia

$$P_{\mathbb{O}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{x_i = x}.$$

In tale spazio il valore atteso della variabile aleatoria identità I è la media empirica dei dati:

$$\begin{aligned} \mu = E(I) &= \sum_{x \in \{x_1, \ldots, x_n\}} x P_{\mathbb{O}}(x) \\ &= \sum_{x \in \{x_1, \ldots, x_n\}} x \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{x_i = x} \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i. \end{aligned}$$

La mediana di I è la mediana empirica $m = m(I)$ tale che $|\{i : x_i \leq m\}| \geq n/2$ e $|\{i : x_i \geq m\}| \geq n/2$. Si noti che le mediane empiriche si possono determinare ordinando (debolmente) le osservazioni x_i, che vengono in questo caso denotate con $\{x_{(1)}, \ldots, x_{(n)}\}$ con $x_{(1)} \leq \cdots \leq x_{(n)}$, e scegliendo poi un valore nell'intervallo (banale se n è dispari) $[x_{(\lfloor \frac{n+1}{2} \rfloor)}, x_{(\lceil \frac{n+1}{2} \rceil)}]$. La moda di I è la moda dei dati, ossia ogni valore $mo(I)$ che massimizza la frequenza di osservazione.

Esempio 30. *Con dati* $5.0, 3.5, 2.3, 5.0, 1.2, 1.6$ *risultano:* $\mu = 3.1, m \in [2.3, 3.5], mo = 5.0$.

Prima di condurre un esperimento è possibile valutare a priori le quantità empiriche appena descritte considerandole come funzioni di variabili aleatorie $X_1, \ldots, X_n$. In particolare la media empirica delle variabili aleatorie diviene

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

di cui si discuteranno varie proprietà nel seguito.

Per concludere la discussione sul valore atteso osserviamo che data una variabile aleatoria X ed una funzione $\Phi : S_X \to \mathbb{R}$, il risultato seguente sul cambiamento di variabili permette di semplificare il calcolo del valore atteso della variabile aleatoria $\Phi(X)$.

Teorema 3. *Data una variabile aleatoria* X *su* (S, P) *e* $\Phi : B \to \mathbb{R}$ *con* $S_X \subseteq B$ *si ha*

$$\begin{aligned} E(\Phi(X)) &= \sum_{\phi \in S_{\Phi(X)}} \phi P(\Phi(X) = \phi) \\ &= \sum_{x \in S_X} \Phi(x) P_X(x) \end{aligned}$$

Dimostrazione. la prima uguaglianza discende dal Lemma 5. Per la seconda si ha

$$\begin{aligned} E(\Phi(X)) &= \sum_{s \in S} \Phi(X(s)) P(s) \\ &= \sum_{x \in S_X} \sum_{s \in S : X(s) = x} \Phi(x) P(s) \\ &= \sum_{x \in S_X} \Phi(x) P_X(x) \end{aligned}$$

□

La prima espressione di $E(\Phi(X))$ data nel teorema richiede la determinazione di $(S_{\Phi(X)}, P_{\Phi(X)})$, mentre per la seconda espressione basta la distribuzione di X che può essere utilizzata per tutte le Φ.

Il valore atteso di una potenza di una variabile aleatoria X viene definito **momento di ordine** k di X ed in accordo con l'ultimo risultato vale $\sum_{x \in S_X} x^k P_X(x)$.

4.3. Indipendenza tra variabili aleatorie

E' stata sollevata nell'ultimo capitolo la questione della dipendenza tra variabili aleatorie. Per definire la mancanza di dipendenza estendiamo il concetto di indipendenza tra eventi. Per comprendere come realizzare questa estensione partiamo dall'esempio seguente.

Per semplificare la notazione, per variabili aleatorie $X_1, , \ldots, X_n$ definite su uno stesso spazio di probabilità (S, P) indicheremo con

$$\begin{aligned} &\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \\ &\qquad = \{s \in S : X_1(s) = x_1 \text{ e } X_2(s) = x_2 \text{ e } \ldots \text{ e } X_n(s) = x_n\}. \end{aligned}$$

Esempio 31. *Se $X_i = \mathbb{I}_{A_i}$, con A_i eventi indipendenti in uno spazio di probabilità (S, P), allora*

$$\begin{aligned} P(\mathbb{I}_{A_1} = 1)P(\mathbb{I}_{A_2} = 1) &= P(A_1)P(A_2) \\ &= P(A_1 \cap A_2) \\ &= P(\mathbb{I}_{A_1}\mathbb{I}_{A_2} = 1) \\ &= P(\mathbb{I}_{A_1} = 1, \mathbb{I}_{A_2} = 1) \end{aligned} \tag{4.8}$$

ma abbiamo anche visto che

$$\begin{aligned} P(\mathbb{I}_{A_1} = 1)P(\mathbb{I}_{A_2} = 0) &= P(A_1)P(A_2^c) \\ &= P(A_1 \cap A_2^c) \\ &= P(\mathbb{I}_{A_1} = 1, \mathbb{I}_{A_2} = 0) \end{aligned} \tag{4.9}$$

e così via, di modo che

$$P(\mathbb{I}_{A_1} = i)P(\mathbb{I}_{A_2} = j) = P(\mathbb{I}_{A_1} = i, \mathbb{I}_{A_2} = j)$$

per ogni $i, j = 0, 1$.

Per definire l'indipendenza si può quindi generalizzare questa proprietà :

Definizione 10. *n variabili aleatorie $X_1, , \ldots, X_n$ definite su uno spazio di probabilità (S, P) sono indipendenti se*

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \tag{4.10}$$

Per $n = 2$, quindi, X_1, X_2 sono indipendenti se $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$ per ogni possibile valore x_1, x_2. Si noti che nella definizione di indipendenza per variabili aleatorie non è stata richiesta la fattorizzazione delle probabilità per sottofamiglie: mostreremo tra poco che questa proprietà è conseguenza della definizione data.

Esempio 32. *Due estrazioni successive X_1 e X_2 dalla tombola sono indipendenti se effettuate con reinserimento e dipendenti se non c'è reinserimento; infatti, nel primo caso $P(X_1 = k, X_2 = m) = \frac{1}{90^2} = P(X_1 = k)P(X_2 = m)$ per ogni $k, m = 1, \ldots, 90$, mentre nel secondo caso $P(X_1 = 1, X_2 = 1) = 0 \neq \frac{1}{90^2} = P(X_1 = 1)P(X_2 = 1)$.*

Lemma 7. *Se $X_1, \ldots, X_n$ sono variabili aleatorie indipendenti allora*
(I) per ogni $A_1, \ldots, A_n$, $A_i \subseteq S_{X_i}$ si ha

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

(II) per ogni $J \subseteq \{1, \ldots, n\}$ le variabili aleatorie X_i, $i \in J$, sono indipendenti,
(III) per ogni $A_1, \ldots, A_n$, $A_i \subseteq S_{X_i}$, gli eventi $(X_i \in A_i) = \{s \in S : X_i(s) \in A_i\}$ sono collettivamente indipendenti.

Dimostrazione. Indichiamo con $\mathbf{x} = (x_1, \ldots, x_n)$ i vettori n dimensionali:

(I) Fissati $A_1, \ldots, A_n$ come nell'ipotesi si ha

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(\cup_{x_1 \in A_1, \ldots, x_n \in A_n} \{\mathbf{X} = \mathbf{x}\}).$$

Gli eventi $\{\mathbf{X} = \mathbf{x}\}$ sono disgiunti per definizione di funzione, quindi per l'indipendenza degli X_i

$$\begin{aligned} P(X_1 \in A_1, \ldots, X_n \in A_n) &= \sum_{x_1 \in A_1, \ldots, x_n \in A_n} P(\mathbf{X} = \mathbf{x}) \\ &= \sum_{x_1 \in A_1, \ldots, x_n \in A_n} \prod_{i=1}^{n} P(X_i = x_i) \\ &= \prod_{i=1}^{n} \sum_{x_1 \in A_1, \ldots, x_n \in A_n} P(X_i = x_i) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \end{aligned}$$

(II) segue da (I) con

$$A_i = \begin{cases} x_i & \text{se } i \in J \\ S_{X_i} & \text{se } i \notin J \end{cases}$$

essendo

$$\begin{aligned} P(X_i = x_i \text{ se } i \in J) &= P(X_i = x_i \text{ se } i \in J \text{ e } X_j \in S_{X_j} \text{ se } j \notin J) \\ &= \prod_{i \in J} P(X_i = x_i) \prod_{j \notin J} P(X_j \in S_{X_j}) \\ &= \prod_{i \in J} P(X_i = x_i) \end{aligned}$$

(III) segue da (II) prendendo $X_i = \mathbb{I}_{A_i}$. $\square$

Quindi la richiesta di fattorizzazione delle probabilità per tutti i valori del codominio delle variabili aleatorie include già la fattorizzazione delle stesse espressioni per sottinsiemi di funzioni.

