

Università di Torino

QUADERNI DIDATTICI

del

Dipartimento di Matematica

A. NEGRO

Elementi di
Calcolo delle Probabilità

Quaderno # 33 - Aprile 2005



Prefazione

Questa breve monografia si propone di introdurre in modo piano e sintetico, ma con rigore logico e grande attenzione ad una definizione precisa delle strutture fondamentali, i primi elementi di Calcolo delle Probabilità per il Corso di laurea in Matematica.

Pertanto si procede nella presentazione di concetti, schemi elementari e variabili aleatorie classiche, che formalizzano esperienze comuni e ipotesi ricorrenti nei modelli fisici, alternandoli con quadri concettuali di ampio orizzonte e con nozioni e risultati avanzati, particolarmente in connessione con gli sviluppi, che si studieranno in corsi successivi, della teoria generale della misura e dell'integrazione e con il confronto di convergenze differenti in spazi di funzioni misurabili.

Molta parte del testo è riservata alla descrizione delle variabili aleatorie discrete e di quelle continue, ma si è cercato di collocare con chiarezza tutti i risultati specifici nella cornice canonica dello studio di variabili generali definite in arbitrari spazi di probabilità, anche per fornire il linguaggio adeguato ad eventuali studi successivi, ad esempio sui processi stocastici e sulle equazioni differenziali stocastiche. Sono quindi presenti vari approfondimenti e complementi più avanzati che difficilmente possono essere proposti nelle lezioni previste e che non costituiscono parte del programma d'esame.

Nel quarto capitolo si introduce un breve collegamento con i primi temi della statistica elementare attraverso esempi di stime intervallari di medie e varianze con prescritto livello di confidenza.

Negli ultimi due capitoli sono presentati e parzialmente dimostrati i teoremi fondamentali del limite centrale e sulle leggi deboli e forti dei grandi numeri.

Questo testo nasce da una riflessione sui principi fondamentali del Calcolo delle Probabilità e da una selezione di temi essenziali da parte di un docente di Analisi matematica, che per vari anni si è trovato a svolgere, non come specialista, attività didattica in un settore disciplinare affine e affascinante. Il testo è una rielaborazione, completamento ed estensione di appunti preparati prima per un corso di Calcolo della Probabilità (I modulo) del vecchio ordinamento e poi per un corso di Calcolo delle Probabilità I della nuova laurea triennale, entrambi tenuti per alcuni anni presso il Corso di laurea in Matematica della Facoltà di Scienze M.F.N. dell'Università di Torino.

Ringrazio vivamente tutti i Colleghi, docenti, ricercatori e dottorandi, del settore disciplinare di Calcolo delle Probabilità, che, con grande pazienza e disponibilità, mi hanno fornito indicazioni essenziali, assistenza continua, utili chiarimenti e cordiale incoraggiamento.

Torino, dicembre 2004

ANGELO NEGRO

Università degli studi di Torino

Indice

1	Eventi e probabilità	5
1.1	<i>Eventi</i>	5
1.2	<i>σ-algebre di insiemi e spazi misurabili</i>	6
1.3	<i>Esempi elementari</i>	7
1.4	<i>Probabilità (misura di probabilità)</i>	8
1.5	<i>Esempi di spazi di probabilità</i>	8
1.6	<i>Probabilità e frequenza</i>	12
1.7	<i>Alcune regole di calcolo delle probabilità</i>	13
1.8	<i>Probabilità condizionale e indipendenza</i>	14
1.9	<i>Esempi di applicazione delle regole elementari di calcolo</i>	18
2	Variabili aleatorie reali discrete	25
2.1	<i>Variabili aleatorie reali</i>	25
2.2	<i>Variabili aleatorie reali discrete</i>	26
2.3	<i>Distribuzione binomiale (Schema di Bernouilli)</i>	27
2.4	<i>I teoremi di DeMoivre-Laplace</i>	28
2.5	<i>Distribuzione multinomiale</i>	31
2.6	<i>Distribuzione geometrica</i>	32
2.7	<i>Distribuzione binomiale negativa (di parametri p ed r)</i>	33
2.8	<i>Valor medio (variabili discrete)</i>	34
2.9	<i>Varianza e deviazione standard (variabili discrete)</i>	36
2.10	<i>Funzione generatrice dei momenti</i>	38
2.11	<i>Esempi di calcolo di medie, varianze e funzioni generatrici</i>	38
2.12	<i>Distribuzione ipergeometrica</i>	42
2.13	<i>Limite della distribuzione ipergeometrica per $N \rightarrow +\infty$</i>	46
2.14	<i>Distribuzione di Poisson qual limite della distribuzione binomiale</i>	46
3	Variabili aleatorie reali di tipo generale e variabili continue	47
3.1	<i>Distribuzione, indipendenza</i>	47
3.2	<i>Valor medio</i>	52
3.3	<i>Variabili aleatorie continue</i>	55
3.4	<i>Funzioni di variabili aleatorie</i>	56
3.5	<i>Varianza</i>	58
3.6	<i>Disuguaglianza di Chebyshev</i>	59

3.7	<i>Funzione generatrice dei momenti</i>	60
3.8	<i>Distribuzione uniforme</i>	61
3.9	<i>Distribuzione di Cauchy</i>	63
3.10	<i>Distribuzione normale o di Gauss</i>	64
3.11	<i>Distribuzione di Maxwell</i>	66
3.12	<i>Distribuzione esponenziale e processi di Poisson</i>	67
4	Densità congiunta e funzioni di più variabili aleatorie reali	72
4.1	<i>Densità congiunta (caso discreto)</i>	72
4.2	<i>Densità congiunta (caso continuo)</i>	72
4.3	<i>Covarianza e correlazione</i>	76
4.4	<i>La distribuzione Gaussiana bivariata</i>	79
4.5	<i>Variabili normali indipendenti</i>	81
4.6	<i>Distribuzioni gamma</i>	83
4.7	<i>Distribuzioni χ^2</i>	84
4.8	<i>Stima simultanea di media e varianza di una variabile normale</i>	86
4.9	<i>Distribuzioni t di Student</i>	89
4.10	<i>Le variabili F e z di Fisher</i>	92
5	Funzioni caratteristiche e teorema del limite centrale	95
5.1	<i>Funzione caratteristica</i>	95
5.2	<i>Il teorema di Lévy-Cramér</i>	98
5.3	<i>Comportamento asintotico della distribuzione binomiale</i>	98
5.4	<i>Il teorema del limite centrale</i>	99
6	La legge dei grandi numeri	103
6.1	<i>La legge debole dei grandi numeri</i>	103
6.2	<i>Convergenza quasi certa, convergenza debole e convergenza in legge</i>	103
6.3	<i>Il teorema di Markov, i suoi corollari e il teorema di Khinchin</i>	105
6.4	<i>La legge forte dei grandi numeri</i>	109
	Bibliografia	120

Capitolo 1

Eventi e probabilità

1.1 *Eventi*

Nel Calcolo delle probabilità si opera su valutazioni numeriche normalizzate (probabilità) di **eventi**. Una famiglia di eventi $\mathcal{E} = \{E_\alpha, E_\beta, E_\gamma, \dots\}$, ai quali verrà assegnata una probabilità, si può pensare come un modello astratto di una famiglia di *risultati* possibili di un *esperimento* \mathcal{S} , essendo la realizzazione effettiva di ciascun risultato soggetta a una misura di incertezza, e quindi “dipendente dal caso”.

Ad una famiglia di eventi si può far corrispondere una famiglia $\mathcal{F} = \{f_\alpha, f_\beta, f_\gamma, \dots\}$ di *proposizioni* che li descrivono e che possono risultare vere o false al termine dell’esperimento \mathcal{S} , “secondo il caso”.

È opportuno studiare famiglie sufficientemente ampie, precisamente, considerando proposizioni che descrivono eventi, tali che valga almeno l’implicazione

$$f, g \in \mathcal{F} \Rightarrow \neg f, f \vee g \text{ (e quindi } f \wedge g \dots) \in \mathcal{F}.$$

I connettivi proposizionali \neg, \wedge, \vee indicano come al solito la negazione, la congiunzione, la disgiunzione (non esclusiva). Più interessante è il caso in cui connettendo un’infinità numerabile di proposizioni si ottengano ancora proposizioni della famiglia, ad esempio:

$$f_1, f_2, f_3, \dots \in \mathcal{F} \Rightarrow f_1 \vee f_2 \vee f_3 \vee \dots \in \mathcal{F}.$$

Le operazioni *f AND g* e *f XOR g* determinano in \mathcal{F} una *struttura di algebra di Boole* (σ -*algebra*, nel caso di ammissibilità di un’infinità numerabile di operazioni). *XOR* funge da somma e *AND* da prodotto, e *f AND f = f*.

(*f AND g* è vera se e solo se *f* e *g* sono entrambe vere, mentre *f XOR g* è vera se e solo se una sola delle due proposizioni *f* e *g* è vera. *f AND g* ha lo stesso significato di $f \wedge g$).

Ricordiamo che \mathcal{A} è un’*algebra di Boole* se e solo se è un’*algebra* e ogni elemento è *idempotente* : $a \in \mathcal{A}$ implica $a^2 = a$. Ne segue che ogni elemento coincide con il proprio opposto

($a = -a$) e che \mathcal{A} risulta necessariamente commutativa ($ab = ba$ per ogni coppia a, b).

Spesso le proposizioni che si considerano, o i risultati dell' esperimento che esse esprimono, sono formulate in termini di *appartenenza di un elemento ω , che codifica gli esiti "elementari dell'esperimento \mathcal{S} , ad un insieme A , sottoinsieme di un insieme di riferimento o "sample space Ω* :

$$f = \text{" } \omega \in A \text{"}.$$

Gli eventi sono allora rappresentati da una famiglia $\mathcal{A} = \{A_\alpha, A_\beta, A_\gamma, \dots\}$ di sottoinsiemi di Ω . È sempre possibile *rappresentare* fedelmente una σ -algebra di Boole con una σ -algebra di sottoinsiemi di un opportuno insieme Ω (*M.H. Stone, 1936*).

La moderna teoria del Calcolo delle probabilità, che si può assumere fondata con la memoria di *A.Kolmogorov* del 1933, ricorre sistematicamente a tale rappresentazione.

Qualche esempio di corrispondenza, indicata con " \rightarrow ", tra proposizioni composte mediante connettivi proposizionali e relativi insiemi costruiti mediante corrispondenti operazioni insiemistiche è fornita dalle formule seguenti. Se

$$f \rightarrow A, g \rightarrow B, f_k \rightarrow A_k,$$

allora:

$$\neg f \rightarrow A^c \text{ (oppure } -A, \text{ oppure } \overline{A})$$

$$f \wedge g \rightarrow A \cap B$$

$$f \text{ XOR } g \rightarrow A \triangle B = (A - B) \cup (B - A)$$

$$f \vee g \rightarrow A \cup B, f_1 \vee f_2 \vee f_3 \vee \dots \rightarrow A_1 \cup A_2 \cup A_3 \cup \dots = \bigcup_{n=1}^{+\infty} A_n$$

1.2 σ -algre di insiemi e spazi misurabili

Sia Ω un insieme non vuoto ed \mathcal{A} una famiglia non vuota di sottoinsiemi di Ω ($\mathcal{A} \subseteq \mathcal{P}(\Omega)$). Si dice che \mathcal{A} è una σ -**algebra** (di sottoinsiemi di Ω) e (Ω, \mathcal{A}) è uno spazio misurabile se valgono le proprietà seguenti:

- 1) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- 2) $A_n \in \mathcal{A}$ per $n = 1, 2, 3, \dots \Rightarrow \bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$.

Ne segue che se A, B, A_n sono elementi di \mathcal{A} , anche

$$\Omega = A \cup A^c, A \cap B = (A^c \cup B^c)^c, \emptyset = \Omega^c,$$

$$A - B, A \triangle B, \bigcap_{n=1}^{+\infty} A_n \text{ sono elementi di } \mathcal{A}.$$

In generale, un'infinità numerabile di operazioni insiemistiche su elementi della σ -algebra produce un insieme ancora appartenente alla σ -algebra.

Se in 2) ci si limita ad unioni finite, e quindi si pretende soltanto che \mathcal{A} sia stabile quando si esegue un numero finito di operazioni insiemistiche, si dice che \mathcal{A} è un'**algebra**.

1.3 Esempi elementari

Si osservi che se Ω è finito ogni algebra è una σ -algebra.

1) Esperimento \mathcal{S} : due tentativi, per ciascuno di essi due possibilità s, f (successo, fallimento).

$$\Omega = \{ss, sf, fs, ff\}$$

$$\mathcal{A}_0 = \{\emptyset, \Omega\} \text{ algebra banale;}$$

$$\mathcal{A}_1 = \{\emptyset, \{ss, sf\}, \{fs, ff\}, \Omega\} :$$

se $A \in \mathcal{A}_1$ ed $\omega \in \Omega$, l'affermazione $\omega \in A$ non fornisce informazioni sul risultato del secondo tentativo;

$$\mathcal{A}_2 = \{\emptyset, \{ss, fs\}, \{sf, ff\}, \Omega\} :$$

se $A \in \mathcal{A}_2$ ed $\omega \in \Omega$, l'affermazione $\omega \in A$ non fornisce informazioni sul risultato del primo tentativo;

$$\mathcal{A}_3 = \{\emptyset, \{ss\}, \{sf, fs\}, \{ff\}, \{ss, ff\}, \{ss, sf, fs\}, \{sf, fs, ff\}, \Omega\} :$$

se $A \in \mathcal{A}_3$ ed $\omega \in \Omega$, l'affermazione $\omega \in A$ fornisce informazioni sul numero di risultati, non sul fatto che essi siano stati ottenuti nel primo o nel secondo tentativo;

2) Esperimento \mathcal{S} : successione infinita di tentativi ripetuti

$$\omega : \mathbf{N}^+ \longrightarrow \{s, f\}$$

[esempio $\omega = (s, f, f, s, s, s, f, \dots)$, ω_n vale s oppure f].

Si può prendere provvisoriamente $\mathcal{A} = \mathcal{P}(\Omega)$. Purtroppo, nella maggior parte dei casi importanti, non sarà possibile assegnare in modo coerente una probabilità ad ogni sottoinsieme di Ω . Converrà allora limitarsi a σ -algebre più piccole, ma tuttavia sufficientemente ampie perchè in esse siano presenti tutti i sottoinsiemi C (detti cilindrici) della forma

$$C = \{ \omega \in \Omega \mid \omega_{k_1} = x_1, \omega_{k_2} = x_2, \dots, \omega_{k_n} = x_n \},$$

dove n è un intero (finito) arbitrario e arbitrari sono sia gli indici $k_1 < k_2 \dots < k_n$ che i valori $x_1, x_2, \dots, x_n \in \{s, f\}$.

L'insieme C è quindi formato da tutte le successioni nelle quali è prescritto il risultato (successo o fallimento) in un numero finito di prove (tentativi) esattamente individuate, mentre il risultato nelle altre prove della successione può essere qualunque.

3) Esperimento \mathcal{S} : selezionare (colpire con un proiettile puntuale) un punto di un disco D (bersaglio) di centro O e raggio R . Si può prendere

$$\Omega = D \text{ ed } \mathcal{A} = \mathcal{L}(\Omega),$$

insieme di tutti i sottoinsiemi A di D dotati di area (nel senso di Lebesgue). Se x è il punto colpito, interessano le proposizioni del tipo $x \in A$.

(Anche in questo caso si vorrebbe $\mathcal{A} = \mathcal{P}(\Omega)$, ma allora non sarebbe possibile assegnare in modo utile e coerente una probabilità ad ogni evento).

1.4 Probabilità (misura di probabilità)

Una **misura di probabilità** P è una valutazione degli eventi con un indice numerico, convenzionalmente compreso tra 0 e 1, dunque un'applicazione

$$P: \mathcal{A} \longrightarrow [0, 1],$$

normalizzata e numerabilmente additiva, cioè tale che

$$P(\Omega) = 1$$

e per ogni successione A_k di eventi disgiunti ($A_i \cap A_j = \emptyset$)

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

[Conviene richiedere l'additività numerabile, invece dell'additività semplice (cioè concernente soltanto unioni finite).]

La probabilità di un evento A , $P(A)$, è un indice quantitativo del “grado di certezza dell'evento (casuale).

Una stessa σ -algebra può essere utilizzata per descrivere esperimenti S differenti ed in essa si possono introdurre diverse misure di probabilità.

Un esperimento S , con risultati “dipendenti dal caso, è descritto da uno **spazio di probabilità** o *spazio di misura con misura normalizzata ad 1*, (Ω, \mathcal{A}, P) , dove \mathcal{A} è una σ -algebra in Ω e P una misura di probabilità definita su \mathcal{A} .

L'esperimento ideale S è un modello (probabilistico) di un esperimento reale (o di più esperimenti reali).

Nella scelta del modello astratto, ed in particolare nella scelta della misura di probabilità, si pongono i seguenti problemi:

- come assegnare le probabilità, ovvero *costruire* la misura P , sulla base di (poche) ipotesi elementari, in modo razionale (coerente) ?
- quale *utilità, efficacia*, nell'analizzare l'esperimento e stimarne i risultati, ha il modello probabilistico?

Tali problemi sono strettamente connessi con

- l'*interpretazione* pratica del concetto di probabilità, che a sua volta può suggerire
- metodi di stime* delle probabilità ideali.

