

Rapporto n. _____ 200

dmsia  unibg.it



**Dipartimento
di Matematica, Statistica,
Informatica e Applicazioni
“Lorenzo Mascheroni”**

UNIVERSITÀ DEGLI STUDI DI BERGAMO



CORSO
DI
STATISTICA
MATEMATICA

LUCIO BERTOLI-BARSOTTI

Indice

I PARTE

Sezione I.1.

- 1. Probabilità classica..... 1
- 2. Il problema di Galileo della somma del punteggio di tre dadi1
- 3. “Anagrammi” di parole con lettere ripetute o meno.....2

Sezione I.2.

- 1. Problemi di conteggio dovuti all’interpretazione della vera molteplicità dei casi osservabili..... 3
- 2. Eventi – una formalizzazione astratta.....4
- 3. Evento “verificato”. Evento certo. Evento impossibile.....5
- 4. Interpretazione logica degli operatori insiemistica.....5
- 5. Eventi incompatibili.....5
- 6. Concezioni probabilistiche.....5

Sezione I.3.

- 1. Gli assiomi del Calcolo.....6
- 2. Primi teoremi.....7

Sezione I.4.

- 1. Una nota sulla formulazione degli assiomi.....7
- 2. Eventi stocasticamente indipendenti.....8
- 3. Probabilità condizionata ad un evento.....8
- 4. Teorema delle probabilità totali e formula di Bayes..... .8

Sezione I.5.

- 1. Formula della probabilità condizionata per successioni di eventi..... 10
- 2. Nota su indipendenza e incompatibilità.....11

3. Variabili casuali.....	11
4. Distribuzioni.....	11
5. Punti di salto di una f.d.r.	12
Sezione I.6.	
1. Corrispondenza fra v.c. e f.d.r.	12
2. V.c. discrete	13
3. V.c. continue.....	14
4. Altre vv.cc.	14
Sezione I.7.	
1. Valore atteso di una variabile casuale	15
2. Giochi equi	15
3. Il paradigma della scommessa e la costruzione della Probabilità soggettiva..	16
4. Il problema della coerenza nell'assegnazione della Probabilità soggettiva....	16
Sezione I.8.	
1. Assegnazione della Probabilità soggettiva e paradosso di S.Pietroburgo.....	17
2. Linearità dell'operatore valore atteso.....	17
3. La varianza di una variabile casuale.....	18
4. Alcune proprietà della varianza	18
5. Momenti non-centrali di una v.c.	18
Sezione I.9.	
1. Momenti centrali di una v.c.	19
2. Momenti centrali in funzione dei momenti non-centrali.....	19
3. Indici di forma della distribuzione	19
4. Momenti fattoriali di una v.c.	19
5. V.c. discrete di particolare interesse applicativo – “schema dell'urna”	20
6. V.c. bernoulliana.....	20
7. V.c. binomiale.....	20
Sezione I.10.	
1. V.c. ipergeometrica.....	21
2. V.c. di Poisson.....	22
3. V.c. geometrica.....	23

Sezione I.11.

1. Un integrale notevole dell'Analisi e alcune formule da esso conseguenti.....	23
2. V.c. normale standard.....	24
3. V.c. normale.....	25

Sezione I.12.

1. La funzione speciale Gamma di Eulero.....	25
2. Una applicazione di calcolo con la Gamma di Eulero: momenti di ordine pari della normale.....	26
3. V.c. gamma.....	26

Sezione I.13.

1. Vv.cc. collegate alla v.c. gamma: v.c. esponenziale e v.c. chi-quadrato.....	27
2. V.c. di Cauchy.....	29
3. V.c. di Laplace (esponenziale doppia).....	29
4. V.c. logistica.....	29
5. V.c. beta.....	30
6. V.c. Weibull	31
7. V.c. t di Student	32
8. V.c. beta-binomiale	33
9. Disuguaglianza di Chebyshev.....	34

Sezione I.14.

1. Vv.cc. indipendenti.....	34
2. Vv.cc. somiglianti.....	35
3. Media e varianza di una somma e di una media di vv.cc. indipendenti e somiglianti.....	36

Sezione I.15.

1. Convergenze stocastiche.....	36
2. Legge dei grandi numeri.....	37
3. Teorema del Limite Centrale.....	37

II PARTE

Sezione II.1.

1. Funzioni di ripartizione multivariate.....	39
2. Un controesempio.....	40
3. Componenti marginali di una v.c. multipla (a componenti indipendenti).....	40
4. Esempio: v.c. multinomiale.....	42

Sezione II.2.

1. Valore atteso di una funzione di una v.c. multipla – Momenti di una v.c. multipla.....	42
2. Matrice di covarianza e matrice di correlazione.....	43
3. Esempio.....	44
4. Distribuzioni condizionate: densità.....	45
5. Distribuzioni condizionate: valori attesi.....	46
6. Esempio.....	46

Sezione II.3.

1. Distribuzione normale bivariata “standard”.....	47
2. Esempio.....	48
3. Distribuzione normale bivariata: densità.....	50
4. Distribuzione normale bivariata: distribuzioni condizionate.....	51
5. Distribuzione normale multivariata*.....	51
6. Distribuzione normale multivariata: distribuzioni condizionate*.....	52

Sezione II.4.

1. Funzione caratteristica.....	53
2. V.c. speciali – funzioni caratteristiche.....	56
3. Teorema del limite centrale (TLC): dimostrazione.....	59

Sezione II.5.

1. Trasformazioni di vv.cc.....	60
2. Esempi di trasformazioni: la trasformazione lineare – famiglia locazione e scala.....	62
3. Esempi di trasformazioni: una trasformazione monotona a tratti.....	63
4. Esempi di trasformazioni: trasformazioni regolari – somma di vv.cc.....	65
5. Esempi di trasformazioni: una particolare trasformazione regolare.....	65

Sezione II.6.

1. Variabile di campionamento.....	66
2. Alcune importanti trasformazioni ottenute a partire da c.c.s. da normale....	67
3. Somma campionaria e media campionaria (varianza nota).....	67
4. Distribuzioni collegate alla varianza campionaria	68
5. Media campionaria (varianza ignota).....	68

Sezione II.7.

1. Inferenza statistica parametrica.....	69
2. Stimatore.....	69
3. Errore quadratico medio.....	70
4. Efficienza.....	70

Sezione II.8.

1. Metodo dei momenti.....	70
2. Massima verosimiglianza: concetto base.....	72
3. Funzione di verosimiglianza	73
4. Aspetti computazionali – approssimazioni numeriche per il calcolo della stima di ML.....	74
5. Invarianza della stima di massima verosimiglianza.....	77

Sezione II.9.

1. Sufficienza.....	78
2. Criteri per la verifica della sufficienza.....	78
3. Completezza.....	80
4. Famiglia esponenziale.....	82
5. Efficienza secondo Rao-Blackwell / Lehmann-Scheffé.....	82
6. Efficienza secondo Rao-Cramér	82

Sezione II.10.

1. Intervalli di confidenza.....	83
2. Intervallo di confidenza per la media di una normale – varianza nota.....	84
3. Intervallo di confidenza per la media di una normale – varianza ignota.....	85
4. Intervallo di confidenza per la varianza di una normale.....	85
5. Intervalli di confidenza asintotici per la media.....	86
6. Intervalli di confidenza asintotici per la varianza.....	86

Premessa

Questa breve monografia è stata pensata come testo a supporto di un corso di Statistica Matematica di 60 ore. La Statistica Matematica è basata sulla Teoria della Probabilità; buona parte di questo lavoro (quanto meno tutta la I Parte) tratta temi che si inquadrano tipicamente in quest'ultima.

L'obiettivo finale del corso è quello di arrivare a introdurre lo studente alle classiche tecniche dell'Inferenza Statistica, in particolare la Stima, *puntuale e intervallare*.

Il testo è suddiviso in sezioni, con enfasi alla ripartizione temporale (piuttosto che tematica) dell'esposizione del docente: le 15 sezioni della prima parte corrispondono infatti a 15 "lezioni", ciascuna composta da 2 ore accademiche, nelle quali può essere articolato un "modulo" da 30 ore.

Nella II Parte le sezioni sono più ricche di contenuti, pertanto alcune di esse sono pensate per essere diluite nell'arco di 2 lezioni.

L'esposizione è relativamente scarna, essendo inteso, in particolare, che il lettore è rimandato:

- a) per gli esercizi, all'apposito riferimento Bertoli-Barsotti (1996);
- b) per le tavole statistiche al programma PQRS scaricabile gratuitamente all'indirizzo internet: <http://www.eco.rug.nl/medewerk/knypstra/pqrs.shtml> .

Lucio Bertoli Barsotti

I PARTE

SEZIONE I.1.

1. Probabilità classica

Si dà per nota la nozione intuitiva di “probabilità” come “grado di certezza” riguardo al verificarsi di un accadimento aleatorio. Il modello di riferimento “ingenuo” cosiddetto della *probabilità classica*, interpreta la probabilità come il rapporto fra il numero di casi favorevoli e il numero di casi possibili – nelle situazioni di simmetria che consentano di ritenere i casi in oggetto come “equiprobabili”. Pur essendo evidente l’inconsistenza, da un punto di vista epistemologico -a rigore-, di tale interpretazione della probabilità (proprio perché questa è definita appoggiandosi a una non-definita nozione di “equiprobabilità”), se ne riconosce la rilevanza: a) per elaborazioni di calcolo riferibili a contesti particolarmente semplici quali quelli tipici dei giochi di sorte – utili a fini didattico-esemplificativi; b) per la costruzione e l’interpretazione di modelli probabilistici complessi.

2. Il problema di Galileo della somma del punteggio di tre dadi

Si considera un problema descritto da Galileo Galilei (1564-1642) nel suo lavoro “*Sopra le scoperte dei dadi*” (per una breve introduzione ad alcuni dei problemi probabilistici classici si veda Bertoli-Barsotti, 1995; un riferimento per un quadro storico completo sulle origini è Hald, 1990; per una analisi del concetto da un punto di vista storico-epistemologico si veda Hacking, 1987), ossia il calcolo della probabilità di ottenere il risultato “9” e quella di ottenere “10”, come somma dei punteggi prodotti nel lancio di tre dadi equi. Si enumerano, ordinatamente, le 6 possibili combinazioni di facce che danno luogo al risultato “9”, e le 6 possibili combinazioni di facce che danno luogo al risultato “10” (tutte le possibili combinazioni di facce sono 56, a meno di permutazioni):

somma 9	somma 10
6,2,1	6,3,1
5,3,1	6,2,2
5,2,2	5,4,1
4,4,1	5,3,2
4,3,2	4,4,2
3,3,3	4,3,3

A questo punto ci si può chiedere: i due risultati hanno la medesima probabilità di uscita? come enumerare i risultati possibili? a meno di permutazioni? Poiché l'indistinguibilità dei dadi è un effetto apparente, ma non sostanziale (è possibile ad es. contrassegnare i dadi per diversificarli...), essendo i dadi oggetti macroscopici differenti (si veda però I.2.1, per un caso più generale), si conclude che *vanno contate tutte le diverse permutazioni* delle terne individuate.

3. "Anagrammi" di parole con lettere ripetute o meno

Il calcolo della probabilità classica porta spesso, dunque, a risolvere problemi di calcolo combinatorio.

Si studia quindi il problema dell'enumerazione di stringhe di simboli che possono o meno ripetersi (ovvero "anagrammi" di parole con lettere ripetute e non).

Una parola di n lettere diverse ha $n!$ anagrammi (*permutazioni*).

Una parola di n lettere di r ($r \leq n$) tipi diversi ha

$$\frac{n!}{s_1!s_2!\dots s_r!}$$

anagrammi, dove s_i è il numero di lettere di tipo i , $i = 1, 2, \dots, r$.

Si dimostra che si contano, in particolare, le *combinazioni* se e solo se si contano gli anagrammi di parole con *due* soli tipi di lettere.

Le combinazioni $\binom{n}{k}$, o coefficienti binomiali, compaiono nel triangolo di

Pascal, calcolate iterativamente riga per riga; si osservi che la somma di tutte le combinazioni in riferimento ad un prefissato numero di oggetti n (totale di riga

n -ma del triangolo) dà la cardinalità dell'insieme delle parti di un insieme di cardinalità n , cioè 2^n .

Esempio - soluzione del problema di Galileo. Per il risultato “9” si trova:

<i>terna</i>	<i>“anagrammi”</i>
6,2,1	3!
5,3,1	3!
5,2,2	3!/2!
4,4,1	3!/2!
4,3,2	3!
3,3,3	1
TOTALE	25

Per il risultato “10” si trova:

<i>terna</i>	<i>“anagrammi”</i>
6,3,1	3!
6,2,2	3!/2!
5,4,1	3!
5,3,2	3!
4,4,2	3!/2!
4,3,3	3!/2!
TOTALE	27

SEZIONE I.2.

1. Problemi di conteggio dovuti all'interpretazione della “vera” molteplicità dei casi osservabili

Il problema di Galileo propone l'enumerazione delle configurazioni diverse ottenibili con r oggetti (dadi) che possono assumere (aleatoriamente e con ugual probabilità) n diversi stati (le diverse facce numerate): in altri termini si tratta di contare le possibili configurazioni ottenibili disponendo in n celle diverse r oggetti (che possiamo pure supporre sperimentalmente indistinguibili all'occhio

dello sperimentatore). Si può analizzare il problema in tre diverse ipotesi così formulate (cfr. Gnedenko, 1992, p. 31-34):

- a) gli oggetti *sono diversi* (quindi, ad es., AAB è diverso da ABA; cdd. statistica di *Maxwell-Boltzman*);
- b) sono possibili ed equiprobabili le configurazioni diverse, a meno dell'ordine (quindi, ad es., AAB è indistinguibile da ABA; cdd. statistica di *Bose-Einstein*);
- c) sono possibili ed equiprobabili le configurazioni diverse, a meno dell'ordine, con al più un oggetto per cella (quindi, ad es., AAB è impossibile, mentre ABC è indistinguibile da CAB, cdd. statistica di *Fermi-Dirac*).

Rispettivamente si trova:

a) Ognuno degli n stati può essere assunto da ciascuno degli r oggetti, quindi si hanno n^r configurazioni.

b) I diversi stati sono ordinatamente elencati con $n - 1$ segni di separazione del tipo “ / ”, mentre gli r oggetti -indistinguibili- da segni del tipo “ * ”. Si contano

quindi $\binom{n-1+r}{r}$ configurazioni.

c) I diversi stati sono ordinatamente elencati con segni “1” e “0”, rispettivamente per indicare se lo stato è “occupato” o meno da qualche oggetto. Poiché ci sono r

segni “1”, si contano $\binom{n}{r}$ configurazioni.

Tutti questi contesti sono fisicamente rilevanti e validi. L'esempio mostra che c'è una fondamentale e ineludibile rilevanza dell'effettivo contesto sperimentale di riferimento, nella determinazione della probabilità del verificarsi di un evento aleatorio.

2. Eventi – una formalizzazione astratta

Si introduce la nozione astratta di esperimento aleatorio \mathcal{E} . Ogni risultato di \mathcal{E} è detto *evento elementare*. L'insieme Ω degli eventi elementari è detto di spazio fondamentale. In senso lato, un *evento* è una “collezione (eventualmente vuota) di eventi elementari”. Quindi, in particolare, un evento elementare è un evento. La definizione va però meglio precisata in termini più rigorosi.

Si considera, innanzi tutto, il paradigma di riferimento più elementare, costituito dal familiare caso in cui Ω ha cardinalità finita. Si può allora definire “evento”

un qualsiasi elemento dell'insieme delle parti di Ω , $\mathcal{P}(\Omega)$. Si noti che, in riferimento ad uno spazio Ω con cardinalità n si hanno, in corrispondenza, 2^n eventi.

Più in generale, se Ω è *numerabile*, allora ogni suo sottoinsieme può considerarsi un evento. Viceversa, se Ω non è *numerabile*, non tutti i suoi sottoinsiemi possono essere considerati eventi (vd. sezione I.4).

3. Evento “verificato”. Evento certo. Evento impossibile

Un evento E si dice “verificato” se il risultato dell'esperimento attua (i.e. soddisfa, rende vero) E nel senso che *appartiene* insiemisticamente ad E .

In tal senso Ω è un evento che è sempre verificato dall'esperimento \mathcal{E} ; perciò si dice anche *evento certo*.

Al contrario, l'insieme vuoto \emptyset non è mai verificato; perciò si dice anche *evento impossibile*.

4. Interpretazione logica degli operatori insiemistici

Si ricordano i significati logici degli operatori insiemistici: unione=“oppure”, intersezione= “e”, complemento (rispetto a Ω)= “non”, differenza=“ma non”.

5. Eventi incompatibili

Due eventi si dicono *incompatibili* se la loro intersezione è l'insieme vuoto. In altri termini, due eventi sono incompatibili se non c'è nessun risultato di \mathcal{E} che possa verificarli congiuntamente.

6. Concezioni probabilistiche

Il problema della determinazione della specifica probabilità di un evento verrà toccato per certi versi solo indirettamente in questo Corso. Quando sarà esplicitamente posto (p.es. in sede di esercizio, di esempio o di problema) questo problema verrà risolto, in questa sede, facendo essenzialmente riferimento a contesti ideali (ad es. quelli tipici dei giochi di sorte -peraltro con utilizzo dadi, mazzi di carte, monete, ecc. “perfetti”- o similari) ragionevolmente passibili di essere affrontati con il solo ricorso alla nozione di probabilità classica.

Tuttavia è il caso di rimarcare che, nel caso generale, tale impostazione, come accennato (cfr. I.1.1), è del tutto insufficiente.

La determinazione della probabilità di un evento, in generale, è operata oggi secondo uno dei due seguenti paradigmi (per approfondimenti si veda Hacking, 1987):

- 1) l'impostazione *frequentista*;
- 2) l'impostazione *soggettivista*.

La prima -detta anche *oggettivista* o *statistica*- assegna la probabilità solo a una ristretta classe di esperimenti aleatori: quelli che possono ritenersi (anche solo "potenzialmente") operativamente *ripetibili* nelle "stesse", ovvero simili, condizioni sperimentali. Il lancio di una moneta, ad es., è comunemente ritenuto un esperimento ripetibile nelle "stesse" condizioni (perciò lo si accetta comunemente come criterio sul quale basare l'operazione di scelta del campo prima di una partita di calcio). La ripetizione, di un esperimento ripetibile, un grande numero N di volte conduce a frequenze relative tendono a stabilizzarsi. Si tratta di una evidenza empirica.

L'idea è che proprio questo valore limite della frequenza relativa, diciamo $n(E)/N$, del verificarsi di un evento aleatorio E possa essere interpretato come la *probabilità* di E . Un numero idealmente determinabile con qualsiasi grado di approssimazione, sperimentalmente "oggettivo".

La seconda impostazione -detta anche *epistemica*- consente di assegnare la probabilità a una classe ben più ampia di esperimenti aleatori (non necessariamente ripetibili), ma perde, in generale, ogni speranza di essere determinata univocamente. Secondo l'impostazione soggettivista il valore della probabilità di un evento E dipende infatti dall'osservatore. Poiché la nozione necessita, per essere meglio illustrata, di quella di "valore atteso", si rimanda per ulteriori particolari al capitolo I.7.

SEZIONE I.3.

1. Gli assiomi del Calcolo

Abbiamo ora tutti gli elementi nozionistici che ci servono per introdurre la assiomatizzazione di Kolmogorov. L'assiomatizzazione di Kolmogorov (1933) consente di definire la probabilità in quanto oggetto matematico dotato di alcune ragionevoli (coerenti, in particolare con il significato intuitivo di tale nozione) proprietà, mettendo fra parentesi il problema della sua determinazione pratica. L'idea è quindi quella di stabilire un certo numero di proprietà matematiche che possano essere (più o meno pienamente) accettate a prescindere dalla specifica concezione probabilistica (cfr. I.2.6).

Si introduce quindi la nozione di probabilità come funzione reale P con dominio su un assegnato *spazio probabilizzabile*, ovvero su una assegnata collezione di sottoinsiemi di Ω .

Facendo in prima battuta riferimento al caso di Ω finito (o al più numerabile), si considera dominio della funzione P l'insieme delle parti di Ω , $\mathcal{P}(\Omega)$.

Si richiede una volta per tutte, convenzionalmente, che tale funzione soddisfi tre proprietà (*assiomatizzazione di Kolmogorov*):

- I) (non-negatività e limitatezza) $P(A) \in [0,1], \forall A \in \mathcal{P}(\Omega)$;
- II) (normalizzazione) $P(\Omega)=1$;
- III) (additività finita) la probabilità dell'unione di un numero finito di eventi incompatibili è la somma delle probabilità di quegli eventi.

Questo terzo assioma può porsi in forma più forte consentendo l'operazione di unione su una infinità numerabile di eventi.

- III*) (additività numerabile) la probabilità dell'unione di una infinità numerabile di eventi incompatibili è la somma delle probabilità di quegli eventi.

2. Primi teoremi

Alcuni semplici risultati si deducono immediatamente dagli assiomi, come teoremi.

In particolare:

- 1) (probabilità del complemento) Se $A \in \mathcal{P}(\Omega)$, allora $P(\bar{A}) = 1 - P(A)$;
- 2) (probabilità dell'insieme vuoto) $P(\emptyset) = 0$;
- 3) (probabilità dell'unione) Se $A, B \in \mathcal{P}(\Omega)$, allora

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

- 4) (monotonicità della probabilità) Se $A, B \in \mathcal{P}(\Omega)$, allora $A \subset B$ implica

$$P(A) \leq P(B);$$

- 5) (continuità) Se $A_1, A_2, \dots \in \mathcal{P}(\Omega)$, allora $A_1 \subset A_2 \subset A_3 \subset \dots$ implica

$$P\left(\bigcup_i A_i\right) = \lim_{i \rightarrow \infty} P(A_i).$$

SEZIONE I.4.

1. Una nota sulla formulazione degli assiomi.

Possiamo chiederci se esistono classi “più semplici” dell’insieme delle parti di Ω , $\mathcal{P}(\Omega)$, che possano costituire un appropriato dominio per la funzione “probabilità” P .

La domanda è giustificata per il fatto che, quando Ω è particolarmente complesso (p.es. quando Ω coincide con un sottoinsieme di \mathfrak{R}), l’insieme $\mathcal{P}(\Omega)$ risulta essere troppo esteso per poter costituire, coerentemente, un dominio per P . La risposta alla domanda è sì: esistono classi più semplici di $\mathcal{P}(\Omega)$. Occorre comunque chiedere che tali classi soddisfino certi requisiti. Per esempio debbono risultare “chiuso” rispetto ad operazioni insiemistiche sui suoi elementi - in virtù del significato logico di queste operazioni (cfr. I.2.4). Precisamente, l’ambiente ideale è la struttura di σ -algebra. Una collezione \mathcal{B} di sottoinsiemi di Ω ha la struttura di σ -algebra quando:

- contiene l’insieme vuoto \emptyset ;
- contiene Ω ;
- è chiusa rispetto all’operazione di complemento (i.e. se $A \in \mathcal{B}$ allora $\bar{A} \in \mathcal{B}$);
- è chiusa rispetto alle operazioni insiemistiche di unione, intersezione e differenza, anche iterate infinite volte (i.e. se $A_1, A_2, \dots \in \mathcal{B}$ allora $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$).

Sono esempi di σ -algebra: a) l’insieme delle parti $\mathcal{P}(\Omega)$; b) l’insieme $\{\emptyset, \Omega\}$; c) l’insieme $\{\emptyset, A, \bar{A}, \Omega\}$.

Gli assiomi del calcolo possono dunque essere ri-formulati nel caso più generale (in cui Ω può anche avere la cardinalità del continuo) considerando “eventi” (ossia i sottoinsiemi di Ω sui quali è definibile la probabilità) solamente tutti e soli gli elementi di una “opportuna” σ -algebra \mathcal{B} .