Vediamo ora che funzioni di variabili aleatorie indipendenti sono ancora indipendenti, nel senso che

TEOREMA 4. *Date variabili aleatorie indipendenti $X_1, \dots, X_n$ definite su uno spazio di probabilità (S, P) e due funzioni $\phi : C \to \mathbb{R}$ e $\psi : D \to \mathbb{R}$ tali che $S_{X_1} \times S_{X_2} \times \cdots \times S_{X_k} \subseteq C$ e $S_{X_{k+1}} \times \cdots \times S_{X_n} \subseteq D$ si ha che $T = \phi(X_1, \dots, X_k)$ e $Z = \psi(X_{k+1}, \dots, X_n)$ sono indipendenti.*

DIMOSTRAZIONE. poiché gli eventi nell'unione che segue sono, come al solito, disgiunti, ed essendo le variabili aleatorie indipendenti si ha:

$$
\begin{aligned}
&P(T = t, Z = z) \\
&\quad = P(s : \phi(X_1(s), \dots, X_k(s)) = t, \\
&\qquad\qquad \psi(X_{k+1}(s), \dots, X_n(s)) = z) \\
&\quad = P\Big(\bigcup_{\substack{\boldsymbol{x}=(x_1,\dots,x_n):\phi(x_1,\dots,x_k)=t,\\ \psi(x_{k+1},\dots,x_n)=z}} \{s : (X_1(s), \dots, X_n(s)) = \boldsymbol{x}\} \\
&\quad = \sum_{\substack{\boldsymbol{x}=(x_1,\dots,x_n):\phi(x_1,\dots,x_k)=t,\\ \psi(x_{k+1},\dots,x_n)=z}} P((X_1, \dots, X_n) = \boldsymbol{x}) \\
&\quad = \sum_{\substack{\boldsymbol{x}=(x_1,\dots,x_n):\phi(x_1,\dots,x_k)=t,\\ \psi(x_{k+1},\dots,x_n)=z}} \prod_{i=1}^{n} P(X_i = x_i) \\
&\quad = \sum_{\substack{(x_1,\dots,x_k):\\ \phi(x_1,\dots,x_k)=t}} \prod_{i=1}^{k} P(X_i = x_i) \sum_{\substack{(x_{k+1},\dots,x_n):\\ \psi(x_{k+1},\dots,x_n)=z}} \prod_{i=k+1}^{n} P(X_i = x_i) \\
&\quad = \sum_{\substack{\boldsymbol{x}=(x_1,\dots,x_k):\\ \phi(x_1,\dots,x_k)=t}} P((X_1, \dots, X_k) = \boldsymbol{x}) \\
&\qquad\qquad \times \sum_{\substack{\boldsymbol{x'}=(x_{k+1},\dots,x_n):\\ \psi(x_{k+1},\dots,x_n)=z}} P((X_{k+1}, \dots, X_n) = \boldsymbol{x'}) \\
&\quad = P(T = t)P(Z = z).
\end{aligned}
$$

$\square$

Questo risultato ha molte conseguenze interessanti, ad esempio la seguente:

Esempio 33. *Giocando $k+1$ partite alla roulette, se X_i è il numero che esce nell'i-sima partita si ha che le variabili $X_1, \ldots, X_k, X_{k+1}$ sono indipendenti. Ogni strategia che, a k fissato, cerchi di determinare su quale numero scommettere alla $k+1$-sima partita osservando i risultati delle partite precedenti è equivalente ad una funzione $\phi(X_1, \ldots, X_k)$ e, per il teorema appena provato, X_{k+1} ne risulta quindi indipendente. Per cui il teorema porta a concludere che non esiste nessuna strategia che dalle prime k partite dia suggerimenti sulla successiva partita.*

Si può ovviamente pensare una strategia che non fissi k a priori (tipo aspettare la prima uscita del 3), ma ne rimandiamo l'analisi in quanto richiede un modello con una infinità di possibili risultati, che tratteremo più avanti.

Una delle proprietà fondamentali delle variabili aleatorie indipendenti si riferisce al valore atteso del prodotto di variabili aleatorie.

Teorema 5. *Se X_i, $i = 1, \ldots, n$ sono variabili aleatorie indipendenti su (S, P) allora*

$$E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i).$$

Dimostrazione. Cominciamo da $n = 2$:

$$\begin{aligned}
E(X_1X_2) &= \sum_{t \in S_{X_1X_2}} tP(X_1X_2 = t) \\
&= \sum_{t \in S_{X_1X_2}} \sum_{x_1 \in S_{X_1}, x_2 \in X_2} x_1x_2P(X_1 = x_1, X_2 = x_2) \\
&= \sum_{x_1 \in S_{X_1}} x_1P(X_1 = x_1) \sum_{t \in S_{X_1X_2}} \sum_{x_2 \in S_{X_2}: x_1x_2 = t} x_2P(X_2 = x_2) \\
&= \sum_{x_1 \in S_{X_1}} x_1P(X_1 = x_1) \sum_{x_2 \in S_{X_2}} x_2P(X_2 = x_2) \\
&= E(X_1)E(X_2)
\end{aligned}$$

poiché la seconda uguaglianza discende dall'additività della probabilità dell'unione disgiunta di insiemi.

Il risultato per n generico discende per induzione su n essendo $\prod_{i=1}^{n-1} X_i$ indipendente da X_n. □

Esercizio 11. *Due variabili aleatorie dipendenti tali che $E(X_1X_2) = E(X_1)E(X_2)$ si possono ottenere assegnando le probablità seguenti. Siano X_1, X_2 a valori in $\{0, 1\}$ e siano $P(X_1 = 0, Y_1 = 0) = \frac{2}{12}$, $P(X_1 =, Y_1 = 0) = \frac{6}{12}$, $P(X_1 = 0, Y_1 = 1) = \frac{1}{12}$ e di conseguenza $P(X_1 = 1, Y_1 = 1) = \frac{3}{12}$. Verificare quanto affermato.*

Definizione 11. *Variabili aleatorie $X_1, \ldots, X_n$ su uno spazio di probabilità che siano indipendenti e tali che $X_i =^d X_j$ per ogni $i, j =$*

$1, \ldots, n$ *si dicono* **indipendenti ed identicamente distribuite** *o* **i.i.d.**.

4.4. Deviazioni dalla media

Il valore atteso $E(X)$ di una variabile aleatoria X dà un'idea di quanto ci possiamo aspettare, ma abbiamo già visto che non è esattamente il valore $E(X)$ che ci possiamo attendere come risultato di una realizzazione di X. Il teorema di De Moivre-Laplace fornisce una stima molto precisa di quanto i valori che si realizzano più spesso differiscono dal valore atteso di una distribuzione di Bernoulli, ma per il caso generale discutiamo ora una prima stima. Ricordiamo inoltre che non abbiamo ancora chiarito il significato del termine $\sqrt{p(1-p)}$.

Lo scarto massimo $\max(X-E(X))$ tende a sopravvalutare lo scarto tipico di una realizzazione e chiaramente lo scarto medio $E(X-E(X))$ non ha significato, essendo identicamente nullo per la linearità del valore atteso. Lo scarto assoluto medio $E(|X-E(X)|)$ è una soluzione migliore.

Esempio 34. *Per $X \sim B(1,p)$ si ha $E(X)=p$ e lo scarto assoluto medio risulta essere $2p(1-p)$; per il risultato del lancio di un dado Y lo scarto assoluto medio è $1,5$.*

Si osservi però che il calcolo dello scarto assoluto medio per variabili con distribuzione $B(n,p)$ risulta laborioso e che comunque non risulta spiegato il termine $\sqrt{p(1-p)}$. Pensandoci, non c'è una ragione evidente per non considerare lo scarto quadratico medio di X

$$SD(X) = \sqrt{E(X-E(X))^2}$$

detta anche **deviazione standard** di X. O anche qualche altra potenza. Tuttavia risulta che

Esempio 35. *Per $X \sim B(1,p)$ lo scarto quadratico medio risulta essere $\sqrt{p(1-p)}$ e per il risultato del lancio di un dado lo scarto quadratico medio è circa $1,70$.*

Si vede da questi esempi che lo scarto quadratico medio non differisce sostanzialmente dallo scarto assoluto medio, ma soprattutto 'spiega' il termine che appare nel teorema di De Moivre-Laplace.

L'espressione $Var(X) = E(X-E(X))^2$ è detta **varianza** di X.

La varianza può essere anche calcolata come segue.

Lemma 8. *Per ogni variabile aleatoria X si ha*

$$Var(X) = E(X^2) - (E(X))^2.$$

Dimostrazione.

$$\begin{aligned} Var(X) &= E(X-E(X))^2 \\ &= E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - (E(X))^2 \end{aligned} \tag{4.11}$$

□

Un importante risultato relativo alla varianza è che risulta additiva per somme di variabili aleatorie indipendenti.

TEOREMA 6. *Per ogni $X_1, \dots X_n$ variabili aleatorie indipendenti su uno spazio di probabilità (S, P) si ha*

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

e

$$SD(\sum_{i=1}^{n} X_i) = \sqrt{\sum_{i=1}^{n} Var(X_i)}.$$

DIMOSTRAZIONE. Iniziamo da $n = 2$.

$$\begin{aligned} Var(X_1 + X_2) &= E(X_1 + X_2)^2 - (E(X_1 + X_2))^2 \\ &= E(X_1)^2 + E(X_2)^2 + 2E(X_1X_2) \\ &\qquad -(E(X_1))^2 - (E(X_2))^2 - 2E(X_1)E(X_2) \\ &= Var(X_1) + Var(X_2) + 2(E(X_1X_2) - E(X_1)E(X_2)) \\ &= Var(X_1) + Var(X_2) \end{aligned}$$

poiché per variabili indipendenti $E(X_1X_2) = E(X_1)E(X_2)$ dal Teorema 5.

Il risultato per n generico si ottiene per induzione essendo $\sum_{i=1}^{n-1} X_i$ indipendente da X_n.

Il risultato per la deviazione standard segue dalla definizione. □

Quindi per variabili i.i.d. $X_1, \dots, X_n$ si ha che $SD(\sum_{i=1}^{n} X_i) = nSD(X_1)$. In particolare se $Z \sim B(n, p)$ $Z = \sum_{i=1}^{n} X_i$ in cui le X_i rappresentano la singola prova e quindi sono Bernoulli $B(1, p)$, per cui $SD(\sum_{i=1}^{n} X_i) = \sqrt{np(1-p)}$.

Quindi la deviazione standard è una delle espressioni che rappresentano la deviazione di una variabile aleatoria dalla sua media, ma nel contempo è anche facile da calcolare per somme di variabili indipendenti.

CAPITOLO 5

PROBABILITÀ NEL CONTINUO

5.1. Variabili aleatorie continue

In molti problemi l'insieme dei valori che possono essere assunti da una variabile aleatoria può avere la cardinalità del continuo. In questo caso la formalizzazione di spazi di probabilità adeguati non è elementare e qui ci limitiamo a trattare un caso particolare in cui l'insieme è $\mathbb{R}$ e la probabilità è determinata da funzioni continue a tratti; questi spazi di probabilità costituiscono la distribuzione di parecchie variabili aleatorie continue che permettono di studiare numerosi problemi. In un certo senso seguiamo la linea storica di studiare prima le variabili aleatorie continue definite dalla loro distribuzione e rimandare la chiarificazione dei fondamenti.