1.5 Esempi di spazi di probabilità

Negli esempi che seguono, tranne l'esempio 3), lo spazio di probabilità Ω è finito e a tutti gli eventi elementari $\{\omega\}$, costituiti da un singolo elemento, si assegna la stessa probabilità. Ne segue che la probabilità di un evento A è uguale al rapporto tra il numero di elementi di A (*casi favorevoli*) e il numero complessivo di elementi di Ω (*casi possibili*). Si parla allora di *probabilità classica* e valutazione delle probabilità si riduce a questioni di cardinalità, che talvolta

non sono semplici e richiedono risultati di calcolo combinatorio.

1) *Roulette (onesta)*: lo 0 è verde, degli altri numeri 18 sono neri e 18 rossi; assumiamo che la pallina possa fermarsi con la stessa probabilità su ogni numero.

$$\Omega = \{0, 1, 2, 3, \dots, 36\}, \mathcal{A} = \mathcal{P}(\Omega), P(\{n\}) = p_n = p,$$

dunque per l'additività:

$$P(\Omega) = \sum_{n=0}^{36} p_n = 37p = 1, \quad p = \frac{1}{37},$$

$$P(A) = \sum_{n \in A} p_n = p \operatorname{card}(A)$$

Così l'evento $R = \{n \in \Omega / n \text{ è "rosso"}\}$ ha probabilità $18/37$.

2) *Due dadi (non truccati)*:

$$\Omega = \{(1, 1), (1, 2), \dots, (2, 1), (2, 2), \dots, (6, 5), (6, 6)\}, \mathcal{A} = \mathcal{P}(\Omega),$$

$$P(\{(m, n)\}) = p_{m,n} = p = \frac{1}{36}.$$

Così l'evento $A = \{(m, n) \in \Omega / m + n = 4\} = \{(1, 3), (2, 2), (3, 1)\}$ ha probabilità $p \operatorname{card}(A) = 3/36 = 1/12$.

3) *Bersaglio circolare*:

$$\Omega = \{X = (x, y) \in \mathbf{R}^2 / x^2 + y^2 \leq R^2\}, \mathcal{A} = \mathcal{L}(\Omega),$$

dunque, supponendo tutti i punti del disco equivalenti ed essendoci in Ω infiniti punti:

$$P(\{X\}) = 0 \text{ e } P(A) = \operatorname{Cost} \operatorname{area}(A), \quad P(\Omega) = \operatorname{Cost} \pi R^2,$$

dunque $\operatorname{Cost} = 1/(\pi R^2)$ e per

$$A = \{X = (x, y) \in \mathbf{R}^2 / x^2 + y^2 \leq r^2\} : P(A) = \frac{\pi r^2}{\pi R^2} = \left(\frac{r}{R}\right)^2.$$

4) *Frequenza di caratteri in una popolazione*:

Ω (*popolazione*) è costituito da un insieme finito di individui; C_1, C_2, \dots, C_l sono *strati* della popolazione (sottoinsiemi disgiunti, la cui unione è Ω , degli individui che possiedono i *caratteri* $1, 2, \dots, l$).

Supponendo di estrarre "a caso un individuo ω della popolazione (tutti gli individui hanno la stessa probabilità di essere estratti), la probabilità che si presenti il carattere k è:

$$P(C_k) = \frac{\operatorname{card}(C_k)}{\operatorname{card}(\Omega)}.$$

5) *Configurazioni in meccanica statistica:*

A) *Statistica di Maxwell-Boltzmann:* n sistemi “distinguibili, ciascuno dei quali può trovarsi in uno qualunque di l stati, indipendentemente dallo stato degli altri sistemi (oppure: n particelle distinguibili che possono assumere l livelli di energia “senza esclusione). Le configurazioni sono applicazioni ω che ad ogni sistema (diciamo il sistema numero j) associano lo stato (diciamo lo stato numero s_j) corrispondente:

$$\omega = (s_1, s_2, \dots, s_n), \text{ dove } 1 \leq s_j \leq l,$$

dunque

$$\text{card}(\Omega) = l^n.$$

Siano $K_1(\omega), K_2(\omega), \dots, K_l(\omega)$ i numeri di occupazione dei vari stati: $K_1(\omega) + K_2(\omega) + \dots + K_l(\omega) = n$. Consideriamo, selezionando “a caso una configurazione qualunque, l’evento

$$A = \{\omega \in \Omega \mid K_1(\omega) = k_1, K_2(\omega) = k_2, \dots, K_l(\omega) = k_l\},$$

dove $k_1 + k_2 + \dots + k_l = n$. Allora

$$P(A) = \frac{n!}{k_1! k_2! \dots k_l!} / l^n,$$

infatti si pensi a n posti (l’ordine con il quale li esaminiamo è indifferente), k_1 dei quali devono essere etichettati con il carattere l_1 (primo stato o livello), k_2 dei quali con $l_2 \dots k_l$ con l_l . Dobbiamo selezionare k_1 posti tra gli n disponibili, quindi k_2 posti tra gli $n - k_1$ rimasti disponibili ..., e dunque abbiamo

$$\binom{n}{k_1} \binom{n - k_1}{k_2} \dots \binom{n - k_1 - \dots - k_{l-2}}{k_{l-1}} = \frac{n!}{k_1! k_2! \dots k_l!}.$$

Si osservi che

$$\sum_{|k|=n} \frac{n!}{k_1! k_2! \dots k_l!} 1^{k_1} 1^{k_2} \dots 1^{k_l} = (1 + 1 + \dots + 1)^n = l^n.$$

B) *Statistica classica con esclusione:* n sistemi “distinguibili l stati, con $l \geq n$, in ciascuno dei quali può trovarsi al più un sistema (oppure: n particelle distinguibili che possono assumere l livelli di energia “con esclusione). Le configurazioni sono applicazioni *iniettive* ω che ad ogni sistema (diciamo il sistema numero j) associano lo stato (diciamo lo stato numero s_j) corrispondente:

$$\omega = (s_1, s_2, \dots, s_n), \text{ dove } 1 \leq s_j \leq l, s_i \neq s_j \text{ se } i \neq j.$$

dunque

$$\text{card}(\Omega) = l(l-1)(l-2)\dots(l-n+1) = \frac{l!}{(l-n)!}.$$

Siano $K_1(\omega), K_2(\omega), \dots, K_l(\omega)$ i numeri di occupazione dei vari stati: $K_1(\omega) + K_2(\omega) + \dots + K_l(\omega) = n$ e ogni $K(\omega) = 0$ oppure 1. Consideriamo, selezionando “a caso una configurazione qualunque, l’evento

$$A = \{\omega \in \Omega | K_1(\omega) = k_1, K_2(\omega) = k_2, \dots, K_l(\omega) = k_l\},$$

dove le k_j sono assegnate e tali che: $k_1 + k_2 + \dots + k_l = n$ e $k_j = 0$ oppure 1. Dunque le $k_j = 1$ sono esattamente n e individuano una selezione di n stati tra gli l possibili, allora le $\omega \in A$ descrivono la disposizione degli n sistemi negli n stati selezionati. Pertanto

$$\text{card}(A) = n! \quad \text{e quindi} \quad P(A) = \frac{n!(l-n)!}{l!},$$

ovvero

$$P(A) = \frac{1}{\binom{l}{n}}.$$

C) *Statistica di Fermi-Dirac*: n sistemi “indistinguibili ed l stati, con $l \geq n$, in ciascuno dei quali può trovarsi al più un sistema (oppure: n particelle indistinguibili che possono assumere l livelli di energia “con esclusione). Le configurazioni sono classi di equivalenza di applicazioni *iniettive* ω che ad ogni sistema (diciamo il sistema numero j) associano lo stato (diciamo lo stato numero s_j) corrispondente. Sono equivalenti applicazioni con la stessa immagine (i sistemi sono indistinguibili) e sono quindi univocamente descritti dall’elenco degli stati occupati, che può essere fatto assegnando un codice k_j , 1 oppure 0 rispettivamente se lo stato j è occupato o non occupato:

$$\begin{aligned} \omega &= [(s_1, s_2, \dots, s_n), \text{ dove } 1 \leq s_j \leq l, s_i \neq s_j \text{ se } i \neq j] = \\ &= (k_1, k_2, \dots, k_l) \text{ dove } k_j = 0 \text{ oppure } 1, k_1 + k_2 + \dots + k_l = n. \end{aligned}$$

dunque

$$\text{card}(\Omega) = \frac{l!}{n!(l-n)!},$$

numero di sottoinsiemi di n elementi scelti tra l possibili, ovvero $\binom{l}{n}$.

Se l’evento A è definito come nel caso precedente, risulta $\text{card}(A) = 1$ e quindi

$$P(A) = \frac{1}{\binom{l}{n}}.$$

D) *Statistica di Bose-Einstein*: n sistemi “indistinguibili, ciascuno dei quali può trovarsi in uno qualunque di l stati (oppure: n particelle indistinguibili che possono assumere l livelli di energia “senza esclusione). Le configurazioni sono classi di equivalenza di applicazioni, *non necessariamente iniettive*, ω che ad ogni sistema (diciamo il sistema numero j) associano lo stato (diciamo

lo stato numero s_j) corrispondente. Sono equivalenti applicazioni con lo stesso numero k_i di sistemi nel medesimo stato i (i sistemi sono indistinguibili):

$$\begin{aligned}\omega &= [(s_1, s_2, \dots, s_n), \text{ dove } 1 \leq s_j \leq l] = \\ &= (k_1, k_2, \dots, k_l) \text{ dove } k_i \geq 0, k_1 + k_2 + \dots + k_l = n.\end{aligned}$$

Le ω si possono mettere in corrispondenza biunivoca con le scritture del tipo

$$pptptpppttptp,$$

che in questo caso ($n = 8$ e $l = 6$) indica 2 particelle nel primo livello, 1 nel secondo, 3 nel terzo, 0 nel quarto, 1 nel quinto e 1 nel sesto. Sono scritture che debbono essere formate da $l + n - 1$ lettere, delle quali $l - 1$ sono la lettera t (tagli tra livelli) e n sono la lettera p . Queste scritture sono tante quanti i sottoinsiemi di $l + n - 1$ elementi (posti) formati da n elementi (lettere p). Dunque

$$\text{card}(\Omega) = \frac{(l + n - 1)!}{n!(l - 1)!},$$

ovvero $\binom{l + n - 1}{n}$.

Se l'evento A è definito come nel caso precedente, risulta $\text{card}(A) = 1$ e quindi

$$P(A) = \frac{1}{\binom{l + n - 1}{n}}.$$

1.6 Probabilità e frequenza

Supponiamo di ripetere un esperimento S n volte: S_1, S_2, S_3, \dots , evitando che risultati precedenti influenzino esperimenti successivi. Si consideri un evento A di probabilità $P(A) = p$; sia F_n (frequenza) il numero di volte che A si verifica nel corso delle n prove indipendenti ed $f_n = F_n/n$ la frequenza relativa.

Possiamo definire un esperimento unico $\Sigma^{(n)} = S_1 \times S_2 \times \dots \times S_n$, che consiste nel ripetere indipendentemente n volte S , e possiamo coerentemente costruire una misura di probabilità $P^{(n)}$ associata a tale esperimento., come si comprenderà meglio in seguito, dopo aver precisato il concetto di indipendenza.

Si può dimostrare che per ogni $\delta > 0$

$$P^{(n)}(|f_n - p| \geq \delta) \leq \frac{p(1-p)}{n\delta^2},$$

che risulta minore di qualunque $\varepsilon > 0$ assegnato, purchè n sia sufficientemente grande, cioè maggiore di un opportuno N dipendente da ε .

Dunque la frequenza relativa “converge in probabilità alla probabilità ideale p (caso particolare della legge debole dei grandi numeri).

In un certo senso un grado di fiducia p per un evento A implica al crescere di n un “sempre maggior grado di fiducia che lo scarto tra la frequenza relativa e la probabilità ideale non superi un livello prefissato δ (comunque piccolo).

1.7 Alcune regole di calcolo delle probabilità

1) $P(\emptyset) = 0$,

infatti $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$.

2) $P(A^c) = 1 - P(A)$,

perchè $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$ e $P(\Omega) = 1$.

3) Se $B \subseteq A$, $B, A \in \mathcal{A}$, allora $P(B) \leq P(A)$ (monotonia).

infatti $A = B \cup (A - B)$ e dunque, essendo B e $A - B$ disgiunti, risulta $P(A) = P(B) + P(A - B) \geq P(B)$. Si ha inoltre $P(B - A) = P(B) - P(A)$.

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

infatti $A \cup B = A \cup (B - A)$, $B = (B - A) \cup (A \cap B)$, essendo le unioni disgiunte, e dunque $P(A \cup B) = P(A) + P(B - A)$, mentre $P(B - A) + P(A \cap B) = P(B)$.

Si osservi che $P(B) = P(B \cap A) + P(B \cap A^c)$ e quindi per esempio $P(B \cap A^c) = P(B) - P(B \cap A)$.

5) Se per $n = 1, 2, \dots$ $A_n \subseteq A_{n+1}$, allora (continuità)

$$P\left(\bigcup_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n),$$

infatti, posto $A_0 = \emptyset$, $A = \bigcup_n A_n = \bigcup_n (A_n - A_{n-1})$, e poichè l'ultima unione è disgiunta, in virtù delle σ -additività:

$$P(A) = \sum_{k=1}^{+\infty} P(A_k - A_{k-1}) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(A_k - A_{k-1}) = \lim_{n \rightarrow +\infty} P(A_n).$$

6) Similmente se per $n = 1, 2, \dots$ $A_n \supseteq A_{n+1}$, allora

$$P\left(\bigcap_n A_n\right) = \lim_{n \rightarrow +\infty} P(A_n),$$

infatti $A_n^c \subseteq A_{n+1}^c$, $\bigcup_n A_n^c = (\bigcap_n A_n)^c$, dunque

$$P\left(\bigcap_n A_n\right) = 1 - P\left(\bigcup_n A_n^c\right) = 1 - \lim_n P(A_n^c) = \lim_n (1 - P(A_n^c)) = \lim_n P(A_n).$$

1.8 Probabilità condizionale e indipendenza

Introduciamo due concetti che avranno un ruolo essenziale negli sviluppi successivi.

Definizione. Siano $A, B \in \mathcal{A}$ e $P(B) \neq 0$: si dice *probabilità condizionale di A dato B* il rapporto

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

Ne segue che $P(A \cap B) = P(A|B)P(B)$.

L'uguaglianza precedente sarebbe ancora vera con $P(B) = 0$, se fosse definito (comunque) $P(A|B)$.

È facile dimostrare il seguente

Teorema. $(\Omega, \mathcal{A}, P(\cdot|B))$ è uno spazio di probabilità. Inoltre, posto $\mathcal{A}_B = \{A \cap B \mid A \in \mathcal{A}\}$ e $P_B(E) = P(E|B)$ per $E \in \mathcal{A}_B$, (B, \mathcal{A}_B, P_B) è uno spazio di probabilità.

Infatti \mathcal{A}_B è una σ -algebra e, se $A = A_1 \cup A_2 \cup \dots$ con A_1, A_2, \dots disgiunti, si avrà $A \cap B = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots$ con $(A_1 \cap B), (A_2 \cap B), \dots$ disgiunti, e quindi

$$\frac{P(A \cap B)}{P(B)} = \sum_k \frac{P(A_k \cap B)}{P(B)},$$

vale a dire

$$P(A|B) = \sum_k P(A_k|B).$$

Ovviamente $P(\Omega|B) = P(B|B) = 1$.

Osservazione. Nel teorema precedente B è fisso. Se al contrario si facesse variare B , tenendo fisso A , non si avrebbe una misura. Infatti l'applicazione $B \rightarrow P(A|B)$ in generale non è additiva. Sia ad esempio

$$B = B_1 \cup B_2, \quad B_1 \cap B_2 = \emptyset, \quad P(B_1) > 0, \quad P(B_2) > 0.$$

Allora

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B_1) + P(A \cap B_2)}{P(B)} = \\ &= P(A|B_1) \frac{P(B_1)}{P(B)} + P(A|B_2) \frac{P(B_2)}{P(B)} < P(A|B_1) + P(A|B_2). \end{aligned}$$

Regola della probabilità totale. Sia B_1, B_2, \dots, B_n una partizione di Ω , allora

$$P(A) = \sum_{k=1}^n P(A|B_k)P(B_k).$$

Infatti $A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)$ è una unione disgiunta e quindi

$$P(A) = \sum_k P(A \cap B_k).$$

La regola vale anche per partizioni B_k numerabili.

Proposizione.

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) .$$

Infatti

$$P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} .$$

La proposizione si generalizza facilmente al caso di n eventi A_1, \dots, A_n .

Teorema di Bayes. Sia C_1, C_2, \dots, C_n una partizione di Ω e $P(D) \neq 0$, allora

$$P(C_k|D) = \frac{P(D|C_k)P(C_k)}{P(D)} = \frac{P(D|C_k)P(C_k)}{\sum_j P(D|C_j)P(C_j)} , \quad k = 1, 2, \dots, n .$$

Infatti il secondo denominatore è $P(D) \neq 0$, per la regola delle probabilità totali, e

$$P(C_k|D)P(D) = P(C_k \cap D) = P(D|C_k)P(C_k) .$$

Lo stesso risultato vale per partizioni numerabili.

Indipendenza di due eventi. Due eventi A e B si dicono indipendenti se

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

Dunque se ad esempio $P(B) \neq 0$ risulta

$$P(A|B) = P(A) .$$

(L'informazione che l'evento B si è verificato non cambia la probabilità che si verifichi l'evento A .)

Osservazione 1. Se A e B sono indipendenti, anche A^c e B , oppure A e B^c , sono indipendenti. Pertanto A^c e B^c sono indipendenti.