2. Eventi stocasticamente indipendenti

Si introduce la nozione di indipendenza stocastica (detta anche “indipendenza” o “indipendenza statistica”): $A, B \in \mathcal{B}$ sono *stocasticamente indipendenti* se e solo se $P(A \cap B) = P(A)P(B)$.

Gli eventi A_1, A_2, \dots, A_n sono indipendenti se, per ogni k -pla, $k = 2, \dots, n$, di eventi distinti scelti fra A_1, A_2, \dots, A_n , vale la condizione di fattorizzazione della probabilità (la probabilità dell’intersezione è il prodotto delle probabilità).

3. Probabilità condizionata ad un evento

Siano $A, B \in \mathcal{B}$. Si definisce probabilità condizionata di A dato (o condizionato) B , e si scrive $P(A|B)$, il rapporto $P(A \cap B) / P(B)$. Nel caso in cui $P(B) = 0$ la probabilità condizionata $P(A|B)$ non è definita.

In particolare si trova $P(A) = P(A|\Omega)$. Quindi la consueta notazione $P(\bullet)$ può ritenersi la forma semplificata della più esplicita notazione $P(\bullet/\Omega)$.

Inoltre, visto il punto precedente, se $P(B) \neq 0$, allora $P(A|B) = P(A)$ se e solo se A e B sono indipendenti.

4. Teorema delle probabilità totali e formula di Bayes

Siano $B_1, B_2, \dots \in \mathcal{B}$, con $B_i \cap B_j = \emptyset$, per ogni $i \neq j$, e $P(B_i) \neq 0$. Allora

$$P(A) = P\left\{\bigcup_i (A \cap B_i)\right\} = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i).$$

Inoltre, poiché $P(A|B_i)P(B_i) = P(B_i|A)P(A)$, si ottiene

$$P(B_i|A) = P(A|B_i)P(B_i) / \left\{ \sum_i P(A|B_i)P(B_i) \right\} \quad (\text{formula di Bayes}).$$

Esempio. Si supponga che una analisi di laboratorio per verificare la presenza di encefalopatia spongiforme bovina (“morbo della mucca pazza”) in un capo di bestiame abbia le seguenti caratteristiche: a) il test risulta positivo con probabilità 0.99 se la mucca è effettivamente malata; b) il test risulta positivo con probabilità 0.003 se la mucca è sana (falso positivo). Si supponga inoltre che

la mucca sia presa a caso in un allevamento collocato in una regione dove l'incidenza della malattia è stata stimata in 1 caso su 10000.

Quale è la probabilità che la mucca sia ammalata se il test è risultato positivo?

Posto B_1 = mucca malata, B_2 = mucca sana, T = il test è positivo, si ha per ipotesi: $P(B_1) = 0.0001$; $P(B_2) = 0.9999$; $P(T/B_1) = 0.99$; $P(T/B_2) = 0.003$.

Allora si trova $P(T) = P(T/B_1)P(B_1) + P(T/B_2)P(B_2) = 0.0030987$, e quindi la probabilità richiesta $P(B_1/T) = P(T/B_1)P(B_1) / P(T) = 0.032$.

Questa probabilità può sembrare sorprendentemente bassa (si tratta comunque di una probabilità della quale è tipicamente difficile fornire una corretta valutazione intuitiva), visto che il test è sicuro "quasi" al 100%. Bisogna tuttavia considerare che, alla luce del risultato del test, la probabilità di malattia è cresciuta dal valore incondizionato di 0.0001, al valore 0.032, ben più alto del primo. Possiamo quindi concludere che il peso dell'evidenza empirica -pur non portando alle conseguenze certe che avremmo idealmente auspicato- ha giocato in maniera, tutto sommato, relativamente forte nell' "aggiornare" il valore della probabilità in studio, facendolo crescere di 320 volte.

SEZIONE I.5.

1. Formula della probabilità condizionata per successioni di eventi

Nello schema degli esperimenti con prove ripetute (si possono citare, come esempi, il "problema della divisione della posta" di Luca Pacioli (1445-1509):

A e B giocano un certo numero di partite, con ugual probabilità di vincere ciascuna partita; si aggiudica l'intera posta in palio chi raggiunge per primo N vittorie. Se allora quando A ha già vinto $a < N$ partite e B ha già vinto $b < N$ partite il gioco viene interrotto, quale è il modo equo di dividere la posta. [R., per $N=6$, $a=5$, $b=3$, si trova 7:1]

e un problema di Huyghens (1629-1695):

A e B lanciano a turno 2 dadi; A deve ottenere 6 e B deve ottenere 7. Vince chi per primo raggiunge il suo obiettivo. Se comincia a lanciare A, chi dei due ha maggior probabilità di vincere? [R. $P(A) = 30/61$],

sono utili le formule che generalizzano al caso di n ($n > 2$) eventi (successioni di eventi) la probabilità dell'intersezione

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

In particolare se gli eventi sono indipendenti allora:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i).$$

2. Nota su indipendenza e incompatibilità

Attenzione a non confondere indipendenza e incompatibilità. Se due eventi -con probabilità positiva- sono indipendenti, allora non sono incompatibili; se sono incompatibili, allora non sono indipendenti. (Naturalmente due eventi possono anche non esser né indipendenti né incompatibili.)

3. Variabili casuali

Un esperimento aleatorio che produce risultati *numerici* definisce una variabile casuale (v.c.; con un termine più tipico in lingua francese si potrebbe parlare di *numero aleatorio*). In altri termini lo spazio Ω deve essere un sottoinsieme di \mathfrak{R}^k (quando $k=1$ la v.c. si dice unidimensionale, altrimenti essa è multidimensionale).

Una v.c. è ben definita quando la probabilità è *assegnata* ad una opportuna classe di sottoinsiemi di $\Omega \subset \mathfrak{R}^k$ con la struttura di σ – algebra (cfr. I.4.1). Tuttavia, si può dimostrare che per definire univocamente una v.c. è sufficiente assegnare la probabilità ad una classe di sottoinsiemi “semplici” di \mathfrak{R}^k (cfr. I.6.1 e II.1.1): gli intervalli. Ciò giustifica la rilevanza di un particolare tipo di funzione che determina la probabilità degli intervalli, e pertanto rappresenta “tutta” la v.c.: la funzione di ripartizione.

4. Distribuzioni

Con riferimento al caso univariato, si dice *distribuzione*, o *funzione di ripartizione* (f.d.r.), una funzione F reale di variabile reale, $F : \mathfrak{R} \rightarrow [0,1]$, con le seguenti proprietà:

- 1) $F(x) \rightarrow 1$ per $x \rightarrow \infty$;
- 2) $F(x) \rightarrow 0$ per $x \rightarrow -\infty$;

- 3) F è monotona non decrescente;
 4) F è continua da destra in ogni punto, ossia:

$$\lim_{\varepsilon \rightarrow 0} F(x + \varepsilon) = F(x), \quad \varepsilon > 0, \quad \forall x \in \mathfrak{R}.$$

5. Punti di salto di una f.d.r.

L'insieme dei punti di salto di una f.d.r. F ha cardinalità al più numerabile. Infatti, poiché F è limitata, essa non può avere più di

- 1 salto di ampiezza compresa in $(\frac{1}{2}, 1]$;
- 2 salti di ampiezza compresa in $(\frac{1}{3}, \frac{1}{2}]$;
- ...
- n salti di ampiezza compresa in $(\frac{1}{n+1}, \frac{1}{n}]$;
- ...

Ora, l'insieme di tutti questi potenziali punti di salto è chiaramente numerabile. Quindi l'insieme dei punti di salto di F -che in esso è contenuto- lo è a fortiori.

SEZIONE I.6.

1. Corrispondenza fra v.c. e f.d.r.

E' ben definita la corrispondenza fra una v.c., diciamo X , e una f.d.r. F . Precisamente, essa è data dalla uguaglianza seguente:

$$P(X \leq x) = F(x). \tag{*}$$

E' importante notare che: **a)** data una v.c. X , la funzione $P(X \leq x)$, definita tramite l'uguaglianza (*), è una f.d.r.. Viceversa, **b)** data una f.d.r. F , l'uguaglianza (*) definisce una v.c. X .

a) Infatti, per le note proprietà della probabilità, si vede che:

1) $P(X \leq x) \rightarrow P(\mathfrak{R}) = 1$ per $x \rightarrow \infty$;

- 2) $P(X \leq x) \rightarrow P(\emptyset) = 0$ per $x \rightarrow \infty$;
- 3) $P(X \leq x) \leq P(X \leq x')$ per ogni $x < x'$ (poiché $(-\infty, x] \subset (-\infty, x']$ e P è monotona)
- 4) $\lim_{\varepsilon \rightarrow 0} P(X \leq x + \varepsilon) = \lim_{\varepsilon \rightarrow 0} \{1 - P(X > x + \varepsilon)\} = 1 - P(X > x) = P(X \leq x)$,
dove $\varepsilon > 0$, e dove per la prima e la terza uguaglianza si applica la proprietà della probabilità del complemento (vd. I.3.2.(1)), mentre per la seconda si applica la proprietà di continuità (vd. I.3.2.(5)).

b) Una giustificazione euristica di quest'implicazione inversa (una dimostrazione rigorosa va oltre gli obiettivi di questo corso) sta nel fatto che la (*) è sufficiente ad esempio per assegnare la probabilità ad *ogni* intervallo $(a, b]$, in quanto

$$P\{X \in (a, b]\} = F(b) - F(a),$$

ma anche ad *ogni* punto x , essendo

$$P(X = x) = P(X \leq x) - P(x < X) = F(x) - \lim_{\varepsilon \rightarrow 0} F(x - \varepsilon) \quad (\varepsilon > 0).$$

A questo punto le proprietà della probabilità consentono di “estendere” univocamente la assegnazione ad un grande numero di altri eventi via via più complessi, ottenuti iterando opportunamente operazioni insiemistiche sugli insiemi di partenza ... (vd. ad es. Zanella, 1980, per una trattazione sistematica).

2. V.c. discrete

Se F è “a gradini”, ossia ha un numero finito o una infinità numerabile di punti di salto, al di fuori dei quali F è *costante*, allora si dice che la v.c. X definita tramite l'uguaglianza (*) è *discreta*.

In questo caso se x è un punto di salto per F , allora

$$F(x) - \lim_{\varepsilon \rightarrow 0} F(x - \varepsilon) > 0 \quad (\varepsilon > 0).$$

Detto S l'insieme dei punti di salto di F la funzione dei salti p :

$$\begin{aligned} p(x) &= F(x) - \lim_{\varepsilon \rightarrow 0} F(x - \varepsilon), & x \in S \\ p(x) &= 0, & x \notin S \end{aligned}$$

è detta *funzione di probabilità* (f.p.). Pertanto:

$$\sum_{x \in \mathcal{S}} p(x) = 1 \quad \text{e} \quad F(x) = \sum_{y \leq x} p(y).$$

3. V.c. continue

Al contrario, se F è continua allora si ha

$$P(X = x) = F(x) - \lim_{\varepsilon \rightarrow 0} F(x - \varepsilon) = 0 \quad (\varepsilon > 0)$$

per ogni $x \in \mathfrak{R}$.

Se, in particolare, F è derivabile per ogni $x \in \mathfrak{R}$, allora è disponibile una utile

rappresentazione di F , $F(x) = \int_{-\infty}^x f(y) dy$, tramite una funzione, f , che diremo

funzione di densità di probabilità (f.d.p.). La f.d.p. f è una funzione non-

negativa e tale che $\int_{-\infty}^{\infty} f(y) dy = 1$.

In questo caso la v.c. X si dirà *assolutamente continua* o, con abuso di linguaggio, più brevemente, *continua*. In questo caso è possibile calcolare la probabilità degli intervalli tramite la formula:

$$P\{X \in (a, b]\} = \int_a^b f(y) dy.$$

Ad esempio, la v.c. con f.d.p. $f(x) = \pi^{-1}(1+x^2)^{-1}$ (cdd. v.c. “di Cauchy”) ha f.d.r. $F(x) = \frac{1}{2} + \pi^{-1} \arctg x$. Per essa si trova, ad es.,

$$P\{X \in (-1, +1)\} = \int_{-1}^{+1} f(y) dy = 0.5.$$

4. Altre vv.cc.

Oltre alle due citate, esistono altre tipologie di f.d.r. (e quindi di variabili casuali). In particolare F potrebbe avere dei punti di salto, non essendo però

“costante” al di fuori di essi. In questo caso F definirebbe una v.c. “di tipo misto” (per esempi vd. ad es. Bertoli-Barsotti, 1996, p.51 e p.62). Infine, si danno casi di f.d.r. continue che però non ammettono l’esistenza di una funzione di densità (con quella rappresentazione integrale). Tuttavia nel seguito il nostro interesse sarà circoscritto esclusivamente alle sole vv.cc. da annoverare unicamente nei casi descritti ai precedenti punti 2 e 3, con le denominazioni introdotte, rispettivamente, di vv.cc. “discrete” e vv.cc. “continue”.

SEZIONE I.7.

1. Valore atteso di una variabile casuale

Si definisce *supporto* di una v.c. il sottoinsieme S di \mathfrak{R} su cui la f.d.p., o la f.p. è diversa da 0. Nel caso continuo S è, in genere, un intervallo aperto (anche illimitato) di \mathfrak{R} . Nel caso discreto, S è un sottoinsieme finito, o al più infinito-numerabile, di \mathfrak{R} .

Si definisce *valore atteso (expectation)*, o *media*, della v.c. X il numero, se esiste,

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_S x \cdot f(x) dx$$

nel caso continuo e

$$E(X) = \sum_{x \in S} x \cdot p(x)$$

nel caso discreto. Non sempre la media $E(X)$ esiste (un controesempio è dato dalla v.c. di Cauchy).

In generale $E(X)$ è diversa dalla *moda* (i.e. il punto di massimo per la f.d.p., o f.p.) e dalla *mediana* (i.e. il punto x_0 tale che -nella notazione del continuo- $P(X \leq x_0) = 0.5$).

2. Giochi equi

Un gioco equo (*fair game*) è un gioco, il cui esito aleatorio è un dare/avere economico, in cui i guadagni attesi controbilanciano le perdite attese, ovvero in

cui la media della variabile casuale “bilancio economico totale” esiste ed è pari a 0.

3. Il paradigma della scommessa e la costruzione della Probabilità soggettiva

Se G e T sono, rispettivamente, il guadagno economico netto nel caso di “vincita”, e la perdita economica netta (in valore assoluto) nel caso di “non-vincita”, in un gioco equo, allora posta $P(\text{“vincita”})=p$, si trova che, per definizione, vale l’uguaglianza $Gp + (-T)(1-p) = 0$. Da essa si trae $G/T = (1-p)/p$, ossia $p = 1/(Q+1)$, dove $Q = G/T$ è detta “quota” (*odds*). Si può dunque introdurre la definizione *soggettivista* della Probabilità. Secondo l’impostazione soggettivista, un “evento aleatorio E ha una certa probabilità di accadere” solo nella misura in cui un “soggetto” osservatore è in una condizione di non-sufficiente conoscenza per esprimersi nei più precisi termini di certezza assoluta. La probabilità è dunque necessariamente una sorta di opinione soggettiva che può essere esplicitata utilizzando il paradigma della scommessa nel modo seguente. Ipotizzando che si possa contrattare sulla quota Q , concernente una scommessa sul verificarsi di E , fino a giungere alla quota minima Q_0 che il soggetto accetterebbe per entrare ancora nel gioco (è inteso che il soggetto non accetterebbe quote inferiori a Q_0 , mentre accetterebbe, a fortiori, quote superiori a Q_0), la probabilità (soggettiva) cercata è

$$p = \frac{1}{Q_0 + 1}.$$

Naturalmente tale valore dipende, per costruzione, dalle aspettative e dalle conoscenze che il soggetto considerato ripone nel verificarsi di E . Ogni altro soggetto produrrebbe, in generale, un valore di p diverso.

4. Il problema della coerenza nell’assegnazione della Probabilità soggettiva

Nell’assegnazione della probabilità a diversi eventi (assegnazioni multiple) di uno spazio fondamentale Ω con il paradigma soggettivista, sorge il problema del soddisfacimento delle “leggi” della probabilità. Anche se la condizione di non negatività è soddisfatta (essendo $Q \in (0, +\infty)$), è tuttavia possibile, nella pratica, che le assegnazioni multiple rivelino *incoerenza* (p.es. può verificarsi $P(A) > P(B)$ mentre $A \subset B$). In ultima analisi il problema della coerenza è risolto se: a) si considerano solo esperimenti con un numero *finito* di eventi

elementari E_1, E_2, \dots, E_k ; b) si attribuisce la probabilità a tutti e soli tali eventi elementari procedendo *per ciascuno* di essi alla determinazione secondo lo schema descritto al punto 3, sotto l'ulteriore condizione $\sum_{i=1}^k P(E_i) = 1$. A questo punto la probabilità di ogni evento potrà essere univocamente determinata in base al terzo assioma (additività finita) e il principio di coerenza sarà senz'altro soddisfatto.

SEZIONE I.8.

1. Assegnazione della Probabilità soggettiva e paradosso di S.Pietroburgo

Il seguente paradosso mostra come l'intuizione (soprattutto di fronte a esperimenti con infiniti risultati possibili) possa dar luogo a contrasti problematici con ciò che può suggerire il calcolo. Si consideri un gioco che, in corrispondenza all'esperimento che consiste nel lancio di una moneta equa, assegni un premio di 2^i euro se il segno "testa" compare per la prima volta all' i -esimo lancio, $i=1,2,\dots$. La domanda è: quale è la tassa che si ritiene equo pagare (ossia l'importo massimo che si è disposti a pagare, oltre il quale c'è la percezione di andare in perdita giocando) per entrare in questo gioco? Per rispondere alla domanda occorre calcolare il guadagno netto "atteso". L'ammontare cercato della tassa dovrebbe essere (in valore assoluto) esattamente pari a tale guadagno (cfr. la nozione di gioco equo). In questo esperimento gli eventi elementari sono le successioni {T}, {C,T}, {C,C,T}, ... , che hanno probabilità, rispettivamente, $1/2^1, 1/2^2, 1/2^3, \dots$. La v.c. X , guadagno netto, è la legge che "ricodifica" -con le stesse probabilità- i suddetti eventi elementari con i numeri, rispettivamente, $2^1, 2^2, 2^3, \dots$ (le vincite corrispondenti). X ha valore atteso

$$E(X) = 2^1(1/2^1) + 2^2(1/2^2) + 2^3(1/2^3) + \dots = 1 + 1 + 1 + \dots = \infty.$$

La media di X è illimitata, eppure è probabile che, per chiunque, la "percezione" di andare in perdita venga raggiunta per valori ben più bassi!

2. Linearità dell'operatore valore atteso

Se sottoponiamo la v.c. X ad un cambio di unità di misura e/o un cambio dell'origine (traslazione), in simboli $aX + b$, ci possiamo chiedere come cambia

il valore atteso. La linearità dell'integrale e della sommatoria implicano in modo diretto la linearità dell'operatore valore atteso. Pertanto possiamo scrivere una volta per tutte -sia per il caso continuo che per il caso discreto- che:
 $E(aX + b) = aE(X) + b$ ($a \neq 0$).

3. L'operatore varianza

A partire da una v.c. assegnata X , dotata di media finita $E(X)$, è possibile definire la seguente v.c. "scarto" (dalla media) $X - E(X)$. A parità di valore atteso, due vv.cc. possono differire nella forma della variabile scarto (si osservi che la variabile scarto ha media 0). Precisamente una può presentare un maggior grado di "dispersione", rispetto alla media, dell'altra. Questo fatto può essere quantificato calcolando il valore atteso del quadrato della v.c. scarto, ossia $E\{(X - E(X))^2\}$, posto che esista. Tale valore è detto *varianza* della v.c. X ed è indicato con $Var(X)$. La radice quadrata della varianza $Var(X)^{1/2}$ è detta *scarto quadratico medio* (o *deviazione standard*).

4. Alcune proprietà dell'operatore varianza

Con semplici passaggi algebrici si dimostrano alcune semplici proprietà dell'operatore Var , utilizzando la proprietà di linearità dell'operatore valore atteso di cui al precedente punto 2. Se la varianza esiste, allora:

- 1) $E\{(X - E(X))^2\} = E(X^2) - E(X)^2$ (formula assai utile per il calcolo operativo).
- 2) $Var(X - E(X)) = Var(X)$ (la varianza di X coincide con quella della v.c. scarto $X - E(X)$).
- 3) $Var(aX + b) = a^2 Var(X)$ ($a \neq 0$).

5. Momenti non-centrali di una v.c.

Le costanti $E(X)$ e $E(X^2)$ incontrate nei punti precedenti possono ritenersi casi particolari della formula più generale $\mu'_r = E(X^r)$, che fornisce, per $r = 1, 2, \dots$, i cosiddetti *momenti* (non-centrali) di ordine r della v.c. X . Per semplicità di notazione, nel caso dell'ordine 1 useremo, equivalentemente, il simbolo μ in luogo di μ'_1 .

SEZIONE I.9.

1. Momenti centrali di una v.c.

La formula $\bar{\mu}_r = E\{(X - \mu)^r\}$, fornisce, per $r = 2, 3, \dots$, i cosiddetti *momenti centrali* di ordine r della v.c. X . Si osservi che $\bar{\mu}_1 = 0$ (media della v.c. scarto).

I momenti centrali di ordine dispari sono nulli se la distribuzione è simmetrica.

2. Momenti centrali in funzione dei momenti non-centrali

Svolgendo la potenza r -esima del binomio $X - \mu$ e sfruttando la linearità dell'operatore valore atteso, è facile esprimere il generico momento centrale di ordine r in funzione dei momenti non centrali (di ordine minore o uguale a r).

In particolare si può scrivere: $Var(X) = \bar{\mu}_2 = \mu'_2 - (\mu'_1)^2$.

3. Indici di forma della distribuzione

I rapporti $\alpha_3 = \bar{\mu}_3 / \sigma^3$ (indice di asimmetria di Fisher) e $\alpha_4 = \bar{\mu}_4 / \sigma^4$ (indice di curtosi) sono utili per misurare, rispettivamente, il grado di simmetria -attorno ad un asse mediano- della distribuzione e la quantità di probabilità assegnata alle zone di coda della distribuzione. (Il riferimento ideale quando si calcolano questi indici è quello della distribuzione di tipo normale -vd.sezione 11-: per essa $\alpha_3 = 0$ e $\alpha_4 = 3$)

4. Momenti fattoriali di una v.c.

Spesso, nel caso di v.c. di tipo discreto è più agevole calcolare i momenti "fattoriali". La formula $E\{X(X-1)(X-2)\dots(X-r+1)\}$, fornisce, per $r = 1, 2, 3, \dots$, i cosiddetti *momenti fattoriali* di ordine r della v.c. X . Per $r = 1$ si ottiene la media μ .

Anche qui è facile esprimere il generico momento fattoriale di ordine r in funzione dei momenti non centrali (di ordine minore o uguale a r). Per es., per $r = 2$ si ha $E\{X(X-1)\} = \mu'_2 - \mu$; possiamo quindi esprimere anche la varianza in funzione dei momenti fattoriali nel modo seguente: $Var(X) = E\{X(X-1)\} + \mu - \mu^2$.

5. V.c. discrete di particolare interesse applicativo – “schema dell’urna”

Alcune vv.cc. discrete rivestono particolare interesse perché sono associate ad esperimenti aleatori frequentemente ricorrenti in ambito applicativo.

In particolare, il modello dell’urna (o *schema dell’urna*) idealizza un tipo di esperimento aleatorio dove si compiono estrazioni di palline da un’urna con diverse possibili composizioni e modalità di estrazione.