Sarebbe possibile utilizzare funzioni con al più una infinità numerabile di punti di discontinuità , ma per semplicità consideriamo il caso in cui questi sono al più un numero finito.

Definizione 12. *Uno* **spazio di probabilità sul continuo** *è una coppia* $(\mathbb{R}, f)$ *con* $\mathbb{R}$ *insieme dei numeri reali ed* $f : \mathbb{R} \to \mathbb{R}$ *una funzione continua tranne al più un numero finito di punti di discontinuità tale che*

(i) $f(x) \geq 0$ *per ogni* $x \in \mathbb{R}$
(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

Si noti che l'integrale nella definizione è ben definito, come integrale improprio. Infatti l'integrale esiste in ogni intervallo $[a, b]$ in cui f è continua; inoltre, detto $D(f) = \{x_1, \ldots, x_n\}$, $x_1 < \cdots < x_n$, l'insieme finito dei punti di discontinuità di f e considerate successioni $\bar{x}_j(0) \searrow_j -\infty$, $x_j(i) \nearrow_j x_i$, $\bar{x}_j(i) \searrow_j x_i$ e $x_j(n) \nearrow_j \infty$ esistono gli integrali in $[\bar{x}_j(i), x_j(i+1)]$, $i = 0, \ldots, n-1$. Per definizione di integrale improprio, (ii) dice che il limite seguente esiste, vale quanto indicato:

$$\lim_{j\to\infty} \sum_{i=0}^{n-1} \int_{\bar{x}_j(i)}^{x_j(i+1)} f(x)dx = 1$$

e non dipende dalle successioni scelte. Allo stesso modo sono definiti gli integrali per ogni intervallo $[a, b]$. Conviene in realtà considerare gli intervalli semiaperti della forma $]a, b]$, $a < b$, $a, b \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ perchè si puo' esprimere un ulteriore intervallo come unione di due disgiunti.

Definizione 13. *In uno spazio di probabilità continuo* $(\mathbb{R}, f)$ *si definiscono* **eventi** *le unioni finite di intervallini semiaperti disgiunti e si indica con* $\mathcal{C}$ *la famiglia di tali insiemi. Per ogni* $A = \cup_{i=1}^{n}]a_i, b_i]$ *si pone*

$$P_f(A) = \sum_{i=1}^{n} \int_{a_i}^{b_i} f(x)dx.$$

Si noti che, benchè strettamente parlando la probabilità di un punto $x \in \mathbb{R}$ non sia stata definita, essa deve essere minore di quella di qualsiasi intervallo che contenga questo punto, ossia deve essere 0.

Si noti inoltre che (i) ed (ii) della definizione di spazio di probabilità continuo sono l'analogo di (1) e (2) nella definizione **??**. L'additività numerabile invece è qui una conseguenza delle proprietà di additività dell'integrazione rispetto al dominio. Verificheremo per semplicità solo l'additività finita.

Lemma 9. *In uno spazio di probabilità* $(\mathbb{R}, f)$ *se* $A, B \in \mathcal{C}$ *sono tali che* $\cup_{I \in A} I = \cup_{J \in B} J$ *allora* $P_f(A) = P_f(B)$.

Dimostrazione. Si noti che l'intersezione di due intervallini semiaperti non disgiunti è un intervallino semiaperto. Se $A = \{I_1, \ldots, I_n\}$ e $B = \{J_1, \ldots, J_k\}$ allora $\{L_{i,j} = I_i \cap J_j, i = 1, \ldots, n, j = 1, \ldots, k\} \in \mathcal{C}$ e per l'additività dell'integrale rispetto al dominio

$$P_f(A) = \sum_{i=1}^{n} \int_{I_i} f(x)dx = \sum_{i=1}^{n} \sum_{j=1}^{k} \int_{L_{i,j}} f(x)dx = \sum_{j=1}^{k} \int_{J_j} f(x)dx = P_f(B)$$

in cui la seconda e la terza uguaglianza dipendono dal fatto che se un intervallo $]a, b]$ è unione finita di intervallini $]a_i, b_i]$ disgiunti allora questi si possono ordinare in modo che $b_i = a_{i+1}$ e

$$\sum_{i=1}^{n} \int_{a_i}^{b_i} f(x)dx = \int_{a}^{b} f(x)dx.$$

□

Uno spazio di probabilità $(\mathbb{R}, f)$ può essere utilizzato per definire la distribuzione di una variabile aleatoria continua.

Definizione 14. *Si dice che una variabile aleatoria continua* X *ha distribuzione* $(\mathbb{R}, f)$ *se per ogni* $A \in \mathcal{C}$ *vale*

$$P(X \in A) = \int_{A} f(x)dx.$$

In tal caso f *è detta la densità di* X.

Si noti che l'espressione di sinistra è soltanto formale in quanto non è stato definito lo spazio di probabilità su cui è definito X e su cui agisce P. Diremo anche che $P(X \in [a, b]) = P(X \in]a, b])$.

Esempio 36. *Una variabile aleatoria con densità* $f(x) = \mathbb{I}_{[0,1]}$ *è detta uniforme in* $[0,1]$. *In generale, la variabile aleatoria con densità* $f(x) = \frac{1}{b-a}\mathbb{I}_{[a,b]}$ *è detta* **uniforme** *in* $[a,b]$ *e si indica* $X \sim U([a,b])$.

Esempio 37. *Una variabile aleatoria con densità* $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ *è detta Gaussiana standard. In generale, la variabile aleatoria con densità* $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{\sigma}}$ *è detta* **gaussiana o normale** *di media* m *e varianza* σ^2 *e si indica* $X \sim N(\mu, \sigma^2)$.

Definiamo ora il valore atteso di una variabile aleatoria continua.

Definizione 15. *Data una variabile aleatoria continua* X *con densità* f *si dice* **valore atteso** *di* f*:*

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

se esiste

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

Con questa definizione valgono tutte le proprietà viste per il valore atteso, la varianza e la deviazione standard di variabili aleatorie.

5.2. Funzioni di distribuzione

Una descrizione delle variabili aleatorie che inizia ad unificare il trattamento delle variabili discrete e continue è la funzione cumulativa o funzione di distribuzione.

Definizione 16. *Si dice funzione* **funzione cumulativa** *o* **funzione di distribuzione** F_X *di una variabile aleatoria* X *la funzione* $F_X : \mathbb{R} \to \mathbb{R}$ *tale che*

$$F_X(t) = P(X \leq t).$$

Tale definizione è valida sia per le variabili aleatorie discrete che continue.

Esempio 38. *La funzione di distribuzione di una variabile aleatoria* $Y \sim U(\{1, \ldots, 6\})$ *vale*

$$F_Y(t) = \mathbb{I}_{t \geq 0}\frac{\lfloor t \rfloor \wedge 6}{6}.$$

Esempio 39. *Se* $Z \sim U([a,b])$ *allora*

$$F_Z(t) = \mathbb{I}_{t \geq a}\frac{t \wedge b - a}{b - a}.$$

Descriviamo ora alcune proprietà delle funzioni di distribuzione, una delle quali differenzia le variabili discrete da quelle continue.

Teorema 7. *Se* X *è una variabile aleatoria (discreta o continua) allora*

(1) F_X *è non decrescente in* t.
(2) $\lim_{t\to-\infty} F_X(t) = 0$
(3) $\lim_{t\to\infty} F_X(t) = 1$
(4) F_X *è continua da destra.*

Inoltre, se X *è una variabile aleatoria discreta allora*
(5a) se $(a, b]$ *non contiene punti di discontinuità di* F_X *allora* F_X *è costante in* $(a, b]$.

Se invece X *è una variabile aleatoria continua con densità* f_X *allora*
(5b) F_X *è continua e* $F'_X(t)$ *esiste per tutti i punti* t *tranne al più un numero finito ed è continua.*

In quest'ultimo caso vale $F'_X(t) = f_X(t)$

Dimostrazione.

(1) Chiaramente $\{s : X(s) \le t\} \subseteq \{s : X(s) \le u\}$ per tutti i $t \le u$ da cui la diseguaglianza sulle probabilità .

(2) Per le variabili discrete

$$\begin{aligned}\lim_{t\to-\infty} F_X(t) &= \lim_{t\to-\infty} P(X \le t)\\ &= \lim_{t\to-\infty} \sum_{x\in S_X : x\le t} P_X(x) = 0\end{aligned}$$

in quanto la sommatoria nell'ultima espressione è il resto di una serie convergente; mentre per le variabili continue

$$\lim_{t\to-\infty} F_X(t) = \lim_{t\to-\infty} \int_{-\infty}^{t} f_X(x)dx = 0$$

per definizione di integrale improprio.

(3) Si ha

$$\lim_{t\to\infty} F_X(t) = 1 - \lim_{t\to-\infty} P(X > t) = 1$$

per gli stessi motivi del punto precedente.

(4) Vale $\lim_{u\searrow t} F_X(u) - F_X(t) = \lim_{u\searrow t} P(t < X \le u) = 0$; per variabili discrete in quanto anche in questo caso, dopo aver riordinato la serie, si tratta del resto di una serie convergente. Per le variabili continue, se f è continua in t allora

$$P(t < X \le u) = \int_t^u f_X(x)dx = \max\{f(x) : x \in [t, u]\}(t - u) \to 0;$$

altrimenti il risultato vale per definizione di integrale improprio in t.

(5) Poiché F_X è non decrescente esiste $\lim_{u\nearrow t} F_X(u) \le F_X(t)$ e

$$\begin{aligned}F_X(t) - \lim_{u\nearrow t} F_X(u) &= \lim_{u\nearrow t} P(u < X \le t)\\ &= P(X = t) - \lim_{u\nearrow t} P(u < X < t) = P(X = t)\end{aligned}$$

per gli stessi motivi di (4).

(5a) Quindi se X è discreta F_X è discontinua in t se e solo se $t \in S_X$ e se $(a, b]$ non contiene punti di discontinuità di F_X allora

$$0 = P(X \in (a, b]) = F_X(b) - F_X(a).$$

(5b) Se invece X è continua allora per il teorema fondamentale del calcolo integrale $F_X(t) = \int_{-\infty}^{t} f_X(x)dx$ è derivabile nei punti in cui f_X è continua e vale $F'_X(t) = f_X(t)$. □

Vediamo ora che queste proprietà caratterizzano le funzioni che possono essere funzioni di distribuzione per qualche variabile aleatoria.