Infatti

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(B)P(A) = (1 - P(A))P(B) = P(A^c)P(B) .$$

Osservazione 2. Se A è indipendente sia da B che da C , allora

i) se $B \supseteq C$ A è indipendente da $B - C$;

ii) se B e C sono disgiunti ($B \cap C = \emptyset$) allora A è indipendente da $B \cup C$. Infatti

$$P(A \cap (B - C)) = P(A \cap B) - P(A \cap C) = P(A)P(B) - P(A)P(C) = P(A)(P(B) - P(C)) ;$$

$$P(A \cap (B \cup C)) = P(A \cap B) + P(A \cap C) = P(A)P(B) + P(A)P(C) = P(A)(P(B \cup C)) .$$

Eventi mutuamente indipendenti. Gli eventi A_1, A_2, \dots, A_n si dicono **mutuamente (congiuntamente) indipendenti** (o semplicemente **indipendenti**, se il contesto non lascia dubbi) se e solo se per ogni $1 \leq k \leq n$ e per ogni scelta di indici $1 \leq j_1 < j_2 < \dots < j_k \leq n$ si ha

$$P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = P(A_{j_1})P(A_{j_2})\dots P(A_{j_k}) .$$

Ovviamente eventi indipendenti sono **a due a due indipendenti**. In generale non vale l'implicazione inversa. Per rendersene conto si consideri il seguente semplice esempio:

Un'urna contiene quattro biglietti, sui quali compaiono rispettivamente i codici 1,2,3 e 123. Si estrae a caso un biglietto (tutti i biglietti hanno la stessa probabilità di essere estratti). Sia A_j l'evento "è comparsa la cifra j ". Allora $P(A_j) = 1/2$, $P(A_j \cap A_k) = 1/4$ per $j \neq k$ e gli eventi A_1, A_2, A_3 sono dunque a due a due indipendenti. Ma essi non sono mutuamente indipendenti perchè $P(A_1 \cap A_2 \cap A_3) = 1/4$ e non $1/8$.

Più in generale gli eventi di un insieme infinito \mathcal{I} si dicono mutuamente indipendenti se e solo se ogni sottoinsieme finito di \mathcal{I} è costituito da eventi mutuamente indipendenti.

Osservazione 1. Se A_1, A_2, \dots, A_n sono mutuamente indipendenti, anche A_1^c, A_2, \dots, A_n sono mutuamente indipendenti.

Infatti

$$P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = P(A_{j_1})P(A_{j_2})\dots P(A_{j_k})$$

se nessuna delle $j_s = 1$. Altrimenti sia ad esempio $j_1 = 1$ e poniamo $B = A_{j_2} \cap \dots \cap A_{j_k}$. Essendo A_1 e B indipendenti, tali sono anche A_1^c e B , dunque

$$P(A_1^c \cap A_{j_2} \cap \dots \cap A_{j_k}) = P(A_1^c \cap B) = P(A_1^c)P(B) = P(A_1^c)P(A_{j_2})\dots P(A_{j_k}) .$$

Più in generale se nella sequenza A_1, A_2, \dots, A_n si sostituisce un numero arbitrario di eventi con i loro complementari si ottiene ancora una sequenza di eventi mutuamente indipendenti.

Osservazione 2. Se A_1, A_2, \dots, A_n sono mutuamente indipendenti, uno qualunque di essi è indipendente da ogni evento che si può ottenere mediante usuali operazioni insiemistiche sugli altri.

Limitiamoci ad un esempio: siano A, B, C , mutuamente indipendenti, allora A è indipendente da $B \cap C$ e da $B \cup C$. Infatti, esplicitando in questo caso particolare il calcolo sottointeso nell'osservazione precedente, quando si affermava che A_1 e B erano indipendenti,

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

e, essendo A, B^c, C^c indipendenti, A è indipendente da $B^c \cap C^c$. Ma allora A è indipendente da $(B^c \cap C^c)^c$, cioè da $B \cup C$.

Se le A_j sono a due a due indipendenti, ma non mutuamente indipendenti, il risultato ingenerale non vale: nell'esempio sopra riportato dell'urna con quattro biglietti (1, 2, 3, 123) si ha

$$P(A_1 \cap (A_2 \cup A_3)) = \frac{1}{4} \neq P(A_1)P(A_2 \cup A_3) = \frac{1}{2} \cdot \frac{3}{4} .$$

Dunque la famiglia \mathcal{I} degli eventi indipendenti da A non è in generale una σ -algebra, mentre si vede che, se A, B_1, B_2, \dots sono mutuamente indipendenti, allora tutti gli eventi della più piccola

σ -algebra alla quale appartengono B_1, B_2, \dots (o come si dice la σ -algebra generata da B_1, B_2, \dots) sono indipendenti da A .

Indipendenza condizionata. A e B si dicono indipendenti dato C , o condizionatamente indipendenti (con C evento condizionante), se e solo se

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

Più in generale, gli eventi A_1, A_2, \dots, A_n sono condizionatamente indipendenti dato C se e solo se la probabilità condizionata dell'intersezione di un numero arbitrario di essi è il prodotto delle probabilità condizionate degli eventi dei quali si fa l'intersezione.

Essendo P e P_C , $P_C(A \cap C) = P(A | C)$, misure di probabilità su Ω e su C , per l'indipendenza condizionata valgono le stesse osservazioni presentate sopra per l'indipendenza (non condizionata).

Ad esempio, se A e B sono condizionatamente indipendenti dato C , allora A^c e B sono condizionatamente indipendenti dato C . Basta osservare che $A \cap C$ e $B \cap C$ sono indipendenti in C e $A^c \cap C$ è il complementare in C di $A \cap C$ e dunque indipendente in C da $B \cap C$:

$$P(A^c \cap B | C) = P_C((A^c \cap C) \cap (B \cap C)) = P_C(A^c \cap C)P_C(B \cap C) = P(A^c | C)P(B | C) .$$

Osservazione. *Eventi indipendenti possono non essere condizionatamente indipendenti ed eventi non indipendenti possono essere condizionatamente indipendenti.*

Si considerino i seguenti semplici esempi, nei quali avremo sempre

$$\Omega = \{ \omega_{i,j} \}_{i,j=1\dots 6} , P(\omega_{i,j}) = \frac{1}{36} \text{ per ogni } i \text{ e } j$$

(lancio di due dadi indipendenti non truccati).

1) Siano

$$A = \{ \omega_{i,j} \mid i = 1 \vee (i > 3 \wedge j = 6) \} , B = \{ \omega_{i,j} \mid (j = 1 \wedge i \leq 3) \vee i = 6 \} , \\ C = \{ \omega_{i,j} \mid i \leq 3 \} , \bar{C} = \{ \omega_{i,j} \mid i > 3 \} .$$

È facile verificare che

$$P(A \cap B) \neq P(A)P(B) , \text{ ma} \\ P(A \cap B | C) = P(A | C)P(B | C) \text{ e } P(A \cap B | \bar{C}) = P(A | \bar{C})P(B | \bar{C}) ,$$

cioè A e B non sono indipendenti, ma lo sono condizionatamente dato C e anche dato \bar{C} .

2) Se invece si pone

$$A = \{ \omega_{i,j} \mid i = 1 \} , B = \{ \omega_{i,j} \mid j = 1 \} , \\ C = \{ \omega_{i,j} \mid \max(i, j) \leq 2 \} ,$$

si vede che A e B sono indipendenti, sono condizionatamente indipendenti dato C , ma non sono condizionatamente indipendenti dato \bar{C} .

3) Se prendiamo

$$C = \{ \omega_{i,j} \mid (i \leq 2 \wedge (j \leq 2 \vee j \geq 5)) \vee (i \geq 5 \wedge j \leq 2) \} ,$$

allora, restando ovviamente A e B indipendenti, essi non sono condizionatamente indipendenti né dato C né dato \overline{C} .

4) Se

$$A = \{ \omega_{i,j} \mid i = 1 \} , B = \{ \omega_{i,j} \mid j \leq 3 \vee j = 6 \} , C = \{ \omega_{i,j} \mid j \leq 4 \} ,$$

allora

$$P(A|C) = P(A|\overline{C}) = P(A) = \frac{1}{6} , P(B) = \frac{2}{3} , P(B|C) = \frac{3}{4} , P(B|\overline{C}) = \frac{1}{2}$$

e quindi A e B sono indipendenti e condizionatamente indipendenti sia dato C sia dato \overline{C} .

In generale, se A e B sono indipendenti e condizionatamente indipendenti, sia dato C , sia dato \overline{C} , purché $0 < P(C) < 1$, allora uno almeno dei due eventi A e B deve essere indipendente da C . Infatti, posto

$$P(A|C) = p_1 , P(A|\overline{C}) = p_2 , P(B|C) = q_1 , P(B|\overline{C}) = q_2 , P(C) = \alpha , \beta = 1 - \alpha ,$$

si trova

$$\begin{aligned} f &= P(A)P(B) = (p_1\alpha + p_2\beta)(q_1\alpha + q_2\beta) , \\ g &= P(A \cap B) = P(A|C)P(B|C)P(C) + P(A|\overline{C})P(B|\overline{C})P(\overline{C}) = p_1q_1\alpha + p_2q_2\beta \\ g - f &= \alpha\beta(p_1 - p_2)(q_1 - q_2) . \end{aligned}$$

Dunque, se $0 < \alpha < 1$, $g = f$ se e solo se $p_1 = p_2$ oppure $q_1 = q_2$. Nel primo caso $P(A) = p_1 = P(A|C)$, cioè A e C sono indipendenti, e nel secondo $P(B) = q_1 = P(B|C)$, cioè B e C sono indipendenti.

1.9 Esempi di applicazione delle regole elementari di calcolo

Presentiamo alcuni problemi che si propongono di illustrare il ruolo delle regole elementari di calcolo.

1) Due urne contengono rispettivamente n_1 ed n_2 palline, delle quali k_1 e k_2 bianche, essendo le rimanenti nere. Si considerano le seguenti due procedure:

i) Si sceglie a caso una delle due urne; da essa si estrae una pallina e si registra se essa è bianca. Si rimette la pallina nell'urna. Si ripetono una seconda volta tutte le operazioni precedenti.

ii) Si sceglie a caso una delle due urne; da essa si estrae una pallina e si registra se essa è bianca. Si rimette la pallina nell'urna. Dalla stessa urna si estrae una seconda volta una pallina e si registra se essa è bianca.

In ogni caso si vince se sono state registrate due estrazioni di pallina bianca. Quale delle due procedure è più conveniente?

Denotiamo con 1 e 2 gli eventi “selezione dell'urna numero 1 oppure 2”, che sono scelte “a caso ovvero ciascuna con probabilità 0.5. Siano $p_1 = k_1/n_1$ e $p_2 = k_2/n_2$ le probabilità di estrazione di una pallina bianca dalle due urne. Indichiamo con B l'evento “registrazione di una pallina bianca in una estrazione e BB l'evento “registrazione di una pallina bianca in entrambe le estrazioni:

i) con la prima procedura abbiamo

$$P(BB) = (P(B|1)P(1) + P(B|2)P(2))^2 = (0.5(p_1 + p_2))^2 = 0.25(p_1 + p_2)^2 ,$$

ii) con la seconda procedura abbiamo

$$P(BB) = P(B|1)^2 P(1) + P(B|2)^2 P(2) = 0.5(p_1^2 + p_2^2).$$

Poichè

$$p_1^2 + p_2^2 \geq 2p_1p_2 \Rightarrow 2(p_1^2 + p_2^2) \geq (p_1 + p_2)^2 ,$$

la seconda procedura è più conveniente.

Se indichiamo con N l'evento estrazione di una pallina nera, otteniamo che anche $P(NN)$ è maggiore con la seconda procedura che con la prima. È facile verificare che la probabilità di ottenere esattamente una pallina nera ed una bianca (l'ordine non ha importanza) è ovviamente maggiore con la prima procedura che con la seconda.

2) Un sistema S è costituito dai componenti S_1, S_2, \dots, S_n . Consideriamo gli eventi, che assumeremo indipendenti:

$$C_k = \{S_k \text{ è difettoso} \}, \quad k = 1, 2, \dots, n$$

e siano dunque indipendenti gli eventi

$$\overline{C}_k = \{S_k \text{ non ha difetti} \}, \quad k = 1, 2, \dots, n.$$

Sia $p_k = P(C_k)$ e $1 - p_k = P(\overline{C}_k)$.

Due tecnici T_1 e T_2 eseguono controlli indipendenti su ciascun componente S_k . Inoltre, supponiamo che gli esiti dei due controlli su S_k siano condizionatamente indipendenti, sia dato C_k , sia dato \overline{C}_k . Consideriamo gli eventi

$$D_{k,j} = \{T_j \text{ segnala un difetto in } S_k \}, \quad j = 1, 2 ,$$

e supponiamo che

$$P(D_{k,j}|C_k) = q_j > 0 , \quad P(D_{k,j}|\overline{C}_k) = 0$$

e quindi

$$P(\overline{D}_{k,j}|\overline{C}_k) = 1 , \quad P(\overline{D}_{k,j}|C_k) = 1 - q_j \text{ e } P(D_{k,j}) = q_j p_k + 0(1 - p_k) = q_j p_k.$$

(Si osservi che $D_{k,1}$ e $D_{k,2}$ non sono indipendenti, se $0 < p_k < 1$).

Se almeno uno dei due tecnici scopre un difetto in almeno uno dei componenti il sistema S viene revisionato e messo in funzione con ritardo (evento E).

Trovare la probabilità dell'evento

$$D = \{S \text{ è difettoso, ma è messo in funzione senza ritardo} \} .$$

Si potrebbero considerare tutti i casi di non affidabilità di S : S può essere difettoso perchè un qualunque sottoinsieme non vuoto (ce ne sono $2^n - 1$) dei suoi componenti è costituito da

componenti tutti difettosi, mentre gli altri componenti sono tutti affidabili. In ciascuna di queste $2^n - 1$ circostanze disgiunte si dovrebbe valutare la probabilità condizionale di non rilevare un difetto.

È più semplice seguire un altro procedimento. Consideriamo l'evento:

$F = \{S \text{ è messo in funzione senza ritardo, abbia o no difetti} \}$,

allora

$$\begin{aligned} E &= \overline{F}, \quad F = \cap_k (\overline{D_{k,1}} \cap \overline{D_{k,2}}), \\ P(F) &= 1 - P(E) = \prod_k P(\overline{D_{k,1}} \cap \overline{D_{k,2}}). \\ P(\overline{D_{k,1}} \cap \overline{D_{k,2}}) &= P(\overline{D_{k,1}} \cap \overline{D_{k,2}} | C_k) P(C_k) + P(\overline{D_{k,1}} \cap \overline{D_{k,2}} | \overline{C_k}) P(\overline{C_k}) = \\ &= P(\overline{D_{k,1}} | C_k) P(\overline{D_{k,2}} | C_k) P(C_k) + P(\overline{D_{k,1}} | \overline{C_k}) P(\overline{D_{k,2}} | \overline{C_k}) P(\overline{C_k}) = \\ &= (1 - q_1)(1 - q_2)p_k + 1 \cdot 1 \cdot (1 - p_k). \end{aligned}$$

Se N è l'evento $\{S \text{ non è difettoso} \}$, allora

$$P(N) = \prod_k (1 - p_k), \quad F = N \cup D \text{ (unione disgiunta) e}$$

$$P(D) = P(F) - P(N),$$

formula che risponde alla domanda proposta.

3) Come nel caso precedente, sia S un sistema costituito dai componenti S_1, S_2, \dots, S_n . Consideriamo gli eventi, che assumeremo indipendenti:

$$C_k = \{S_k \text{ è difettoso} \}, \quad k = 1, 2, \dots, n$$

e

$$\overline{C_k} = \{S_k \text{ non ha difetti} \},$$

Sia $p_k = P(C_k)$ e $1 - p_k = P(\overline{C_k})$. (Per esempio potrebbe essere $p_k = 0.01$).

Il sistema è controllato da n dispositivi automatici T_k , $k = 1, 2, \dots, n$ che operano indipendentemente. T_k controlla soltanto S_k e blocca tale componente se rileva un difetto. Sia

$$R_k = \{T_k \text{ rivela un difetto in } S_k\}$$

e siano note le probabilità condizionali di R_k rispetto agli eventi C_k e $\overline{C_k}$:

$$P(R_k | C_k) = q_k, \quad P(R_k | \overline{C_k}) = e_k,$$

(ad esempio $q_k = 0.99$ e $e_k = 0.005$) che qualificano la validità del dispositivo di controllo. q_k grande indica elevata probabilità di rilevare un difetto esistente, e_k piccola indica lieve probabilità di segnalare per errore un difetto inesistente.

Perché si verifichi l'evento $A = \{ \text{il sistema si arresta} \}$ è sufficiente che si verifichi almeno uno degli eventi R_k .

i) Si chiede di calcolare la probabilità di arresto.

Si osservi che

$$A = \cup_k R_k, \quad \overline{A} = \cap_k \overline{R_k},$$

quindi

$$P(A) = 1 - P(\bar{A}) = 1 - \prod_k P(\bar{R}_k) ,$$

$$P(\bar{R}_k) = P(\bar{R}_k|C_k)P(C_k) + P(\bar{R}_k|\bar{C}_k)P(\bar{C}_k) = (1 - q_k)p_k + (1 - e_k)(1 - p_k).$$

ii) Si chiede inoltre di calcolare la probabilità che il componente S_k sia difettoso, sapendo che il sistema si è arrestato.