6. V.c. bernoulliana

Se si compie una estrazione da un’urna contenente due soli tipi, diciamo A e B, di palline, e p è la proporzione di palline di tipo A nell’urna (se $n(A)$ è il numero di palline di tipo A nell’urna e $n(B)$ è il numero di palline di tipo B nell’urna, $p = n(A) / [n(A) + n(B)]$), la v.c. X che esprime il *numero di palline di tipo A* osservate nell’ estrazione ha supporto $S = \{0,1\}$. X è detta v.c. bernoulliana. Essa associa la probabilità p al numero 1 e $(1-p)$ al numero 0. E’ facile vedere che X ha media p e varianza $p(1-p)$. La varianza è massima per $p=1/2$.

7. V.c. binomiale

Se si compiono n ($n \geq 1$) estrazioni dall’urna, in condizioni di indipendenza stocastica (i.e. con reimmissione), la v.c. X che esprime il *numero di palline di tipo A* osservate in tali n estrazioni ha supporto $S = \{0,1,2,\dots,n\}$. X è una v.c. binomiale. Si scrive $X \sim Bi(n, p)$; nel caso $n = 1$ si ha la v.c. bernoulliana. Precisamente, la probabilità di osservare x palline di tipo A nelle n estrazioni è

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

Si vede facilmente che $\sum_{x=0}^n p(x) = \sum_{x=0}^n [p + (1-p)]^n = 1^n = 1$. Inoltre:

$$\begin{aligned} E(X) &= \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \\ &= np \cdot \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-1-(x-1))!} p^{x-1} (1-p)^{n-1-(x-1)} = np. \end{aligned}$$

(l'ultimo passaggio sostituendo $y = x - 1$ e $m = n - 1$). Con passaggi simili, si trova $Var(X) = E\{X(X-1)\} + np - (np)^2 = np(1-p)$ (si noti l'impiego dei momenti fattoriali).

Si può trovare poi: $\bar{\mu}_3(X) = npq(q-p)$ e $\bar{\mu}_4(X) = 3(npq)^2 + npq(1-6pq)$. Quindi per n che tende all'infinito l'indice di curtosi $\alpha_4(X)$ tende al valore 3, mentre $\alpha_3(X)$ tende a 0.

SEZIONE I.10.

1. V.c. ipergeometrica

Se si compiono n estrazioni senza reimmissione (ovvero "in blocco") da un'urna contenente due soli tipi, diciamo A e B, di palline, e $n(A) = a$ ($1 \leq a \leq N - 1$) è il numero di palline di tipo A nell'urna e $n(B) = N - a$ è il numero di palline di tipo B nell'urna, la v.c. X che esprime il numero di palline di tipo A osservate nelle n ($n \geq 1$) estrazioni è detta ipergeometrica. E' facile ottenere la f.p. (cfr. I.1.3)

$$p(x) = \frac{\binom{a}{x} \binom{N-a}{n-x}}{\binom{N}{n}}.$$

Ora, i valori possibili di x sono necessariamente non-negativi e non superiori a n ; ma x non può superare neanche a ; infine, se n supera $N - a$ qualche pallina di tipo A deve essere estratta. In conclusione $p(x)$ è definita sul supporto:

$$\max\{0, n - (N - a)\} \leq x \leq \min\{n, a\}.$$

La v.c. ipergeometrica ha media $E(X) = n \frac{a}{N}$ (in piena analogia con la media delle binomiale) e varianza $Var(X) = n \frac{a}{N} \left(1 - \frac{a}{N}\right) \frac{N-n}{N-1}$.

2. V.c. di Poisson

Si considera un esperimento nel quale si osserva il ripetersi di un certo “accadimento” casuale, in un certo “intervallo” di tempo o di spazio, con una assegnata frequenza media λ ($\lambda > 0$).

L’ “accadimento” osservato può essere, ad esempio, l’emissione di una particella radioattiva da una sorgente di materiale radioattivo in un certo intervallo di tempo; oppure la presenza di impurità su un wafer di silicio (qui la collocazione dell’impurità è di tipo spaziale, essendo l’intervallo la superficie del wafer). Si vuole determinare la probabilità di osservare x accadimenti nel generico intervallo. Si suddivide l’intervallo in n sottointervallini uguali, sufficientemente piccoli perché ciascun sottointervallino contenga nessuno o al più 1 accadimento. Si suppone infine che gli eventi {sottointervallino-vuoto} e {sottointervallino-pieno} siano indipendenti per ogni n e che la probabilità dell’evento {sottointervallino-pieno} non dipenda dalla particolare collocazione del sottointervallino all’interno dell’intervallo (risultando quindi λ/n). Si trova pertanto, applicando la formula della f.p. binomiale,

$$p(x) = \binom{n}{x} \cdot \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}.$$

Allora, per n abbastanza grande, rimanendo x fissato, è valida l’approssimazione

seguinte: $p(x) = \frac{\lambda^x}{x!} \cdot \frac{n(n-1) \cdot \dots \cdot (n-x+1)}{n^x} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x} \cong \frac{\lambda^x}{x!} \cdot e^{-\lambda}$, dove

$x \in \{0, 1, 2, 3, \dots\}$, che definisce la f.p. della cosiddetta v.c. di Poisson.

(Quest’approssimazione è buona nella misura in cui n è grande rispetto ad x ; ciò implica, in particolare, che la f.p. della binomiale è ben approssimata, per ogni x , nella misura in cui il rapporto λ/n è piccolo). Da

$$\begin{aligned} E\{X(X-1) \cdot \dots \cdot (X-r+1)\} &= \sum_{x=0}^{\infty} x(x-1) \cdot \dots \cdot (x-r+1) \frac{\lambda^x}{x!} e^{-\lambda} = \\ &= \lambda^r \sum_{x=r}^{\infty} \frac{\lambda^{x-r}}{(x-r)!} e^{-\lambda} = \lambda^r \end{aligned}$$

si trae, fra l’altro: $E(X) = \lambda$, $Var(X) = E\{X(X-1)\} + \lambda - \lambda^2 = \lambda$.

Inoltre si può trovare: $\bar{\mu}_3(X) = \lambda$ e $\bar{\mu}_4(X) = 3\lambda^2 + \lambda$. Quindi al crescere di λ l’indice di curtosi $\alpha_4(X)$ tende al valore 3, mentre $\alpha_3(X)$ tende a 0.

3. V.c. geometrica

Se si compiono estrazioni da un'urna -con una proporzione p di palline di tipo A e $q = 1 - p$ di palline di tipo B- con reimmissione, la v.c. X che esprime il *numero di palline di tipo B* estratte prima di osservare la prima pallina di tipo A, è detta *geometrica*. La sua f.p. è, evidentemente, $p(x) = (1 - p)^x p$, dove $x \in \{0, 1, 2, \dots\}$.

Si trova facilmente $E(X) = \frac{q}{p}$, $E(X^2) = \frac{(2-p)q}{p^2}$ e $Var(X) = \frac{q}{p^2}$.

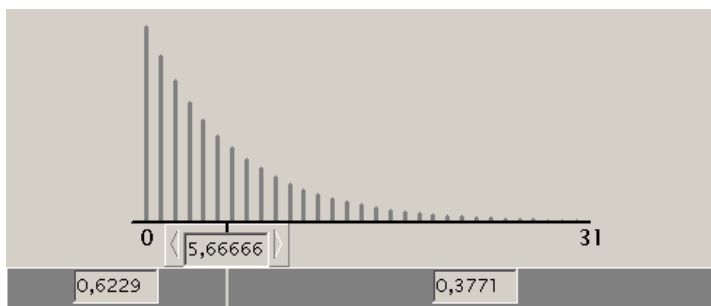


Figura. Funzione di probabilità della v.c. geometrica con $p = 0.15$. (In evidenza la media e le probabilità di stare al di sotto e al di sopra di essa)
[Grafica PQRS (Knypstra, 2002)]

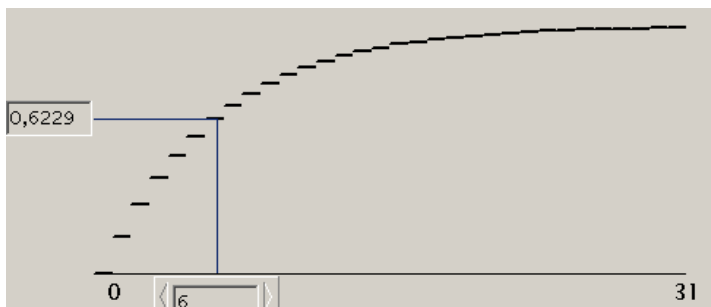


Figura. F.d.r. della v.c. geometrica con $p = 0.15$. (In evidenza la media e la probabilità cumulata corrispondente)
[Grafica PQRS (Knypstra, 2002)]

SEZIONE I.11.

1. Un integrale notevole dell'Analisi e alcune formule da esso conseguenti

Si introduce il seguente integrale notevole:

$$(II) \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

(sugg. per il calcolo: si calcola considerando l'equivalente integrale $\int_{-\infty}^{+\infty} e^{-x^2} dx \cdot \int_{-\infty}^{+\infty} e^{-y^2} dy = \pi$ e sostituendo le coordinate polari alla variabile (x, y)).

Da esso si ricava $\int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$ che, sostituendo $x^2 = t$, a sua volta equivale a:

$$(I2) \int_0^{+\infty} t^{-1/2} \cdot e^{-t} dt = \sqrt{\pi}.$$

Integrando quest'ultimo per parti, si ricava infine:

$$(I3) \int_0^{+\infty} t^{1/2} \cdot e^{-t} dt = \frac{\sqrt{\pi}}{2}.$$

2. V.c. normale standard

La f.d.p. della v.c. *normale* ha la sua piena giustificazione "costruttiva" in un teorema che verrà introdotto successivamente (Teorema del limite centrale). Tuttavia, tenendo conto di (I1), è provato che la funzione

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad -\infty < x < +\infty.$$

è positiva ed ha integrale 1 su \mathfrak{R} ; è questa la f.d.p. della v.c. *normale standard*. Si tratta di una funzione "pari" (simmetrica rispetto all'asse delle ordinate) con punto di massimo in 0. Pertanto moda, mediana e media coincidono con 0. Di più, tutti i momenti di ordine dispari (esistono e) sono nulli: $\mu'_{2k+1} = 0$. Per il calcolo della varianza si ottiene facilmente, visto (I3),

$$\text{Var}(X) = M(X^2) = 2 \int_0^{+\infty} x^2 \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 2 \int_0^{+\infty} t^{1/2} \cdot \frac{e^{-t}}{\sqrt{2\pi}} \sqrt{2} dt = 1.$$

I valori più interessanti della f.d.r. (tipicamente indicata con il simbolo Φ)

$\int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$ sono tabulati. Naturalmente dalle tavole si possono ricavare, in

particolare, i valori dei *percentili* della v.c. X (il percentile $p \cdot 100$ -esimo, $p \in (0,1)$ (fissato al centesimo), della v.c. X è il valore x_p tale che $P(X \leq x_p) = p$).

3. V.c. normale

Con la sostituzione $x = \frac{t-a}{b}$ sotto il segno di integrale, dove $a \in \Re$, $b > 0$,

l'uguaglianza $\int_{-\infty}^{+\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1$ diventa, equivalentemente, $\int_{-\infty}^{+\infty} \frac{e^{-\frac{1}{2}(\frac{t-a}{b})^2}}{\sqrt{2\pi}} \cdot \frac{1}{b} dt = 1$.

La funzione integranda è pertanto la più generale espressione della f.d.p. di una v.c. di tipo normale. La nuova f.d.p. è arricchita -rispetto alla versione standard- di due parametri, uno di locazione (a) e uno di scala (b). La v.c. normale, diciamo Y , con tale f.d.p. ha, evidentemente, media a e varianza b^2 (cfr. I.8.2 e I.8.4). Usando quindi, in luogo di a e b i più evocativi simboli μ e σ , si indicherà il fatto che Y è di tipo normale con media μ e varianza σ^2 con la scrittura: $Y \sim N(\mu, \sigma^2)$. La densità associata a Y ha quindi forma:

$$f_Y(y) = \frac{e^{-(y-\mu)^2/2}}{\sqrt{2\pi\sigma^2}}, \quad -\infty < x < +\infty.$$

SEZIONE I.12.

1. La funzione speciale Gamma di Eulero

Si introduce ora la funzione speciale Γ , detta *gamma di Eulero*,

$$\int_0^{+\infty} t^{\alpha-1} \cdot e^{-t} dt = \Gamma(\alpha), \quad \alpha > 0.$$

Si vede facilmente, per integrazione elementare, che $\Gamma(1) = \Gamma(2) = 1$. Inoltre,

per $\alpha \geq 1$ si trova, integrando per parti, $\int_0^{+\infty} t^{\alpha-1} \cdot e^{-t} dt = \int_0^{+\infty} \frac{t^{(\alpha+1)-1}}{\alpha} \cdot e^{-t} dt$.

Ma allora: $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$ ($\alpha \geq 1$). In particolare si ha: $\Gamma(3) = \Gamma(2 + 1) = 2 \cdot \Gamma(2) = 2!$, $\Gamma(4) = \Gamma(3 + 1) = 3 \cdot \Gamma(3) = 3!$, ecc.. In generale: $\Gamma(n + 1) = n!$, $n = 1, 2, \dots$ (la funzione gamma di Eulero generalizza il fattoriale a tutti i numeri reali positivi).

Inoltre si è già mostrato (v. I.11.1.(I2)) che: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$; pertanto

$$\Gamma(\frac{3}{2}) = \frac{1}{2} \cdot \Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}, \quad \Gamma(\frac{5}{2}) = \frac{3}{2} \cdot \Gamma(\frac{3}{2}) = \frac{3}{2} \cdot \frac{\sqrt{\pi}}{2}, \quad \text{ecc.. In generale:}$$

$$\Gamma(k + \frac{1}{2}) = 1 \cdot 3 \cdot \dots \cdot (2k - 1) \frac{\sqrt{\pi}}{2^k}, \quad k = 1, 2, \dots$$

2. Una applicazione di calcolo con la Gamma di Eulero: momenti di ordine pari della normale

Si consideri la v.c. normale standard $X \sim N(0, 1)$. Si trovano i momenti di ordine

pari $\mu'_{2k} = \bar{\mu}_{2k} = 2 \int_0^{+\infty} x^{2k} \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$; con la sostituzione $x^2/2 = t$ si può scrivere

$$\mu'_{2k} = \frac{2^k}{\sqrt{\pi}} \int_0^{+\infty} t^{(k+1/2)-1} \cdot e^{-t} dt = \frac{2^k}{\sqrt{\pi}} \Gamma(k + \frac{1}{2}) = 1 \cdot 3 \cdot \dots \cdot (2k-1).$$

In particolare se $Y \sim N(\mu, \sigma^2)$, allora $\alpha_4 = E\{(Y - \mu)^4\} / \sigma^4 = 3$, per ogni μ e σ^2 . Tale valore di curtosi, 3, è pertanto caratteristico del modello normale.

3. V.c. gamma

Per definizione della Gamma di Eulero, la funzione $f(x) = \frac{x^{\alpha-1} \cdot e^{-x}}{\Gamma(\alpha)}$, $x > 0$, è

positiva ed ha integrale 1 su \mathfrak{R}^+ per ogni $\alpha > 0$; è questa la f.d.p. della v.c. cosiddetta di tipo *gamma*. La costante α è interpretabile come *parametro di forma* di questa famiglia di vv.cc..

Con la solita sostituzione $x = \frac{t-a}{b}$, dove $a \in \mathfrak{R}$, $b > 0$, si ottiene la nuova

famiglia di f.d.p. $f(t) = \frac{(t-a)^{\alpha-1} \cdot e^{-(t-a)/\beta}}{\beta^\alpha \Gamma(\alpha)}$, $t > a$, con parametri di forma

(α), locazione (a) e scala (β). La f.d.p. è più spesso utilizzata nella forma

$$f(y) = \frac{y^{\alpha-1} \cdot e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad y > 0,$$

con parametro di scala, $\beta > 0$, e di forma, $\alpha > 0$.

Si indica il fatto che Y è di tipo gamma con tale f.d.p. con la scrittura: $Y \sim G(\alpha, \beta)$. Per tale v.c., con facile integrazione si trova:

$$E(Y^k) = \int_0^\infty y^k \frac{y^{\alpha-1} \cdot e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} dy = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \beta^k = \alpha(\alpha+1) \cdot \dots \cdot (\alpha+k-1) \beta^k.$$

In particolare $E(Y) = \alpha\beta$ e $Var(Y) = \alpha\beta^2$.

SEZIONE I.13.

1. vv.cc. collegate alla v.c. gamma: v.c. esponenziale e v.c. chi-quadrato

Se $Y \sim G(\alpha, \beta)$ con $\alpha = 1$, allora Y è una v.c. di tipo *esponenziale*. Indichiamo ciò con la scrittura $Y \sim E(\beta)$. La f.d.p. è pertanto

$$f_Y(y) = \frac{e^{-y/\beta}}{\beta}, \quad 0 < y < +\infty.$$

mentre la f.d.r. è $F(y) = 1 - e^{-y/\beta}$, per $y > 0$ e zero altrove.

Se $Y \sim G(\alpha, \beta)$ con $\alpha = 1/2$ e $\beta = 2$, allora Y è una v.c. di tipo *chi-quadrato* con un grado di libertà. Indichiamo ciò con la scrittura $Y \sim \chi_1^2$. Se $Y \sim G(\alpha, \beta)$ con $\alpha = k/2$ e $\beta = 2$, allora Y è una v.c. di tipo *chi-quadrato* con k gradi di libertà. Indichiamo ciò con la scrittura $Y \sim \chi_k^2$. La f.d.p. è pertanto

$$f_Y(y) = \frac{y^{\frac{k}{2}-1} \cdot e^{-y/2}}{\sqrt{2^k} \Gamma(\frac{k}{2})}, \quad 0 < y < +\infty.$$

Per la v.c. $Y \sim \chi_k^2$ si trova: $E(Y^r) = k(k+2) \cdot \dots \cdot (k+2(r-1))$. In particolare,
 $E(Y) = k$, $Var(Y) = 2k$, $\alpha_3(Y) = \sqrt{\frac{8}{k}}$, $\alpha_4(Y) = 3 + \frac{12}{k}$. Al crescere di k
 l'indice di curtosi $\alpha_4(Y)$ tende a 3, mentre $\alpha_3(Y)$ tende a 0.

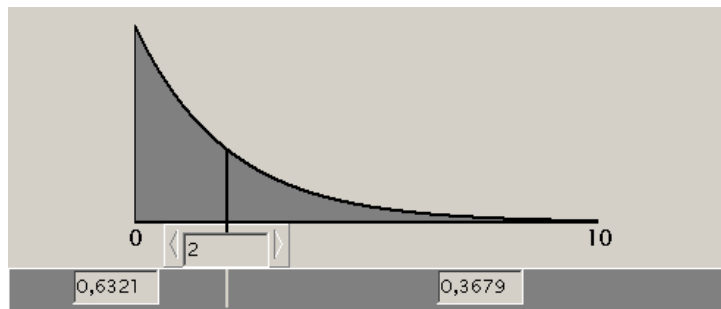
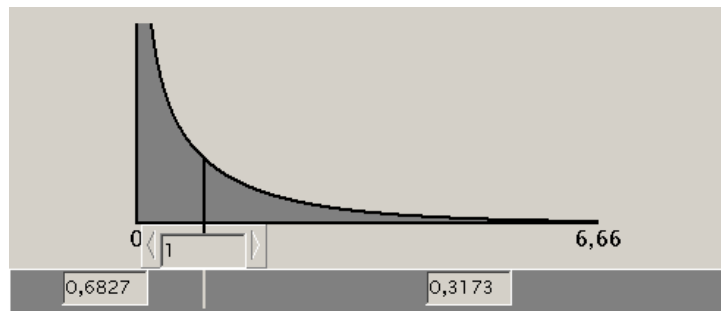


Figura. Funzioni di densità della v.c. chi-quadrato con 1, 2 e 8 gradi di libertà.
 (In evidenza la media e le probabilità di stare al di sotto e al di sopra di essa)
 [Grafica PQRS (2002)]

2. v.c. di Cauchy

La v.c. di Cauchy è definita dalla f.d.p. $f(x) = \pi^{-1}(1+x^2)^{-1}$, $-\infty < x < +\infty$, ovvero dalla f.d.r. $F(x) = 2^{-1} + \pi^{-1} \arctg x$. Con la sostituzione $x = \frac{y-a}{b}$, dove $a \in \mathfrak{R}$, $b > 0$, si ottiene la famiglia di f.d.p. $f(y) = (1 + (\frac{y-a}{b})^2)^{-1} \cdot \frac{1}{\pi b}$. La v.c. di Cauchy non ha momenti.

3. v.c. di Laplace (esponenziale doppia)

La v.c. dotata di f.d.p. $f(x) = \frac{1}{2} e^{-|x|}$, $-\infty < x < \infty$, è detta di Laplace (o esponenziale doppia). Con la sostituzione $x = \frac{y-a}{b}$, dove $a \in \mathfrak{R}$, $b > 0$, si ottiene la famiglia di f.d.p.

$$f(y) = \frac{1}{2b} e^{-\left|\frac{y-a}{b}\right|}, \quad -\infty < y < \infty;$$

media e varianza sono: $E(Y) = a$, $Var(Y) = 2b^2$. I momenti centrali di ordine dispari sono nulli; quelli di ordine pari danno $\bar{\mu}_{2k}(Y) = b^{2k} (2k)!$; quindi $\alpha_3(Y) = 0$, $\alpha_4(Y) = 4! b^4 / (2b^2)^2 = 6$ (ipernormalità).

4. v.c. logistica

Se X ha f.d.p. $f(x) = \frac{e^x}{(1+e^x)^2}$, $-\infty < x < \infty$, è detta logistica. Con la sostituzione $x = \frac{y-a}{b}$, dove $a \in \mathfrak{R}$, $b > 0$, si ottiene la famiglia di f.d.p.

$$f(y) = e^{-\frac{y-a}{b}} (1 + e^{-\frac{y-a}{b}})^{-2} \frac{1}{b}, \quad -\infty < y < +\infty$$

con f.d.r. $F(y) = (1 + e^{-\frac{y-a}{b}})^{-1}$; media e varianza sono: $E(Y) = a$, $Var(Y) = \frac{\pi^2}{3} b^2$. In particolare, per $b = \frac{\sqrt{3}}{\pi}$ e $a = 0$ si ottiene una v.c. molto prossima alla normale standard (il massimo scostamento fra le f.d.r. è: $\max_z |F(z) - \Phi(z)| = 0.0228$). La curtosi ($\bar{\mu}_4 / \bar{\mu}_2^2$) vale 4.2 (ipernormalità).

5. v.c. beta

La v.c. X che ha f.d.p.

$$f_x(x) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1,$$

$\beta > 0$, $\alpha > 0$, dove $B(\cdot, \cdot)$ indica la funzione speciale beta -definita in funzione della gamma da $B(\cdot, \cdot) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$ - è detta v.c. di tipo beta.

Si trovano i momenti: $\mu'_k(X) = \frac{\Gamma(\alpha + k)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\alpha + \beta + k)}$.

In particolare $E(X) = \frac{\alpha}{\alpha + \beta}$, $Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$.