Lemma 10. *Se una funzione soddisfa (1)-(4) e (5a) allora è la funzione di distribuzione di una variabile aleatoria discreta; se invece soddisfa (1)-(4) e (5b) è la funzione di distribuzione di una variabile continua.*

Dimostrazione. Da (1) F è non decrescente quindi in ogni punto esiste il limite sinistro $F(t^-)$ e destro $F(t^+)$ e da (2) e (3) è limitata. Per questo l'insieme S dei punti di discontinuità di F è al più numerabile: infatti un punto di discontinuità è caratterizzato dal fatto che $F(t^-) \neq F(t^+)$ e di punti tali che $F(t^+) - F(t^-) \geq 1/n$ ce ne sono al più n; l'unione su n di tali insiemi dà tutti i punti di discontinuità . Inoltre segue da (4) che $F(t) = F(t^+)$.

Se vale (5a) possiamo porre $P(x) = F(t) - F(t^-) > 0$ per ogni $x \in S$. Se indichiamo con $\ldots, x_{-n}, \ldots, x_0, \ldots, x_n, \ldots$ i punti di S ordinati, allora dalla (5a) $F(x_{n-1}) = F(x_{n-1}^+) = F(x_n^-)$ per cui da (2) e (3) segue che:

$$\begin{aligned} \sum_{x \in S} P(x) &= \sum_{N \in \mathbb{Z}} F(x_n) - F(x_n^-) \\ &= \sum_{N \in \mathbb{Z}} F(x_n) - F(x_{n-1}) \\ &= \lim_{n \to \infty} (F(x_n) - F(x_{-n})) = 1. \end{aligned} \tag{5.12}$$

Risulta poi

$$\sum_{x \in S, x \leq t} P(x) = \lim_{n \to \infty} (F(x) - F(x_{-n})) = F(x). \tag{5.13}$$

Se invece vale (5b) si può porre $f(t) = F'(t)$ per tutti i t in cui F è derivabile ed un valore qualunque altrove. Dalla (1) $f \geq 0$ e dalle (2) e (3)

$$\int_{-\infty}^{\infty} f(t)dt = \lim_{t \to \infty} (F(t) - F(-t)) = 1.$$

Infine,

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

per cui F è la funzione di distribuzione di una variabile aleatoria di densità f.

□

Un esempio importante di variabile aleatoria continua ha densità

$$f(t) = \lambda e^{-\lambda t} \mathbb{I}_{t \geq 0}$$

detta **distribuzione esponenziale** $exp(\lambda)$. Si noti che se $X \sim exp(\lambda)$ allora

$$E(X) = \int_{-\infty}^{\infty} t\lambda e^{-\lambda t} dt = 1/\lambda,$$

$$Var(X) = \int_{-\infty}^{\infty} t^2 \lambda e^{-\lambda t} dt - E(X)^2 = 1/(\lambda)^2,$$

e $SD(X) = 1/\lambda$. Inoltre

$$F_X(t) = e^{-\lambda t}$$

e

$$P(X > t + T | X > t) = e^{-\lambda T}$$

una proprietà che si esprime come perdita di memoria. Per questa caratteristica la distribuzione esponenziale è usata come modello per il tempo di vita di componenti elettronici.

In particolare questo modello potrebbe essere usato per analizzare i tempi di funzionamento dei componenti elettronici di cui all'esempio 2. Per completare l'analisi di questo esempio si devono però considerare le somme di tempi di vita indipendenti, ossia somme di variabili aleatorie continue indipendenti e questo verrà sviluppato nei prossimi capitoli.

Definizione 17. *Si dice che due variabili aleatorie continue X ed Y sono uguali in distribuzione, e si indica $X =^d Y$, se $F_X(t) = F_Y(t)$ per ogni t.*

5.3. Trasformazioni di densità

Descriviamo come cambiano le densità di variabili aleatorie prendendo una funzione di forma semplice di una data variabile aleatoria.

Teorema 8 (Trasformazione di densità). *Data una variabile aleatoria X continua con densità f_X ed una funzione $\phi : \mathbb{R} \to \mathbb{R}$ continua e strettamente monotona con inversa $x = \phi^{-1}(y)$ si ha:*

$$f_Y(y) = f_X(\phi^{-1}(y)) \left| \frac{d}{dy} \phi^{-1}(y) \right|$$

per ogni y tranne al più i punti in cui f_Y e $f_X \circ \frac{d}{dy}\phi^{-1}$ sono discontinui, che sono al più un numero finito di punti.

Dimostrazione. Supponiamo ϕ crescente. Per ogni $a, b \in \mathbb{R}$, $a < b$ si ha, effettuando nell'integrazione della terza riga il cambiamento di variabili $y = \phi(x)$ con inversa $x = \phi^{-1}(y)$ e quindi $dx = \frac{d}{dy}\phi^{-1}(y)dy$:

$$\begin{aligned} \int_a^b f_Y(y)dy &= P(Y \in]a,b]) \\ &= P(\phi^{-1}(Y) \in]\phi^{-1}(a), \phi^{-1}(b)]) \\ &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f_X(x)dx \\ &= \int_a^b f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y)dy. \end{aligned} \tag{5.14}$$

Se a è un punto di continuità per f_Y e per $f_X \circ \phi^{-1}$ allora per il teorema del valor medio, se $[a, b]$ è piccolo a sufficienza da non contenere punti di discontinuità di f_Y e per $f_X \circ \phi^{-1}$:

$$\frac{1}{(b-a)}\int_a^b f_Y(y)dy = f_Y(\alpha(a,b))$$

e

$$\frac{1}{(b-a)}\int_a^b f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y) = f_X(\phi^{-1}(\beta(a,b)))\frac{d}{dy}\phi^{-1}(\beta(a,b))$$

con $\alpha(a,b), \beta(a,b) \in [a,b]$; poiché l'uguaglianza (5.14) vale per ogni intervallo $]a,b]$ si ha:

$$\begin{aligned} f_Y(a) &= \lim_{b\to a^+} f_Y(\alpha(a,b)) \\ &= \lim_{b\to a^+} f_X(\phi^{-1}(\beta(a,b)))\frac{d}{dy}\phi^{-1}(\beta(a,b)) \\ &= f_X(\phi^{-1}(a)\frac{d}{dy}\phi^{-1}(a) \end{aligned}$$

quindi $f_Y(y) = f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y)$ per ogni y tranne al più i punti in cui f_Y e $f_X \circ \frac{d}{dy}\phi^{-1}$ sono discontinui, che sono al più un numero finito di punti.

Se invece ϕ è decrescente è facile verificare che tutti gli intervalli di integrazione risultano invertiti e quindi per riportare l'usuale espressione degli estremi di integrazione dal minore al maggiore si cambia il segno in $\frac{d}{dy}\phi^{-1}$. □

Osservazione 3. *La trasformazione delle densità è facile da ricordare pensando alla densità di X come $f_X(x)dx$ per cui se $Y = \phi(X)$ con inversa $X = \phi^{-1}(Y)$ allora si considera il cambiamento di variabili $y = \phi(x)$ con inversa $x = \phi^{-1}(y)$ e quindi $dx = \frac{d}{dy}\phi^{-1}(y)dy$ per cui il risultato precedente si ottiene sostituendo formalmente i valori x e dx così ottenuti.*

Esempio 40. *Se* $X \sim N(0,1)$ *e* $Y = \sigma X + \mu$ *allora* $X = \phi^{-1}(Y) = \frac{Y-\mu}{\sigma}$ *e*

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

Si noti che per le proprietà del valore atteso

$$E(Y) = E(\sigma X + \mu) = \mu$$

e

$$Var(Y) = E((\sigma X + \mu - \mu)^2) = \sigma^2 E(X) = \sigma^2$$

così che $SD(Y) = \sigma$. *Una tale variabile aleatoria* Y *è detta normale con media* μ *e varianza* σ^2, *che si indica con* $Y \sim N(\mu, \sigma^2)$ *ed ha quindi la densità sopra indicata.*

5.4. Simulazione di variabili aleatorie continue

Simulare una variabile aleatoria X vuol dire generare uno o più numeri che siano prodotti con le probabilità della distribuzione di X. Per variabili finite basta inserire in una scatola un numero sufficiente di bigliettini (cosa non sempre fattibile, ma qui trascuriamo i dettagli). Per variabili continue la cosa richiede una certa attenzione. Per esempio, scegliere un punto su di un intervallo unitario a caso senza privilegiare nessun punto è un modo per simulare una variabile uniforme U in $[0,1]$. Nei fogli di calcolo generalmente è disponibile una funzione che genera un tale valore (spesso mediante un metodo numerico noto come generatore alle congruenze lineari - si veda per esempio la voce Random number generator su Wikipedia) richiamata dalla funzione RAND oppure CASUALE.

Una volta che un valore u di U è disponibile si può simulare qualunque variabile aleatoria continua X dalla sua funzione di distribuzione F_X prendendo $x = F_X^{-1}(u)$. Infatti, la distribuzione di $F_X^{-1}(U)$ soddisfa:

$$P(F_X^{-1}(U) \leq t) = P(U \leq F_X(t)) = F_X(t)$$

ossia la distribuzione di $F_X^{-1}(U)$ coincide con quella di X.

I fogli di calcolo forniscono le inverse delle funzioni di distribuzione più comuni. Per esempio,viene indicata talora con INVNORM la funzione che fornisce l'inversa della funzione di distribuzione della Normale $N(\mu, \sigma^2)$.

CAPITOLO 6

DISEGUAGLIANZE E LIMITI DI SOMME DI VARIABILI ALEATORIE

6.1. Diseguaglianze

La varianza dà un'idea della deviazione tipica dalla media, ma ora vediamo qualche risultato che dia una stima di questa deviazione.

LEMMA 11 (Diseguaglianza di Markov). *Per ogni variabile aleatoria $X \geq 0$ non negativa e per ogni $a > 0$ si ha*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

.