Per calcolare $P(C_k|A)$ si può ricorrere al Teorema di Bayes, calcolando quindi in primo luogo $P(A|C_j)$. Poiché

$$A = R_j \cup (\bar{R}_j \cap \cup_{i \neq j} R_i) .$$

unione disgiunta, e

$$P((\bar{R}_j \cap \cup_{i \neq j} R_i) \cap C_j) = P(\bar{R}_j \cap C_j)P(\cup_{i \neq j} R_i)$$

per l'indipendenza dei difetti e dei controlli relativi a componenti diversi. Allora

$$P(A|C_j) = P(R_j|C_j) + P(\bar{R}_j|C_j)P(\cup_{i \neq j} R_i) = P(R_j|C_j) + P(\bar{R}_j|C_j)(1 - P(\cap_{i \neq j} \bar{R}_i)) .$$

Dunque

$$P(A|C_j) = q_j + (1 - q_j)(1 - \prod_{i \neq j} P(\bar{R}_i)) =$$

$$= q_j + (1 - q_j)(1 - \prod_{i \neq j} [(1 - q_i)p_i + (1 - e_i)(1 - p_i)]) .$$

Naturalmente

$$P(C_k|A) = \frac{P(A|C_k)p_k}{\sum_{j=1}^n P(A|C_j)p_j} .$$

4) Si considera un carattere C che può presentarsi in due varianti A e B (ad esempio occhi azzurri o bruni). C è determinato dalla copia di geni g_1, g_2 ereditati dal padre P e rispettivamente dalla madre M , che possono presentarsi nella forma a oppure b . Se $g_1 = a$ e anche $g_2 = a$ si presenta la variante A , in tutti gli altri casi si presenta la variante B (A recessiva, B dominante).

Per quanto concerne la popolazione dei genitori le quattro configurazioni possibili hanno probabilità assegnate:

$$P(4 \stackrel{def}{=} aa) = p_4 , P(3 \stackrel{def}{=} ab) = p_3 , P(2 \stackrel{def}{=} ba) = p_2 , P(1 \stackrel{def}{=} bb) = p_1 .$$

(Ad esempio potrebbe essere $p_j = 0.25$ per $j = 1, 2, 3, 4$).

Supponiamo che P o M si trovino, indipendentemente, in una delle configurazioni possibili e supponiamo inoltre che la probabilità di ereditare il primo o il secondo gene, sia da P che da M , siano eguali e quindi pari a 0.5.

Supponendo di sapere che nel figlio il carattere C ha assunto la versione A , calcolare la probabilità (condizionale) che entrambi i genitori si trovino nella configurazione 4.

Consideriamo le 16 coppie ij di configurazioni possibili i per P e j per M ($i, j = 1, 2, 3, 4$). Pensando di usare il Teorema di Bayes si devono calcolare le probabilità condizionali $P(AF|ij)$, dove AF è l'evento "il figlio F presenta la versione A del carattere C , conoscendo già le $P(ij) = p_i p_j$ (1/16 per esempio).

$$P(AF|1k) = P(AF|k1) = 0, \quad k = 1, 2, 3, 4, \quad P(AF|44) = 1,$$

$$P(AF|j4) = P(AF|4j) = \frac{1}{2} \text{ se } j = 2, 3,$$

$$P(AF|ij) = \frac{1}{4} \text{ se } i, j = 2, 3.$$

Dunque

$$P(44|AF) = \frac{P(AF|44)p_4^2}{\sum_{i,j} P(AF|ij)p_i p_j}.$$

Nell'esempio

$$P(44|AF) = \frac{1 \cdot \frac{1}{16}}{(7 \cdot 0 + 1 \cdot 1 + 4 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4}) \frac{1}{16}} = \frac{1}{4}$$

è il risultato cercato.

5) *Trasmissione di lettere:* siano disponibili n (ad esempio 21) caratteri c_1, c_2, \dots, c_n e si considerino trasmissioni di parole casuali T formate da l lettere casuali (ad esempio 5) $T_1 T_2 \dots T_l$, ciascuna delle quali possa essere uno qualunque degli n caratteri disponibili. Per ogni lettera trasmessa T_j , vi sia la stessa probabilità ($1/n$) di essere uno qualunque dei caratteri disponibili. Le lettere siano trasmesse indipendentemente, quindi per la parola trasmessa T vi sia la stessa probabilità ($1/N$) di essere una qualunque delle $N = n^l$ possibili stringhe t di l caratteri $t_1 t_2 \dots t_l$. La trasmissione delle lettere sia soggetta a disturbi. Si indichi con R la parola ricevuta formata dalle lettere $R_1 R_2 \dots R_l$. Si indichino con

$$\{T = t\} \quad \text{e} \quad \{R = r\}$$

gli eventi: "è stata trasmessa la stringa t ed "è stata ricevuta la stringa r .

Si supponga che, per ogni carattere, in qualunque posizione, la probabilità di trasmissione senza errore sia p (ad esempio 0.9), indipendentemente dai caratteri presenti nelle altre posizioni e dalla trasmissione corretta o meno delle altre lettere; quindi per ogni j :

$$P(R_j = T_j) = \sum_{k=1}^n P(R_j = c_k | T_j = c_k) P(T_j = c_k) = n \cdot p \cdot \frac{1}{n} = p.$$

In caso di errore venga ricevuto uno qualunque degli altri caratteri, ciascuno con probabilità $1/(n-1)$. Trovare la probabilità che sia stata trasmessa la stringa θ sapendo che è stata ricevuta la stringa ρ (ad esempio $\theta = \text{ROTTTO}$, $\rho = \text{ROSSO}$).

In questo caso possiamo utilizzare la formula di Bayes nella versione più semplice:

$$P(T = \theta | R = \rho) = \frac{P(R = \rho | T = \theta) P(T = \theta)}{P(R = \rho)}.$$

Si vede facilmente che

$$P(T = \theta) = \frac{1}{N} = \frac{1}{n^l}, \quad P(R = \rho) = \left(\frac{1}{n}p + \frac{n-1}{n}(1-p)\frac{1}{n-1}\right)^l = \frac{1}{n^l},$$

$$P(R = \rho|T = \theta) = p^{l-i}\left(\frac{1-p}{n-1}\right)^i = P(T = \theta|R = \rho).$$

dove i è il numero di caratteri diversi tra θ e ρ .

Nel caso dell'esempio $P(T = \theta|R = \rho) = (0.9)^3(0.1)^220^{-2} = 0.000018225$.

6) *Ago di Buffon. In un piano è presente una griglia costituita da infinite rette parallele (fili) $r_j, j \in \mathcal{Z}$, con distanza costante tra rette adiacenti pari a $2l$. Un segmento (ago) di lunghezza l viene lanciato "a caso sul piano, parallelamente ad esso. Trovare la probabilità che l'ago colpisca uno dei fili.*

Utilizziamo un riferimento cartesiano con l'asse x coincidente con il filo r_j immediatamente "sotto la posizione dell'ago nell'istante in cui questo colpisce il piano. La posizione dell'ago è determinata dall'intero j , dalle coordinate x, y dell'estremo A più "basso dell'ago e dall'angolo ϑ formato dal segmento AB con il semiasse positivo delle x .

$$j \in \mathcal{Z}, \quad -\infty < x < +\infty, \quad 0 < y \leq 2l, \quad 0 \leq \vartheta < \pi.$$

Interpretiamo la dizione "lancio a caso come assunzione di totale equivalenza delle posizioni possibili, nel senso di equivalenza dei punti (j, x, y, ϑ) e più precisamente ancora nel senso che:

$$\Omega = \mathcal{Z} \times \mathcal{R} \times \Omega_r, \quad \Omega_r =]0, 2l] \times [0, \pi[, \quad P(\mathcal{Z} \times \mathcal{R} \times E) = \frac{\text{area}(E)}{\text{area}(\Omega_r)} \quad \text{se } E \subseteq \Omega_r,$$

supponendo E dotato di area nel senso di Lebesgue: $E \in \mathcal{L}(\Omega_r)$.

Ovviamente $\text{area}(\Omega_r) = 2\pi l$ ed il filo immediatamente "superiore all'asse x è colpito se e solo se l'ordinata di B supera $2l$. Ci interessa dunque l'evento

$$\mathcal{Z} \times \mathcal{R} \times E, \quad E = \{y + l \cdot \sin(\vartheta) \geq 2l\} = \{y \geq 2l - l \cdot \sin(\vartheta)\}$$

per il quale

$$\text{area}(E) = \int_0^\pi (2l - (2l - l \cdot \sin(\vartheta)))d\vartheta = \int_0^\pi l \cdot \sin(\vartheta)d\vartheta = 2l.$$

Dunque

$$P(\mathcal{Z} \times \mathcal{R} \times E) = \frac{2l}{2l\pi} = \frac{1}{\pi}.$$

Capitolo 2

Variabili aleatorie reali discrete

2.1 Variabili aleatorie reali

Una variabile aleatoria reale X è una funzione che assume valori reali “dipendenti dal caso e per la quale è possibile assegnare una probabilità ad ogni proposizione di una famiglia sufficientemente ampia di affermazioni coinvolgenti la funzione X .”

Più precisamente, sia (Ω, \mathcal{A}, P) uno spazio di probabilità, che immaginiamo associato ad un esperimento S . Una funzione $X : \Omega \rightarrow \mathbf{R}$ assume un valore $X(\omega)$ dipendente dal risultato ω dell’esperimento S . È importante poter assegnare una probabilità alle proposizioni “ X assume valori compresi tra a e b , dove a e b sono due numeri reali arbitrari. Consideriamo allora soltanto funzioni X tali che per ogni $a \leq b \in \mathbf{R}$ l’insieme $\{\omega \in \Omega \mid a \leq X(\omega) < b\}$ sia un evento, cioè un insieme di \mathcal{A} . Essendo \mathcal{A} una σ -algebra, si può vedere che, in tal caso, sono eventi tutti gli insiemi $\{\omega \in \Omega \mid X(\omega) \in B\}$, dove B è un qualunque insieme **Boreliano** di \mathbf{R} . Si dice allora che la funzione X è **misurabile**. I Boreliani costituiscono, per definizione, la più piccola σ -algebra alla quale appartengono tutti gli aperti, che nel caso di \mathbf{R} si possono sempre presentare come unione numerabile di intervalli (volendo disgiunti). D’altra parte si vede che è sufficiente chiedere che, per ogni $c \in \mathbf{R}$ l’insieme $\{\omega \in \Omega \mid X(\omega) < c\}$ sia un evento. Arriviamo allora alla seguente

Definizione. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Si dice **variabile aleatoria (v.a.) reale** ogni funzione

$$X : \Omega \longrightarrow \mathbf{R}$$

misurabile, cioè tale che

$$\forall c \in \mathbf{R} \quad \{\omega \in \Omega \mid X(\omega) < c\} \in \mathcal{A}.$$

L’insieme sopra indicato si scriverà più semplicemente $\{X < c\}$.

Sarebbe equivalente chiedere che $\{X > c\}$ oppure $\{X \leq c\}$, o $\{X \geq c\}$, o ancora $X^{-1}(]a, b])$ appartengano ad \mathcal{A} , ovvero siano eventi.

2.2 Variabili aleatorie reali discrete

Si dice che X è **discreta** se può assumere solo un numero finito o al più un'infinità numerabile di valori distinti. È certamente questo il caso se Ω è finito o numerabile.

Se x_k sono i valori distinti che la variabile aleatoria X può assumere, si dice **densità di X** la funzione

$$f_X(x_k) = p_k^X = P(A_k), \quad A_k = \{\omega \in \Omega \mid X(\omega) = x_k\}.$$

Si dice **distribuzione (cumulativa) di X** la funzione della variabile reale c

$$F_X(c) = P(\{X < c\}).$$

Indipendenza di variabili aleatorie discrete.

Definizione: le v.a. X e Y , che possono assumere i valori x_k e y_l , si dicono **indipendenti** se per ogni coppia (k, l) risulta:

$$p_{k,l}^{XY} = P(X = x_k \wedge Y = y_l) = P(X = x_k)P(Y = y_l) = p_k^X p_l^Y.$$

Osservazione. Se X e Y sono indipendenti, anche $X - c$ e $Y - d$, dove c e d sono costanti, sono indipendenti.

Se X, Y, \dots, Z sono v.a. discrete che possono assumere i valori x_k, y_l, \dots, z_m , si dice **densità congiunta** di X, Y, \dots, Z la funzione

$$f_{X,Y,\dots,Z}(x_k, y_l, \dots, z_m) = P(X = x_k \wedge Y = y_l \wedge \dots \wedge Z = z_m).$$

Talvolta scriveremo più brevemente $f_{X,Y,\dots,Z}(k, l, \dots, m)$.

Le v.a. $X_1, X_2, \dots, X_n, \dots$ si dicono **mutuamente indipendenti** se, fissati comunque gli indici $1 \leq j_1 < j_2 < \dots < j_m$ e i valori ammissibili $x_{k_1}, x_{k_2}, \dots, x_{k_m}$, per gli eventi $\{X_{j_1} = x_{k_1}\}, \{X_{j_2} = x_{k_2}\}, \dots, \{X_{j_m} = x_{k_m}\}$ risulta:

$$\begin{aligned} P(\{X_{j_1} = x_{k_1}\} \cap \{X_{j_2} = x_{k_2}\} \dots \cap \{X_{j_m} = x_{k_m}\}) &= \\ = P(\{X_{j_1} = x_{k_1}\})P(\{X_{j_2} = x_{k_2}\}) \dots P(\{X_{j_m} = x_{k_m}\}). \end{aligned}$$

Dunque, scelti arbitrariamente l'intero n e n valori ammissibili x_1, x_2, \dots, x_n , gli eventi $\{X_1 = x_1\}, \{X_2 = x_2\}, \dots, \{X_n = x_n\}$ sono mutuamente indipendenti.

Si vede immediatamente che X e Y sono indipendenti se e solo se la densità congiunta $f_{XY}(k, l)$ è il prodotto delle densità $f_X(k)$ e $f_Y(l)$.

2.3 Distribuzione binomiale (Schema di Bernouilli)

Si considerano n prove (*mutuamente*) *indipendenti* (un esperimento S viene ripetuto n volte, essendo i risultati di ciascuna prova indipendenti dai risultati delle altre). In ogni prova un determinato evento E può verificarsi con probabilità p (*successo: s*) o non verificarsi con probabilità $q = 1 - p$ (*insuccesso o fallimento: f*).

Come insieme di riferimento Ω consideriamo l'insieme delle scritte o stringhe del tipo $\omega = sffssfsff$ (in questo esempio $n = 10$), che indica un successo nella prima prova, un fallimento nella seconda e terza prova, Essendo Ω finito possiamo prendere $\mathcal{A} = \mathcal{P}(\Omega)$ e introdurre una misura di probabilità ponendo

$$P(\{\omega \mid \omega_k = s\}) = p, \quad k = 1, 2, \dots, n,$$

dove ω_k indica la k -esima lettera (componente di ω , coerentemente con l'ipotesi fatta sulle singole prove, e assumendo che gli eventi $\{\omega \mid \omega_k = s\}$ siano mutuamente indipendenti. Si trova allora per un arbitrario ω , scrivendo brevemente $P(\omega)$ in luogo di $P(\{\omega\})$, e per un arbitrario evento A :

$$P(\omega) = p^{K(\omega)} q^{n-K(\omega)}, \quad P(A) = \sum_{\omega \in A} P(\omega),$$

dove $K(\omega)$ è il numero di successi in ω , cioè il numero di caratteri s nella stringa ω . Se invece dei simboli s ed f si usassero le cifre 0 ed 1, indicando sempre con ω_k la k -esima componente di ω , si avrebbe

$$K(\omega) = \omega_1 + \omega_2 + \dots + \omega_n.$$

K è una variabile aleatoria discreta che può assumere i valori $0, 1, 2, \dots, n$. La probabilità che K assuma un valore determinato k è

$$P(\{\omega \in \Omega \mid K(\omega) = k\}) = P(K = k) = C_n^k p^k q^{n-k},$$

infatti vi sono C_n^k elementi di Ω che hanno esattamente k successi e ciascuno di essi ha la stessa probabilità $p^k q^{n-k}$.

Con C_n^k indichiamo l'usuale coefficiente binomiale

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

che conta il numero di sottoinsiemi costituiti esattamente da k elementi di un insieme di n elementi.

L'insieme degli eventi ω con k componenti uguali a 1 (s) è infatti in corrispondenza biunivoca con l'insieme dei sottoinsiemi S di $\{1, 2, \dots, n\}$ formati esattamente da k elementi. Ogni ω si può considerare come la funzione caratteristica di un tale sottoinsieme S .

Se siamo interessati solo al numero di successi, lo spazio campione può essere *ridotto* all'insieme $\Omega^* = 0, 1, 2, \dots, n$, con le probabilità

$$f(k) = C_n^k p^k q^{n-k}.$$

Si osservi che

$$f(k) \geq 0 \quad \text{e} \quad \sum_{k=0}^n f(k) = (p+q)^n = 1.$$

f è la densità della variabile aleatoria K . La *distribuzione (cumulativa)* F della variabile aleatoria K :

$$F(c) = P(\{K < c\}),$$

è data da

$$F(k) = \sum_{j=0}^k f(j) = \sum_{j=0}^k C_n^j p^j q^{n-j}$$

ed è detta *distribuzione binomiale* .