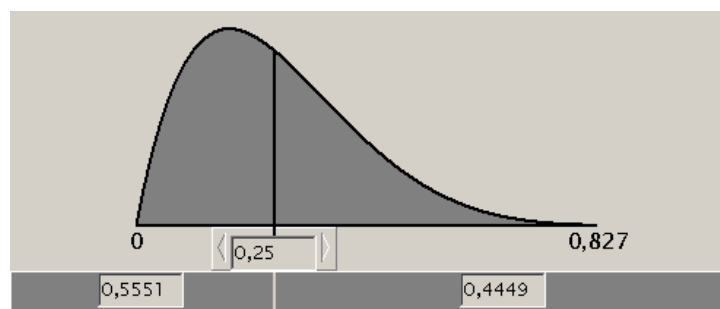


Figura. Funzione di densità della v.c. beta con parametri $\alpha = 2$ e $\beta = 6$. (In evidenza la media e le probabilità di stare al di sotto e al di sopra di essa) [Grafica PQRS (Knypstra, 2002)]

6. v.c. Weibull

Se Y è esponenziale, $Y \sim E(1)$, e $Y = \left(\frac{X-a}{b}\right)^c$, con $b, c > 0$, allora la v.c. X ha la distribuzione di Weibull. Si trova

$$f_X(x) = \exp\left\{-\left(\frac{x-a}{b}\right)^c\right\} \cdot \frac{c}{b^c} (x-a)^{c-1}, \quad x > a.$$

La Weibull è spesso adoperata nelle parametrizzazioni a due e ad un solo parametro. Nella parametrizzazione con due parametri ($a = 0$) e posto $\beta = 1/b^c$ si ha anche $f_X(x) = \beta c x^{c-1} e^{-\beta x^c}$, $x > 0$. Con questa parametrizzazione la f.d.r. è $F_X(x) = 1 - e^{-\beta x^c}$. Nella parametrizzazione (standard) con un solo parametro ($a = 0, b = 1$) si ha

$$f_X(x) = c x^{c-1} e^{-x^c}, \quad x > 0.$$

Per quest'ultima si trova $EX^k = \Gamma\left(\frac{k}{c} + 1\right)$.

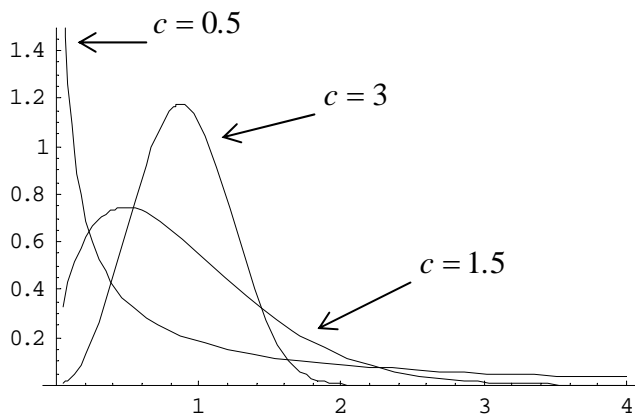


Figura. Funzioni di densità della v.c. Weibull standard con parametro c uguale a 0.5, 1.5 e 3.

7. v.c. *t* di Student

Se $X_1 \sim N(0,1)$ e $X_2 \sim \chi_k^2$ sono vv.cc. indipendenti, allora la v.c. $X = \frac{X_1}{\sqrt{X_2/k}}$ ha f.d.p.

$$f_X(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k} \Gamma\left(\frac{k}{2}\right) \left(\frac{x^2}{k} + 1\right)^{\frac{k+1}{2}}}, \quad -\infty < x < +\infty$$

ed è chiamata *t* di Student con k gradi di libertà. Per $k = 1$ si ottiene la v.c. di Cauchy.

Si trova $EX = 0$. Quindi $Var(x) = EX^2$; per l'indipendenza di X_1 e X_2 (cfr.

I.14.1) si trova $EX^2 = kEX_1^2EX_2^{-1} = kEX_2^{-1} = \frac{k}{k-2}$, per $k > 2$. In generale i

momenti di ordine dispari sono nulli; per quelli di ordine pari si trova invece $EX^r = k^{r/2} \frac{1 \cdot 3 \cdot \dots \cdot (r-1)}{(k-r)(k-r+2)\dots(k-2)}$, per $k > r$. L'indice di curtosi risulta

quindi $\alpha_4(X) = 3 \cdot \frac{k-2}{k-4}$ (per $k > 4$). Al crescere di k l'indice di curtosi tende a 3.

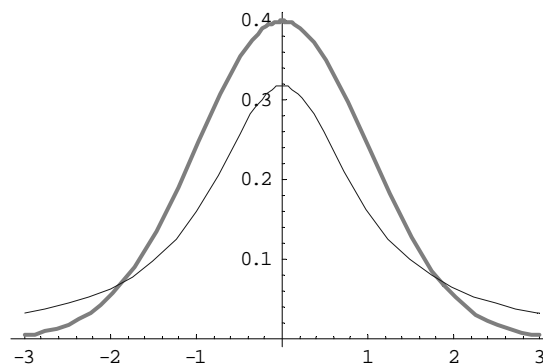


Figura. Confronto fra le funzioni di densità della v.c. normale standard (tratto più spesso) e della Cauchy (*t* di Student con 1 grado di libertà).

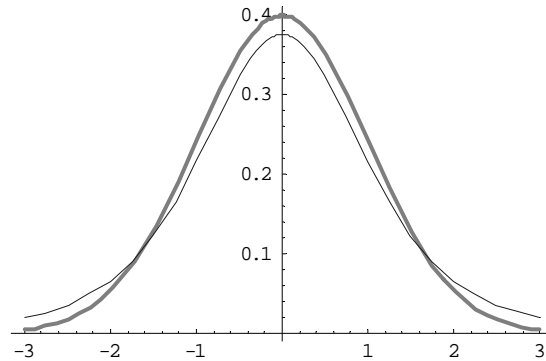


Figura . Confronto fra le funzioni di densità della v.c. normale standard (tratto più spesso) e della t di Student con 4 gradi di libertà.

8. v.c. beta-binomiale

La v.c. (di tipo discreto) X che ha f.d.p.

$$p_x(x) = \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)}, \quad x = 0, 1, 2, \dots, n$$

si dice beta-binomiale. Ricorrendo allo schema dell'urna, questa v.c. si può descrivere nel modo seguente: essa rappresenta il numero di successi nella estrazione bernoulliana di n palline da un'urna la cui composizione è caratterizzata da una proporzione p di palline vincenti che è a sua volta una v.c., e precisamente una v.c. di tipo beta con parametri α e β .

Si trova: $E(X) = n\pi$ e $Var(X) = n\pi(1-\pi)\frac{n\phi+1}{\phi+1}$, dove $\pi = \frac{\alpha}{\alpha+\beta}$ e

$\phi = \frac{1}{\alpha+\beta}$ (si ottiene il modello binomiale per $\alpha, \beta \rightarrow \infty$ tali che

$\frac{\alpha}{\alpha+\beta} = \text{costante}$).

9. Disuguaglianza di Chebyshev

A parità di media e varianza, due vv.cc. possono ancora esser differenti. Tuttavia, esse restano accomunate dal fatto che i valori di probabilità in regioni di coda sono, per entrambe, limitati secondo il vincolo stabilito dal seguente importante risultato (*disuguaglianza di Chebyshev*): se X è una qualsiasi v.c. dotata di momento secondo, con media μ e varianza σ^2 , e t è una costante positiva fissata, allora $P\{|X - \mu| \geq t\} \leq \frac{\sigma^2}{t^2}$. Infatti (con la notazione per una v.c. continua),

$$\sigma^2 = \int_S (x - \mu)^2 f(x) dx + \int_{S'} (x - \mu)^2 f(x) dx \leq \int_{S'} (x - \mu)^2 f(x) dx \leq t^2 \int_{S'} f(x) dx = t^2 P\{|X - \mu| \geq t\},$$

dove $S = \{x : |x - \mu| < t\}$ e $S' = \{x : |x - \mu| \geq t\}$ ($S \cup S' = \mathfrak{R}$ e $S \cap S' = \emptyset$).

SEZIONE I.14.

1. Vv.cc. indipendenti

X_1, \dots, X_n è una collezione di vv.cc. *indipendenti* se e solo se vale la seguente condizione di fattorizzazione della probabilità

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i) \text{ per ogni } (x_1, \dots, x_n) \in \mathfrak{R}^n,$$

ciò che può anche scriversi

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

per ogni $(x_1, \dots, x_n) \in \mathfrak{R}^n$, interpretando, per definizione,

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n),$$

in analogia al caso univariato.

(Di fatto la condizione di fattorizzazione della f.d.r. equivale alla condizione di fattorizzazione della f.d.p. (caso continuo) o della f.p. (caso discreto): rispettivamente

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) \text{ e } p_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i).$$

Esempio. Due vv.cc. X_1 e X_2 discrete uniformi su $\{0,1\}$ (cfr. l'esperimento del lancio di una moneta equa ripetuto 2 volte) hanno una f.d.r. che è il prodotto delle due f.d.r. $F_{X_i}(x_i)$, $i = 1,2$, così definite:

$$F_{X_i}(x_i) = 0 \text{ per } x_i < 0; F_{X_i}(x_i) = 0.5 \text{ per } 0 \leq x_i < 1; F_{X_i}(x_i) = 1 \text{ per } x_i \geq 1.$$

Pertanto si trova la f.d.r. $F_{X_1, X_2}(x_1, x_2)$ così definita: $F_{X_1, X_2}(x_1, x_2) = 0$ per $\min\{x_i\} < 0$; $F_{X_1, X_2}(x_1, x_2) = 0.25$ per $0 \leq x_i < 1$, $i = 1,2$; $F_{X_1, X_2}(x_1, x_2) = 0.5$ per $x_1 \geq 1$, $0 \leq x_2 < 1$ e per $x_2 \geq 1$, $0 \leq x_1 < 1$; $F_{X_1, X_2}(x_1, x_2) = 1$ per $x_i \geq 1$, $i = 1,2$.

2. Vv.cc. somiglianti

X_1, \dots, X_n è una collezione di vv.cc. *somiglianti* se e solo se hanno tutte la stessa f.d.r., ossia $F_{X_i}(z) = F_{X_j}(z)$ per ogni i e j , $i, j = 1, 2, \dots, n$, e per ogni $z \in \mathfrak{R}$.

3. Media e varianza di una somma e di una media di vv.cc. indipendenti e somiglianti

Siano X_1, \dots, X_n *indipendenti e somiglianti* (si dice anche che sono *i.i.d.*, ossia indipendenti e identicamente distribuite), con media μ e varianza σ^2 finite. E' facile dimostrare le seguenti proprietà:

a) Sia $S_n = \sum_{i=1}^n X_i$ la v.c. *somma (somma campionaria)*. Allora:

$$M(S_n) = n\mu \text{ e } \text{Var}(S_n) = n\sigma^2.$$

(Infatti -con notazione del continuo- si trova:

$$\begin{aligned} M(S_n) &= \int \dots \int (x_1 + \dots + x_n) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \int \dots \int x_1 f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n + \dots + \int \dots \int x_n f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \int x_1 f(x_1) dx_1 \dots \int f_{X_n}(x_n) dx_n + \dots + \\ &= \int f_{X_1}(x_1) dx_1 \dots \int f_{X_{n-1}}(x_{n-1}) dx_{n-1} \int x_n f_{X_n}(x_n) dx_n = n\mu \end{aligned}$$

b) Sia $\bar{X}_n = \sum_{i=1}^n X_i / n$ la v.c. *media (campionaria)*. Allora (cfr. I.8.2, I.8.4):

$$M(\bar{X}_n) = \mu \text{ e } \text{Var}(\bar{X}_n) = \sigma^2 / n.$$

SEZIONE I.15.

1. Convergenze stocastiche

Se Y_1, Y_2, \dots è una successione di vv.cc., si dice che:

A) Y_n converge *in probabilità* alla v.c. Y , e si scrive $Y_n \xrightarrow{p} Y$, se

$$\lim_{n \rightarrow \infty} P\{|Y_n - Y| < \varepsilon\} = 1, \forall \varepsilon > 0;$$

B) Y_n converge *in distribuzione* alla v.c. Y , e si scrive $Y_n \xrightarrow{d} Y$, se, dette

F_1, F_2, \dots ed F le corrispondenti funzioni di ripartizione, si verifica che $\lim_{n \rightarrow \infty} F_n(t) = F(t)$, per ogni punto t di continuità di F .

In particolare, le convergenze “in probabilità” e “in distribuzione” possono essere considerate nell’importante caso particolare in cui Y è “degenere” in un punto c (i.e. $P\{Y = c\} = 1$). In tal caso si equivalgono.

Si osservi che in generale vale l’implicazione $Y_n \xrightarrow{p} Y \Rightarrow Y_n \xrightarrow{d} Y$: la convergenza in distribuzione è più debole della convergenza in probabilità.

Si dà un controesempio a quest'implicazione. Sia $\Omega = \{\omega_1, \omega_2\}$, con $P(\omega_1) = P(\omega_2) = 0.5$. Siano Y_1, Y_2, \dots ed Y indipendenti e somiglianti, con $Y(\omega_1) = 0$ e $Y(\omega_2) = 1$. Essendo somiglianti, Y_n ed Y hanno la medesima funzione di ripartizione, quindi, banalmente, la convergenza " \xrightarrow{d} " è verificata.

Viceversa, come funzioni di $\omega \in \Omega$, esse corrispondono a valori numerici (0 e 1 appunto) che possono (con una probabilità che non diventa infinitesima, ma che resta maggiore di un $\delta > 0$) non risultare "vicini", né tanto meno coincidere, al divergere di n . E' facile vedere infatti che $\forall n$ la probabilità che Y_n ed Y corrispondano a valori di distanza 1 (i.e. $P\{|Y_n - Y| = 1\}$) rimane sempre uguale a 0.5.

2. Legge dei grandi numeri

Siano X_1, \dots, X_n vv.cc. indipendenti e somiglianti, con media e varianza finite, diciamo, rispettivamente $\mu = M(X_i)$ e $\sigma^2 = \text{Var}(X_i)$. Dalla disuguaglianza di Chebyshev, applicata alla v.c. media campionaria $Y_n = \bar{X}_n$, si trae

$$P\{|Y_n - M(Y_n)| \geq \varepsilon\} \leq \frac{\text{Var}Y_n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}, \text{ ovvero il limite:}$$

$$P\{|Y_n - \mu| \geq \varepsilon\} \rightarrow 0, \text{ per } n \rightarrow \infty \text{ (cosiddetta Legge dei grandi numeri).}$$

Questa convergenza può quindi anche essere scritta: $Y_n \xrightarrow{p} \mu$.

3. Teorema del Limite Centrale

Se X_1, X_2, \dots è una successione di vv.cc. i.i.d., con media μ e varianza σ^2 finite, allora:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1).$$

La tesi del teorema si può leggere dicendo che la v.c. S_n è "asintoticamente normale" con media $n\mu$ e varianza $n\sigma^2$. In simboli si scrive

$$S_n \text{ è } AN(n\mu, n\sigma^2).$$

In termini pratici, ciò significa che per n abbastanza grande la distribuzione della v.c. S_n può essere approssimata con quella di una v.c. di tipo normale, $N(n\mu, n\sigma^2)$.

Si osservi che la tesi può essere scritta equivalentemente:

$$\frac{\bar{X}_n - \mu}{(\sigma/\sqrt{n})} \xrightarrow{d} Z \sim N(0,1),$$

ovvero $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} Z \sim N(0, \sigma^2)$. In simboli si può scrivere che:

$$\bar{X}_n \text{ è } AN(\mu, \sigma^2/n).$$

II PARTE

SEZIONE II.1.

1. Funzioni di ripartizione multivariate

La classe delle funzioni $F : \mathfrak{R}^n \rightarrow \mathfrak{R}$, tali che F è

- 1) a incrementi positivi; i.e. $\sum_{k=0}^n (-1)^k \sum_{\delta \in \Delta(k,n)} F(\delta) \geq 0$, dove, nell'insieme delle 2^n n -ple (d_1, \dots, d_n) dove $d_i = a_i$ oppure $d_i = b_i$, per $i = 1, \dots, n$, $\Delta(k, n)$ rappresenta l'insieme delle $\binom{n}{k}$ n -ple che contengono esattamente k segni di tipo "a";
- 2) continua da destra in ciascuno dei suoi argomenti, ossia:

$$\lim_{\varepsilon \rightarrow 0} F(x_1, \dots, x_i + \varepsilon, \dots, x_n) = F(x_1, \dots, x_i, \dots, x_n) \quad (\varepsilon > 0);$$

- 3) $F(x_1, \dots, x_n) \rightarrow 1$, per $\min\{x_i\} \rightarrow \infty$ (ossia $F(\infty, \dots, \infty) = 1$),
 $F(x_1, \dots, x_n) \rightarrow 0$ per $x_i \rightarrow -\infty$, per ciascun $i \in \{1, \dots, n\}$ fissato;

è la classe delle f.d.r. (distribuzioni) multivariate. Con questo si estende la nozione di f.d.r. univariata.

La f.d.r. $F_X(x_1, \dots, x_n)$ individua univocamente la distribuzione della *variabile casuale multipla* (o *vettore casuale*) $X = (X_1, \dots, X_n)$, nel senso che

$$P(X_1 \leq x_1; X_2 \leq x_2; \dots; X_n \leq x_n) = F_X(x_1, \dots, x_n).$$

2. Un controesempio

La funzione $F(x_1, x_2) = 1$ per $x_2 \geq -x_1$ e 0 altrove rispetta le richieste 2) e 3), ma non è a incrementi positivi.

E' infatti possibile trovare (vd. Figura) 4 punti di \mathcal{R}^2 , (a_1, a_2) , (b_1, a_2) , (a_1, b_2) e (b_1, b_2) tali che:

$$\sum_{k=0}^2 (-1)^k \sum_{\delta \in \Delta(k,2)} F(\delta) = F(a_1, a_2) + F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) = -1 < 0.$$

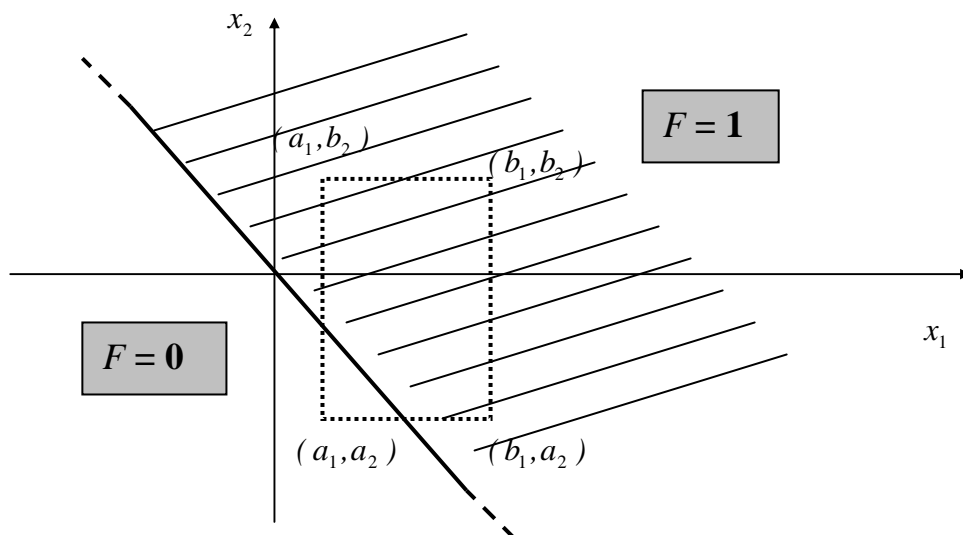


Figura. Un controesempio di funzione che non è f.d.r..

3. Componenti marginali di una v.c. multipla (a componenti indipendenti)

Nell'espressione della funzione di distribuzione multivariata $F_X(x_1, \dots, x_n)$, associata alla v.c. multipla $X = (X_1, \dots, X_n)$, le combinazioni alternative a quella

$(x_1, \dots, x_n) = (\infty, \dots, \infty)$ e a quella in cui $x_i < \infty$ per ogni i , $i = 1, \dots, n$, sono $2^n - 2$, e definiscono le distribuzioni delle componenti *marginali* di X .
 La v.c. $Y = (X_{\alpha(1)}, \dots, X_{\alpha(k)})$, dove $\alpha(1), \dots, \alpha(k)$ è una scelta di k numeri entro l'insieme $\{1, 2, \dots, n\}$ è una marginale di X se è definita dalla funzione di distribuzione

$$F_Y(y_1, \dots, y_k) := F_X(\infty, \infty, \dots, y_1, \infty, \dots, y_2, \infty, \dots, y_k, \infty, \dots)$$

dove y_i , $i = 1, \dots, k$, occupa la coordinata $\alpha(i)$.

La v.c. $X = (X_1, \dots, X_n)$ è a componenti *indipendenti* se e solo se

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i) \text{ per ogni } (x_1, \dots, x_n) \in \mathfrak{R}^n.$$

Come nel caso univariato esiste la possibilità di esprimere la probabilità in funzione di una f.d., $f_X = f_X(x_1, \dots, x_n)$. Nel caso continuo la probabilità di un evento B si ottiene per integrazione di f_X :

$$P(X \in B) = \int \dots \int_B f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Quindi, assegnata la f.d. $f_X(x_1, \dots, x_n)$, la f.d.r. corrispondente può essere ottenuta

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_X(y_1, \dots, y_n) dy_1 \dots dy_n.$$

In particolare, assegnata la f.d., la f.d.r. $F_Y(y_1, \dots, y_k)$ della marginale $Y = (X_{\alpha(1)}, \dots, X_{\alpha(k)})$ può essere ottenuta per integrazione

$$F_Y(y_1, \dots, y_k) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_X(x_1, \dots, x_n) dx_1 \dots dx_n,$$

i.e. integrazione nell'intervallo da $-\infty$ a y_i nella coordinata che occupa il posto $\alpha(i)$, $i = 1, \dots, k$, e integrazione nell'intervallo da $-\infty$ a $+\infty$ nelle altre coordinate.

4. Esempio: v.c. multinomiale

Generalizzando l'esperimento di estrazione dall'urna, se si compiono n ($n \geq 1$) estrazioni dall'urna contenente k tipi, A_1, A_2, \dots, A_k , di palline, e p_i è la proporzione di palline di tipo A_i nell'urna, $i = 1, \dots, k$, allora la probabilità di osservare x_1 palline di tipo A_1 e x_2 palline di tipo A_2 , ... , e x_k palline di tipo A_k , è

$$p_X(x_1 \dots x_{k-1}) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k},$$

dove $p_k = 1 - \sum_{i=1}^{k-1} p_i$ e $x_k = n - \sum_{i=1}^{k-1} x_i$.

La v.c. $X = (X_1, \dots, X_{k-1})$ ottenuta è detta multinomiale.

Si dimostra in particolare che le componenti marginali X_i di X sono binomiali $Bi(n, p_i)$.

SEZIONE II.2.

1. Valore atteso di una funzione di una v.c. multipla – Momenti di una v.c. multipla

Sia $X = (X_1, \dots, X_n)$ una v.c. multipla. Si può estendere la nozione di valore atteso (cfr. I.7.1) a quello di una funzione di X , $Y = g(X)$. In simboli si scrive $Eg(X)$, intendendo con ciò il risultato del calcolo seguente:

$$Eg(X) = \int \dots \int g(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n$$

con notazione di una v.c. X continua, ovvero

$$Eg(X) = \sum_{x_1} \dots \sum_{x_n} g(x_1, \dots, x_n) p_X(x_1, \dots, x_n)$$

con notazione di una v.c. X discreta.

Sono particolarmente interessanti alcuni casi particolari per g :

a) Nel caso $Y = g(X) = X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}$, $k = 1, 2, \dots$, il valore atteso di $g(X)$, se esiste, si chiama *momento misto* (o *momento congiunto*) di ordine (k_1, \dots, k_n) della v.c. X . In simboli

$$\mu'_{k_1, \dots, k_n}(X) = E(X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}).$$

b) Nel caso $Y = g(X) = (X_1 - EX_1)^{k_1} \dots (X_n - EX_n)^{k_n}$, $k = 1, 2, \dots$, il valore atteso di $g(X)$, se esiste, si chiama *momento misto* (o *momento congiunto*) *centrale* di ordine (k_1, \dots, k_n) della v.c. X . In simboli

$$\bar{\mu}_{k_1, \dots, k_n}(X) = E\{(X_1 - EX_1)^{k_1} \dots (X_n - EX_n)^{k_n}\}.$$

In particolare, per $k_i = k_j = 1$, $k_r = 0$, $i, j, r \in \{1, \dots, n\}$, $i \neq j \neq r$, si ha la *covarianza* fra X_i e X_j :

$$E\{(X_i - EX_i) \cdot (X_j - EX_j)\} = Cov(X_i, X_j),$$

che si scrive anche $\sigma(X_i, X_j)$.