DIMOSTRAZIONE. Essendo $X \geq 0$, per ogni $a > 0$ si ha:

$$\begin{aligned} E(X) &= \sum_{x \in S_X} x P_X(x) \\ &\geq \sum_{x \in S_X, x \geq a} x P_X(x) \geq a P(X \geq a) \end{aligned}$$

□

Da questo segue

COROLLARIO 2 (Diseguaglianza di Chebyshev). *Per ogni variabile aleatoria X finita e per ogni $a > 0$ si ha:*

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2},$$

ossia

$$P(\left|\frac{X - E(X)}{SD(X)}\right| \geq a) \leq \frac{1}{a^2}.$$

DIMOSTRAZIONE. Essendo $(|X - E(X)|)^2 = (X - E(X))^2 \geq 0$, dal Lemma 11 segue che per ogni $a > 0$ si ha:

$$P(|X - E(X)| \geq a) = P((X - E(X))^2 \geq a^2) \leq \frac{Var(X)}{a^2}$$

□

Queste diseguaglianze non sono troppo accurate, come si vede dall'esempio seguente, anzi talvolta sono banali in quanto il maggiorante è maggiore di 1.

Esempio 41. *Se X è il risultato del lancio di un dado allora*

$$\frac{1}{3} = P(X \geq 5) \leq \frac{1}{5}3,5 = 0,7$$

dal Lemma 11 e

$$\frac{1}{3} = P(|X - 3,5| \geq 2,5) \leq \frac{1}{(2,5)^2}Var(X) \approx 0.47$$

dal Corollario 2.

6.2. Limiti di somme di variabili aleatorie

Ora ci possiamo porre il problema di valutare cosa succeda quando si fanno molte prove. Il primo risultato dice che se le variabili sono indipendenti la media empirica dei dati converge al valore atteso della variabile (nel senso specificato).

Teorema 9 (Legge (debole) dei grandi numeri). *Per variabili aleatorie finite i.i.d. $X_1, X_2, \ldots$ su si ha che per ogni $\alpha > 0$*

$$\lim_{n\to\infty} P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > \alpha) = 0.$$

Dimostrazione. Dalla diseguaglianza di Chebyshev si ha:

$$\begin{aligned} P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > \alpha) &= P(\left|\sum_{i=1}^{n} X_i - nE(X_1)\right| > n\alpha) \\ &\leq \frac{1}{n^2\alpha^2}Var(\sum_{i=1}^{n} X_i) \\ &= \frac{nVar(X_1)}{n^2\alpha^2} \to_{n\to\infty} 0 \end{aligned}$$

□

Esempio 42. *Se X_i sono i risultati di lanci indipendenti di un dado allora*

$$P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > 10^{-10}) \to_{n\to\infty} 0$$

anzi

$$P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right| > 10^{-10}) \leq \frac{2,9 \times 10^{20}}{n}.$$

Si noti che quest'ultima stima ha senso solo per $n \geq 2,9 \times 10^{20}$; tuttavia ha poco senso dare troppa importanza al valore quantitativo di queste stime a causa della loro scarsa accuratezza.

Il secondo risultato riguarda una stima più accurata di quanto possa deviare la media empirica dal valore atteso in relazione alla deviazione standard. Di questo risultato omettiamo la dimostrazione (troppo complessa). Cominciamo dall'osservare che una variabile aleatoria può essere resa adimensionale facendo una standardizzazione:

DEFINIZIONE 18. *Data una variabile aleatoria X, si dice* **variabile standardizzata** *o* **versione standardizzata di** X *l'espressione*

$$\frac{X - E(X)}{SD(X)}$$

.

Il risultato seguente riguarda quindi le somme di variabili indipendenti standardizzate:

TEOREMA 10 (Teorema Centrale del Limite). *Per variabili aleatorie i.i.d. $X_1, \ldots, X_n$ su uno spazio di probabilità finito vale per ogni $a \in \mathbb{R}$*

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i - nE(X_1)}{\sqrt{nVar(X_1)}} \leq a\right) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Questo risultato permette di stimare la probabilità che il numero di successi in prove indipendenti sia minore di un certo valore. Infatti, se $T_n \sim B(n,p)$ ossia tale che $T_n = \sum_{i=1}^n X_i$, $X_i \sim B(1,p)$, si ha

$$\begin{aligned}
\sum_{j \leq np + a\sqrt{np(1-p)}} \mathcal{P}(j,n,p) &= \sum_{j \leq np + a\sqrt{np(1-p)}} P(\sum_{i=1}^n X_i = j) \\
&= P\left(\sum_{i=1}^n X_i \leq np + a\sqrt{np(1-p)}\right) \\
&= P\left(\sum_{i=1}^n X_i \leq nE(X_1) + a\sqrt{nVar(X_1)}\right) \\
&= P\left(\frac{\sum_{i=1}^n X_i - nE(X_1)}{\sqrt{nVar(X_1)}} \leq a\right) \\
&= P\left(\frac{(\sum_{i=1}^n X_i) - E(\sum_{i=1}^n X_i)}{\sqrt{Var(\sum_{i=1}^n X_i)}} \leq a\right).
\end{aligned}$$

La conclusione, a cui si può arrivare senza menzionare i dettagli del teorema, è la seguente:

i valori assunti da una somma di variabili aleatorie indipendenti distano solitamente (ossia con grande probabilità) da $nE(X_1)$ non più di $2\sqrt{n}SD(X_1)$, ossia con grande probabilità

$$|\sum_{i=1}^n X_i - nE(X_1)| \leq 2\sqrt{n}SD(X_1).$$

Nel problema 1 delle monete, si ha quindi $n = 1000$, $p = 1/2$, $E(X_1) = 1/2$, $SD(X_1) = 1/2$ e quindi l'errore tipico della somma, ossia del numero di successi, è $2\sqrt{1000}/2 \approx 31,6$.

Una seconda conseguenza riguarda la distanza tra la media empirica ed il valore atteso, che risulta, dividendo per n, circa $2SD(X_1)/\sqrt{n}$ ossia con grande probabilità

$$|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)| \leq 2SD(X_1)/\sqrt{n}.$$

Si noti che ora il fattore $\sqrt{n}$ è al denominatore, quindi la media empirica approssima sempre meglio il valore atteso.

CAPITOLO 7

VARIABILI ALEATORIE CONGIUNTE

7.1. Variabili aleatorie congiunte

In molti problemi, come quando abbiamo parlato di indipendenza di variabili aleatorie o linearità del valore atteso, si considerano più variabili aleatorie contemporaneamente. In questo caso conviene pensarle come un vettore aleatorio.

Definizione 19. *Dato uno spazio di probabilità (discreta)* (S, P), *un* **vettore aleatorio** $\boldsymbol{X} = (X_1, \ldots, X_n)$ *n-dimensionale è una funzione* $\boldsymbol{X}\colon S \to \mathbb{R}^n$ *tale che* $s \to \boldsymbol{X}(s) = ((X_1(s), \ldots, X_n(s))$.

Si possono ripetere ora molte definizioni e proprietà delle variabili aleatorie, che si ottengono semplicemente sostituendo $\boldsymbol{X}$ ad X.

Definizione 20. *La distribuzione di un vettore aleatorio* $\boldsymbol{X} = (X_1, \ldots, X_n)$, *detta anche* **distribuzione congiunta** *delle* X_i, *è la coppia* $(S_{\boldsymbol{X}}, P_{\boldsymbol{X}})$ *con*

$$S_{\boldsymbol{X}} = \{\boldsymbol{x} \in \mathbb{R}^n |\ \textit{esiste } s \in S : \boldsymbol{X}(s) = \boldsymbol{x}\}$$

e

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P\{s \in S : \boldsymbol{X}(s) = \boldsymbol{x}\}$$

Si noti che per l'usuale proprietà delle funzioni, per ogni $B \subseteq S_{\boldsymbol{X}}$

$$P_{\boldsymbol{X}}(B) = \sum_{\boldsymbol{x} \in B} P_{\boldsymbol{X}}(\boldsymbol{x}).$$

Esercizio 12. *Se* $\boldsymbol{X}$ *è un vettore aleatorio discreto, verificare che* $(S_{\boldsymbol{X}}, P_{\boldsymbol{X}})$ *è uno spazio di probabilità discreto.*

In un vettore aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$ ciascuna delle componenti X_i è una variabile aleatoria, con una sua distribuzione (S_{X_i}, P_{X_i}), ed è interessante ed utile in vari problemi studiare la relazione tra queste e la distribuzione congiunta $(S_{\boldsymbol{X}}, P_{\boldsymbol{X}})$. Le (S_{X_i}, P_{X_i}) sono dette **distribuzioni marginali**.

Per semplificare la notazione assumeremo che $S_{\boldsymbol{X}} = \mathbb{R}^n$ ponendo $P_{\boldsymbol{X}}(\boldsymbol{x}) = 0$ se $\boldsymbol{x}$ non era originalmente appartenente al codominio di $\boldsymbol{X}$.

Lemma 12. *Per ogni un vettore aleatorio* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *vale*

$$P_{X_1}(x_1) = \sum_{(x_2, \ldots, x_n) \in \mathbb{R}^{n-1}} P_{\boldsymbol{X}}(x_1, \ldots, x_n).$$

Dimostrazione. Per ogni $x_1 \in \mathbb{R}$, gli eventi

$$A_{(x_2,\ldots,x_n)} = \{s \in S : \boldsymbol{x}(s) = (x_1, \ldots, x_n)\} \subseteq S$$

sono disgiunti, per cui

$$\begin{aligned}\sum_{(x_2,\ldots,x_n)\in\mathbb{R}^{n-1}} P_{\boldsymbol{X}}(x_1, \ldots, x_n) &= \sum_{(x_2,\ldots,x_n)\in\mathbb{R}^{n-1}} P(A_{(x_2,\ldots,x_n)}) \\ &= P(\cup_{(x_2,\ldots,x_n)\in\mathbb{R}^{n-1}} A_{(x_2,\ldots,x_n)}) \\ &= P(s : X_1(s) = x_1) = P_{X_1}(x_1)\end{aligned}$$

□

Naturalmente il Lemma precedente si applica ad ogni proiezione o marginale X_i, per cui dalle congiunte si determinano le marginali.