2.4 I teoremi di DeMoivre-Laplace

Teorema del limite locale (DeMoivre-Laplace). Sia $P_n(k)$ la densità binomiale e, con $0 < p < 1$ e $q = 1 - p$, fissato $x \in]a, b[$, sia k l'intero definito da

$$k \leq np + x\sqrt{npq} < k + 1 .$$

$k = k(n) = k(n, x)$ è quindi la parte intera di $np + x\sqrt{npq}$ e, posto

$$x(n) = \frac{k - np}{\sqrt{npq}} ,$$

la differenza $x - x(n)$ (non negativa e $\leq (npq)^{-\frac{1}{2}}$) è $O(n^{-\frac{1}{2}})$.

Allora:

$$\lim_{n \rightarrow +\infty} \sqrt{npq} P_n(k(n, x)) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) ,$$

uniformemente per $x \in]a, b[$.

Osservazione. Più in generale, se $k(n)$ è una successione tale che $x(n)$ si mantiene limitata in $]a, b[$, allora

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{npq} P_n(k(n))}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x(n)^2}{2}\right)} = 1$$

uniformemente, nel senso che dato $\varepsilon > 0$ il rapporto dista da 1 meno di ε per n maggiore di una soglia dipendente soltanto da ε, a e b .

Dimostrazione del teorema. Ricordiamo in primo luogo la **formula di Stirling**

$$n! = \sqrt{2\pi n} n^n \exp(-n) \exp(\theta_n) , \quad |\theta_n| \leq \frac{1}{12n} .$$

Si ha

$$k = np + x(n)\sqrt{npq}, \quad n - k = nq - x(n)\sqrt{npq} \rightarrow +\infty$$

per $a < x < b$, con a, b fissi .

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{\sqrt{2\pi n}}{\sqrt{2\pi k}\sqrt{2\pi(n-k)}} \frac{n^n \exp(-n) \exp(\theta_n - \theta_k - \theta_{n-k}) p^k q^{n-k}}{k^k (n-k)^{n-k} \exp(-k) \exp(-(n-k))},$$

dunque

$$P_n(k) = \left(\frac{n}{2\pi k(n-k)}\right)^{\frac{1}{2}} \frac{n^n p^k q^{n-k} \exp(\theta)}{k^k (n-k)^{n-k}},$$

dove $\theta = \theta_n - \theta_k - \theta_{n-k}$ e quindi

$$|\theta| \leq \frac{1}{12} \left(\frac{1}{n} + \frac{1}{k} + \frac{1}{n-k} \right) \leq \frac{1}{12n} \left(1 + \frac{1}{p + a\sqrt{pq/n}} + \frac{1}{q - b\sqrt{pq/n}} \right).$$

Dunque $\theta \rightarrow 0$ per $n \rightarrow +\infty$, uniformemente in $a < x < b$ e quindi $\exp(\theta) \rightarrow 1$ uniformemente.

$$\begin{aligned} B_n &= \left(\frac{n}{k(n-k)}\right)^{\frac{1}{2}} = \left(\frac{n}{np(1+x(n)\sqrt{q/np})nq(1-x(n)\sqrt{p/nq})}\right)^{\frac{1}{2}} = \\ &= \frac{1}{\sqrt{npq}} (1+x(n)\sqrt{\frac{q}{np}})^{-\frac{1}{2}} (1-x(n)\sqrt{\frac{p}{nq}})^{-\frac{1}{2}} \end{aligned}$$

e quindi $\sqrt{npq}B_n \rightarrow 1$ uniformemente in x . Inoltre

$$\begin{aligned} \ln A_n &= \ln \frac{n^n p^k q^{n-k}}{k^k (n-k)^{n-k}} = \ln \left(\frac{np}{k}\right)^k + \ln \left(\frac{nq}{n-k}\right)^{n-k} = \\ &= -k \ln \frac{k}{np} - (n-k) \ln \frac{n-k}{nq} = \\ &= -(np+x(n)\sqrt{npq}) \ln(1+x(n)\sqrt{\frac{q}{np}}) - (nq-x(n)\sqrt{npq}) \ln(1-x(n)\sqrt{\frac{p}{nq}}) = \\ &= -(np+x(n)\sqrt{npq}) \left(x(n)\sqrt{\frac{q}{np}} - \frac{x(n)^2 q}{2np} + O(n^{-\frac{3}{2}})\right) - (nq-x(n)\sqrt{npq}) \left(-x(n)\sqrt{\frac{p}{nq}} - \frac{x(n)^2 p}{2nq} + O(n^{-\frac{3}{2}})\right) = \\ &= -x(n)\sqrt{npq} + \frac{x(n)^2 q}{2} + O(n^{-\frac{1}{2}}) - x(n)^2 q + O(n^{-\frac{1}{2}}) + \dots \\ &+ x(n)\sqrt{npq} + \frac{x(n)^2 p}{2} + O(n^{-\frac{1}{2}}) - x(n)^2 p + O(n^{-\frac{1}{2}}) + \dots = \\ &= -\frac{x(n)^2}{2} + O(n^{-\frac{1}{2}}) = -\frac{x^2}{2} + O(n^{-\frac{1}{2}}) \rightarrow -\frac{x^2}{2} \end{aligned}$$

uniformemente in x e dunque

$$A_n / \exp\left(-\frac{x^2}{2}\right) \rightarrow 1.$$

Pertanto

$$\sqrt{npq}P_n(k) = \frac{\exp(\theta)}{\sqrt{2\pi}} \sqrt{npq}B_n A_n \rightarrow \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad q.e.d.$$

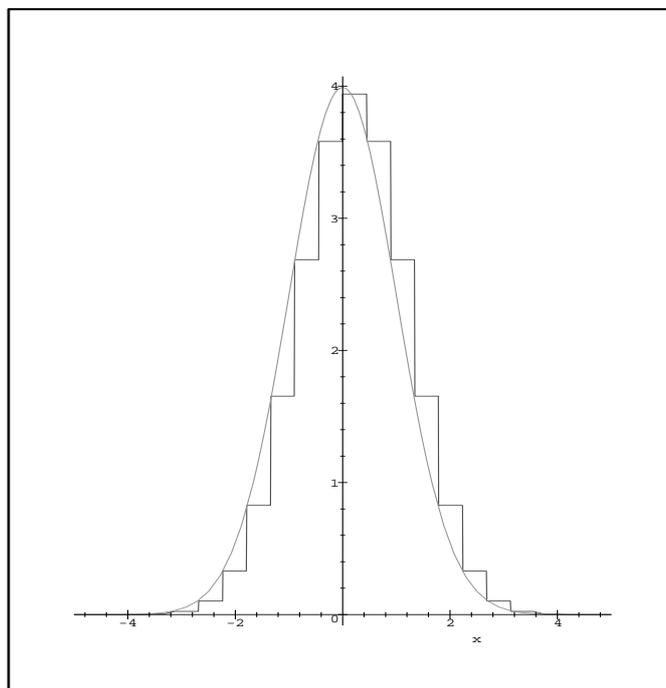


Figura 2.1: Densità binomiale normalizzata, con $n = 20$ e $p = 0.5$, e Normale standard

Nella Figura 1 sono riportati, in scale opportune, i grafici della funzione costante a tratti $\sqrt{npq}P_n(k(n, x))$, con $p = 0.5$ e $n = 20$, e della funzione limite $\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$. I tratti hanno ampiezza $1/\sqrt{npq} \cong 0.447$.

Valore massimo di $P_n(k)$: dobbiamo studiare le disuguaglianze

$$C_n^{k-1} p^{k-1} q^{n-k+1} \leq C_n^k p^k q^{n-k} \geq C_n^{k+1} p^{k+1} q^{n-k-1},$$

equivalenti alle disuguaglianze

$$\frac{q}{p} \leq \frac{n-k+1}{k}, \quad \frac{q}{p} \geq \frac{n-k}{k+1},$$

dalle quali si ricava

$$np - q \leq k \leq np + p.$$

Vi possono essere dunque uno o due argomenti k per i quali $P_n(k)$ assume il valore massimo; per $k \leq np + p$ $P_n(k)$ è crescente, mentre per $np - q \leq k$ $P_n(k)$ è decrescente.

Teorema del limite integrale di DeMoivre-Laplace. Sia K il numero aleatorio di successi in n prove indipendenti, in ciascuna delle quali la probabilità di successo è p , con $0 < p < 1$. Fissati $a < b$, si ha per $n \rightarrow +\infty$

$$P\left(a \leq \frac{K - np}{\sqrt{npq}} < b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt .$$

Dimostrazione. Si deve studiare il comportamento asintotico della probabilità dell'evento $E_n = \{K \in A_n\}$, dove

$$A_n = \left\{k \mid a \leq x_k(n) = \frac{k - np}{\sqrt{npq}} < b\right\}$$

e, posto $i = \min A_n$, $j = \max A_n$, si introduca la funzione

$$f_n(x) = \begin{cases} 0 & x < x_{i-1} \vee x \geq x_{j+1} \\ \sqrt{npq}P_n(k) & x_k \leq x < x_{k+1}, \quad i \leq k < j \end{cases} ,$$

osservando che

$$k = np + x_k \sqrt{npq}, \quad x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$$

e, se $a^* < a < b < b^*$, per il teorema del limite locale

$$f_n(x) = \sqrt{npq}P_n(k) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{uniformemente su }]a^*, b^*[.$$

Allora, essendo $x_{i-1} < a \leq x_i < x_j < b \leq x_{j+1}$

$$\begin{aligned} P(E_n) &= \sum_{k=i}^j \sqrt{npq}P_n(k)(x_{k+1} - x_k) = \\ &= \int_a^b f_n(x)dx - \int_a^{x_i} f_n(x)dx - \int_{x_j}^b f_n(x)dx . \end{aligned}$$

Inoltre $|a - x_i|, |b - x_j| \leq 1/\sqrt{npq} \rightarrow 0$. Per la convergenza uniforme delle f_n si trova il risultato, che non sarebbe difficile estendere a $a = -\infty \vee b = +\infty$. *q.e.d.*

2.5 Distribuzione multinomiale

Si considerano ancora n prove indipendenti. In ogni prova si possono presentare gli eventi disgiunti E_1, E_2, \dots, E_l con probabilità p_1, p_2, \dots, p_l :

$$p_1 + p_2 + \dots + p_l = 1 \quad \text{e ovviamente } 0 \leq p_j \leq 1 .$$

Allora la probabilità che si presentino nelle n prove esattamente $k_1 = k_1(n)$ occorrenze di E_1 ... $k_l = k_l(n)$ occorrenze di E_l ($\sum_j k_j = n$) è data da

$$P_n(k_1, k_2, \dots, k_l) = \frac{n!}{k_1!k_2!\dots k_l!} p_1^{k_1} p_2^{k_2} \dots p_l^{k_l} ,$$

come si dimostra facilmente con un percorso ad l stadi. Indicando con K_j la v.a. che conta il numero di esiti E_j nelle n prove, si ha:

$$P(K_1 = k_1) = \frac{n!}{k_1!(n - k_1)!} p_1^{k_1} (1 - p_1)^{n - k_1} ,$$

$$P(K_1 = k_1 \wedge K_2 = k_2) = P(K_1 = k_1) P(K_2 = k_2 | K_1 = k_1) .$$

Ma se esattamente k_1 prove hanno avuto esito E_1 , le probabilità condizionali degli altri esiti in ognuna delle prove rimanenti è $\tilde{p}_j = p_j / (1 - p_1)$. Infatti, se E_j^m è l'evento "si è verificato E_j nella prova m -esima e $j \neq 1$, $P(E_j^m | \overline{E_1^m}) / P(\overline{E_1^m}) = P(E_j^m) / P(\overline{E_1^m})$. Si noti che $\sum_j \tilde{p}_j = 1$. Dunque

$$P(K_2 = k_2 | K_1 = k_1) = \frac{(n - k_1)!}{k_2!(n - k_1 - k_2)!} \tilde{p}_2^{k_2} (1 - \tilde{p}_2)^{n - k_1 - k_2} .$$

Osservando che $1 - \tilde{p}_2 = (1 - p_1 - p_2) / (1 - p_1)$, si ottiene allora

$$P(K_1 = k_1 \wedge K_2 = k_2) = \frac{n!}{k_1! k_2! (n - k_1 - k_2)!} p_1^{k_1} p_2^{k_2} (1 - p_1 - p_2)^{n - k_1 - k_2} ,$$

e continuando in questo modo si giunge alla formula che si voleva dimostrare.

Se $q_j = 1 - p_j$ e $p_j, q_j \neq 0$, date l successioni $k_j(n)$ tali che

$$\sum_{j=1}^l k_j(n) = n$$

e tali che le i loro valori normalizzati

$$x_j = x_j(n) = \frac{k_j - np_j}{\sqrt{np_j q_j}} ,$$

che soddisfino la relazione

$$\sum_j x_j \sqrt{p_j q_j} = 0 ,$$

si mantengano limitati, diciamo in $]a, b[$, risulta

$$\lim_{n \rightarrow +\infty} \sqrt{n^{l-1} p_1 p_2 \dots p_l} P_n(k_1, k_2, \dots, k_l) / \frac{1}{(2\pi)^{(l-1)/2}} e^{-(q_1 x_1^2 + q_2 x_2^2 + \dots + q_l x_l^2)/2} = 1 ,$$

uniformemente, nel senso che dato $\varepsilon > 0$ il rapporto dista da 1 meno di ε per n maggiore di una soglia dipendente soltanto da ε, a e b .

Per $n = 2$ si ritrova il risultato di DeMoivre-Laplace.

2.6 Distribuzione geometrica

Come nello schema di Bernoulli si considerano prove indipendenti con probabilità p di successo, ed ovviamente $q = 1 - p$ di fallimento in ogni prova. Questa volta il numero delle prove

non è prefissato. Si procede finchè non si ottiene un primo successo.

Prendiamo come spazio campione l'insieme di tutte le possibili "storie":

$$\Omega = \{s, fs, ffs, fffs, \dots\}$$

e per ogni ω sia $N(\omega)$ il numero di caratteri in ω . Dunque in ω vi sono $N(\omega) - 1$ lettere f e l'ultima lettera è s . Poniamo, in considerazione dell'indipendenza delle prove:

$$p(k) = P(N = k) = P(ff\dots fs) = q^{k-1}p, \text{ dove le } f \text{ sono } k - 1.$$

Si osservi che

$$p(k) \geq 0, \sum_{k=1}^{+\infty} pq^{k-1} = p \sum_{j=0}^{+\infty} q^j = \frac{p}{1-q} = 1.$$

La variabile aleatoria N , numero di prove per un primo successo ("tempo di attesa" di un primo successo), ha per densità $f(k) = p(k) = pq^{k-1}$ e la distribuzione corrispondente:

$$F(n) = P(N < n) = \sum_{k=1}^{n-1} pq^{k-1}$$

si dice *distribuzione geometrica* (si dice anche che N è una *variabile aleatoria geometrica*).

2.7 Distribuzione binomiale negativa (di parametri p ed r)

Come nel caso della distribuzione geometrica, si considerano prove indipendenti con probabilità p di successo, ed ovviamente $q = 1 - p$ di fallimento in ogni prova. Il numero delle prove non è prefissato. Si procede finchè non si ottengono esattamente r successi.

Sia ad esempio $r = 3$. Sia

$$\Omega = \{sss, fsss, sfss, ssfs, fffss, fsfss, \dots\},$$

l'insieme delle scritte ω con 3 (in generale r) caratteri s , uno dei quali nell'ultima posizione, e con $n - 3$ (in generale $n - r$) caratteri f . Poniamo, in considerazione dell'ipotesi di indipendenza:

$$P(\omega) = p^r q^{n-r}.$$

Vi sono C_{n-1}^{r-1} scritte di questo tipo, perchè esse sono tante quanti i sottoinsiemi di $r - 1$ posizioni selezionate arbitrariamente tra le prime $n - 1$ posizioni, nelle quali scrivere il carattere s . Dunque la densità di N è:

$$f(n) = P(\{\omega \in \Omega \mid N(\omega) = n\}) = C_{n-1}^{r-1} p^r q^{n-r}.$$

N si dice *variabile binomiale negativa di parametri p e r* e la sua distribuzione si dice *distribuzione binomiale negativa*.

2.8 Valor medio (variabili discrete)

Consideriamo in primo luogo il caso in cui Ω stesso è finito o numerabile (e ogni $\{\omega\}$ sia un evento). Possiamo allora definire il valor medio di una v.a., necessariamente discreta, nel modo seguente.

Definizione. Si dice **valor medio** (o anche **valore atteso** o **speranza matematica**) della v.a. discreta X la quantità

$$E(X) = \sum_{\omega \in \Omega} X(\omega)p(\omega),$$

dove $p(\omega) = P(\{\omega\})$.

Se Ω non è finito si pretende che la serie sia assolutamente convergente, perchè $E(X)$ non deve dipendere dall'ordine dei termini della serie:

$$\sum_{\omega \in \Omega} |X(\omega)|p(\omega) < +\infty.$$

In altre parole la famiglia di numeri $\{X(\omega)\}_{\omega \in \Omega}$ è sommabile.

Se non vi è convergenza assoluta si dice che X non è dotata di media.