Qualora vi sia indipendenza delle componenti di X allora vale la fattorizzazione

$$E\{(X_i - EX_i) \cdot (X_j - EX_j)\} = E(X_i - EX_i) \cdot E(X_j - EX_j).$$

Quindi la covarianza è nulla.

2. Matrice di covarianza e matrice di correlazione

Sia $X = (X_1, \dots, X_n)$ una v.c. multipla. La matrice (di dimensione $n \times n$) delle covarianze

$$\Sigma = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \dots & \text{Cov}(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

è detta *matrice di covarianza* (o *matrice di varianza-covarianza*). Si tratta di una matrice simmetrica definita non-negativa. Ovviamente $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$, $i = 1, \dots, n$.

La covarianza ha un limite massimo dato dalla seguente disuguaglianza (disuguaglianza di Cauchy-Schwarz):

$$|\text{Cov}(X_i, X_j)| \leq \sqrt{\text{Var}(X_i)} \sqrt{\text{Var}(X_j)},$$

con uguaglianza se e solo se $X_i = a + bX_j$ per qualche a e b .

Il rapporto $\rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)} \sqrt{\text{Var}(X_j)}}$ è detto *coefficiente di correlazione lineare* fra X_i e X_j . Esso vale 1, in valore assoluto, se e solo se sussiste la relazione lineare $X_i = a + bX_j$ per qualche a e b .

Alla matrice di covarianza corrisponde la *matrice di correlazione*, ottenuta con $\rho(X_i, X_j)$ in luogo di $\text{Cov}(X_i, X_j)$.

3. Esempio

Se $X = (X_1, \dots, X_{k-1})$ è una v.c. multinomiale, allora si trovano, in particolare, i seguenti momenti

$$E(X_i) = np_i,$$

$$\text{Var}(X_i) = np_i(1 - p_i),$$

$$E(X_i X_j) = n(n-1)p_i p_j$$

$$\text{Cov}(X_i, X_j) = -np_i p_j.$$

4. Distribuzioni condizionate: densità

Sia $F_X(x_1, \dots, x_n)$ la funzione di distribuzione associata alla v.c. multipla $X = (X_1, \dots, X_n)$. Si fa riferimento ai due casi euclidei “discreto” e “continuo”. In analogia alla definizione di probabilità condizionata fra *eventi*, si può introdurre la nozione di *variabile casuale* condizionata. Fissati i valori di k ($1 \leq k \leq n-1$) componenti di X

$$X_{\alpha(1)} = x_{\alpha(1)}, X_{\alpha(2)} = x_{\alpha(2)}, \dots, X_{\alpha(k)} = x_{\alpha(k)},$$

è definita la v.c. $(n-k)$ -dimensionale

$$(X_{\beta(1)}, \dots, X_{\beta(n-k)}) \text{ dato } Y = y$$

(dove $X_{\beta(1)}, \dots, X_{\beta(n-k)}$ sono le componenti complementari di $Y = (X_{\alpha(1)}, \dots, X_{\alpha(k)})$ in X e $y = (x_{\alpha(1)}, \dots, x_{\alpha(k)})$) e si scrive $(X_{\beta(1)}, \dots, X_{\beta(n-k)} / X_{\alpha(1)} = x_{\alpha(1)}, \dots, X_{\alpha(k)} = x_{\alpha(k)})$, intendendo che, per ogni evento A ,

$$P(A/y) = \int_{A/y} dF_{X/y},$$

per una opportuna distribuzione $F_{X/y}$.

Precisamente si dimostra che tale distribuzione è definita dalla densità ottenuta come rapporto fra la densità congiunta e quella marginale di Y , sia nel caso discreto che nel caso continuo:

$$f_{X_{\beta(1)}, \dots, X_{\beta(n-k)} / X_{\alpha(1)}, \dots, X_{\alpha(k)}}(x_{\beta(1)}, \dots, x_{\beta(n-k)} / y) = \frac{f_X(x_1, \dots, x_n)}{f_{X_{\alpha(1)}, \dots, X_{\alpha(k)}}(x_{\alpha(1)}, \dots, x_{\alpha(k)})}.$$

(Ciò è ovvio nel caso discreto in applicazione della definizione di probabilità condizionata. Nel caso continuo si può invece presentare la seguente argomentazione. Si consideri ad es. la v.c. doppia (X, Y) con densità f_{XY} , e sia

$$R = \{x / x \in (x_0, x_0 + \Delta x)\}, S = \{y / y \in (y_0, y_0 + \Delta y)\}.$$

Allora

$$P\{(X, Y) \in R \times S\} \approx f_{XY}(x_0, y_0) \Delta x \Delta y \text{ e } P\{X \in R\} \approx f_X(x_0) \Delta x$$

con approssimazione buona nella misura in cui Δx e Δy sono piccoli. Allora

$$P(R \times S / R) = \frac{P(R \times S)}{P(R)} \approx \frac{f_{XY}(x_0, y_0) \Delta x \Delta y}{f_X(x_0) \Delta x} = \frac{f_{XY}(x_0, y_0)}{f_X(x_0)} \Delta y$$

che giustifica l'interpretazione del rapporto $g(y) = \frac{f_{XY}(x_0, y)}{f_X(x_0)}$ in qualità di

funzione di densità della v.c. condizionata Y / X .)

Se X è a componenti indipendenti, per la fattorizzazione della densità si ottiene la marginale:

$$f_{X_{\beta(1)}, \dots, X_{\beta(n-k)} / X_{\alpha(1)}, \dots, X_{\alpha(k)}}(x_{\beta(1)}, \dots, x_{\beta(n-k)} / y) = f_{X_{\beta(1)}, \dots, X_{\beta(n-k)}}(x_{\beta(1)}, \dots, x_{\beta(n-k)}).$$

5. Distribuzioni condizionate: valori attesi

Assegnata la v.c. condizionata $(X_{\beta(1)}, \dots, X_{\beta(n-k)} / X_{\alpha(1)} = x_{\alpha(1)}, \dots, X_{\alpha(k)} = x_{\alpha(k)})$ è ben posta la definizione di valore atteso:

$$\begin{aligned} E\{g(X_{\beta(1)}, \dots, X_{\beta(n-k)}) / X_{\alpha(1)} = x_{\alpha(1)}, \dots, X_{\alpha(k)} = x_{\alpha(k)}\} = \\ = \int g(X_{\beta(1)}, \dots, X_{\beta(n-k)}) dF_{X_{\beta(1)}, \dots, X_{\beta(n-k)} | Y}. \end{aligned}$$

6. Esempio

La v.c. doppia (X, Y) ha densità $f_{XY} = 8xy$ per $0 \leq y \leq x \leq 1$ (zero altrove). Si trova $f_X(x) = 4x^3$, $0 \leq x \leq 1$ (zero altrove).

Quindi $f_{Y/X=x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{2y}{x^2}$, $0 \leq y \leq x$, con x fissato, $0 \leq x \leq 1$. Allora

il valore atteso di Y dato $X = x$ è

$$E(Y / X = x) = \int_0^x y \cdot \frac{2y}{x^2} dy = \frac{2}{3} x.$$

SEZIONE II.3.

Definizione. La v.c. $X = (X_1, \dots, X_n)$ è una normale multipla se e solo se ogni combinazione lineare $a'X$ delle sue componenti è una v.c. normale (univariata).

1. Distribuzione normale bivariata “standard”

La f.d. della v.c. $Z = (Z_1, Z_2)$ normale bivariata “standard” (cioè con componenti a media 0 e varianza 1) è

$$f_Z(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right\},$$

con supporto \mathfrak{R}^2 , $-1 < \rho < +1$.

Integrando in z_2 e z_1 si trovano, rispettivamente, le due densità marginali

$f_Z(z_1) = \frac{e^{-z_1^2/2}}{\sqrt{2\pi}}$ e $f_Z(z_2) = \frac{e^{-z_2^2/2}}{\sqrt{2\pi}}$. Pertanto: $Z_i \sim N(0,1)$, $i = 1,2$. In particolare

si ha: $EZ_1 = EZ_2 = 0$; $Var(Z_1) = Var(Z_2) = 1$. Con calcolo diretto si trova inoltre: $EZ_1 Z_2 = Cov(Z_1, Z_2) = \rho$, parametro che può essere quindi interpretato come il coefficiente di correlazione lineare fra Z_1 e Z_2 .

Si scrive $Z \sim N(0,0,1,1,\rho)$, o anche $Z \sim N_2(0, \Sigma)$, dove qui 0 rappresenta il vettore delle medie e

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

la matrice di varianza-covarianza.

Si trova $|\Sigma| = 1 - \rho^2$. Inoltre, essendo

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix},$$

si vede che vale l'uguaglianza $\frac{1}{(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2) = Z'\Sigma^{-1}Z$.

Quindi, in forma matriciale, la f.d. si può scrivere, più compattamente:

$$f_Z(z_1, z_2) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp\{-Z'\Sigma^{-1}Z/2\}.$$

Le linee di livello della f.d., ovvero i luoghi dei punti sul piano (z_1, z_2) di uguale densità di probabilità sono *ellissi* del tipo

$$(z_1^2 - 2\rho z_1 z_2 + z_2^2) = c$$

(c cost.), quindi aventi assi ruotati rispetto a quelli del piano cartesiano (z_1, z_2) . In forma canonica (cioè ruotando opportunamente gli assi cartesiani), l'ellisse risulta avere equazione, nel nuovo piano cartesiano (u_1, u_2) , del tipo: $(1 + \rho)u_1^2 + (1 - \rho)u_2^2 = d$ (d cost.).

2. Esempio

Le vv.cc. X, Y, Z , sono normali bivariate standard, rispettivamente con matrici di varianza covarianza $\begin{pmatrix} 1 & 1/3 \\ 1/3 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1/\sqrt{3} \\ 1/\sqrt{3} & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$. Le densità sono:

$$f_X(x_1, x_2) = \frac{3}{4\pi\sqrt{2}} \exp\left\{-\frac{9}{16}\left(x_1^2 - \frac{2}{3}x_1x_2 + x_2^2\right)\right\}$$

$$f_Y(y_1, y_2) = \frac{\sqrt{3}}{2\pi\sqrt{2}} \exp\left\{-\frac{3}{4}\left(y_1^2 - \frac{2}{\sqrt{3}}y_1y_2 + y_2^2\right)\right\}$$

$$f_Z(z_1, z_2) = \frac{1}{2\pi\sqrt{0.19}} \exp\left\{-\frac{1}{0.38}\left(z_1^2 - 1.8z_1z_2 + z_2^2\right)\right\}$$

Le ellissi contenenti regioni di piano di probabilità 0.2, 0.5, 0.8 sono rappresentate nella Figura 1.

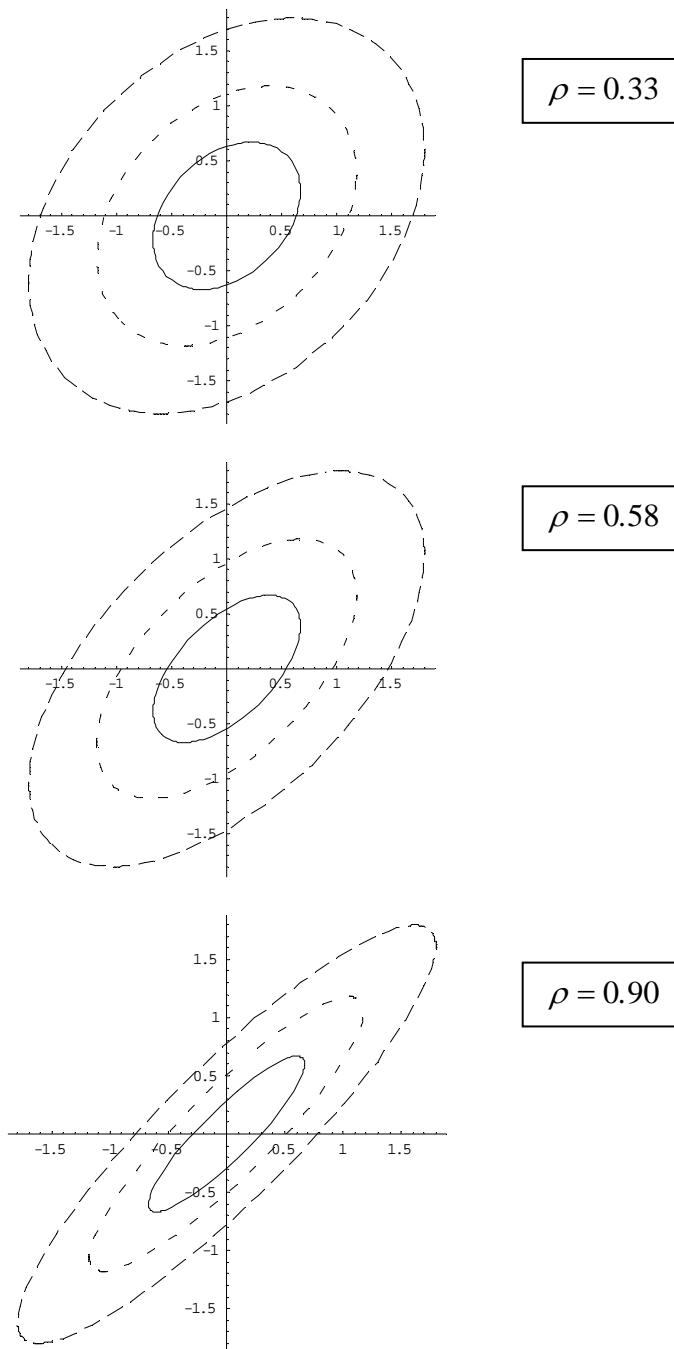


Figura 1. Regioni di piano di probabilità 0.2 (linea continua), 0.5 e 0.8, per X, Y, Z .

3. Distribuzione normale bivariata: densità

Con la trasformazione lineare $Z_i = \frac{X_i - \mu_i}{\sigma_i}$, dove $\mu_i \in \mathfrak{R}$, $\sigma_i > 0$, $i = 1, 2$, si generalizza la famiglia di funzioni di densità del §1, ottenendo

$$f_X(x_1, x_2) = f_Z\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}\right) \cdot \frac{1}{\sigma_1 \sigma_2} =$$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left[\frac{x_1 - \mu_1}{\sigma_1}\right]^2 - 2\rho\left[\frac{x_1 - \mu_1}{\sigma_1}\right]\left[\frac{x_2 - \mu_2}{\sigma_2}\right] + \left[\frac{x_2 - \mu_2}{\sigma_2}\right]^2\right)\right\}$$

Si scrive $X \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, oppure $X \sim N_2(\mu, \Sigma)$, dove qui μ rappresenta il vettore delle medie e Σ la matrice di varianza-covarianza.

Le marginali sono le vv.cc. normali: $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$.

Si trova la covarianza diciamo

$$\sigma_{12} = E(X_1 - \mu_1)(X_2 - \mu_2) = Cov(X_1, X_2) = \rho\sigma_1\sigma_2.$$

Pertanto, la matrice di varianza-covarianza è

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

dove σ_{11} sta per σ_1^2 e σ_{22} sta per σ_2^2 . Allora $\sigma_1\sigma_2\sqrt{1-\rho^2} = \sqrt{|\Sigma|}$.

Inoltre si trova $\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1/\sigma_{11} & -\rho/\sqrt{\sigma_{11}\sigma_{22}} \\ -\rho/\sqrt{\sigma_{11}\sigma_{22}} & 1/\sigma_{22} \end{pmatrix}$. Allora si vede che

vale l'uguaglianza:

$$\frac{1}{(1-\rho^2)}\left(\left[\frac{x_1 - \mu_1}{\sigma_1}\right]^2 - 2\rho\left[\frac{x_1 - \mu_1}{\sigma_1}\right]\left[\frac{x_2 - \mu_2}{\sigma_2}\right] + \left[\frac{x_2 - \mu_2}{\sigma_2}\right]^2\right) = (X - \mu)' \Sigma^{-1} (X - \mu)$$

Quindi in forma compatta la densità è

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(X - \mu)' \Sigma^{-1}(X - \mu)\right\}.$$

X è a componenti indipendenti *se e solo se* $\rho = 0$ (ovvero Σ è diagonale).

4. Distribuzione normale bivariata: distribuzioni condizionate

In corrispondenza di $X \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ si trovano le due vv.cc. condizionate $X_1/X_2 = x_2$ e $X_2/X_1 = x_1$. In particolare, il rapporto $f_X(x_1, x_2)/f_{X_1}(x_1)$ definisce la f.d. di una v.c. di tipo normale. Precisamente

$$X_2/X_1 = x_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1), \sigma_{22}(1 - \rho^2)\right),$$

dove $x_2 = \mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1)$ è l'equazione della retta (funzione) di regressione nel piano (x_1, x_2) . La non dipendenza della varianza condizionata $(\sigma_{22}(1 - \rho^2))$ dallo specifico valore (x_1) della variabile condizionante riflette la condizione di *omoschedasticità* del modello.

In particolare, per la normale standard Z , si ha $Z_2/Z_1 = z_1 \sim N(x_1\rho, 1 - \rho^2)$ (la retta di regressione ha equazione $x_2 = x_1\rho$).

Simmetricamente, si trova: $X_1/X_2 = x_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11}(1 - \rho^2)\right)$, e

per la normale standard Z , si ha $Z_1/Z_2 = z_2 \sim N(x_2\rho, 1 - \rho^2)$.

5. Distribuzione normale multivariata*

La v.c. $X = (X_1, \dots, X_n)$ è una v.c. normale multivariata n -dimensionale se ha f.d.

$$f_X(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(X - \mu)' \Sigma^{-1}(X - \mu)\right\}.$$

Si scrive $X \sim N_n(\mu, \Sigma)$, dove $\mu = (\mu_1, \dots, \mu_n)$ è il vettore delle medie e $\Sigma = (\sigma_{ij})$ una qualsiasi matrice simmetrica definita positiva ($\Sigma \neq 0$).

Poiché si trova $E(X_i - \mu_i)(X_j - \mu_j) = \text{Cov}(X_i, X_j) = \sigma_{ij}$, la matrice Σ è matrice di varianza-covarianza.

Ogni v.c. marginale di X , $Y = (X_{\alpha(1)}, \dots, X_{\alpha(k)})$ (dove $\alpha(1), \dots, \alpha(k)$ è una scelta di k numeri entro l'insieme $\{1, 2, \dots, n\}$) è una v.c. normale multivariata k -dimensionale.

La v.c. $X = (X_1, \dots, X_n)$ è a componenti indipendenti se e solo se Σ è diagonale.

6. Distribuzione normale multivariata: distribuzioni condizionate*

Sia $X \sim N_n(\mu, \Sigma)$, con partizione $X = (X_1, X_2)$, $X_1 \sim N_p(\mu_1, \Sigma_{11})$,

$$X_2 \sim N_q(\mu_2, \Sigma_{22}) \text{ (v. B §2), } p + q = n, \mu = (\mu_1, \mu_2), \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Allora, in analogia al caso univariato la v.c. condizionata $X_2 / X_1 = x_1$ è una v.c. normale q -dimensionale del tipo

$$N_q(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

($x_2 = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$ è l'iperpiano di regressione).

La non dipendenza della matrice di varianza-covarianza condizionata ($\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$) dallo specifico valore (x_1) della variabile condizionante riflette la condizione di *omoschedasticità* del modello.

Simmetricamente si ottiene la distribuzione p -dimensionale della v.c. $X_1 / X_2 = x_2$.

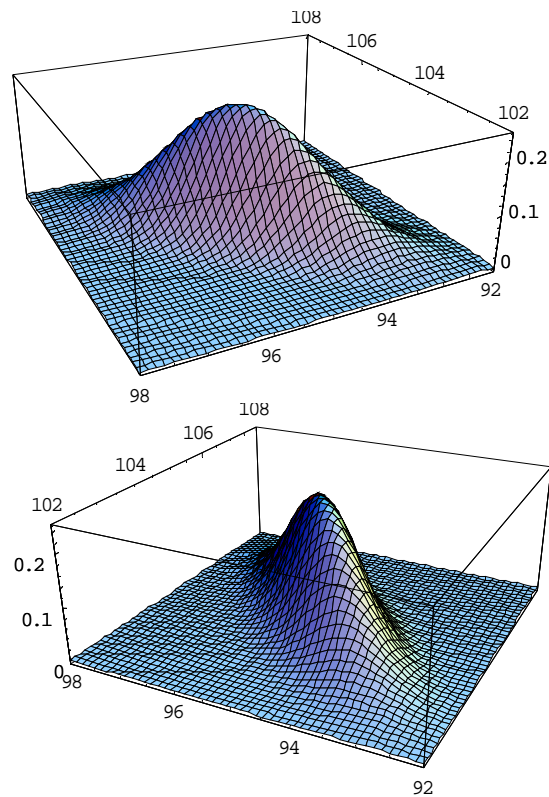


Figura 2. F.d. della v.c. $X \sim N(105, 95, 1450, 3875, 0.77)$

SEZIONE II.4.

1. Funzione caratteristica

E' particolarmente importante il valore atteso di una funzione di variabile complessa. Si tratta della *funzione caratteristica* (f.c.) ψ_X associata alla v.c. X :

$$\psi_X(t) = E(\exp\{itX\}) = E \cos(tX) + i \cdot E \sin(tX), \quad t \in (-\infty, \infty).$$

E' una funzione con dominio \mathfrak{R} e codominio il corpo complesso.

Nel caso di v.c. multipla $X = (X_1, \dots, X_n)$, la definizione di f.c. diventa $\psi_X(t) = E(\exp\{i\langle t, X \rangle\})$, essendo $t = (t_1, \dots, t_n)$ e dove $\langle t, x \rangle$ indica il prodotto scalare dei vettori t e x , cioè $\sum_{i=1}^n t_i \cdot x_i$.

Valgono le seguenti proprietà:

a) $E(\exp\{itX\})$ esiste sempre, infatti $|\psi_X(t)| \leq E[\cos tX + i \sin tX] = 1$. In particolare $\psi_X(0) = 1$.

b) Se $Y = a + bX$, si deduce $\psi_Y(t) = \exp\{iat\} \cdot \psi_X(bt)$.

c) Se $Z = X + Y$, con X e Y indipendenti (vd. §7):

$$\begin{aligned} \psi_Z(t) &= E \exp(itZ) = E \exp(it(X + Y)) = E \exp(itX) \cdot E \exp(itY) \\ &= \psi_X(t) \cdot \psi_Y(t); \end{aligned}$$

Quindi in generale, se X_1, X_2, \dots, X_n sono vv.cc. indipendenti e somiglianti, e $Z = \sum X_i$, $\psi_Z(t) = \psi_{X_1}(t)^n$.

d) (*Calcolo dei momenti di una v.c.*) Si dimostra l'implicazione seguente: se il momento EX^k esiste, allora esiste (continua) la derivata k -ma $\psi_X^{(k)}(t)$ della f.c., e vale in particolare l'uguaglianza

$$\psi_X^{(j)}(0) = i^j \cdot EX^j, \text{ per } j = 1, \dots, k.$$

e) Vale infine il seguente risultato: se F e $\{F_n\}$ sono funzioni di ripartizione e, rispettivamente, ψ e $\{\psi_n\}$ le corrispondenti f.c. $\forall n = 1, 2, \dots$, allora $F_n \rightarrow F$ se e solo se $\psi_n(t) \rightarrow \psi(t) \quad \forall t \in \mathfrak{R}$.