Esercizio 13. *Dimostrare tramite un esempio che l'opposto non è vero e vi sono distribuzioni congiunte diverse che danno luogo alle stesse marginali.*

L'indipendenza delle variabili aleatorie si può esprimere in termini del rapporto tra le distribuzioni congiunta e marginali, nel senso che le componenti di un vettore aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$ sono indipendenti se e solo se

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} P_{X_i}(x_i)$$

per tutti gli $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Quindi nel caso indipendente le marginali permettono la ricostruzione della distribuzione congiunta.

Se di una o più variabili conosciamo il valore assunto abbiamo delle distribuzioni condizionali.

Definizione 21. *Dato un vettore aleatorio* $\boldsymbol{X} = (X_1, \ldots, X_n)$, *per* $k = 1, \ldots, n$, *se* $(x_1, \ldots, x_k)$ *è tale che* $P(X_1 = x_1, \ldots, X_k = x_k) \neq 0$, *si dice* **distribuzione condizionata o condizionale** *di* $X_{k+1}, \ldots, X_n$ *dato che* $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$ *lo spazio di probabilità costituito da*

$$\begin{aligned}&S_{X_{k+1},\ldots,X_n|(X_1,\ldots,X_k)=(x_1,\ldots,x_k)} \\ &\quad = \{(x_{k+1}, \ldots, x_n) : (x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \in S_{\boldsymbol{X}}\}\end{aligned}$$

e

$$\begin{aligned}&P_{X_{k+1},\ldots,X_n|X_1,\ldots,X_k}((x_{k+1}, \ldots, x_n)|(x_1, \ldots, x_k)) \\ &\quad = \frac{P_{\boldsymbol{X}}(\boldsymbol{x})}{P(X_1 = x_1, \ldots, X_k = x_k)}\end{aligned}$$

Naturalmente si può prendere $S_{X_{k+1},\ldots,X_n|(X_1,\ldots,X_k)=(x_1,\ldots,x_k)} = \mathbb{R}^{n-k}$.

Esercizio 14. *Verificare che la coppia definita nella definizione precedente è uno spazio di probabilità .*

Naturalmente la medesima definizione si poteva dare permutando gli indici. Dal teorema delle probabilità totali si vede che

$$\begin{aligned} &P(X_1 = x_1, \ldots, X_n = x_n) \\ &= \sum_{(x_1,\ldots,x_k)\in S_{X_1,dots,X_k}} P_{X_{k+1},\ldots,X_n|X_1,\ldots,X_k}((x_{k+1},\ldots,x_n)|(x_1,\ldots,x_k)) \\ &\qquad \cdot P(X_1 = x_1, \ldots, X_k = x_k). \end{aligned}$$

Per $k = 1$ questa osservazione indica come ricostruire la distribuzione congiunta dalla conoscenza delle distribuzioni condizionate e della relativa marginale.

Definizione 22. *Il valore atteso calcolato rispetto alla probabilità condizionale di una variabile aleatoria date le altre si chiama* **valore atteso condizionale** *e si denota*

$$E(X_n|(X_1,\ldots,X_{n-1}) = (x_1,\ldots,x_{n-1}).$$

7.2. Vettori aleatori continui

Anche per il caso continuo introduciamo i vettori aleatori, naturalmente attraverso le loro densità congiunta. Per semplicità ammettiamo insiemi di possibili discontinuità lineari.

Definizione 23. *Uno spazio di probabilità continua in $\mathbb{R}^n$ è una coppia $(\mathbb{R}^n, f)$ in cui $f : \mathbb{R}^n \to \mathbb{R}$ è una funzione continua tranne al più un insieme costituito da un numero finito di iperpiani tale che*

(i) $f(\boldsymbol{x}) \geq 0$ per ogni $\boldsymbol{x}\in \mathbb{R}^n$.

(ii) $\int_{\mathbb{R}^n} f(\boldsymbol{x})d\boldsymbol{x} = 1$

Si dice che un vettore aleatorio continuo $\boldsymbol{X}= (X_1,\ldots,X_n)$ n-dimensionale ha distribuzione $(\mathbb{R}^n, f_{\boldsymbol{X}})$ se per ogni plurirettangolo $A = \prod_{i=1}^n (a_i, b_i]$ si ha

$$P(\boldsymbol{X} \in A) = \int_A f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}.$$

La distribuzione di un vettore aleatorio $\boldsymbol{X}= (X_1,\ldots,X_n)$ è detta anche **distribuzione congiunta** *delle X_i e la funzione $f_{\boldsymbol{X}}(\boldsymbol{x})$ è detta* **densità congiunta**.

Si noti che gli integrali utilizzati in questa definizione sono integrali di Riemann (eventualmente impropri) in dimensione n e sono ben definiti essendo le funzioni coinvolte continue su insiemi normali. Si possono ripetere ora molte definizioni e proprietà dei vettori aleatori discreti.

Definizione 24. *In un vettore aleatorio continuo $\boldsymbol{X}= (X_1,\ldots,X_n)$, ciascuna delle sue componenti X_i è una variabile aleatoria continua, con una sua distribuzione $(\mathbb{R}, f_{X_i})$ ottenuta come* **distribuzione**

marginale *da*

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f(\boldsymbol{x}) dx_2 \cdot \ldots \cdot dx_{n-1}.$$

Naturalmente la definizione precedente si applica ad ogni proiezione o marginale X_i, per cui anche nel caso continuo dalle congiunte si determinano le distribuzioni marginali.

Esempio 43. *Sia $f = c \cdot \mathbb{I}_T$ in cui T è il triangolo*

$$\{\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 - x_1\}.$$

Dalla (ii) della definizione di spazio continuo si ha

$$\begin{aligned} 1 &= \int_{\mathbb{R}^2} f(\boldsymbol{x}) d\boldsymbol{x} \\ &= c \int_{\mathbb{R}} (\int_{\mathbb{R}} \mathbb{I}_T(\boldsymbol{x}) dx_2) dx_1 \\ &= c \int_0^1 (\int_0^{1-x_1} dx_2) dx_1 \\ &= c \int_0^1 (1 - x_1) dx_1 = c/2 \end{aligned} \tag{7.15}$$

da cui $c = 2$, poichè essendo i domini di integrazione normali e l'integranda continua l'integrale si può calcolare tramite integrali ripetuti. La marginale di X_1 si ottiene da

$$f_{X_i}(x_1) = \int_{\mathbb{R}^{n-1}} f(\boldsymbol{x}) dx_2 = 2(1 - x_1).$$

Data una funzione $\Phi : \mathbb{R}^n \to \mathbb{R}$ ed un vettore aleatorio $\boldsymbol{X}$, la composizione $\Phi(\boldsymbol{X})$ è una variabile aleatoria. Senza determinarne la densità possiamo fin da ora calcolarne il valore atteso rispetto alla distribuzione di $\boldsymbol{X}$ come

$$E(\Phi(\boldsymbol{X})) = \int_{\mathbb{R}^n} \Phi(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}.$$

Lemma 13. *Dato il vettore aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$ si ha*

$$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i).$$

Dimostrazione. Utilizzando l'induzione su n, è sufficiente dimostrare l'asserzione per $n = 2$:

$$\begin{aligned} E(X_1 + X_2) &= \int_{\mathbb{R}^2} (x_1 + x_2) f(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\mathbb{R}} (\int_{\mathbb{R}} x_1 f(\boldsymbol{x}) dx_2) dx_1 + \int_{\mathbb{R}} (\int_{\mathbb{R}} x_2 f(\boldsymbol{x}) dx_1) dx_2 \\ &= \int_{\mathbb{R}} x_1 f_{X_1}(x_1) dx_1 + \int_{\mathbb{R}} x_2 f_{X_2}(x_2) dx_2 = E(X_1) + E(X_2) \end{aligned}$$

□

L'indipendenza delle variabili aleatorie continue si può esprimere in termini del rapporto tra densità congiunta e marginali, nel senso che le componenti di un vettore aleatorio $\boldsymbol{X}= (X_1, \ldots, X_n)$ sono indipendenti se e solo se

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

per tutti $\boldsymbol{x}= (x_1, \ldots, x_n) \in \mathbb{R}^n$. Anche per le variabili continue le marginali permettono la ricostruzione della distribuzione congiunta quando le variabili siano indipendenti.

LEMMA 14. *Se le variabili aleatorie continue X_i, $i = 1, \ldots, n$ sono indipendenti allora*

$$E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i).$$

DIMOSTRAZIONE.

$$\begin{aligned} E(\prod_{i=1}^{n} X_i) &= \int_{\mathbb{R}^n} \prod_{i=1}^{n} x_i f(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} \prod_{i=1}^{n} (x_i f_{X_i}(x_i)) d\boldsymbol{x} \\ &= \prod_{i=1}^{n} \int_{\mathbb{R}^n} x_i f_{X_i}(x_i) dx_i = \prod_{i=1}^{n} E(X_i) \end{aligned}$$

□

Anche per le variabili continue è possibile definire le distribuzioni condizionali.

DEFINIZIONE 25. *Dato un vettore aleatorio $\boldsymbol{X}= (X_1, \ldots, X_n)$, per $k = 1, \ldots, n$, se $(x_1, \ldots, x_k)$ è tale che la densità marginale soddisfa $f_{(X_1,\ldots,X_k)}(x_1, \ldots .x_k) \neq 0$, si dice* **distribuzione condizionata** *di $X_{k+1}, \ldots, X_n$ dato che $(X_1, \ldots, X_k) = (x_1, \ldots, x_k)$ lo spazio di probabilità costituito da $\mathbb{R}^{n-k}$ con densità*

$$\begin{aligned} & f_{(X_{k+1},\ldots,X_n)|(X_1,\ldots,X_k)}((x_{k+1}, \ldots, x_n)|(x_1, \ldots, x_k)) \\ & \qquad = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{(X_1,\ldots,X_k)}(x_1, \ldots .x_k)} \end{aligned}$$

DEFINIZIONE 26. *Il valore atteso calcolato rispetto alla probabilità condizionale di una variabile aleatoria date le altre si chiama* **valore atteso condizionale** *e si denota*

$$E(X_n|(X_1, \ldots, X_{n-1}) = (x_1, \ldots, x_{n-1})).$$

Concludiamo con il risultato relativo alla trasformazione di densità per i vettori aleatori.