Siano x_k i valori distinti che la v.a. X può assumere e sia $A_k = \{\omega \mid X(\omega) = x_k\}$. Se $f(k)$ è la probabilità del valore x_k di X , cioè se $f(k) = P(A_k) = f_X(x_k)$, dove f_X è la densità di X , risulta:

$$E(X) = \sum_{\omega \in \Omega} X(\omega)p(\omega) = \sum_{k=0}^{+\infty} \sum_{\omega \in A_k} X(\omega)p(\omega) = \sum_{k=0}^{+\infty} x_k P(A_k) = \sum_{k=0}^{+\infty} x_k f(k).$$

Se $H(x)$ è una funzione reale di variabile reale, possiamo considerare la variabile aleatoria $H(X)$ definita da

$$H(X)(\omega) = H(X(\omega)) \text{ per ogni } \omega \in \Omega.$$

Il suo valor medio risulta

$$E(H(X)) = \sum_{\omega \in \Omega} H(X(\omega))p(\omega).$$

Se $f(k)$ è la densità di X e $g(j) = P(H(X) = h_j) = h_j$ è la densità di H , dove h_j sono i valori distinti che H può assumere, allora

$$\begin{aligned} E(H(X)) &= \sum_j h_j g(j) = \sum_{\omega \in \Omega} H(X(\omega))p(\omega) = \sum_k \sum_{\omega \in A_k} H(X(\omega))p(\omega) = \\ &= \sum_k H(x_k) \sum_{\omega \in A_k} p(\omega) = \sum_k H(x_k) f(k). \end{aligned}$$

Linearità del valor medio:

$$E(cX) = cE(X), \quad E(X + Y) = E(X) + E(Y),$$

infatti

$$\sum_{\omega} (cX(\omega) + Y(\omega))p(\omega) = c \sum_{\omega} X(\omega)p(\omega) + \sum_{\omega} Y(\omega)p(\omega).$$

Si osservi che per una variabile costante c si ha $E(c) = c$. Se ne deduce che per ogni v.a. X

$$E(X - E(X)) = 0 .$$

Positività e monotonia del valor medio:

$$X \geq 0 \Rightarrow E(X) \geq 0 \quad , \quad X \geq Y \Rightarrow E(X) \geq E(Y) .$$

$X \geq 0$ significa che $X(\omega) \geq 0$ per tutte le $\omega \in \Omega$, tranne eventualmente quelle di un sottoinsieme di probabilità nulla. In modo analogo si interpreta la disuguaglianza $X \geq Y$. La prima implicazione è conseguenza immediata della definizione (tutti i termini $X(\omega)p(\omega)$ della somma che fornisce $E(X)$ sono non negativi). La seconda implicazione è immediata perché $X - Y \geq 0$ e il valor medio è lineare e positivo.

Valor medio del prodotto:

in generale il valor medio del prodotto è diverso dal prodotto dei valori medi di due v.a. Ma nel caso particolare di v.a. **indipendenti** si ha invece

$$E(XY) = E(X)E(Y).$$

Infatti gli eventi $A_{kl} = \{ \omega \mid X(\omega) = x_k \wedge Y(\omega) = y_l \}$ costituiscono una partizione di Ω e quindi, poiché $p_{k,l}^{XY} = p_k^X p_l^Y$:

$$\begin{aligned} E(XY) &= \sum_{\omega \in \Omega} X(\omega)Y(\omega)p(\omega) = \sum_{kl} \sum_{\omega \in A_{kl}} X(\omega)Y(\omega)p(\omega) = \\ &= \sum_k \sum_l x_k y_l p_{k,l}^{XY} = \left(\sum_k x_k p_k^X \right) \left(\sum_l y_l p_l^Y \right) = E(X)E(Y). \end{aligned}$$

Nel caso di uno spazio Ω generale (non necessariamente numerabile), per una v.a. discreta X che assume i valori x_k con probabilità $f(k) = f_X(x_k)$, dove f_X è la densità di X , cioè $f(k) = P(A_k)$, dove $A_k = \{ \omega \mid X(\omega) = x_k \}$ **si definisce il valore atteso mediante la formula:**

$$E(X) = \sum_{k=0}^{+\infty} x_k P(A_k) = \sum_{k=0}^{+\infty} x_k f(k),$$

se la serie converge assolutamente, e quindi per essa vale la proprietà commutativa.

Se si dovesse considerare soltanto la v.a. X , si potrebbe sostituire Ω con lo spazio ridotto Ω^* i cui elementi ω^* sono semplicemente gli interi k o i valori distinti x_k , associando a ciascuno di essi la probabilità $f(k)$, riconducendosi in tal modo al caso di uno spazio campione numerabile.

Molto spesso tuttavia è importante considerare simultaneamente più v.a. X, Y, Z, \dots e non sempre è opporuno o possibile ricondursi ad un unico spazio campione numerabile.

Anche nel caso generale si controlla la validità delle proprietà fondamentali del valore atteso sopra considerate (linearità, positività e quindi monotonia). Ad esempio, se X assume i valori x_k sui sottoinsiemi A_k (che formano una partizione di Ω) e Y assume i valori y_l sui sottoinsiemi B_l (che formano una partizione di Ω), considerando la partizione più fine costituita dagli insiemi $C_{kl} = A_k \cap B_l$, si ha:

$$\begin{aligned} E(X + Y) &= \sum_{k,l} (x_k + y_l) P(C_{kl}) = \sum_k x_k \sum_l P(C_{kl}) + \sum_l y_l \sum_k P(C_{kl}) = \\ &= \sum_k x_k P(A_k) + \sum_l y_l P(B_l) = E(X) + E(Y), \end{aligned}$$

per la σ -additività di P .

La composizione e scomposizione delle somme è lecita, ed $X + Y$ ammette valor medio, in virtù della convergenza assoluta delle serie che definiscono $E(X)$ e $E(Y)$.

2.9 Varianza e deviazione standard (variabili discrete)

Definizione. Se la v.a. X è dotata di media $E(X) = \mu_X$, si pone

$$V(X) = \sigma_X^2 = E((X - \mu_X)^2).$$

Se $V(X)$ è finita, si dice che essa è la **varianza di X** e σ_X si dice **deviazione standard, o scarto quadratico medio, di X** .

σ_X è un indice importante di dispersione dei valori di X intorno alla sua media.

Regole elementari di calcolo .

$$V(\lambda X) = \lambda^2 V(X)$$

$$V(X) = E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - E(X)^2$$

Disuguaglianza di Markov. Sia $X \geq 0$ (i suoi valori x_k sono ≥ 0), sia $\mu = E(X)$ e c un numero reale positivo arbitrario, allora

$$P(X \geq c\mu) \leq \frac{1}{c}.$$

Dimostrazione. Sia f la distribuzione di X :

$$\mu = \sum_k x_k f(k) \geq \sum_{k \in \{k | x_k \geq c\mu\}} x_k f(k) \geq c\mu \sum_{k \in \{k | x_k \geq c\mu\}} f(k) = c\mu P(X \geq c\mu).$$

Dividendo per $c\mu$ si ottiene il risultato. *q.e.d.*

Disuguaglianza di Chebyshev. Sia X una v.a. dotata di media μ e varianza σ^2 e sia $\lambda > 0$ arbitrario:

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Dimostrazione. Posto $Y = (X - \mu)^2$, risulta $Y \geq 0$ e $E(Y) = \sigma^2$ e per la disuguaglianza di Markov

$$P(|X - \mu| \geq \lambda\sigma) = P(Y \geq \lambda^2\sigma^2) \leq \frac{1}{\lambda^2}. \quad q.e.d.$$

La varianza ovviamente non è in generale un'operazione lineare. Abbiamo già visto che:

$$V(\lambda X) = \lambda^2 V(X),$$

inoltre, se X e Y sono v.a. con media μ_X e μ_Y , allora

$$\begin{aligned} V(X + Y) &= E((X + Y) - (\mu_X + \mu_Y))^2 = \\ &= E((X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)) = V(X) + V(Y) + 2E((X - \mu_X)(Y - \mu_Y)) \end{aligned}$$

L'ultimo addendo, diviso per 2, si dice **covarianza** delle v.a. X e Y e si indica con $Cov(X, Y)$.

La varianza della somma di due v.a. è uguale alla somma delle loro varianze se e solo se la loro covarianza si annulla:

$$Cov(X, Y) = 0 \Leftrightarrow V(X + Y) = V(X) + V(Y),$$

Una condizione sufficiente affinché ciò avvenga è che le v.a. siano **indipendenti**. Infatti in tal caso sono indipendenti $X - E(X)$ e $Y - E(Y)$ e dunque

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(X - \mu_X)E(Y - \mu_Y) = 0.$$

Se le v.a. X_1, X_2, \dots, X_n sono a **due a due indipendenti** (non è necessaria la mutua indipendenza), si ha

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n).$$

In particolare, se le v.a. X_1, X_2, \dots, X_n sono a due a due indipendenti ed hanno stessa media μ e varianza σ^2 , risulta

$$\begin{aligned} E(X_1 + X_2 + \dots + X_n) &= n\mu, \quad V(X_1 + X_2 + \dots + X_n) = n\sigma^2, \\ E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) &= \mu, \quad V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}. \end{aligned}$$

I risultati riguardanti le medie non richiedono l'indipendenza, che invece è presupposta (quale condizione sufficiente, ma non necessaria) per dedurre i risultati precedenti sulle varianze. Si osservi che la v.a. X *media (aritmetica)* delle v.a. X_k ha lo stesso *valor medio* μ , ma *varianza* σ^2/n ridotta di un fattore $1/n$ e quindi tendete a 0 per $n \rightarrow +\infty$.

2.10 Funzione generatrice dei momenti

Sia t un parametro reale. Si pone

$$m_X(t) = E(e^{tX}) = E\left(\sum_{j=0}^{+\infty} \frac{t^j}{j!} X^j\right),$$

se il valor medio di e^{tX} esiste per $|t| \leq h$ (basta che esista per $\pm h$), con h sufficientemente piccolo. La funzione $m_X(t)$ si dice **funzione generatrice dei momenti** della v.a. X . Si può dimostrare che è lecito permutare l'operatore E con la serie esponenziale, e risulta:

$$E(e^{tX}) = \sum_{j=0}^{+\infty} \frac{t^j}{j!} E(X^j),$$

dunque $m_X(t)$ è sviluppabile in serie di Taylor nell'intorno dell'origine e

$$E(X^k) = D_t^k m_X(t)|_{t=0}.$$

I risultati precedenti si dimostrano facilmente ricorrendo alla teoria generale dell'integrazione. Infatti

$$E(X) = \sum_{\omega \in \Omega} X(\omega) p(\omega) = \int_{\Omega} X dP$$

e quindi, posto $\Omega_+ = \{ X \geq 0 \}$, $\Omega_- = \{ X < 0 \}$, si ha

$$E(e^{\pm hX}) < +\infty \Rightarrow \int_{\pm\Omega} e^{\pm hX} dP < +\infty \Rightarrow \int_{\Omega} e^{h|X|} dP < +\infty.$$

Ma allora

$$|X|^k \leq \frac{k!}{h^k} e^{h|X|} \Rightarrow E(|X|^k) = \int_{\Omega} |X|^k dP < +\infty$$

e pertanto esistono i momenti di ogni ordine. Inoltre, essendo

$$\sum_{k=0}^N \frac{t^k X^k}{k!} \rightarrow e^{tX}, \quad \left| \sum_{k=0}^N \frac{t^k X^k}{k!} \right| \leq e^{h|X|}$$

per ogni ω , in virtù del teorema sulla convergenza dominata, si trova

$$E(e^{tX}) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} E(|X|^k).$$

2.11 Esempi di calcolo di medie, varianze e funzioni generatrici

Variabili $K(\omega)$ binomiali.

$$E(K) = \sum_{k=0}^n k C_n^k p^k q^{n-k} = np.$$

Si consideri infatti la funzione ausiliaria

$$\phi(x, y) = (x + y)^n = \sum_k C_n^k x^k y^{n-k},$$

$$x \frac{\partial \phi}{\partial x} = nx(x + y)^{n-1} = \sum_k k C_n^k x^k y^{n-k}.$$

Posto $x = p$, $y = q$ si trova il risultato.

Più facilmente si può scrivere $K = K_1 + K_2 + \dots + K_n$, dove K_j è una v.a. che prende i valori 1 e 0 con probabilità p e q , secondo che nella prova j -esima vi sia un successo o un fallimento. Evidentemente $E(K_j) = 1p + 0q = p$ e quindi $E(K) = \sum_j E(K_j) = np$.

Poichè $K_j^2 = K_j$ e quindi $E(K_j^2) = p$, si ha $V(K_j) = E(K_j^2) - E(K_j)^2 = p - p^2 = pq$. Ma le v.a. K_j sono indipendenti e quindi $V(K) = npq$.

Si potrebbe anche procedere nel modo seguente:

$$x^2 \frac{\partial^2 \phi}{\partial x^2} = n(n-1)x^2(x+y)^{n-2} = \sum_k k(k-1)C_n^k x^k y^{n-k},$$

E posto $x = p$, $y = q$

$$n(n-1)p^2 = E(K^2) - E(K), \quad V(K) = E(K^2) - E(K)^2 =$$

$$n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq.$$

Per la deviazione standard si trova $\sigma_K = \sqrt{npq}$.

Per quanto concerne la funzione generatrice:

$$m_K(t) = \sum_{i=0}^n e^{ti} C_n^i p^i q^{n-i} = \sum_{i=0}^n C_n^i (pe^t)^i q^{n-i} = (pe^t + q)^n.$$

Dunque

$$m_X(0) = (p + q)^n = 1,$$

$$Dm_K(t)|_{t=0} = n(pe^t + q)^{n-1} pe^t|_{t=0} = np, \quad E(K) = np,$$

$$D^2 m_K(t)|_{t=0} = n(n-1)(pe^t + q)^{n-2} p^2 e^{2t} + n(pe^t + q)^{n-1} pe^t|_{t=0} = n(n-1)p^2 + np = E(K^2),$$

$$\sigma^2 = E(K^2) - E(K)^2 = n^2 p^2 - np^2 + np - n^2 p^2 = np(1-p) = npq.$$

Frequenza.

Se consideriamo la *frequenza (relativa)* di successo in n prove: K/n , abbiamo

$$E\left(\frac{K}{n}\right) = p, \quad V\left(\frac{K}{n}\right) = \frac{pq}{n}.$$

Applicando la disuguaglianza di Chebychev, per ogni $\varepsilon > 0$ si ha

$$P(\{|\frac{K}{n} - p| \geq \varepsilon\}) \leq \frac{pq}{\varepsilon^2 n}.$$

Basta infatti ricavare λ dalla relazione

$$\varepsilon = \lambda\sigma = \lambda\sqrt{pq/n}.$$

Dunque per $n \rightarrow +\infty$ tale probabilità tende a 0. Questa è una prima formulazione di una *legge (debole) dei grandi numeri*.

Gioco equo.

Se ogni prova ha un costo C e in caso di successo si ottiene un pagamento uguale ad aC , la v.a. *guadagno complessivo*: $G = aCK - nC$ ha un valor medio

$$E(G) = nC(ap - 1).$$

Dunque il “gioco è vantaggioso, equo o svantaggioso per chi sostiene il costo C in ogni prova, secondo che sia $a > 1/p$, $a = 1/p$ oppure $a < 1/p$.

Variabile $N(\omega)$ geometrica.

La sua funzione generatrice dei momenti è data da

$$m_N(t) = \sum_{k=1}^{+\infty} e^{tk} q^{k-1} p = \frac{p}{q} \sum_{k=1}^{+\infty} (qe^t)^k = \frac{pe^t}{1 - qe^t},$$

per t sufficientemente piccolo ($qe^t < 1$). Dunque

$$\begin{aligned} m_N(t)|_{t=0} &= \frac{p}{1-q} = 1, \\ Dm_N(t)|_{t=0} &= \left(\frac{pe^t}{1-qe^t} + \frac{pe^t qe^t}{(1-qe^t)^2} \right) |_{t=0} = \\ &= \frac{pe^t}{(1-qe^t)^2} |_{t=0} = \frac{p}{(1-q)^2} = \frac{1}{p} = E(N). \end{aligned}$$

Dunque il tempo medio di attesa per un primo successo è il reciproco della probabilità di successo in una singola prova.

$$\begin{aligned} D^2 m_N(t)|_{t=0} &= \left(\frac{pe^t}{(1-qe^t)^2} + \frac{2p q e^{2t}}{(1-qe^t)^3} \right) |_{t=0} = \frac{1}{p} + \frac{2q}{p^2} = \frac{1+q}{p^2} = E(N^2), \\ \sigma^2 &= \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}. \end{aligned}$$

Variabile $N(\omega)$ binomiale negativa (di parametri n e r).

a) Metodo analitico. La distribuzione di N è data da

$$F(n+1) = \sum_{k=r}^n C_{k-1}^{r-1} p^r q^{k-r} = p^r \sum_{j=0}^{n-r} C_{j+r-1}^{r-1} q^j,$$

avendo posto $j = k - r$ ed essendo ovviamente $n \geq r$.

Conviene allora considerare la funzione ausiliaria

$$H(z) = \sum_{j=0}^{+\infty} C_{j+r-1}^{r-1} z^j = \frac{1}{(r-1)!} D^{r-1} \sum_{j=0}^{+\infty} z^{j+r-1} = \frac{1}{(r-1)!} D^{r-1} \frac{z^{r-1}}{1-z},$$

infatti $D^{r-1} z^{j+r-1} = (j+r-1)(j+r-2)\dots(j+1)z^j$. Allora, posto $s = r - 1$, vista la formula di Leibniz:

$$H(z) = \frac{1}{s!} D^s (z^s (1-z)^{-1}) = \frac{1}{s!} \sum_{i=0}^s C_s^i s(s-1)\dots(s-i+1)(s-i)\dots 1 \frac{z^{s-i}}{(1-z)^{s-i+1}},$$

essendo $D^i z^s = s(s-1)\dots(s-i+1)z^{s-i}$ e $D^j (1-z)^{-1} = 1.2\dots j(1-z)^{-1-j}$, pertanto

$$H(z) = (1-z)^{-1} \sum_{i=0}^s C_s^i 1^i \left(\frac{z}{1-z}\right)^{s-i} = (1-z)^{-1} \left(1 + \frac{z}{1-z}\right)^s = (1-z)^{-1-s} = (1-z)^{-r}.$$

Dunque $H(q) = (1-q)^{-r} = (p)^{-r}$ e $F(+\infty) = p^r (p)^{-r} = 1$.