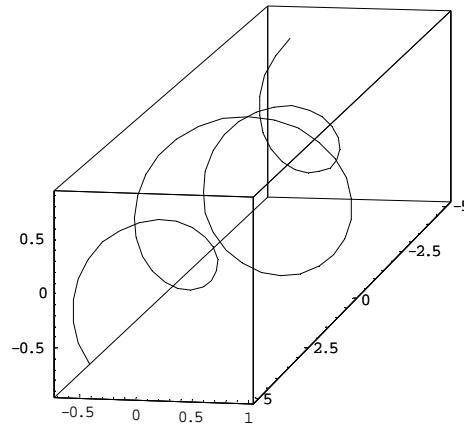


Figura. F.c. associata alla v.c. X che dà probabilità $1/3$ al punto $x = 1$ e $2/3$ al punto $x = 2$.

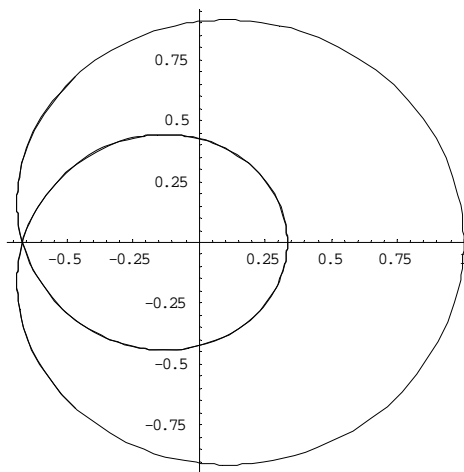


Figura. Proiezione sul piano complesso (codominio) della f.c. associata alla v.c. X che dà probabilità $1/3$ al punto $x = 1$ e $2/3$ al punto $x = 2$.

2. V.c. speciali – funzioni caratteristiche

a) BINOMIALE

DENSITÀ
$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, 2, \dots, n$

FUNZIONE CARATTERISTICA
$\psi_X(t) = (1-p + p \cdot e^{it})^n$

La v.c. binomiale è evidentemente riproduttiva: se $Z = X + Y$, con $X \sim Bi(n, p)$ e $Y \sim Bi(m, p)$ indipendenti, si trova:

$$\psi_Z(t) = (q + p \cdot e^{it})^n (q + p \cdot e^{it})^m = (q + p \cdot e^{it})^{n+m}.$$

Allora $Z \sim Bi(n + m, p)$.

b) MULTINOMIALE

DENSITÀ
$p_X(x_1, \dots, x_{k-1}) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \cdot \dots \cdot p_k^{x_k}$ $p_k = 1 - \sum_{i=1}^{k-1} p_i, \quad x_k = n - \sum_{i=1}^{k-1} x_i$

FUNZIONE CARATTERISTICA
$\psi_X(t_1, \dots, t_{k-1}) =$ $= \left(1 - \sum_{j=1}^{k-1} p_j + \sum_{j=1}^{k-1} p_j \cdot e^{it_j} \right)^n$

d) POISSON

DENSITÀ
$p_X(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$ $x = 0, 1, 2, \dots$

FUNZIONE CARATTERISTICA
$\psi_X(t) = \exp\{\lambda(e^{it} - 1)\}$

La v.c. di Poisson è riproduttiva: infatti, se $Z = X + Y$, con $X \sim \mathcal{P}(\lambda_1)$ e $Y \sim \mathcal{P}(\lambda_2)$ indipendenti, allora:

$$\psi_Z(t) = \exp\{\lambda_1(e^{it} - 1)\} \exp\{\lambda_2(e^{it} - 1)\} = \exp\{(\lambda_1 + \lambda_2)(e^{it} - 1)\}.$$

Quindi $Z \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

e) GEOMETRICA

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

FUNZIONE CARATTERISTICA
$\psi_X(t) = \frac{p}{q} \cdot \frac{1}{1 - qe^{it}}$ $(q = 1 - p)$

Per il calcolo si veda ad es. Bertoli-Barsotti (1996, p. 45).

f) NORMALE standard

DENSITÀ
$f_X(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ $-\infty < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = e^{-t^2/2}$

Per il calcolo si veda ad es. Bertoli-Barsotti (1996, pp. 86-89).

g) NORMALE

DENSITÀ
$f_X(x) = \frac{e^{-(x-\mu)^2/2}}{\sqrt{2\pi\sigma^2}}$ $-\infty < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = e^{-\sigma^2 t^2/2} \cdot e^{it\mu}$

I momenti centrali di ordine pari sono $\bar{\mu}_{2k} = 1 \cdot 3 \cdot \dots \cdot (k-3)(k-1)\sigma^{2k}$; in particolare, $Var(X) = \sigma^2$, $\alpha_3(X) = 0$, $\alpha_4(X) = 3$.

h) GAMMA

DENSITÀ
$f_X(x) = \frac{x^{\alpha-1} \cdot e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$
$0 < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = \frac{1}{(1 - i\beta t)^\alpha}$

Per il calcolo si veda ad es. Bertoli-Barsotti (1996, pp. 186-188).

i) ESPONENZIALE

DENSITÀ
$f_X(x) = \frac{e^{-x/\beta}}{\beta}$
$0 < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = \frac{1}{1 - i\beta t}$

l) CHI QUADRATO

DENSITÀ
$f_X(x) = \frac{x^{\frac{k}{2}-1} \cdot e^{-x/2}}{\sqrt{2^k} \Gamma(\frac{k}{2})}$
$0 < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = \frac{1}{(1 - i 2 t)^{k/2}}$

m) CAUCHY

DENSITÀ
$f_X(x) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a}{b}\right)^2}$
$-\infty < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = e^{-b t } \cdot e^{ita}$

Per il calcolo si veda ad es. Bertoli-Barsotti (1996, pp. 129-133).

o) LAPLACE (ESPONENZIALE DOPPIA)

DENSITÀ
$f_X(x) = \frac{1}{2b} e^{-\left \frac{x-a}{b}\right }$
$-\infty < x < +\infty$

FUNZIONE CARATTERISTICA
$\psi_X(t) = \frac{e^{ita}}{1 + t^2 b^2}$

I momenti centrali di ordine dispari sono nulli; quelli di ordine pari danno $\bar{\mu}_{2k}(X) = k! b^{2k}$; quindi $\alpha_3(X) = 0$, $\alpha_4(X) = 4! b^4 / (2b^2)^2 = 6$ (ipernormalità).

3. Teorema del limite centrale (TLC): dimostrazione

Enunciato. Se X_1, X_2, \dots è una successione di vv.cc. i.i.d., X_1 con media μ e varianza σ^2 finite, allora

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1).$$

Equivalentemente: $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$; ovvero $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Dimostrazione. Il ragionamento è il seguente:

i) poiché $i \cdot E(X_1 - \mu) = 0 = \psi_{X_1 - \mu}^{(1)}(0)$ e $i^2 \cdot E(X_1 - \mu)^2 = -\sigma^2 = \psi_{X_1 - \mu}^{(2)}(0)$, uno sviluppo di Taylor in un intorno dello 0 dà

$$\psi_{X_1 - \mu}(t) = 1 + (-\sigma^2 + \varepsilon(t)) \cdot \frac{t^2}{2},$$

dove $\varepsilon(t) \rightarrow 0$, per $t \rightarrow 0$ (essendo $\psi^{(2)}$ continua);

ii) posto $U_n = \sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$, si ha $\psi_{U_n}(t) = [\psi_{X_1 - \mu}(\frac{t}{\sigma\sqrt{n}})]^n$;

iii) sostituendo lo sviluppo trovato in (i), si ottiene

$$\psi_{U_n}(t) = \left(1 + \frac{-\frac{t^2}{2} + \frac{t^2}{2\sigma^2} \cdot \varepsilon\left(\frac{t}{\sigma\sqrt{n}}\right)}{n} \right)^n,$$

che converge a $\exp\{-\frac{t^2}{2}\}$, che è la funzione caratteristica della v.c. normale standard.

Si conclude che $\psi_{U_n}(t) \rightarrow \psi(t) = \exp\{-\frac{t^2}{2}\}$ implica (vd. §1(e)) la convergenza in distribuzione $U_n \xrightarrow{d} Z \sim N(0,1)$, come volevasi.

SEZIONE II.5.

1. Trasformazioni di vv.cc.

a) Sia X una v.c. con f.d.r. F_X e f.d. f_X sul supporto A . Sia $g: \mathfrak{R} \rightarrow \mathfrak{R}$, $Y = g(X)$ e $B = g(A)$.

a.1) Sia g monotona (i.e. invertibile) con inversa h . Allora

$$F_Y(y) = F_X(h(y)) \text{ se } g \text{ è crescente;}$$

$$F_Y(y) = 1 - F_X(h(y)) \text{ se } g \text{ è decrescente.}$$

Se F_X è continua, allora $f_Y(y) = f_X(h(y)) \cdot |h'(y)|$, $y \in B$ (nel caso discreto è semplicemente $f_Y(y) = f_X(h(y))$, $y \in B$).

a.2) Sia g , monotona a tratti; i.e. $A = \cup A_i$, con $A_i \cap A_j = \emptyset$ e con $g^{(i)} = g|_{A_i}$, restrizioni di g nei diversi tratti A_i (a valori zero altrove), monotone; siano $h^{(i)}$ le corrispondenti inverse, definite su $B_i = g^{(i)}(A_i)$. Allora, se F_X è continua:

$$f_Y(y) = \sum_i f_X(h^{(i)}(y)) \cdot |h^{(i)'}(y)| \cdot I_{B_i}, \quad y \in \cup B_i.$$

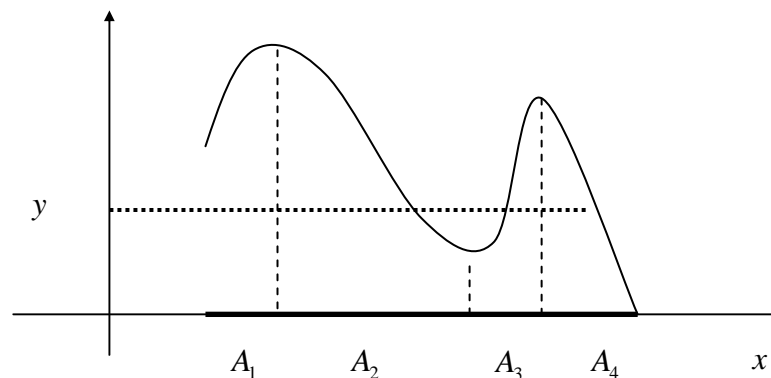


Figura. Trasformazione monotona a tratti.

b) Sia X una v.c. multipla con f.d.r. F_X e f.d. f_X sul supporto A . Sia $Y = g(X)$ e $B = g(A)$, $g : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$. Allora g è espressa tramite il vettore di funzioni

$$y_1 = g_1(x_1, \dots, x_n), \quad y_2 = g_2(x_1, \dots, x_n) \quad \dots \quad y_n = g_n(x_1, \dots, x_n)$$

ed ha inversa h , espressa tramite il vettore di funzioni

$$x_1 = h_1(y_1, \dots, y_n), \quad x_2 = h_2(y_1, \dots, y_n) \quad \dots \quad x_n = h_n(y_1, \dots, y_n)$$

e jacobiano (determinante della matrice delle derivate parziali di h rispetto a y_1, \dots, y_n)

$$J = J(y_1, \dots, y_n) = \det \begin{pmatrix} \partial h_1 / \partial y_1 & \dots & \partial h_1 / \partial y_n \\ \dots & \dots & \dots \\ \partial h_n / \partial y_1 & \dots & \partial h_n / \partial y_n \end{pmatrix}.$$

Sia g monotona ($|J| \neq 0$). Se F_X è continua allora:

$$f_Y(y_1, \dots, y_n) = f_X(h_1(y_1, \dots, y_n), h_2(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n)) \cdot |J|, \quad y \in B.$$

c) Sia X una v.c. multipla con f.d.r. F_X e f.d. f_X sul supporto A . Sia $Y = g(X)$ e $B = g(A)$, $g: \mathfrak{R}^n \rightarrow \mathfrak{R}^k$, $1 \leq k < n$. Si dice che g è *regolare* se esiste un completamento φ di g (i.e. una nuova funzione di X) tale che la nuova trasformazione g^* definita da:

$$y_1 = g_1(x_1, \dots, x_n), \dots, y_k = g_k(x_1, \dots, x_n), y_{k+1} = \varphi_1(x_1, \dots, x_n), \dots, \\ y_n = \varphi_{n-k}(x_1, \dots, x_n)$$

risulti invertibile (eventualmente a tratti, vd. punto b) precedente). Sia $B^* = g^*(A)$.

Considerando per semplicità di notazione il caso in cui g^* risulta monotona, con inversa h^* , e F_X continua, si ha:

$$f_{Y_1 \dots Y_k}(y_1, \dots, y_k) = \int \dots \int f_X(h_1^*(y_1, \dots, y_n), \dots, h_n^*(y_1, \dots, y_n)) \cdot |J| dy_{k+1} \dots dy_n,$$

$$y \in B^*.$$

2. Esempi di trasformazioni: la trasformazione lineare – famiglia locazione e scala

Sia X una v.c. con f.d.r. F_X e f.d. f_X (rispetto a un'assegnata misura dominante σ -finita) sul supporto A . Sia $g: \mathfrak{R} \rightarrow \mathfrak{R}$ e $Y = g(X) = a + bX$ (trasformazione

lineare), dove $b \neq 0$, e $B = g(A)$. Se F_X è continua, allora

$$f_Y(y) = f_X\left(\frac{y-a}{b}\right) \cdot \left|\frac{1}{b}\right|, \quad y \in B.$$

Spesso nelle applicazioni la trasformazione lineare è introdotta per significare un cambio di “locazione” (introduzione della costante a) e/o di “scala” (introduzione della costante b , qui necessariamente non negativa) per la v.c. X . In tal caso, se F_X è continua, allora $f_Y(y) = \frac{1}{b} \cdot f_X\left(\frac{y-a}{b}\right)$, $y \in B$ (mentre $F_Y(y) = F_X\left(\frac{y-a}{b}\right)$). Al variare di a e b , Y descrive una famiglia di vv.cc., detta *famiglia locazione e scala* (*).

Ad esempio, sia X la v.c. con f.d. $f_X(x) = \pi^{-1}(1+x^2)^{-1}$ (avente, pertanto, f.d.r. $F(x) = 2^{-1} + \pi^{-1} \arctg x$). Allora si ottiene la famiglia locazione e scala definita dalla seguente densità: $f(y) = \left(1 + \left(\frac{y-a}{b}\right)^2\right)^{-1} \cdot \frac{1}{\pi b}$, $a \in \mathfrak{R}$, $b > 0$.

[(*) Si tratta di una importante tipologia di modelli distribuzionali. Una famiglia locazione e scala è ovviamente *chiusa* rispetto alla iterazione della operazione di trasformazione lineare. E' altresì chiusa rispetto all'inversione di una trasformazione lineare (l'inversa di $a + bx$ è $\frac{y-a}{b}$). Ha quindi la struttura algebrica di *gruppo*. Una famiglia locazione e scala è indipendente dalla distribuzione di partenza, scelta per generarla]

3. Esempi di trasformazioni: una trasformazione monotona a tratti

Sia X la v.c. continua descritta dalla f.d. $f_X = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$, sul supporto $A = \mathfrak{R}$. Sia

$g : \mathfrak{R} \rightarrow \mathfrak{R}$ e $Y = g(X) = X^2$, $B = g(A) = \mathfrak{R}^+$. La funzione g è monotona a tratti, con $A = A_1 \cup A_2$, dove $A_1 = (-\infty, 0]$, $A_2 = (0, \infty)$ e con inverse $h^{(1)}(y) = -\sqrt{y}$ e $h^{(2)}(y) = \sqrt{y}$, entrambe definite per $y \in \mathfrak{R}^+$. Allora:

$$f_Y(y) = f_X(-\sqrt{y}) \cdot \left| \frac{1}{-2\sqrt{y}} \right| + f_X(\sqrt{y}) \cdot \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{\sqrt{y}} \frac{e^{-y/2}}{\sqrt{2\pi}}, \quad y \in \mathfrak{R}^+.$$

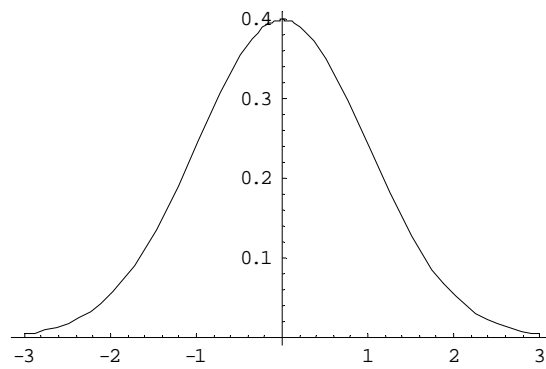


Figura. F.d. della v.c. X .

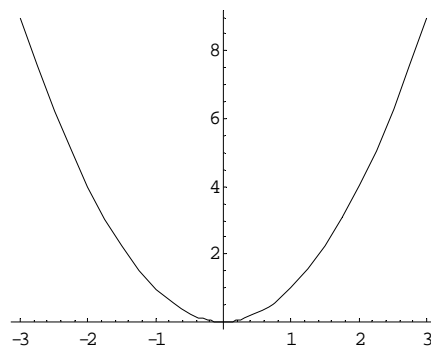


Figura. Trasformazione $Y = X^2$

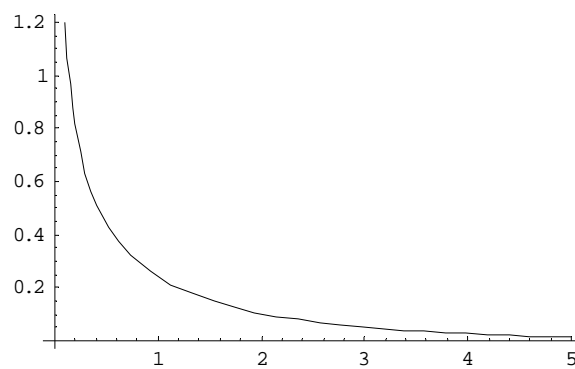


Figura. F.d. della v.c. Y

4. Esempi di trasformazioni: trasformazioni regolari – somma di vv.cc.

Sia $X = (X_1, X_2)$ una v.c. doppia con f.d.r. F_X e f.d. f_X sul supporto A . Sia $g : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ la trasformazione somma delle componenti di X : $Y_1 = g(X) = X_1 + X_2$, $B = g(A)$. La trasformazione si può completare, ad esempio, con la funzione $\varphi(X) = X_2$. Allora g^* definita da

$$y_1 = x_1 + x_2; \quad y_2 = x_2$$

è biunivoca, con inversa h^* definita da: $x_1 = y_1 - y_2$; $x_2 = y_2$.

Se, in particolare, F_X è continua, lo jacobiano è: $J = \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1$.

Quindi la medesima formula $f_Y(y_1, y_2) = f_X(y_1 - y_2, y_2)$, $y \in B$, vale sia nel caso discreto che nel caso continuo.

Si può dunque ottenere la f.d. della marginale Y_1 (v.c. somma) con le formule:

$$f_{Y_1}(y_1) = \int f_X(y_1 - y_2, y_2) dy_2, \quad \text{nel caso continuo;}$$

$$f_{Y_1}(y_1) = \sum_{y_2} f_X(y_1 - y_2, y_2), \quad \text{nel caso discreto.}$$

Se X è a componenti indipendenti, vale la particolarizzazione seguente (cdd. prodotto di convoluzione di f_{X_1} e f_{X_2}):

$$f_{Y_1}(y_1) = \int f_{X_1}(y_1 - y_2) f_{X_2}(y_2) dy_2, \quad \text{nel caso continuo;}$$

$$f_{Y_1}(y_1) = \sum_{y_2} f_{X_1}(y_1 - y_2) f_{X_2}(y_2), \quad \text{nel caso discreto.}$$

5. Esempi di trasformazioni: una particolare trasformazione regolare

Sia $X = (X_1, X_2)$ una v.c. doppia a componenti *indipendenti* dotate,

rispettivamente, delle densità $f_{X_1} = \frac{e^{-x_1^2/2}}{\sqrt{2\pi}}$ (normale standard) e

$f_{X_2} = \frac{1}{\sqrt{x_2}} \frac{e^{-x_2/2}}{\sqrt{2\pi}}$ (chi-quadrato con un grado di libertà) sul supporto

$A = (-\infty, \infty) \times (0, \infty)$. Sia $g : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ la trasformazione $Y_1 = g(X) = \frac{X_1}{\sqrt{X_2}}$,
dove $B = g(A) = (-\infty, \infty)$. Questa si può completare, ad esempio, con la
funzione $\varphi(X) = X_2$. Allora g^* definita da

$$y_1 = x_1 / \sqrt{x_2}; \quad y_2 = x_2,$$

è biunivoca, con inversa h^* definita da: $x_1 = y_1 \sqrt{y_2}$, $x_2 = y_2$,
 $y \in B^* = (-\infty, \infty) \times (0, \infty)$.

Lo jacobiano è $J = \det \begin{pmatrix} \sqrt{y_2} & \frac{y_1}{2\sqrt{y_2}} \\ 0 & 1 \end{pmatrix} = \sqrt{y_2}$. Quindi:

$$\begin{aligned} f_Y(y_1, y_2) &= f_X(y_1 \sqrt{y_2}, y_2) \cdot \sqrt{y_2} = \frac{e^{-y_1^2 y_2 / 2}}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{y_2}} \frac{e^{-y_2 / 2}}{\sqrt{2\pi}} \cdot \sqrt{y_2} \\ &= \frac{1}{2\pi} \exp\left\{-\frac{y_2}{2}(1 + y_1^2)\right\}, \end{aligned}$$

$y \in B^* = (-\infty, \infty) \times (0, \infty)$. Si può dunque ottenere la f.d. della marginale Y_1
integrando in y_2 :

$$f_{Y_1}(y_1) = \int f_Y(y_1, y_2) dy_2 = \int_0^{\infty} \frac{1}{2\pi} \exp\left\{-\frac{y_2}{2}(1 + y_1^2)\right\} dy_2 = \pi^{-1} (1 + y_1^2)^{-1},$$

$y_1 \in \mathfrak{R}$.

Si riconosce la f.d. della v.c. di Cauchy.

SEZIONE II.6.

1. Variabile di campionamento

Sia X_1, X_2, \dots una successione di vv.cc. i.i.d.: si tratta della successione che
idealizza la replicazione -nelle medesime condizioni- di un determinato

esperimento aleatorio (si parla allora di campionamento casuale semplice; c.c.s.); $X = (X_1, \dots, X_n)$ è detta *variabile di campionamento*. Una osservazione $x = (x_1, \dots, x_n)$ dalla v.c. X è detta *campione casuale semplice* (c.c.s.) di dimensione n .

2. Alcune importanti trasformazioni ottenute a partire da c.c.s. da normale

Sia X_1, X_2, \dots, X_n una successione di vv.cc. i.i.d., $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$. Sono particolarmente interessanti le distribuzioni *esatte* di alcune trasformazioni di X che coinvolgono le vv.cc.

$$S_n = \sum_{i=1}^n X_i \text{ (somma campionaria);}$$

$$\bar{X}_n = \sum_{i=1}^n X_i / n \text{ (media campionaria);}$$

$$s_n^{*2} = \sum_{i=1}^n (X_i - \mu)^2 / n \text{ (varianza campionaria – caso di media nota);}$$

$$s_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n - 1) \text{ (varianza campionaria).}$$

Preliminarmente, si osservi che:

Teorema 1. $E\bar{X}_n = \mu$ (cdd. “non-distorsione” della media campionaria)
[l’uguaglianza prescinde dalla condizione di normalità e di indipendenza]

Teorema 2. $Es_n^2 = \sigma^2$ (cdd. “non-distorsione” della varianza campionaria)
[l’uguaglianza prescinde dalla condizione di normalità]

Teorema 3. Se X_1, X_2, \dots, X_n è una successione di vv.cc. i.i.d., $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$, allora le vv.cc. \bar{X}_n e s_n^2 sono *indipendenti*.