TEOREMA 11 (Trasformazione di densità per vettori aleatori). *Data un vettore aleatorio* $\boldsymbol{X}$ *n dimensionale continuo con densità* $f_{\boldsymbol{X}}$ *ed un'applicazione* $\phi : \mathbb{R}^n \to \mathbb{R}^n$ *continua e tale che il determinante dello Jacobiano* $\det(J(\phi(\boldsymbol{X})) \neq 0$, *con* $J(\phi(\boldsymbol{X})) = (\frac{\partial \phi_i}{\partial x_j})_{i,j=1,\dots,n}$, *quindi invertibile con inversa* $\boldsymbol{x} = \phi^{-1}(\boldsymbol{y})$ *si ha:*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\phi^{-1}(\boldsymbol{y}))|\det(J(\phi^{-1}(\boldsymbol{Y}))|$$

in tutti i punti di continuità di $f_{\boldsymbol{Y}}$.

DIMOSTRAZIONE. Per ogni plurirettangolo D si ha, effettuando nell'integrazione della terza riga il cambiamento di variabili $\boldsymbol{y} = \phi(\boldsymbol{x})$ con inversa $\boldsymbol{x} = \phi^{-1}(\boldsymbol{y})$:

$$\begin{aligned} \int_D f_{\boldsymbol{Y}}(\boldsymbol{y})d\boldsymbol{y} &= P(\boldsymbol{Y} \in D) \\ &= P(\phi^{-1}(\boldsymbol{Y}) \in \phi^{-1}(D)) \\ &= \int_{\phi^{-1}(D)} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\ &= \int_D f_{\boldsymbol{X}}(\phi^{-1}(\boldsymbol{y}))|\det(J(\phi^{-1}(\boldsymbol{Y}))|d\boldsymbol{y}. \end{aligned} \tag{7.16}$$

La conclusione segue per il teorema del valor medio come nel caso delle variabili aleatorie continue. □

7.3. Correlazione

Vogliamo introdurre ora una misura della dipendenza di variabili aleatorie, limitandoci al caso di due variabili per semplicità . Abbiamo visto che nel caso di variabili aleatorie indipendenti il valore atteso del prodotto si fattorizza, per cui viene naturale di studiare la quantità seguente:

DEFINIZIONE 27. *Dato un vettore aleatorio* (X, Y), *ossia due variabili aleatorie e la loro distribuzione congiunta, tali che* $E(X^2), E(Y^2) < \infty$, *si dice* **covarianza** *di* X *ed* Y *il valore*

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Si noti che vale $Cov(X, X) = Var(X)$.

LEMMA 15.

$$Cov(X, Y) = E((X - E(X)(Y - E(Y)).$$

Dimostrazione. Se $Cov(X,Y)$ esiste allora $E(X^2), E(Y^2) < \infty$ ed anche $E(|XY|), E(|X|), E(|Y|) < \infty$, il che implica che

$$\begin{aligned} E(|X-E(X)||Y-E(Y)|) &\leq E(|XY|+|X||E(Y)|+|Y||E(X)| \\ &\quad +|E(X)||E(Y)|) \\ &\leq E(|XY|)+3E(|X|)E(|Y|) < \infty \end{aligned}$$

per cui $E((X-E(X))(Y-E(Y))$ esiste. Ora

$$\begin{aligned} E((X-E(X))(Y-E(Y)) &\leq E(XY)-XE(Y)-YE(X) \\ &\quad +E(X)E(Y) \\ &\leq E(XY)+E(X)E(Y) < \infty \end{aligned}$$

□

Dimensionalmente $Cov(X,Y)$ è il prodotto delle dimensioni di X ed Y. Per ottenere un numero puro dividiamo per il prodotto delle deviazioni standard, o equivalentemente, consideriamo la correlazione tra le variabili standardizzate.

Definizione 28. *Dato un vettore aleatorio (X,Y) tale che $E(X^2) < \infty$ e $E(Y^2) < \infty$, si dice* **correlazione** *di X ed Y il valore*

$$r = r(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}.$$

Si ricorda che la variabili standardizzata è definita, per una variabile aleatoria discreta X con secondo momento finito, da $\bar{X} = \frac{X-E(X)}{SD(X)}$ e che valgono $E(\bar{X}) = 0$ e $Var(\bar{X}) = E(\bar{X}) = 1 = SD(\bar{X})$

Lemma 16. $r = E(\bar{X}\bar{Y}) = Cov(\bar{X}\bar{Y})$

Dimostrazione. Dalle proprietà di $\bar{X}$ e $\bar{Y}$ si ha

$$\begin{aligned} Cov(\bar{X}\bar{Y}) &= E(\bar{X}\bar{Y}) \\ &= E((\frac{X-E(X)}{SD(X)})(\frac{Y-E(Y)}{SD(Y)})) \\ &= \frac{Cov(X,Y)}{SD(X)SD(Y)} \end{aligned}$$

□

Oltre ad essere un numero puro la correlazione è limitata:

Teorema 12. *Per ogni vettore aleatorio (X,Y) con secondo momento finito si ha*

$$-1 \leq r(X,Y) \leq 1;$$

inoltre $r = 1$ se e solo se $Y = cX + d$ con $c > 0$ ed $r = -1$ se e solo se $Y = -cX + d$ con $c > 0$.

Dimostrazione. La dimostrazione che segue può essere letta in due modi, prima considerando solo i simboli in alto poi quelli in basso nelle coppie $\pm$ e $\mp$. Si ha che

$$\begin{aligned} 0 \leq E(\bar{X} \pm \bar{Y})^2 &= E(\bar{X}^2) + E(\bar{Y}^2) \pm 2E(\bar{X}\bar{Y}) \\ &= 2 \pm 2r(X,Y). \end{aligned} \tag{7.17}$$

Questo implica che

$$-1 \leq r(X,Y) \leq 1;$$

inoltre, se $r(X,Y) = \mp 1$ vale il segno di uguaglianza in (7.17) per cui $E(\bar{X} \pm \bar{Y})^2 = 0$. Poiché $(\bar{X} \pm \bar{Y})^2 \geq 0$ ne discende che $\bar{X} = \mp \bar{Y}$ e di conseguenza

$$Y = \mp \frac{SD(Y)}{SD(X)} X - \frac{SD(Y)}{SD(X)} E(X) + E(Y) = \mp cX + d$$

con $c = \frac{SD(Y)}{SD(X)} \geq 0$. $\square$

STATISTICA

Come già discusso, si indica con statistica l'insieme delle attività , ricerche e risultati finalizzati all'analisi dei risultati e dei dati ottenuti da esperimenti che in qualche modo si possono definire casuali. La statistica è generalmente suddivisa in parte descrittiva, che si occupa dell'elaborazione dei dati e parte inferenziale, che invece assume che i dati siano il risultato di un qualche meccanismo descrivibile in termini probabilistici, e quindi utilizza tutto quanto abbiamo visto finora per l'analisi dei dati stessi.

CAPITOLO 8

STATISTICA DESCRITTIVA

La statistica descrittiva si occupa del rapporto tra gli avvenimenti con risultati altamente non predicibili del mondo reale e la loro rappresentazione in termini scientificamente utilizzabili. Ad esempio, quando si si fanno osservazioni o si registrano i risultati di esperimenti la rappresentazione delle informazioni ottenute in termini trasmissibili ed analizzabili richiede il prendere molte decisioni spesso altamente opinabili. Per esempio, se vi trovaste ad analizzare i dati relativi a migliaia di pazienti per studiare certe patologie o certe terapie, spesso questi sarebbero sono scarabocchiati su fogli da medici diversi e vanno letti (e qui già ci vorrebbe l'aiuto di un medico), interpretati, trasferiti in forma unificata indovinando possibili differenze tra i diversi redattori e spesso rendendo numeriche delle valutazioni qualitative. Poi i dati vanno accuratamente trascritti in un data base, a quel punto analizzati per cercare delle indicazioni significative; una volta che ci siano delle osservazioni apparentemente sensate queste vanno giustificate e successivamente presentate.

E' chiaro che tutto questo è un processo lungo e delicato e che vi sono innumerevoli articoli e trattati che lo studiano. Noi qui ci occuperemo brevemente solo di un aspetto particolare: supponendo di avere dati già in forma numerica ci poniamo il problema di generare un grafico o alcuni valori rappresentativi.

Supponiamo quindi di avere un *campione* $x_1, \ldots, x_n \in \mathbb{R}^n$ e poniamo

$$x_{min} = \min\{x_i, i = 1, \ldots, n\}$$

e

$$x_{max} = \max\{x_i, i = 1, \ldots, n\}$$

così che $[x_{min}, x_{max}]$ è il range del campione.

8.1. L'istogramma

L'*istrogramma* dei dati, detto usualmente istogramma, è una rappresentazione grafica dei dati che rappresenta frequenze attraverso le aree. Descriveremo ora una modalità di realizzazione degli istogrammi, che come si vedrà prevede comunque numerose scelte arbitrarie.

Dividiamo $[\lfloor x_{min} \rfloor, \lfloor x_{max} \rfloor + 1]$ in intervallini scegliendo

$$\lfloor x_{min} \rfloor = d_0 < d_1 < \cdots < d_m = \lfloor x_{max} \rfloor + 1$$

e formando intervallini semiaperti a destra (scegliendoli semiaperti a destra si semplificano leggermente le definizioni):

$$[d_k, d_{k+1}[$$

per $k = 0, \dots, m-1$. L'istogramma è la funzione che per $x \in [d_k, d_{k+1}[$ vale

$$I(x) = \frac{|\{i : x_i \in [d_k, d_{k+1}[\}|}{n(d_{k+1} - d_k)}.$$

Esercizio 15. *Con dati* $0, 1, 2, 2, 1, 1, 0, 2, 1$ *si disegnino gli istogrammi ottenuti con la scelta* $d_0 = 0, d_1 = 1, d_2 = 3$ *e con la scelta* $d_0 = 0, d_1 = 1, d_2 = 2, d_3 = 3$.