La funzione $H(z)$ può essere utilizzata per calcolare le caratteristiche della v.a. N , ad esempio

$$\begin{aligned} E(N) &= \sum_{k=r}^{+\infty} k C_{k-1}^{r-1} p^r q^{k-r} = p^r \sum_{j=0}^{+\infty} (j+r) C_{j+r-1}^{r-1} q^j = \\ &= p^r (qH'(q) + rH(q)) = p^r (rq(1-q)^{-r-1} + r(1-q)^{-r}) = r \left(\frac{q}{p} + 1\right) = \frac{r}{p}. \end{aligned}$$

Analogamente per la funzione generatrice dei momenti:

$$m_N(t) = \sum_{k=r}^{+\infty} e^{kt} C_{k-1}^{r-1} p^r q^{k-r} = p^r e^{rt} \sum_{j=0}^{+\infty} C_{j+r-1}^{r-1} (qe^t)^j = p^r e^{rt} H(qe^t) = \left(\frac{pe^t}{1-qe^t}\right)^r.$$

Con un semplice esercizio di derivazione si trova

$$E(N) = Dm_N(t)|_{t=0} = \frac{r}{p}, \quad E(N^2) = D^2m_N(t)|_{t=0} = \frac{r(r+q)}{p^2},$$

pertanto

$$V(N) = \sigma^2 = E(N^2) - E(N)^2 = \frac{rq}{p^2}.$$

b) *Decomposizione in v.a. più semplici.*

La variabile binomiale negativa N può essere scritta come somma di r v.a. geometriche indipendenti:

$$N = N_1 + N_2 + \dots + N_r ,$$

dove N_1 è il tempo di attesa (numero di prove) per un primo successo, N_2 l'ulteriore tempo di attesa (ulteriore numero di prove) per un secondo successo... Chiaramente le N_k sono indipendenti ed hanno distribuzione geometrica. Siamo in presenza di una situazione dove non vi è memoria del passato (né influenza del futuro sul presente): la v.a. tempo di attesa per il prossimo successo non è condizionata dalla storia delle prove precedenti (da come si è giunti all'ultimo successo).

Allora

$$E(N) = E(N_1) + E(N_2) + \dots + E(N_r) = r\left(\frac{1}{p}\right) = \frac{r}{p} ,$$

$$V(N) = V(N_1) + V(N_2) + \dots + V(N_r) = r\left(\frac{q}{p^2}\right) = \frac{rq}{p^2} ,$$

$$m_N(t) = E(e^{Nt}) = E(e^{N_1t + N_2t + \dots + N_rt}) =$$

$$E(e^{N_1t} e^{N_2t} \dots e^{N_rt}) = \left(\frac{pe^t}{1 - qe^t}\right)^r.$$

2.12 Distribuzione ipergeometrica

Si considera una popolazione di N individui, dei quali esattamente r hanno una determinata caratteristica χ , mentre gli altri $s = N - r$ individui non hanno χ . Si estrae un campione di n individui: si possono estrarre gli elementi del campione successivamente, ma quelli già estratti non debbono essere più presi in considerazione per le estrazioni successive (“estrazioni senza reimmissione”). Tutti gli individui della popolazione possono essere estratti con ugual probabilità. Più precisamente: in ogni estrazione tutti gli individui rimasti possono essere estratti con la stessa probabilità. Risulta dall'analisi seguente che è equivalente procedere ad una “estrazione in blocco”, supponendo che, qualsiasi $k \leq N$, tutti i campioni di k individui possano essere estratti con la stessa probabilità.

Si considera la v.a. X che assume quale valore il numero di individui del campione con la proprietà χ . I valori possibili x di X sono vincolati da

$$x \leq n , x \leq r , \text{ e dunque } x \leq \min(n, r) ;$$

$$0 \leq x , y = n - x \leq s = N - r , \text{ e dunque } \max(0, n - s) \leq x .$$

Per l'estrazione in blocco è evidente che, considerando i casi possibili e quelli favorevoli:

$$P(X = x) = \frac{C_r^x C_s^y}{C_N^n} ,$$

formula che fornisce la densità della v.a. ipergeometrica X .

Infatti C_N^n è il numero dei campioni possibili; i campioni favorevoli si costruiscono ad esempio prima selezionando x individui tra gli r con la caratteristica χ , cosa che si può fare in C_r^x modi diversi e poi completando il campione con y individui selezionati tra gli s privi della caratteristica χ , cosa che si può fare in C_s^y modi diversi. Per ciascuna delle prime C_r^x selezioni si ha sempre lo stesso numero C_s^y di seconde selezioni possibili; dunque il numero totale di campioni favorevoli si ottiene moltiplicando i due coefficienti binomiali. (Il numero di campioni favorevoli si può mettere in corrispondenza biunivoca con $R \times S$, dove R è l'insieme dei sottoinsiemi formati da x elementi di un insieme di r elementi, e S è l'insieme dei sottoinsiemi formati da y elementi di un insieme di s elementi.)

Se si conviene che $C_n^k = 0$ per $k > n$, la formula precedente è valida per qualunque valore di x .

Esempio. Si estraggono casualmente da un mazzo di $N = 52$ carte $n = 5$ carte. La probabilità di avere tra le 5 carte esattamente 2 assi è

$$\frac{C_4^2 C_{48}^3}{C_{52}^5} \sim 0.0339.$$

Non è facile calcolare le caratteristiche della v.a. X direttamente dalla densità. Meglio scomporre X in v.a. semplici che si presentano naturalmente nel

Caso dell'estrazioni successive senza reinserimento.

Introduciamo allora uno spazio di riferimento Ω analogo a quello introdotto per lo schema di Bernoulli

$$\Omega = \{\omega \mid \omega = (\omega_1, \omega_2, \dots, \omega_n), \text{ con al più } r \omega_k = 1 \text{ e al più } s \omega_k = 0\}.$$

Si immagina quindi di estrarre successivi elementi dalla popolazione, senza riimmissione, indicando con $\omega_k = 1, 0$ il fatto che alla k -esima estrazione si è selezionato un individuo con, o senza, la caratteristica χ . Si può assegnare, in modo coerente con le ipotesi fatte, una probabilità ad ogni ω ; ad esempio se $n = 5$:

$$P(\{(0, 1, 1, 0, 0)\}) = \frac{s}{N} \frac{r}{N-1} \frac{r-1}{N-2} \frac{s-1}{N-3} \frac{s-2}{N-4} = P(\{(1, 0, 0, 1, 0)\}) = \dots$$

Infatti, introducendo le v.a. $X_k(\omega) = \omega_k$ e gli eventi $A_k = \{X_k = 1\}$, $A_k^c = \{X_k = 0\}$:

$$\begin{aligned} P(\{(0, 1, 1, 0, 0)\}) &= P(A_1^c \cap A_2 \cap A_3 \cap A_4^c \cap A_5^c) = P(A_1^c)P(A_2 \cap A_3 \dots A_5^c | A_1^c) = \\ &= P(A_1^c)P(A_2 | A_1^c)P(A_3 \cap \dots A_5^c | A_1^c \cap A_2) = \dots = \frac{s}{N} \frac{r}{N-1} \dots \end{aligned}$$

Nei calcoli precedenti si utilizza ripetutamente la **regola di calcolo**

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

che si verifica immediatamente, ricordando la definizione di probabilità condizionata

$$\frac{P(A \cap B \cap C)}{P(C)} = \frac{P(A \cap C)}{P(C)} \frac{P(B \cap (A \cap C))}{P(A \cap C)} .$$

Si vede immediatamente che la probabilità dell'esempio non dipende dall'ordine delle ω_k . Quindi ogni permutazione degli indici $1, 2, \dots, n$ induce una bijezione $\pi : \Omega \rightarrow \Omega$ tale che $P(\omega) = P(\pi(\omega))$. Allora le v.a. $X_k(\omega) = \omega_k$ hanno tutte la stessa distribuzione, quella ad esempio di X_1 :

$$P(X_1 = 1) = \frac{r}{N}, \quad P(X_1 = 0) = \frac{s}{N}$$

e quindi

$$E(X_1) = \frac{r}{N}, \quad E(X_1^2) = E(X_1) = \frac{r}{N} .$$

Infatti, se ad esempio π scambia ω_j con ω_k e lascia fisse le altre componenti, posto $\omega' = \pi(\omega)$, si ha

$$P(X_j = 1) = \sum_{\omega \in \{\omega | \omega_j = 1\}} P(\omega) = \sum_{\omega \in \{\omega | \omega_j = 1\}} P(\pi(\omega)) = \sum_{\omega' \in \{\omega' | \omega'_k = 1\}} P(\omega') = P(X_k = 1) .$$

I campioni ω con lo stesso numero x di individui con la caratteristica χ sono in numero di C_n^x , dunque

$$P(X = x) = C_n^x \frac{r(r-1)\dots(r-x+1)s(s-1)\dots(s-y+1)}{N(N-1)\dots(N-n+1)},$$

che coincide con il valore precedentemente calcolato per altra via, come si verifica facilmente ($x + y = n, r + s = N$):

$$\begin{aligned} \frac{\binom{r}{x} \binom{s}{y}}{\binom{N}{n}} &= \frac{n!(N-n)!}{N!} \frac{r!}{x!(r-x)!} \frac{s!}{y!(s-y)!} = \\ &= \frac{n!}{x!(n-x)!} \frac{r(r-1)\dots(r-x+1)s(s-1)\dots(s-y+1)}{N(N-1)\dots(N-n+1)}. \end{aligned}$$

Analogamente, sempre per la permutabilità delle ω_k a parità di probabilità, per $j \neq k$

$$P(X_j X_k = 1) = P(X_1 X_2 = 1) = \frac{r}{N} \frac{r-1}{N-1} .$$

Allora essendo $X = X_1 + X_2 + \dots + X_n$, risulta

$$E(X) = \sum_k E(X_k) = \frac{nr}{N},$$

$$E(X^2) = E\left(\sum_k X_k^2 + \sum_{i \neq j} X_i X_j\right) = n \frac{r}{N} + n(n-1) \frac{r}{N} \frac{r-1}{N-1},$$

$$V(X) = \sigma^2 = E(X^2) - E(X)^2 = \frac{nrs}{N^2} \frac{N-n}{N-1}.$$

Infatti

$$\begin{aligned} & n \frac{r}{N} + n(n-1) \frac{r}{N} \frac{r-1}{N-1} - \frac{n^2 r^2}{N^2} = \\ &= \frac{nr}{N^2(N-1)} (N(N-1) + N(n-1)(r-1) - (N-1)nr) = \\ &= \frac{nr}{N^2(N-1)} (N^2 - N + Nnr - Nr - Nn + N - Nnr + nr) = \\ &= \frac{nr}{N^2(N-1)} (N(N-r) - n(N-r)). \end{aligned}$$

Più in generale si consideri la situazione seguente: *una popolazione di N individui è divisa in l strati, rispettivamente di S_1, S_2, \dots, S_l elementi: $S_1 + S_2 + \dots + S_l = N$. Si estrae "a caso un campione ω di n elementi ($n \leq N$). Lo spazio di riferimento Ω è in questo caso l'insieme di tutti i possibili campioni:*

$$\text{card}(\Omega) = \binom{N}{n}.$$

Dicendo che si procede ad una scelta casuale di un campione si intende che tutti i possibili campioni sono equiprobabili:

$$P(\{\omega\}) = p = 1 / \binom{N}{n}.$$

Sia $X_j(\omega)$ la v.a. che conta il numero di elementi di ω appartenenti allo strato j -esimo :

$$X_1(\omega) + X_2(\omega) + \dots + X_l(\omega) = n.$$

Dati dei numeri $0 \leq x_j \leq S_j$ tali che $x_1 + x_2 + \dots + x_l = n$ si chiede la probabilità $P(A)$ che, eseguita l'estrazione, risulti $X_j(\omega) = x_j$ per ogni j , dove

$$A = \{\omega \mid X_j(\omega) = x_j \quad j = 1, 2, \dots, l\}.$$

È sufficiente contare il numero di elementi ω di A . Ogni campione $\omega \in A$ si può costruire con la procedura seguente: si selezionano arbitrariamente x_1 elementi dal primo strato, x_2 elementi dal secondo strato ... x_l elementi dal l -esimo strato. Questa procedura produce una sola volta tutti gli elementi di A , quindi:

$$\text{card}(A) = \binom{S_1}{x_1} \binom{S_2}{x_2} \dots \binom{S_l}{x_l}.$$

La probabilità richiesta è dunque:

$$\begin{aligned} P(A) &= \frac{\binom{S_1}{x_1} \binom{S_2}{x_2} \dots \binom{S_l}{x_l}}{\binom{N}{n}} = \\ &= \left(\frac{n!}{x_1! x_2! \dots x_l!} \frac{(N-n)!}{(S_1-x_1)! (S_2-x_2)! \dots (S_l-x_l)!} \right) / \frac{N!}{S_1! S_2! \dots S_l!}. \end{aligned}$$

2.13 Limite della distribuzione ipergeometrica per $N \rightarrow +\infty$

Se facciamo crescere il numero N di individui della popolazione, mantenendo il rapporto $r/N = p$ costante, e quindi costante il rapporto $(N-r)/N = q$, al limite troviamo la distribuzione binomiale, di parametri n e p :

$$C_n^x \frac{r(r-1)\dots(r-x+1)s(s-1)\dots(s-y+1)}{N(N-1)\dots(N-n+1)} \rightarrow C_n^x p^x q^y ,$$

$$E(X) = n \frac{r}{N} \rightarrow np , \quad V(X) = \frac{nrs}{N^2} \frac{N-n}{N-1} \rightarrow npq ,$$

restando n ed x costanti, e quindi costante $y = n - x$.

2.14 Distribuzione di Poisson qual limite della distribuzione binomiale

Facciamo tendere a $+\infty$ il numero delle prove indipendenti, facendo simultaneamente tendere a 0 la probabilità di successo nella singola prova, in modo che il prodotto $np = \lambda$ resti costante:

$$\begin{aligned} f_n(k) &= C_n^k p^k q^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow \phi(k) = \frac{\lambda^k}{k!} e^{-\lambda} , \quad k = 0, 1, 2, \dots \end{aligned}$$

La funzione $\phi(k)$ è una densità

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

Una variabile aleatoria X che abbia tale densità si dice *variabile di Poisson*. È facile calcolarne le caratteristiche:

$$m_X(t) = E(e^{tX}) = \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t-1)} ,$$

$$m_X(0) = 1 ,$$

$$Dm_X(t)|_{t=0} = \lambda e^t m_X(t)|_{t=0} = \lambda ,$$

$$D^2 m_X(t)|_{t=0} = (\lambda e^t)^2 m_X(t)|_{t=0} + \lambda e^t m_X(t)|_{t=0} = \lambda^2 + \lambda ,$$

dunque

$$E(X) = \lambda ,$$

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda , \quad \sigma_X = \sqrt{\lambda} .$$

Capitolo 3

Variabili aleatorie reali di tipo generale e variabili continue

3.1 Distribuzione, indipendenza

Ricordiamo che, se (Ω, \mathcal{A}, P) uno spazio di probabilità, si dice **variabile aleatoria reale (v.a.)** ogni funzione

$$X : \Omega \longrightarrow \mathbf{R}$$

misurabile, cioè tale che

$$\forall c \in \mathbf{R} \{ \omega \in \Omega \mid X(\omega) < c \} \in \mathcal{A}.$$

Definizione. La funzione reale di variabile reale

$$F(x) = P(X < x)$$

si dice **distribuzione** della v.a. X .

L'insieme $\{ X < x \}$ non decresce al crescere di x e quindi F è una funzione monotona non decrescente. Inoltre

$$\emptyset = \bigcap_{n \geq 1} \{ X < -n \} \quad , \quad \Omega = \bigcup_{n \geq 1} \{ X < n \}$$

e per le proprietà di continuità della misura di probabilità

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

Essendo $F(x)$ una funzione monotona, ammette in ogni punto limite destro $F(x+0)$ e limite sinistro $F(x-0)$ e l'insieme dei punti di discontinuità (salti) è al più numerabile. Con la definizione adottata, F risulta continua a sinistra: $F(x) = F(x-0)$. Infatti per ogni successione di numeri positivi a_n tendenti a 0 in modo decrescente

$$F(x) = P(\{X < x\}) = P(\bigcup_n \{X < x - a_n\}) = \lim_n P(\{X < x - a_n\}) = \lim_n F(x - a_n) ,$$

per la monotonia della successione $\{X < x - a_n\}$ e la continuità di P .

Alcuni autori adottano la definizione: $F(x) = P(\{X \leq x\})$. In tal caso F risulta continua a destra: $F(x) = F(x + 0)$.