Si può ora dire qualcosa di più preciso riguardo la forma delle distribuzioni in gioco.

3. Somma campionaria e media campionaria (varianza nota)

Si trova:

a) $S_n \sim N(n\mu, n\sigma^2)$;

- b) $\bar{X}_n \sim N(\mu, \sigma^2 / n)$;
 c) $\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} =: Z \sim N(0,1)$.

4. Distribuzioni collegate alla varianza campionaria

Si ha inoltre:

- d) $Z^2 \sim \chi_1^2 \equiv G(\frac{1}{2}, 2)$;
 e) $W := \sum_{i=1}^n Z_i^2 \sim \chi_n^2 \equiv G(\frac{n}{2}, 2)$; quindi, se $Y = \sigma Z$, allora
 $\sum_{i=1}^n Y_i^2 \sim G(\frac{n}{2}, 2\sigma^2)$;
 f) $\sum_{i=1}^n Z_i^2 / n \sim G(\frac{n}{2}, \frac{2}{n})$; quindi $\sum_{i=1}^n Y_i^2 / n \sim G(\frac{n}{2}, \frac{2\sigma^2}{n})$;
 g) $U := \sum_{i=1}^n (Z_i - \bar{Z}_n)^2 \sim \chi_{n-1}^2 \equiv G(\frac{n-1}{2}, 2)$ (cfr. punto e);
 h) $s_n^{*2} = \sum_{i=1}^n (X_i - \mu)^2 / n = \frac{\sigma^2}{n} W \sim G(\frac{n}{2}, \frac{2\sigma^2}{n})$; quindi $Es_n^{*2} = \sigma^2$ e
 $Var(s_n^{*2}) = \frac{2\sigma^4}{n}$;
 i) $s_n^2 = \frac{\sigma^2}{n-1} U \sim G(\frac{n-1}{2}, \frac{2\sigma^2}{n-1})$; quindi $Es_n^2 = \sigma^2$ e $Var(s_n^2) = \frac{2\sigma^4}{n-1}$.

5. Media campionaria (varianza ignota)

Dal precedente punto (i) si ottiene che la v.c. $\frac{s_n^2}{\sigma^2}$ è distribuita come un chi-quadrato con $n-1$ gradi di libertà diviso per $n-1$.

Inoltre l'indipendenza di \bar{X}_n e s_n^2 e il punto c) implicano che:

$$l) \frac{\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}}{\sqrt{s_n^2 / \sigma^2}} = \frac{\bar{X}_n - \mu}{s_n / \sqrt{n}}$$

è distribuita come una t di Student con $n - 1$ gradi

di libertà; in particolare questa statistica ha quindi media 0 e varianza

$$\frac{n-1}{n-3} \quad (n > 3).$$

SEZIONE II.7.

1. Inferenza statistica parametrica

Nell'impostazione inferenziale "parametrica" si fissa l'attenzione su una determinata classe di modelli probabilistici specificati dal valore di un parametro incognito.

Precisamente, sia $\mathcal{P} = \{P_g\}_{g \in \Theta}$ una famiglia di distribuzioni di probabilità, ossia variabili casuali, con $g \in \Theta$ parametro reale.

Il riferimento naturale è al campionamento casuale semplice da una v.c. appartenente alla famiglia \mathcal{P} ; $X = (X_1, \dots, X_n)$ è la corrispondente variabile di campionamento. Il problema generale è quello di trarre informazioni, sulla base di X , sul valore del parametro incognito.

Lo scopo inferenziale della Stima Parametrica è quello di determinare, *stimare*, il valore preciso ("puntuale") del parametro -incognito ma fissato- entro Θ , sulla base di una osservazione (c.c.s.) dalla v.c. di campionamento.

Nell'impostazione cosiddetta Bayesiana, questo obiettivo viene raggiunto tenendo anche conto di una distribuzione di probabilità assegnata *a priori* (ossia, prima dell'operazione sperimentale di campionamento casuale semplice) sullo spazio Θ .

Nell'impostazione classica non si fa invece uso di tale distribuzione.

2. Stimatore

Si dice *informatore* (o *informatore statistico* o *statistica*) una funzione della variabile di campionamento X .

Si dice *stimatore* ogni informatore $\mathcal{G}^* = \mathcal{G}^*(X)$ destinato a rimpiazzare il valore incognito del parametro \mathcal{G} .

Se $E(\mathcal{G}^*) = \mathcal{G}$, \mathcal{G}^* si dice stimatore *non-distorto* di \mathcal{G} . Più in generale, se l'interesse è rivolto a una funzione del parametro, diciamo $\eta = g(\mathcal{G})$, allora $\eta^* = \eta^*(X)$ è uno stimatore non-distorto di $\eta = g(\mathcal{G})$, se si verifica $E\eta^* = \eta = g(\mathcal{G})$. La funzione $b(\eta) = E\eta^* - \eta$ è detta *distorsione* (“*bias*”), ovvero *errore sistematico*.

3. Errore quadratico medio

L'accuratezza di uno stimatore $\eta^* = \eta^*(X)$ del parametro η è ragionevolmente determinabile in base al valore atteso

$$MSE_{\eta^*}(\eta) = E(\eta^* - \eta)^2,$$

detto *errore quadratico medio* (*MSE*, “*mean squared error*”). Si ha:

$$MSE_{\eta^*}(\eta) = \text{Var}(\eta^*) + \{b(\eta)\}^2.$$

Nella classe degli stimatori non-distorti, se η è scalare, $MSE_{\eta^*}(\eta)$ coincide con la varianza dello stimatore, $\text{Var}(\eta^*)$; se η è vettoriale, $MSE_{\eta^*}(\eta)$ coincide con la matrice di varianza-covarianza dello stimatore.

4. Efficienza

Nella classe degli stimatori non-distorti di un parametro è scalare, si dice *efficiente* quello che ha minima varianza, per ogni valore del parametro.

SEZIONE II.8.

1. Metodo dei momenti

Si possono dare diverse formulazioni del metodo, non necessariamente equivalenti.

a) Se X_1, \dots, X_n sono vv.cc. i.i.d. ed esiste finito il momento k -mo $\mu'_k(X_1) := E(X_1^k)$, allora la soluzione del sistema -posto che questo ammetta un'unica soluzione- nell'incognita \mathcal{G} ,

$$g_i(\mu'_1, \dots, \mu'_k) = h_i(\mathcal{G}) = g_i(m'_1, \dots, m'_k), \quad i = 1, 2, \dots, k,$$

dove h_i sono funzioni continue e $m'_r = \sum_{i=1}^n \frac{x_i^r}{n}$, $r = 1, 2, \dots, k$, fornisce, per k opportuno (tipicamente la dimensione del parametro \mathcal{G}), lo stimatore secondo il metodo dei momenti $\tilde{\mathcal{G}}_n = \tilde{\mathcal{G}}_n(m'_1, \dots, m'_k)$.

In particolare, si possono ad esempio considerare le equazioni del sistema:

$$\mu'_r = \mu'_r(\mathcal{G}) = m'_r, \quad r = 1, 2, \dots, k.$$

b) Alternativamente, e più in generale, si possono risolvere le equazioni del sistema -posto che questo ammetta un'unica soluzione e che i valori attesi esistano-:

$$Eg_r(X_1) = \frac{\sum_{i=1}^n g_r(x_i)}{n}, \quad r = 1, 2, \dots, k,$$

dove g_i sono funzioni continue.

Evidentemente, il metodo fornisce diverse versioni dello stimatore.

Nel caso di una v.c. multipla le notazioni in a) vanno eventualmente aggiornate considerando i momenti misti.

Lo stimatore secondo il metodo dei momenti è consistente, per la legge dei grandi numeri e per le proprietà delle trasformazioni di successioni convergenti.

Esempio

Sia $X \sim G(3, \beta)$ e X_1, \dots, X_n un c.c.s. da X . Sono stimatori secondo il metodo dei momenti le soluzioni delle seguenti equazioni alternative:

$$i) \mu'_2 - \mu^2 = 3\beta^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2;$$

$$ii) EX = 3\beta = \frac{\sum_{i=1}^n x_i}{n};$$

$$iii) \text{poiché } EX^{-1} = 1/(2\beta), \text{ si pone } \frac{1}{2\beta} = \frac{\sum_{i=1}^n x_i^{-1}}{n}.$$

2. Massima verosimiglianza: concetto base

Il metodo della massima verosimiglianza (*maximum likelihood*, ML) è intuitivamente giustificabile, nel caso parametrico discreto, secondo il paradigma della “probabilità a posteriori”, definita secondo la formula di Bayes.

P.es. si vuol decidere, fra un numero k di possibili valori parametrici $\{\mathcal{G}_1, \dots, \mathcal{G}_k\}$, quale sia quello che genera l’osservazione x . Se non è data informazione a priori su $\{\mathcal{G}_1, \dots, \mathcal{G}_k\}$ (diversamente l’impostazione è bayesiana), allora si può intendere $P(\mathcal{G}_i) = P(\mathcal{G}_j)$, nel raffronto delle formule della “probabilità a posteriori”, dato x

$$P(\mathcal{G}_i/x) = \frac{P(x/\mathcal{G}_i)P(\mathcal{G}_i)}{\sum_{r=1}^k P(x/\mathcal{G}_r)P(\mathcal{G}_r)} \quad \text{vs} \quad P(\mathcal{G}_j/x) = \frac{P(x/\mathcal{G}_j)P(\mathcal{G}_j)}{\sum_{r=1}^k P(x/\mathcal{G}_r)P(\mathcal{G}_r)}$$

$\mathcal{G}_i, \mathcal{G}_j \in \{\mathcal{G}_1, \dots, \mathcal{G}_k\}$. La stima di ML è il valore di \mathcal{G} maggiormente probabile, dato x , ossia ogni valore $\hat{\mathcal{G}}$ per il quale la “probabilità a posteriori” $P(\mathcal{G}/x)$ è massima; ma questa differisce solo per una costante dalla “verosimiglianza” $P(x/\mathcal{G})$, poiché il denominatore è lo stesso per ogni $\mathcal{G} \in \{\mathcal{G}_1, \dots, \mathcal{G}_k\}$, quindi si tratta di trovare $\hat{\mathcal{G}}$ tale che:

$$\max_{\mathcal{G}} P(x/\mathcal{G}) = P(x/\hat{\mathcal{G}}).$$

$\hat{\mathcal{G}}$ è una funzione di x : cdd. stimatore di ML.

Con leggero abuso interpretativo, nel caso continuo il metodo si estende alla determinazione del massimo di $f_X(x/\mathcal{G})$, la densità della v.c. di campionamento X , intesa come funzione di \mathcal{G} , x fissato.

3. Funzione di verosimiglianza

Evidentemente ciò che differenzia $P(\mathcal{G}/x)$ e $P(x/\mathcal{G})$ è una funzione di x : il problema di massimizzare $P(x/\mathcal{G})$, ovvero $P(\mathcal{G}/x)$, è quindi proponibile a meno di costanti o funzioni della sola x . Una funzione $L(\mathcal{G}) \propto f_X(x/\mathcal{G})$ (proporzionalità a meno di costanti o funzioni della sola x) è detta *funzione di verosimiglianza*, mentre $l(\mathcal{G}) = \log L(\mathcal{G})$ è detta *funzione di log-verosimiglianza*. La funzione $l'(\mathcal{G})$ è la *score function*. L'equazione (sistema di equazioni) $l'(\mathcal{G}) = 0$ è detta *equazione di verosimiglianza*.

Qualora x stia per il c.c.s., di dimensione n , (x_1, \dots, x_n) , si può scrivere:

$$L_n(\mathcal{G}) = \prod_{i=1}^n f_{x_i}(x_i/\mathcal{G})$$
 (uguaglianza a meno di costanti o funzioni della sola x);

$$l_n(\mathcal{G}) = \sum_{i=1}^n \log f_{x_i}(x_i/\mathcal{G})$$
 (uguaglianza a meno di costanti o funzioni della sola x).

Nei casi regolari, $\hat{\mathcal{G}}$ è una soluzione della equazione di verosimiglianza, ossia

$$l'(\hat{\mathcal{G}}) = 0$$

(un semplice controesempio è dato dalla stima della media nel caso della v.c. di Laplace; un altro controesempio è dato dalla stima del parametro a per l'ipergeometrica). Tuttavia, se $l'(\tilde{\mathcal{G}}) = 0$ occorre verificare la condizione necessaria $l''(\tilde{\mathcal{G}}) \leq 0$, perché il punto di stazionarietà $\tilde{\mathcal{G}}$ sia punto di massimo per l .

Nei casi regolari, se \mathcal{G}_0 è il vero valore del parametro, si trova:

- i. $E\{l(\mathcal{G})\} \leq E\{l(\mathcal{G}_0)\}$;
- ii. $E\{l'(\mathcal{G}_0)\} = 0$;
- iii. $E\{l''(\mathcal{G}_0)\} < 0$.

La funzione $I(\mathcal{G}) = -E\{l''(\mathcal{G})\}$ è detta *informazione di Fisher* (è una matrice nel caso di \mathcal{G} vettoriale). Più questa è grande, più è facile distinguere \mathcal{G}_0 da valori ad esso vicini. L'informazione di Fisher dipende dalla parametrizzazione scelta: se $\mathcal{G} = g(\tau)$, allora $I(\tau) = I(g(\tau)) \cdot (g'(\tau))^2$.

L'informazione di Fisher è additiva rispetto all'operazione di c.c.s.: se $I_1(\mathcal{G})$ è l'informazione di Fisher per un campione di dimensione 1, si trova quindi:

$$I_n(\mathcal{G}) = n I_1(\mathcal{G}).$$

Esempio

Per $X \sim N(0, \mathcal{G})$ si ottiene $I(\mathcal{G}) = -E\{l''(\mathcal{G})\} = \frac{1}{2\mathcal{G}^2}$. Pertanto

$$I_n(\mathcal{G}) = n I(\mathcal{G}) = \frac{n}{2\mathcal{G}^2}$$

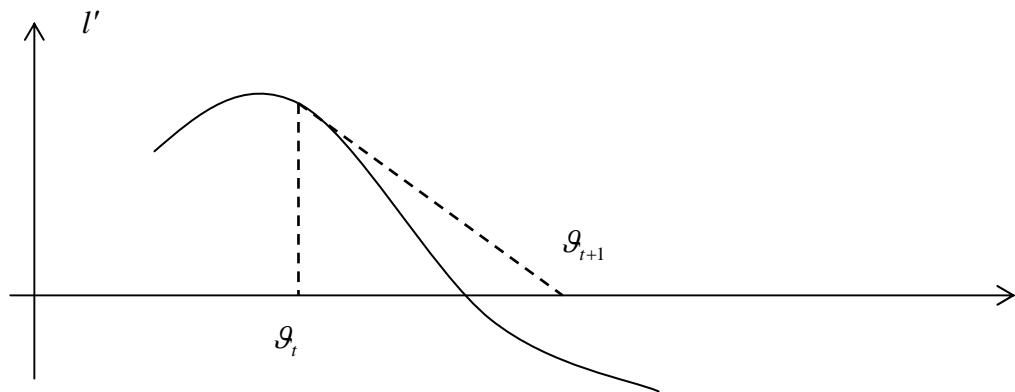
è l'informazione di Fisher per un c.c.s. di dimensione n da X . Inoltre, poiché $\mathcal{G} = \sigma^2$, si ha $I(\sigma) = \frac{1}{2(\sigma^2)^2} \cdot (2\sigma)^2 = \frac{2}{\sigma^2}$.

4. Aspetti computazionali – approssimazioni numeriche per il calcolo della stima di ML

a) Metodo di Newton-Raphson

Dallo sviluppo in serie di Taylor della *score function* l'

$$0 = l'(\hat{\mathcal{G}}) \cong l'(\mathcal{G}) + (\hat{\mathcal{G}} - \mathcal{G})l''(\mathcal{G})$$



si ricava $\hat{\theta} = \theta - l'(\theta)[l''(\theta)]^{-1}$, che può essere usato iterativamente per ottenere $\hat{\theta}$:

$$\theta_{t+1} = \theta_t - l'(\theta_t)[l''(\theta_t)]^{-1}.$$

b) Metodo del Fisher-scoring

Si sostituisce, nella formula di Newton-Raphson, il valore atteso di $l''(\theta_t)$ in luogo di $l''(\theta_t)$, ossia:

$$\theta_{t+1} = \theta_t + l'(\theta_t)[I_n(\theta_t)]^{-1}.$$

c) Regula falsi

Si prendono $a_0 < b_0$ tali che $l'(a_0) > 0$ e $l'(b_0) < 0$. Per un opportuno x_1 vale la proporzione:

$$\frac{x_1 - a_0}{l'(a_0)} = \frac{b_0 - x_1}{-l'(b_0)}$$

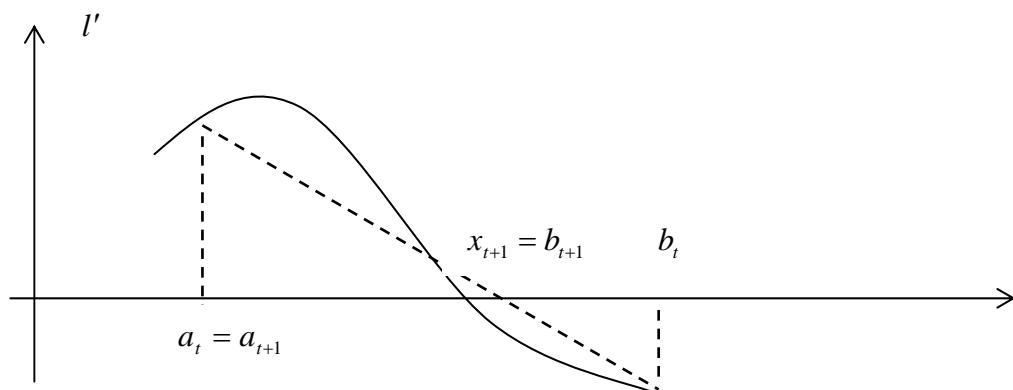
che dà $x_1 = \frac{a_0 l'(b_0) - b_0 l'(a_0)}{l'(b_0) - l'(a_0)}$. Si itera il processo un altro passo ponendo

$$a_1 = x_1 \text{ e } b_1 = b_0 \text{ se } l'(x_1) > 0,$$

oppure

$$a_1 = a_0 \text{ e } b_1 = x_1 \text{ se } l'(x_1) < 0,$$

e così via i passi successivi, finché $b_i - a_i$ è sufficientemente piccolo. Il massimo relativo si approssima con $(b_i + a_i)/2$.



Esempio

Da un c.s.s. di dimensione n da $X \sim N(0, \vartheta)$ si ottiene, posto $m_2 = \sum x_i^2 / n$:

$$l_n = -\frac{n}{2} \log \vartheta - \frac{n}{2\vartheta} m_2;$$

$$l'_n = -\frac{n}{2\vartheta} + \frac{n}{2\vartheta^2} m_2;$$

$$l''_n = \frac{n}{2\vartheta^2} - \frac{n}{\vartheta^3} m_2.$$

Quindi $I_n(\vartheta) = -E \left\{ \frac{n}{2\vartheta^2} - \frac{n}{\vartheta^3} m_2 \right\} = \frac{n}{2\vartheta^2}$ (essendo $E(\sum x_i^2 / n) = \vartheta$).

Per ogni valore di partenza ϑ_1 , il Fisher scoring dà:

$$\mathcal{G}_2 = \mathcal{G}_1 + \frac{n}{2\mathcal{G}_1} \left(\frac{m_2}{\mathcal{G}_1} - 1 \right) \frac{2\mathcal{G}_1^2}{n},$$

che converge in un passo a $\hat{\mathcal{G}}$ (che vale m_2).

Il metodo di Newton-Raphson dà invece:

$$\mathcal{G}_2 = \mathcal{G}_1 + \frac{\frac{n}{2\mathcal{G}_1^2}(\mathcal{G}_1 - m_2)}{\frac{n}{2\mathcal{G}_1^3}(\mathcal{G}_1 - 2m_2)},$$

che, in particolare, porta a un processo iterativo non convergente, se si sceglie $\mathcal{G}_1 > m_2$.

5. Invarianza della stima di massima verosimiglianza

La stima di ML è facilmente estendibile nelle riparametrizzazioni del modello. Se $\hat{\mathcal{G}}$ è uno stimatore di ML di \mathcal{G} e $\tau = h(\mathcal{G})$, allora $h(\hat{\mathcal{G}})$ è uno stimatore di ML di τ .

E' da notare che se h è biunivoca il passaggio è ovvio, mentre se h non è biunivoca occorre considerare la *funzione di verosimiglianza indotta* $L(\tau) := \sup_{\{\mathcal{G} | h(\mathcal{G}) = \tau\}} L(\mathcal{G})$.

Esempio

Da un c.s.s. di dimensione n da $X \sim N(0, \mathcal{G})$ si ottiene $\hat{\mathcal{G}} = \sum x_i^2 / n$. Allora lo stimatore di ML di $\sigma = \sqrt{\mathcal{G}}$ è $\hat{\sigma} = \sqrt{\sum x_i^2 / n}$.

SEZIONE II.9.

Abbiamo visto che nella classe degli stimatori non-distorti è ragionevole pensare di operare confronti in base alle rispettive varianze. Si dice efficiente lo stimatore

che, nella classe degli stimatori non distorti, presenta -uniformemente, ossia per ogni valore del parametro- minima varianza.

Vi sono due criteri per verificare l'efficienza di uno stimatore.

- I) Il criterio di Rao-Cramér;
- II) Il criterio di Lehmann-Scheffé.

Per quest'ultimo è necessario introdurre le importanti nozioni di sufficienza e completezza.

1. Sufficienza

Detto in modo rozzo, la sufficienza di un informatore è la proprietà di preservazione -nella trasformazione che esso definisce- della informazione contenuta nel campione.

Precisamente:

sia $\mathcal{P} = \{P_{\vartheta}\}_{\vartheta \in \Theta}$ una famiglia di distribuzioni di probabilità dipendente da un parametro reale ϑ , $\vartheta \in \Theta$. L'informatore T è *sufficiente* per il parametro ϑ (ovvero per la famiglia \mathcal{P}) se per ogni evento A fissato nello spazio campionario, la distribuzione di probabilità condizionata $g_A(t) = P(X \in A/t)$ esiste e non dipende da ϑ .

- Se S e T sono due informatori e $S = g(T)$, con g biunivoca, allora S e T sono *equivalenti*. In particolare, se T è sufficiente allora anche S è sufficiente.
- Uno stimatore si dice *sufficiente* se è un informatore sufficiente.

2. Criteri per la verifica della sufficienza

La verifica della sufficienza può essere facilmente effettuata invocando uno dei due seguenti principi, stabiliti con i Teoremi 1 e 2.

Teorema 1. Sia X variabile di campionamento. L'informatore T è *sufficiente* per il parametro \mathcal{G} se e solo se il rapporto $\frac{f_X(x)}{f_T(t)}$ non dipende da \mathcal{G} .