Osservazione 4. *Attraverso l'istogramma possiamo capire e mostrare dove i dati si concentrano di più di meno, anche se l'arbitrarietà della scelta degli intervalli e del comportamento agli estremi degli intervalli lascia spazio a molti fraintendimenti e possibili rappresentazioni falsate.*

Osservazione 5. *Se cominciamo ad immaginare, come faremo nella statistica inferenziale, che i dati siano realizzazioni di un qualche meccanismo casuale, ad esempio variabili aleatorie continue indipendenti, allora una delle cose che possono essere suggerite dall'istogramma è la forma della densità di tali variabili aleatorie.*

8.2. Elementi di Statistica Descrittiva

Per descrivere sinteticamente i dati si utilizzano:
(I) **valor medio empirico** $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$;
(II) **varianza empirica** $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})$;
(III) **SD empirica** $SD = \sigma_n = \sqrt{\sigma_n^2}$;
(IV) **momento empirico** k**-simo** $m_k = \frac{1}{n}\sum_{i=1}^n x_i^k$;
per motivi che non discuteremo qui si utilizzano spesso σ_{n-1}^2 e σ_{n-1}, definiti come i precedenti con $n-1$ al posto di n.

Per dati accoppiati $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^n$, quindi riferiti a due variabili aleatorie per ogni individuo della popolazione, si considera la
(v) **correlazione empirica** $r = \frac{1}{n}\sum_{i=1}^n x_i y_i$

CAPITOLO 9

STATISTICA INFERENZIALE

Immaginiamo ora che i dati siano realizzazioni di un qualche meccanismo casuale descrivibile nell'ambito del calcolo delle probabilità. Consideriamo un caso semplice, in cui i dati del campione che stiamo considerando siano realizzazioni di variabili aleatorie indipendenti ed identicamente distribuite, discrete o continue a seconda che i dati osservati siano numeri interi o meno.

Esempio 44. *Possiamo immaginare che i dati dell'esempio 15 di cui sono stati fatti gli istogrammi nel paragrafo precedente siano realizzazioni di variabili i.i.d.* $B(2,p)$*, ossia ognuno dei dati corrisponda al numero di successi in 2 prove indipendenti ognuna con probabilità di successo* p*; oppure che siano* $B(n,p)$*.*

9.1. Stima di parametri: il metodo dei momenti

Una volta identificata la distribuzione delle variabili aleatorie spesso rimangono dei parametri da determinare, come ad esempio la probabilità di successo p nell'ultimo esempio o uno o più parametri di una distribuzione continua. A questo scopo sono stati sviluppati vari metodi, e noi ne presenteremo uno chiamato metodo dei momenti.

In generale consideriamo che la distribuzione dipenda da l parametri $\theta_1, \dots, \theta_l$. Vogliamo determinare dei valori

$$(\hat{\theta}_1, \dots, \hat{\theta}_l) = (\hat{\theta}_1(x_1, \dots, x_n), \dots, \hat{\theta}_l(x_1, \dots, x_n))$$

che siano stime ragionevoli dei valori veri dei parametri.

Si considerano allora i momenti k-simi della distribuzione $E(X^k) = E_{\theta_1,\dots,\theta_l}(X^k)$ e si uguagliano ai momenti k-simi empirici corrispondenti, formando un numero di equazioni sufficiente a permettere di determinare la stima dei parametri ponendo $(\hat{\theta}_1, \dots, \hat{\theta}_l)$ uguale alle soluzioni del sistema composto di tali equazioni $m_k = E(X^k)$.

Esempio 45. *Con i dati dell'esercizio 15 si può stimare la probabilità* p *della binomiale con il metodo dei momenti. Ora se* $X \sim B(2,p)$ *allora* $E(X) = 2p$ *ed* $m_1 = \bar{x} = 10/9$*, per cui* $\hat{p} = 10/18$*.*

Esempio 46. *Se invece interpretiamo i dati* $-2, 3, 1$ *come realizzazioni di variabili aleatorie i.i.d. continue* $X_i \sim U([-a,a])$ *allora*

$f_{X_i}(x) = \frac{1}{2a}\mathbb{I}_{[-a,a]}(x)$ *e* $E(X_i) = 0$ *per simmetria e*

$$E(X_i^2) = \frac{1}{2a}\int_{-a}^{a} x^2 dx = \frac{a^2}{3}$$

per cui $\hat{a} = \sqrt{14} \approx 3.74$. *Se* $\hat{a}$ *fosse risultato minore di* x_{max} *ne avremmo concluso che o il metodo di stima dei parametri non era adeguato o le assunzioni fatte non costituivano un buon modello per i dati.*

9.2. Test di ipotesi: lo z-test

Valutare se le ipotesi fatte siano un buon modello per i dati è naturalmente una questione assai complessa. Qui discuteremo uno schema che in alcuni casi offre qualche elemento di valutazione. Filosoficamente, si tratta della possibilità di evidenziare o meno dai dati elementi che contraddicano statisticamente il complesso delle ipotesi assunte. Nessuna conclusione verrà tratta in positivo: o le ipotesi sono contraddette o non lo sono (almeno sulla base dell'analisi fatta di questi dati).

Lo schema che presentiamo è detto test di ipotesi e si realizza come segue:

(0) Si fissa un livello del test, detto α, tipo $\alpha = 0.05$ oppure $\alpha = 0.01$.

(1) Si descrive l'ipotesi che si vuole testare, detta Ipotesi Nulla H_0. Tipicamente in questo si dovrebbe descrivere l'insieme di tutti i passi che costituiscono l'esperimento. Di solito si indica anche in dettaglio una ipotesi alternativa, chiamata H_1 che si testa in contrasto con H_0, ma per semplicità assumiamo che H_1 sia semplicemente la negazione di H_0.

(2) Si determina il valore di una certa funzione dei dati e dei parametri ottenuti da H_0 che si possa considerare come la realizzazione di una variabile aleatoria di cui sia nota, almeno approssimativamente, la distribuzione. Nel test che discutiamo qui questo valore sarà indicato con z.

(3) Si stima la probabilità che la variabile aleatoria Z di cui z è una realizzazione devii da quanto atteso per più di quanto osservato. Tale probabilità è detta valore p del test.

(4) Si confrontano p e α. Se $p < \alpha$ si rigetta H_0 ed altrimenti non si rigetta.

Come esempio consideriamo lo z-test in cui si valuta l'ipotesi attraverso il confronto tra la differenza della media campionaria da quella attesa con la deviazione standard per dati di variabili aleatorie normali. Lo useremo in forma approssimata utilizzando il Teorema Centrale del Limite.

Da H_0 ricaviamo il valore atteso $E(X)$ e la deviazione standard $SD(X)$: il CLT ci dice che

$$Z = \frac{\sum_{i=1}^{n} X_i - E(X)}{\sqrt{nSD(X)}}$$

ha una distribuzione approssimativamente $N(0,1)$ quando le X_i siano i.i.d. tali che $E(X_i^2) < \infty$. Quindi conosciamo approssimativamente la distribuzione di Z e nel test valutiamo quindi la sua realizzazione

$$Z = \frac{\sum_{i=1}^{n} x_i - E(X)}{\sqrt{nSD(X)}}.$$

Esempio 47. *Ritornando all'esempio 1, se esce testa 459 volte su 1000 lanci di una moneta, allora si esegue un test di ipotesi relativamente alla correttezza della moneta ponendo:*

(0) $\alpha = 0.02$

(1) H_0*= la moneta è equilibrata e ben mescolata ad ogni lancio; i risultati sono correttamente registrati.*

(2) Da H_0 *i risultati sono relativi a prove i.i.d.* $B(1000, 1/2)$*. Per cui* $E(\sum_{i=1}^{1000} X_i) = 1/2$ *e* $SD(\sum_{i=1}^{1000} X_i) = \sqrt{1000}/2$ *e*

$$z = \frac{459 - 500}{\sqrt{1000}/2} \approx -2.59$$

(3) Il valore p si determina dall'approssimazione normale come

$$\int_{-\infty}^{-2.59} \frac{\sqrt{2\pi}}{e^{-\frac{x^2}{2}}} \approx 0.0048$$

(4) Poichè $0.0048 = p < \alpha = 0.02$ *si rigetta* H_0*, ossia si conclude che l'insieme delle ipotesi fatte non è consistente con i dati osservati.*

Osservazione 6. *Avevamo già visto che questo numero di teste non era accettabile; la terminlogia del test di ipotesi ha permesso di chiarire bene la procedura e le conclusioni.*

Ora possiamo concludere l'analisi dell'esempio 2 attraverso uno z-test approssimato.

Esempio 48. *(0)* $\alpha = 0.02$

(1) Un modello adeguato è assumere che il tempo di funzionamento di ogni pezzo i è una variabile aleatoria X_i *indipendente con distribuzione* $exp(\lambda)$*. Quindi poniamo*

H_0 *= I tempi di vita sono realizzazioni di variabili aleatorie i.i.d. esponenziali di parametro* λ *tale che il tempo medio di funzionamento previsto corrisponde al valor medio della distribuzione, ossia* $\frac{1}{\lambda} = 1000$*, quindi* $\lambda = \frac{1}{1000}$*.*

(2) Da H_0 *si ottiene che* $E(X) = 1000$ *e* $SD(X) = 1000$*. Il tempo totale di funzionamento corrisponde a* $\sum_{i=1}^{100} X_i$ *e dal teorema centrale*

del limite conosciamo approssimativamente la distribuzione di

$$z = \frac{90200 - 100 \cdot 1000}{\sqrt{100} \cdot 1000} \approx -0.98$$

(3) Dall'approssimazione normale

$$p \sim \int_{-\infty}^{-0,98} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \sim 16,35\%. \tag{9.18}$$

(4) Poichè $p \geq \alpha$ ne concludiamo che l'ipotesi nulla non è invalidata dai dati.

Quindi un tempo totale di funzionamento di 90000 ore era perfettamente in linea con la durata media prevista.

Esercizio 16. *Verificare che con una soglia dell'1% si dovrebbe considerare falsa la dichiarazione sulla durata media se il tempo totale di funzionamento fosse minore di 76700 ore.*

Esercizio 17. *Perchè questa volta si tende a considerare inappropriato un risultato ancora perfettamente adeguato?*