Esempio

Sia X variabile di c.c.s. di dimensione n da una v.c. di tipo normale con media μ e varianza σ^2 . Si considera l'informatore

$$T = (T_1, T_2) \quad \text{con } T_1 = \frac{\sum X_i}{n}, \quad T_2 = \sum (X_i - T_1)^2.$$

E' noto che $T_1 \sim N(\mu, \frac{\sigma^2}{n})$, $\frac{T_2}{\sigma^2} \sim \chi_{(n-1)}^2$ e che T_1 e T_2 sono indipendenti (vd. Sezione 6). Pertanto è nota la densità di $T = (T_1, T_2)$:

$$\begin{aligned} f_T(t) &= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sigma} \exp\left\{-\frac{n(t_1 - \mu)^2}{2\sigma^2}\right\} \cdot \frac{t_2^{(n-1)/2-1} \exp\left\{-\frac{t_2}{2\sigma^2}\right\}}{(2\sigma^2)^{(n-1)/2} \Gamma\left(\frac{n-1}{2}\right)} = \\ &= \exp\left\{-\frac{t_2}{2\sigma^2} - \frac{n(t_1 - \mu)^2}{2\sigma^2}\right\} \cdot \frac{t_2^{\frac{n-3}{2}} \sqrt{n}}{\Gamma\left(\frac{n-1}{2}\right) \cdot \sqrt{\pi} \cdot \sigma^n \cdot 2^{n/2}} \end{aligned}$$

Quindi si ha il rapporto (sostituendo $t_1 = \frac{\sum x_i}{n}$, $t_2 = \sum (x_i - t_1)^2$)

$$\begin{aligned} \frac{f_X(x)}{f_T(t)} &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum (x_i - t_1)^2 + n \cdot (t_1 - \mu)^2 \right]\right\} / f_T(t) = \\ &= \frac{t_2^{-\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right)}{\sqrt{n} \cdot \pi^{(n-1)/2}} \end{aligned}$$

che non dipende da $\mathcal{G} = (\mu, \sigma^2)$. Pertanto T è sufficiente per il parametro.

Teorema 2. (Criterio di Neyman-Fisher) Sia X variabile di campionamento. L'informazione T è sufficiente per il parametro \mathcal{G} se e solo se si può scrivere:

$$f_x(x) = f_x(x; \mathcal{G}) = h(x) \cdot g(t(x); \mathcal{G})$$

dove $h(x) > 0$ è indipendente da \mathcal{G} e g dipende da x solo attraverso $t(x)$.

Esempio

Sia X variabile di c.c.s. di dimensione n da una v.c. di tipo normale con media 0 e varianza σ^2 . Si considera l'informazione $T = \sum (X_i - \frac{\sum X_i}{n})^2$. Si trova

$$\begin{aligned} f_x(x) &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum (x_i - \bar{x}_n)^2 + n \cdot (\bar{x}_n)^2 \right]\right\} = \\ &= \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{t(x)}{2\sigma^2} - \frac{n \cdot (\bar{x}_n)^2}{2\sigma^2} - n \cdot \log \sigma\right\}. \end{aligned}$$

Si applica il criterio con $g(\bullet) = \exp(\bullet)$. T non è sufficiente perché g non dipende da x solo attraverso $t(x)$.

3. Completezza

Sia $\mathcal{P} = \{P_\mathcal{G}\}_{\mathcal{G} \in \Theta}$ una famiglia di distribuzioni di probabilità -ovvero una famiglia di variabili casuali $X = X(\mathcal{G})$ - dipendente da un parametro reale (k -dimensionale) \mathcal{G} , $\mathcal{G} \in \Theta$.

\mathcal{P} è completa se per ogni funzione g (per cui esiste l'integrale) l'uguaglianza

$$Eg(X) = 0$$

implica $g(x) = 0$ per ogni x . Si noti che, poiché X dipende da \mathcal{G} , è implicitamente richiesto che il valore atteso sia nullo $\forall \mathcal{G} \in \Theta$.

Un informatore T è completo se costituisce una famiglia di distribuzioni completa.

Esempio (famiglia completa)

Sia \mathcal{P} la famiglia di distribuzioni di probabilità di tipo binomiale $Bi(n, \vartheta)$.

La condizione $Eg(X) = 0 \quad \forall \vartheta \in (0,1)$ è

$$\sum_{i=0}^n g_i \binom{n}{i} \vartheta^i (1-\vartheta)^{n-i} = 0 \quad \forall \vartheta \in (0,1)$$

che equivale a

$$\sum_{i=0}^n g_i \binom{n}{i} \xi^i = 0 \quad \forall \vartheta \in (0,1)$$

con $\xi = \frac{\vartheta}{1-\vartheta}$.

L'equazione ha *al più* n soluzioni ξ_1, \dots, ξ_n , i.e. $\vartheta_i = \frac{\xi_i}{\xi_i + 1}$, $i = 1, \dots, n$, e non può essere soddisfatta simultaneamente *per ogni* $\vartheta \in (0,1)$, se non nel caso degenero in cui $g_i = 0$, $i = 0, 1, \dots, n$.

Esempio (informatore sufficiente non completo)

X è una variabile di c.c.s. di dimensione n da una v.c. (X_1, X_2) di tipo normale bivariata $N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$. L'informatore statistico

$$T = \left(\sum_i X_{1i}, \sum_i X_{2i}, \sum_i X_{1i}^2, \sum_i X_{2i}^2 \right)$$

è sufficiente, ma $E\left\{ \sum_i (X_{1i} - X_{2i}) \right\} = 0$, con $\sum_i (X_{1i} - X_{2i})$ non identicamente nulla per ogni $\vartheta = (\mu, \sigma_1, \sigma_2)$. Quindi la famiglia non è completa.

4. Famiglia esponenziale

Una famiglia $\mathcal{P} = \{P_g\}_{g \in \Theta}$ di distribuzioni di probabilità si dice *esponenziale* k -dimensionale se ogni distribuzione P_g ammette densità (rispetto ad una comune misura dominante assegnata) del tipo

$$\exp\left\{\sum_{i=1}^k \eta_i T_i(x) - A(\eta)\right\} h(x)$$

$\eta = (\eta_1, \dots, \eta_k)$ è il cdd. parametro naturale, dove sia gli η_i che i T_i sono da intendere fra loro linearmente indipendenti (scrittura in forma minimale).

$\mathcal{E}(\eta)$ spazio parametrico naturale è un insieme convesso.

Se \mathcal{E} contiene un rettangolo k -dimensionale la famiglia è *di rango pieno*.

- $T = (T_1, \dots, T_k)$ è sufficiente.
- Se la famiglia è di rango pieno $T = (T_1, \dots, T_k)$ è altresì completo.

5. Efficienza secondo Rao-Blackwell / Lehmann-Scheffé

Si fa riferimento per semplicità al solo scalare.

Teorema. Se $\mathcal{G}^* = \mathcal{G}^*(X)$ è uno stimatore del parametro \mathcal{G} :

- a) sufficiente,
- b) completo,
- c) non distorto,

allora è *efficiente* (secondo Lehmann-Scheffé) – ed essenzialmente unico.

6. Efficienza secondo Rao-Cramér - caso scalare

Si fa riferimento per semplicità al caso scalare.

Teorema. (*Disuguaglianza di Rao-Cramér*) Nelle ipotesi di regolarità per la definizione di $I(\mathcal{G})$ (possibilità di derivare due volte sotto il segno di integrale la log-verosimiglianza), se T è un informatore definito sulla variabile di campionamento $X = (X_1, \dots, X_n)$, tale che $E(T^2) < \infty$, allora

$$\text{Var}(T) \geq \frac{\left[\frac{\partial}{\partial \vartheta} E(T) \right]^2}{I_n(\vartheta)} \quad (\text{cdd. lim. inf. di Rao-Cramér}).$$

- In particolare, se T è uno stimatore non distorto di ϑ (i.e. $ET = \vartheta$), allora $\text{Var}(T) \geq I_n(\vartheta)^{-1}$. Se poi $\text{Var}(T) = I_n(\vartheta)^{-1}$ allora T è *efficiente* (secondo Rao-Cramér).

SEZIONE II.10.

1. Intervalli di confidenza

In alternativa alla stima puntuale, si può fare inferenza con la *stima per intervalli* (cdd. anche *intervalli di confidenza* o *regioni di confidenza*).

L'obiettivo è quello di indicare, in corrispondenza di un campione, un intervallo (più in generale una regione) nel quale -con un grado di fiducia prestabilito- il vero valore del parametro incognito è contenuto. L'esito dell'inferenza è quindi meno "preciso" di quello determinato con la stima puntuale, ma più completo, giocando simultaneamente un controllo sulla misura del grado di confidenza con il quale la localizzazione del parametro può esser ritenuta corretta. E' da considerare anche che il controllo sulla dimensione dell'intervallo e sulla misura del grado di confidenza può essere perfezionato grazie alla manipolazione della dimensione campionaria.

In riferimento al campionamento casuale semplice da una v.c. $X = (X_1, \dots, X_n)$ appartenente alla famiglia $\mathcal{P} = \{P_\vartheta\}_{\vartheta \in \Theta}$, $\vartheta \in \Theta$ parametro reale, si può costruire un intervallo di confidenza per ϑ se si trova una funzione Q_n (detta *quantità pivotale*) di X e di ϑ la cui distribuzione non dipende da ϑ stesso. Per essa si può scrivere infatti, per γ fissato, $0 < \gamma < 1$:

$$P\{q_1 \leq Q_n \leq q_2\} = \gamma$$

per opportuni valori q_1 e q_2 . Esplicitando le disuguaglianze in funzione del parametro ϑ si può equivalentemente scrivere:

$$P\{a_n \leq \mathcal{G} \leq b_n\} = \gamma$$

per opportune funzioni di X , $a_n = a_n(X_1, \dots, X_n)$ e $b_n = b_n(X_1, \dots, X_n)$. Allora l'intervallo aleatorio (a_n, b_n) è un *intervallo di confidenza* al $\gamma \cdot 100$ per cento; γ è il *livello di confidenza*; a_n e b_n sono i *limiti di confidenza* dell'intervallo.

Se la famiglia di distribuzioni in gioco è di tipo discreto l'intervallo sarà determinato in funzione del raggiungimento di un prefissato livello di confidenza "minimo":

$$P\{a_n \leq \mathcal{G} \leq b_n\} \geq \gamma.$$

Naturalmente, a parità di livello di confidenza l'obiettivo è quello di ottenere l'intervallo di confidenza "più piccolo". Quindi, caso per caso, la quantità pivotale Q_n e i limiti di confidenza andranno scelti opportunamente.

2. Intervallo di confidenza per la media di una normale – varianza nota

Se $X_i \sim N(\mu, \sigma_0^2)$, $i = 1, 2, \dots, n$, sono i.i.d. con σ_0^2 nota, allora $\bar{X}_n \sim N(\mu, \sigma_0^2/n)$. Quindi:

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

è una quantità pivotale. La condizione $P\{q_1 \leq Q_n \leq q_2\} = \gamma$ è soddisfatta per infinite scelte di q_1 e q_2 . Quella che minimizza il range (q_1, q_2) consiste nel prendere $q_2 = -q_1 = \xi_\gamma$, dove ξ_γ è il quantile $\frac{1+\gamma}{2} \cdot 100$ -esimo di Z . Si ricava l'intervallo di confidenza per μ :

$$\left(\bar{x}_n - \xi_\gamma \frac{\sigma_0}{\sqrt{n}}, \bar{x}_n + \xi_\gamma \frac{\sigma_0}{\sqrt{n}} \right),$$

che ha lunghezza $2\xi_\gamma \frac{\sigma_0}{\sqrt{n}}$.

3. Intervallo di confidenza per la media di una normale – varianza ignota

Se $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$, sono i.i.d., allora

$$Q_n = \frac{\bar{X}_n - \mu}{s_n / \sqrt{n}} \sim t_{(n-1)}.$$

Quindi Q_n è una quantità pivotale. La condizione $P\{q_1 \leq Q_n \leq q_2\} = \gamma$ è soddisfatta per infinite scelte di q_1 e q_2 . Quella che minimizza il range (q_1, q_2) consiste nel prendere $q_2 = -q_1 = \xi_\gamma$, dove ξ_γ è il quantile $\frac{1+\gamma}{2} \cdot 100$ -esimo della t di Student con $(n-1)$ gradi di libertà $t_{(n-1)}$.

Si ricava l'intervallo di confidenza per μ :

$$\left(\bar{x}_n - \xi_\gamma \frac{s_n}{\sqrt{n}}, \bar{x}_n + \xi_\gamma \frac{s_n}{\sqrt{n}} \right),$$

che ha lunghezza $2\xi_\gamma \frac{s_n}{\sqrt{n}}$.

4. Intervallo di confidenza per la varianza di una normale

Se $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$, sono i.i.d., allora

$$Q_n = s_n^2 \cdot \frac{n-1}{\sigma^2} \sim \chi_{n-1}^2.$$

Quindi Q_n è una quantità pivotale. La condizione $P\{q_1 \leq Q_n \leq q_2\} = \gamma$ è soddisfatta per infinite scelte di q_1 e q_2 . Una soluzione approssimata (valida nella misura in cui n non è troppo piccolo) si ottiene considerando per q_1 e q_2 , rispettivamente, i quantili $\frac{1-\gamma}{2} \cdot 100$ -esimo e $\frac{1+\gamma}{2} \cdot 100$ -esimo del χ_{n-1}^2 .

Si ricava l'intervallo di confidenza per σ^2 :

$$\left(\frac{s_n^2}{q_2}(n-1), \frac{s_n^2}{q_1}(n-1) \right).$$

5. Intervalli di confidenza asintotici per la media

Se X_1, X_2, \dots, X_n sono vv.cc. i.i.d. qualsiasi, con media μ e varianza σ^2 finite, la successione \bar{X}_n è $AN(\mu, \sigma^2/n)$. Ciò consente di determinare un intervallo di confidenza approssimato, per n abbastanza grande, della media μ :

$$\left(\bar{x}_n - \xi_\gamma \frac{\tilde{\sigma}_n}{\sqrt{n}}, \bar{x}_n + \xi_\gamma \frac{\tilde{\sigma}_n}{\sqrt{n}} \right),$$

dove $\tilde{\sigma}_n$ è una opportuna stima dello s.q.m. σ , e dove ξ_γ è il quantile $\frac{1+\gamma}{2} \cdot 100$ -esimo della normale standard. Ovviamente, nel caso normale questo intervallo approssima direttamente quello “esatto” ottenuto al §3.

Esempio

Se X_1, X_2, \dots, X_n sono vv.cc. i.i.d. bernoulliane con parametro ignoto p , \bar{X}_n è $AN\left(p, \frac{p(1-p)}{n}\right)$. Allora

$$\left(\bar{x}_n - \xi_\gamma \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}, \bar{x}_n + \xi_\gamma \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \right)$$

è l’intervallo di confidenza approssimato, a livello γ , per il parametro p . (Qui si è utilizzata la stima di massima verosimiglianza dello s.q.m., impiegando la proprietà di invarianza di questo tipo di stimatore)

6. Intervalli di confidenza asintotici per la varianza

Se X_1, X_2, \dots, X_n sono vv.cc. i.i.d. qualsiasi, con media μ e varianza σ^2 finite, la successione s_n^2 è $AN\left(\sigma^2, \frac{\bar{\mu}_4 - \sigma^4}{n}\right)$.

In particolare, in ipotesi di normalità s_n^2 è $AN\left(\sigma^2, \frac{2\sigma^4}{n}\right)$.

Ciò consente di determinare un intervallo di confidenza approssimato (eventualmente con utilizzo di stimatori ausiliari come nel precedente paragrafo), per n abbastanza grande, della varianza σ^2 , usando le seguenti quantità pivotali

$$\text{(caso generale)} \quad \frac{s_n^2 - \sigma^2}{\sqrt{\frac{\mu_4 - \sigma^4}{n}}} \approx N(0,1);$$

$$\text{(caso normale)} \quad \frac{s_n^2 - \sigma^2}{\sqrt{\frac{2\sigma^4}{n}}} \approx N(0,1).$$

In questo ultimo caso si ricava, in particolare, l'intervallo

$$\left(\frac{s_n^2}{1 + \xi_\gamma \sqrt{\frac{2}{n}}}, \frac{s_n^2}{1 - \xi_\gamma \sqrt{\frac{2}{n}}} \right)$$

dove ξ_γ è il quantile $\frac{1+\gamma}{2} \cdot 100$ -esimo della normale standard. Ovviamente questo intervallo approssima direttamente quello “esatto” ottenuto al §4.

Bibliografia

- BERTOLI-BARSOTTI, L., 1995, *Probabilità. Aspetti storici e assiomatizzazione*, ISU, Milano.
- BERTOLI-BARSOTTI, L., 1996, *Problemi e complementi di calcolo delle probabilità ed inferenza statistica*, ISU, Milano.
- GNEDENKO, B.V., 1992, *Teoria della Probabilità*, Editori Riuniti.
- HACKING, I., 1987, *L'emergenza della probabilità*, Il Saggiatore.
- HALD, A., 1990, *A history of probability and statistics and their applications before 1750*, Wiley, New York.
- KNYPSTRA, S., 2002, *PQRS – vers. 3.2* [software statistico per PC; scaricabile dal sito <http://www.eco.rug.nl/medewerk/knypstra/pqrs.html>].
- KOLMOGOROV, A.N. 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Erg. Math., vol. 2, n. 3, Springer-Verlag, Berlin.
- ZANELLA, A., 1980, *Argomenti di Statistica Metodologica. La struttura del modello probabilistico*, Cleup, Padova.

Indice Analitico

A

Asintotica normalità 37
Assiomi 6, 8
Additività, assioma di 7, 17
Asimmetria 19

B

Bayes, formula di 9, 72
Bernoulliana, v.c. 20, 33
Bertoli-Barsotti 1, 15, 57, 58, 59, 88
Beta, funzione 30
Beta, v.c. 30, 33
Beta-binomiale, v.c. 33
Binomiale, v.c. 20, 21, 22, 33, 56, 81
Bose-Einstein, statistica di 4

C

Campione casuale semplice (c.c.s.) 67, 69, 71, 73, 74, 79, 80, 81
Cauchy, v.c. di 14, 15, 29, 32, 44, 59, 66
Cauchy-Schwarz, disuguaglianza di 44
Chebyshev, disuguaglianza di 34, 37
Chi-quadrato, v.c. 27, 65, 68
Coefficiente di correlazione 44, 47
Coerenza 16, 17
Combinazioni 2
Continuità 7
Convergenza in distribuzione 36, 60
Convergenza in probabilità 36
Covarianza 43, 44, 47, 48, 50, 52, 70
Curtosi 19, 21, 22, 26, 28, 30, 32

E

Equazione di verosimiglianza 73
Errore quadratico medio 70
Esponenziale, v.c. 27, 29, 31, 58
Esponenziale, famiglia 82
Eventi incompatibili 5, 11
Eventi indipendenti 9, 11
Evento 4
Evento certo 5
Evento elementare 4
Evento impossibile 5
Evento verificato 5

F

Famiglia completa 80, 81
Fermi-Dirac, statistica di 4
Fisher-scoring, metodo del 75, 76
Funzione caratteristica 53, 56, 57, 58, 59, 60
Funzione di densità condizionata 45, 46
Funzione di ripartizione 11, 12, 14, 15, 24, 27, 29, 30, 31, 35, 37, 39, 40, 41, 60, 61, 62, 63, 65
Funzione di ripartizione multivariata 39, 41, 61, 62, 65
Funzione di verosimiglianza 73, 77

G

Galileo 1, 3
Gamma, funzione 25, 26
Gamma, v.c. 26, 27, 30, 58
Geometrica, v.c. 23, 57
Gioco equo 15, 16, 17

H

Hald 1, 88
Hacking 1, 5, 88
Huyghens 10

I

Indipendenti, eventi vd. eventi indipendenti

Indipendenti, variabili casuali vd. variabili casuali indipendenti

Informatore statistico 69, 70, 78, 79, 80, 81, 82

Informatore sufficiente 78, 79, 80, 81, 82

Informatore completo 80, 81, 82

Informazione di Fisher 74

Insieme delle parti, 5

Intervallo di confidenza 83, 84, 85, 86, 87

Ipergeometrica, v.c. 21, 73

K

Kolmogorov 6, 7

L

Laplace, v.c. 29, 59, 73

Lehmann-Scheffé 78, 82

Legge dei Grandi Numeri 37, 71

Limiti di confidenza 84

Livello di confidenza 84

Locazione e scala, famiglia 62, 63

Logistica, v.c. 29

M

Massima verosimiglianza, metodo della 72, 77, 86

Matrice di correlazione 43, 44

Matrice di covarianza 43, 44, 47, 50, 52, 70

Matrice di varianza-covarianza vd. matrice di covarianza

Maxwell-Boltzman, statistica di 4

Media (v. anche valore atteso) 15, 16, 18, 19, 20, 21, 22, 24, 25, 29, 30, 34, 35,
36, 37, 47, 59, 67, 68, 69, 73, 79, 80, 84, 85, 86

Media campionaria 36, 37, 67, 68

Mediana 15, 24

ML (maximum likelihood) 72, 73, 74, 77

Moda 15, 24

Momenti di una v.c. 18, 24, 26, 29, 30, 32, 34, 42, 43, 44, 54, 58, 59, 71

Momenti centrali 19

Momenti fattoriali 19, 21

Momenti, metodo dei 70, 71

Momenti misti 43

Momenti non-centrali 19

Monotonicità 7

Multinomiale, v.c. 41, 42, 44, 56

N

Newton-Raphson, metodo di 74, 75, 77

Neyman-Fisher, criterio 80

Normale, v.c. 19, 24, 25, 26, 30, 37, 38, 51, 57, 60, 65, 67, 79, 80, 84, 85, 86, 87

Normale bivariata, v.c. 47, 50, 51, 81

Normale multipla, v.c. 47, 51, 52

Numerabile, insieme 5

O

Omoschedasticità 51, 52

P

Pacioli 10

Parametro 69

Permutazioni 1, 2

Poisson, v.c. 22, 56, 57

Probabilità classica 1, 5

Probabilità condizionata 9, 10, 45, 78

Probabilità soggettiva 16, 17

Probabilità totali, formula delle 9

Q

Quantità pivotale, metodo della 83, 84, 85, 87

Quota 16

R

Rao-Cramér (criterio, lim inf, disuguaglianza) 78, 82, 83

Regula falsi 75

S

σ – algebra 8, 11

S.Pietroburgo, paradosso di 17

Scarto quadratico medio 18

Schema dell'urna 20

Scommessa 16

Somma campionaria 35, 67
Spazio fondamentale 4
Spazio probabilizzabile 7
Stimatore 69, 70, 71, 77, 78, 82, 83, 86
Stimatore efficiente 70, 77, 82, 83
Stimatore non-distorto 70
Stimatore di ML 73, 77

T

t di Student, v.c. 32, 69, 85
Teorema del Limite Centrale 24, 37, 59
Trasformazione di variabile casuale 60, 61, 62, 63, 64, 65, 67, 71
Trasformazione monotona a tratti 61, 63
Trasformazione regolare 62, 65

V

Valore atteso 15, 17, 18, 19, 42, 43, 46, 53, 70, 75, 80
Valore atteso condizionato 46
Variabile casuale 11
Variabile casuale (assolutamente) continua 14, 34, 42, 63
Variabile casuale discreta 13, 43
Variabile casuale di tipo misto 15
Variabile casuale multipla 39, 40, 42, 43, 45, 47, 54, 61, 62, 71
Variabili casuali indipendenti 34, 35, 37, 41, 46, 51, 52, 54, 56, 57, 65, 67, 79, 82
Variabili casuali i.i.d. 35, 37, 59, 66, 67, 71, 84, 85, 86
Variabili casuali somiglianti 35
Variabile di campionamento 66, 67, 69, 73, 79, 80, 82, 83
Varianza 18, 19, 20, 21, 24, 25, 29, 30, 34, 35, 37, 44, 47, 48, 50, 51, 52, 59, 67, 68, 69, 70, 78, 79, 80, 84, 85, 86, 87
Varianza campionaria 67, 68

W

Weibull, v.c. 31

Z

Zanella 13, 88