La distribuzione determina univocamente la probabilità di tutti gli eventi concernenti la v.a. Più precisamente vale la

Proposizione. *Se X e Y hanno la stessa distribuzione F , per ogni Boreliano B di \mathbf{R} ($B \in \mathcal{B}$) risulta*

$$P(\{\omega \mid X(\omega) \in B\}) = P(\{\omega \mid Y(\omega) \in B\}) .$$

* * *

Dimostrazione. La tesi si può scrivere nella forma $P(X^{-1}(B)) = P(Y^{-1}(B))$. Consideriamo allora la famiglia di sottoinsiemi di \mathbf{R}

$$\mathcal{P} = \{ A \mid P(X^{-1}(A)) = P(Y^{-1}(A)) \} .$$

Essa è una σ -algebra, perché se $A, A_n \in \mathcal{P}$, con le A_n disgiunte, si ha

$$\begin{aligned} P(X^{-1}(A^c)) &= P(X^{-1}(A)^c) = 1 - P(X^{-1}(A)) = \\ &= 1 - P(Y^{-1}(A)) = P(Y^{-1}(A)^c) = P(Y^{-1}(A^c)) , \\ P(X^{-1}(\cup_n A_n)) &= P(\cup_n X^{-1}(A_n)) = \sum_n P(X^{-1}(A_n)) = \\ &= \sum_n P(Y^{-1}(A_n)) = P(\cup_n Y^{-1}(A_n)) = P(Y^{-1}(\cup_n A_n)) . \end{aligned}$$

Gli intervalli $] - \infty, c[$ appartengono per ipotesi a \mathcal{P} . Ma \mathcal{B} è la minima σ -algebra alla quale tali intervalli appartengono e quindi $\mathcal{B} \subseteq \mathcal{P}$. *q.e.d.*

& & &

La misura μ definita sui Boreliani di \mathbf{R} da

$$\mu(B) = P(X^{-1}(B))$$

è la **legge** di X . Alcuni autori chiamano μ distribuzione di X e allora F funzione di distribuzione.

Definizione. X e Y si dicono **indipendenti** se e solo se, prefissati comunque gli intervalli $[a, b[$ e $[c, d[$, gli eventi

$$B = \{a \leq X < b\} \quad \text{e} \quad C = \{c \leq Y < d\}$$

sono indipendenti ($P(B \cap C) = P(B)P(C)$).

Si può vedere che è sufficiente chiedere che gli eventi

$$U(x) = \{X < x\} \quad \text{e} \quad V(y) = \{Y < y\}$$

siano indipendenti per arbitrari varori reali di x e y . (Si ossrvi che ad esempio $B = U(b) - U(a)$ e $U(a) \subseteq U(b)$).

* * *

Ricordiamo che, data una famiglia \mathcal{F} di sottoinsiemi, si indica con $\sigma(\mathcal{F})$ la più piccola σ -algebra contenete \mathcal{F} . Essa viene detta **la σ -algebra generata da \mathcal{F}** . La definizione è corretta perchè si vede facilmente che:

- 1) Se $\{\mathcal{A}_t\}_{t \in T}$, dove T è un insieme arbitrario, è una collezione di σ -algre, allora $\bigcap_{t \in T} \mathcal{A}_t$ è ancora una σ -algebra.
 - 2) Esiste almeno una σ -algebra contenete \mathcal{F} , ad esempio $\mathcal{P}(X)$.
- Dunque $\sigma(\mathcal{F})$ è l'intersezione di tutte le σ -algre contenenti \mathcal{F} .

Sia $X : \Omega \rightarrow S$ una applicazione ovunque definita e \mathcal{A} una σ -algebra in S . Allora si vede facilmente che anche $X^{-1}(\mathcal{A})$ è una σ -algebra in Ω .

Sia $X : \Omega \rightarrow S$ una applicazione ovunque definita. Allora si può dimostrare che

$$\sigma(X^{-1}(\mathcal{F})) = X^{-1}(\sigma(\mathcal{F})) .$$

In uno spazio topologico i Boreliani sono gli elementi della σ -algebra generata dagli aperti (o, equivalentemente, dai chiusi)

Se \mathcal{B} è la famiglia dei Boreliani di \mathbf{R} e $X : \Omega \rightarrow \mathbf{R}$ è una v.a. reale, la σ -algebra generata da X è definita da

$$\sigma(X) = \sigma(X^{-1}(\mathcal{B})) = X^{-1}(\sigma(\mathcal{B})) = X^{-1}(\mathcal{B}) .$$

Si dice che due famiglie di eventi \mathcal{F} e \mathcal{G} sono indipendenti se e solo se ogni $F \in \mathcal{F}$ e ogni $G \in \mathcal{G}$ sono indipendenti: $P(F \cap G) = P(F)P(G)$.

Proposizione. Due v.a. reali X e Y sono indipendenti se e solo se $\sigma(X)$ e $\sigma(Y)$ sono indipendenti.

Premettiamo alla dimostrazione alcuni importanti risultati.

Definizione. Una famiglia \mathcal{F} di sottoinsiemi di Ω si dice che è una λ -classe se e solo se $\Omega \in \mathcal{F}$ e:

- 1) $F, G \in \mathcal{F}$, $F \subseteq G \Rightarrow G - F \in \mathcal{F}$;
- 2) $F_n \in \mathcal{F}$, $F_j \cap F_k = \emptyset$ ($j \neq k$) $\Rightarrow \bigcup_n F_n \in \mathcal{F}$.

In luogo di 2) si può invece chiedere

$$2^*) F_n \in \mathcal{F}, F_n \subseteq F_{n+1} \Rightarrow \bigcup_n F_n \in \mathcal{F} .$$

Infatti gli insiemi $F_{n+1} - F_n \in \mathcal{F}$ per 1) e la loro unione è disgiunta e coincide con quella degli F_n .

Proposizione. Se $\{\mathcal{F}_t\}_{t \in T}$, dove T è un insieme arbitrario, è una collezione di λ -classi, allora $\bigcap_{t \in T} \mathcal{F}_t$ è ancora una λ -classe.

La dimostrazione consiste in verifiche immediate.

Si indica con $\lambda(\mathcal{F})$ la λ -classe generata da \mathcal{F} , cioè la più piccola λ -classe contenente \mathcal{F} .

Teorema. Sia \mathcal{C} una collezione arbitraria di eventi. La famiglia \mathcal{F} di tutti gli eventi che sono indipendenti da ogni elemento $C \in \mathcal{C}$ è una λ -classe.

Dimostrazione. Basta vedere che la famiglia \mathcal{F}_C degli eventi indipendenti da C è una λ -classe, perchè

$$\mathcal{F} = \bigcap_{C \in \mathcal{C}} \mathcal{F}_C .$$

In primo luogo osserviamo che Ω è indipendente da C . Siano poi $G, F, F_n \in \mathcal{F}_C$, con $F \subseteq G$ e gli F_n disgiunti:

$$\begin{aligned} P((G - F) \cap C) &= P(G \cap C - F \cap C) = P(G \cap C) - P(F \cap C) = \\ &= P(G)P(C) - P(F)P(C) = P(G - F)P(C) ; \\ P((\cup_n F_n) \cap C) &= P(\cup_n (F_n \cap C)) = \sum_n P(F_n \cap C) = \\ &= \sum_n P(F_n)P(C) = P(\cup_n F_n)P(C) \quad q.e.d.. \end{aligned}$$

Definizione. Una famiglia di eventi \mathcal{F} si dice π -sistema se e solo se è chiusa per intersezioni finite:

$$F, G \in \mathcal{F} \Rightarrow F \cap G \in \mathcal{F} .$$

Proposizione. Se una famiglia di eventi \mathcal{F} è sia una λ -classe che un π -sistema, allora essa è una σ -algebra.

Dimostrazione. Essendo \mathcal{F} una λ -classe si ha

$$A \in \mathcal{F} \Rightarrow A^c = \Omega - A \in \mathcal{F} .$$

Inoltre, se $A_n \in \mathcal{F}$, definendo

$$S_1 = A_1 , S_{n+1} = S_n \cup (A_{n+1} - A_{n+1} \cap S_n) ,$$

si vede per induzione che $S_n \in \mathcal{F}$ (sono presenti soltanto intersezioni finite, differenze tra un insieme ed un suo sottoinsieme e unioni disgiunte). Ma allora

$$\cup_n A_n = \cup_n S_n \in \mathcal{F} ,$$

perchè le S_n sono una successione monotona. *q.e.d.*

Teorema (sulle λ -classi o sulle classi monotone). Se \mathcal{F} è un π -sistema allora $\lambda(\mathcal{F}) = \sigma(\mathcal{F})$.

Dimostrazione. Sia

$$\mathcal{H} = \{ H \in \lambda(\mathcal{F}) \mid \forall F \in \mathcal{F} \ H \cap F \in \lambda(\mathcal{F}) \} .$$

Ovviamente $\lambda(\mathcal{F}) \supseteq (\mathcal{H} \supseteq \mathcal{F})$. Inoltre \mathcal{H} è una λ -classe:

$$G, H \in \mathcal{H} , H \subseteq G \Rightarrow G - H \in \lambda(\mathcal{F}) \text{ e}$$

$$\forall F \in \mathcal{F} \ (G - H) \cap F = G \cap F - H \cap F \in \lambda(\mathcal{F}) ,$$

quindi $G - H \in \mathcal{H}$;

$$H_n \in \mathcal{H} , H_n \subseteq H_{n+1} \Rightarrow \cup_n H_n \in \lambda(\mathcal{F}) , \forall F \in \mathcal{F} \ (\cup_n H_n) \cap F = \cup_n (H_n \cap F) \in \lambda(\mathcal{F}) .$$

Allora $\mathcal{H} = \lambda(\mathcal{F})$. Sia poi

$$\mathcal{K} = \{ K \in \mathcal{H} = \lambda(\mathcal{F}) \mid \forall H \in \lambda(\mathcal{F}) \ K \cap H \in \lambda(\mathcal{F}) \} .$$

Per la prima parte della dimostrazione $\mathcal{K} \supseteq \mathcal{F}$. Inoltre \mathcal{K} è una λ -classe:

$$M, K \in \mathcal{K} , K \subseteq M \Rightarrow M - K \in \lambda(\mathcal{F}) \text{ e}$$

$$\forall H \in \lambda(\mathcal{F}) \ (M - K) \cap H = M \cap H - K \cap H \in \lambda(\mathcal{F}) ,$$

quindi $M - K \in \mathcal{K}$;

$$K_n \in \mathcal{K} , K_n \subseteq K_{n+1} \Rightarrow \cup_n K_n \in \lambda(\mathcal{F}) , \forall H \in \lambda(\mathcal{F}) \ (\cup_n K_n) \cap H = \cup_n (K_n \cap H) \in \lambda(\mathcal{F}) .$$

Allora $\mathcal{K} = \lambda(\mathcal{F})$ e, poiché \mathcal{K} è un π -sistema, \mathcal{K} è una σ -algebra contenete \mathcal{F} . D'altra parte ovviamente $\lambda(\mathcal{F}) \subseteq \sigma(\mathcal{F})$ (tutti gli insiemi della prima classe devono essere presenti nella seconda) e finalmente $\mathcal{K} = \lambda(\mathcal{F}) = \sigma(\mathcal{F})$. *q.e.d.*

Possiamo ora dimostrare che se due v.a. (reali) X e Y sono indipendenti allora $\sigma(X)$ e $\sigma(Y)$ sono indipendenti.

Dimostrazione. Sia \mathcal{R} la famiglia degli intervalli aperti e semi-infiniti inferiormente di \mathbf{R} , precisamente la famiglia degli insiemi $] -\infty, c[$, con c reale arbitrario. $\sigma(\mathcal{R}) = \mathcal{B}$, cioè i Boreliani di \mathbf{R} sono generati da \mathcal{R} . Ovviamente \mathcal{R} è un π -sistema e tali sono dunque le famiglie $X^{-1}(\mathcal{R})$ e $Y^{-1}(\mathcal{R})$, indipendenti per ipotesi. Sia

$$\mathcal{Y} = \{ F \mid F \text{ è indipendente da } X^{-1}(\mathcal{R}) \} .$$

Essendo \mathcal{Y} una λ -classe contenete il π -sistema $Y^{-1}(\mathcal{R})$, si ha

$$\mathcal{Y} \supseteq \lambda(Y^{-1}(\mathcal{R})) = \sigma(Y^{-1}(\mathcal{R})) = Y^{-1}(\sigma(\mathcal{R})) = Y^{-1}(\mathcal{B}) = \sigma(Y) .$$

Dunque $X^{-1}(\mathcal{R})$ e $\sigma(Y)$ sono indipendenti. Sia ora

$$\mathcal{X} = \{ E \mid E \text{ è indipendente da } \sigma(Y) \} .$$

Essendo \mathcal{X} una λ -classe contenete il π -sistema $X^{-1}(\mathcal{R})$, si ha

$$\mathcal{X} \supseteq \lambda(X^{-1}(\mathcal{R})) = \sigma(X^{-1}(\mathcal{R})) = X^{-1}(\sigma(\mathcal{R})) = X^{-1}(\mathcal{B}) = \sigma(X) .$$

Dunque $\sigma(X)$ e $\sigma(Y)$ sono indipendenti. *q.e.d.*

& \$ &

Più in generale:

Definizione. Le v.a. X_1, X_2, \dots, X_n si dicono **indipendenti (mutuamente)** se e solo se **prefissati comunque gli intervalli** $[a_1, b_1[, [a_2, b_2[, \dots, [a_n, b_n[$, **gli eventi**

$$B_1 = \{a_1 \leq X_1 < b_1\} , B_2 = \{a_2 \leq X_2 < b_2\} , \dots B_n = \{a_n \leq X_n < b_n\}$$

sono mutuamente indipendenti.

È sufficiente chiedere che per valori reali arbitrari $x_1 \dots x_n$ gli eventi

$$U(x_1) = \{X_1 < x_1\} \dots U(x_n) = \{X_n < x_n\}$$

siano mutuamente indipendenti.

* * *

Le famiglie $\mathcal{F}_j, j = 1, 2, \dots, n$, si dicono mutuamente indipendenti se per ogni scelta di $A_j \in \mathcal{F}_j$ gli eventi A_1, A_2, \dots, A_n risultano mutuamente indipendenti.

Le v.a. (reali) X_1, X_2, \dots, X_n sono mutuamente indipendenti se e solo se le σ -algebre da esse generate $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)$ sono mutuamente indipendenti.

& & &

Osservazione. Per alcuni risultati è sufficiente che le v.a. X_j siano a due a due indipendenti e non necessariamente indipendenti (mutuamente).

3.2 Valor medio

Definizione. Sia X una v.a. Si dice **valor medio o speranza matematica** $E(X)$ di X l'integrale di X su Ω rispetto alla misura di probabilità P .

Tale integrale, se esiste, è definito come limite di somme (o serie) integrali. Per esempio, posto

$$A_n^k = \left\{ \omega \in \Omega \mid \frac{k}{n} \leq X(\omega) < \frac{k+1}{n} \right\} .$$

si ha

$$E(X) = \int_{\Omega} X dP = \lim_n \sum_{k \in \mathcal{Z}} \frac{k}{n} P(A_n^k) .$$

Le serie che compaiono nella formula devono essere *assolutamente convergenti* (ricordiamo che la convergenza assoluta è condizione necessaria e sufficiente per l'indipendenza della somma dall'ordine dei termini della serie: non vogliamo che la definizione di valor medio dipenda dall'ordine con il quale i valori approssimanti (k/n nell'esempio) sono presi in esame).

Si indica l'integrale anche con la notazione

$$\int_{\Omega} X(\omega) dP(\omega) .$$

Se X_n è la v.a. discreta che assume il valore k/n sull'insieme A_n^k (le A_n^j , $j \in \mathbf{Z}$, costituiscono una *partizione* di Ω), X_n converge uniformemente a X su Ω . Infatti per ogni ω risulta $|X_n(\omega) - X(\omega)| < \frac{1}{n}$.

Più in generale possiamo considerare una qualunque successione di v.a. discrete X_n tali che

1) $X_n \rightarrow X$ uniformemente in Ω :

$$\forall \varepsilon > 0 \exists \nu \forall n > \nu \forall \omega \in \Omega \quad |X_n(\omega) - X(\omega)| < \varepsilon ;$$

2) X_n ammette valor medio: $E(X_n) = \sum_k x_n^k P(B_n^k)$,

dove x_n^k è il valore costante assunto da X_n su B_n^k , le B_n^k costituiscono una partizione di Ω e si chiede che la serie converga *assolutamente* . Allora:

$$E(X) = \int_{\Omega} X dP = \lim_n E(X_n).$$

Infatti è facile verificare che:

a) la convergenza uniforme delle X_n implica la convergenza delle $E(X_n)$, in quanto

$$|E(X_n) - E(X_m)| \leq \int_{\Omega} |X_n - X_m| dP \leq \varepsilon \text{ se } n, m \geq \nu \text{ opportuno} ,$$

essendo X_n di Cauchy; dunque $E(X_n)$ è di Cauchy in \mathbf{R} e quindi converge ad un limite l ;

b) il limite non dipende dalla successione approssimante, perchè se Y_n converge uniformemente ad X ed $E(Y_n)$ converge a m , anche la successione mista $X_1, Y_1, X_2, Y_2, \dots$ converge uniformemente a X ; pertanto la successione $E(X_1), E(Y_1), \dots$ deve convergere e allora necessariamente a $l = m$.

È inoltre chiaro che se X è discreta e prende i valori x^k sugli eventi B^k , il suo valor medio può essere definito sia direttamente, sia come integrale:

$$E(X) = \sum_k x^k P(B^k) = \int_{\Omega} X dP ,$$

basta infatti pensare X come limite uniforme della successione approssimante costante X, X, \dots

Le proprietà di linearità si conservano nel passaggio al limite e dunque

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y) .$$

Inoltre, se indichiamo, ricorrendo al consueto abuso di linguaggio, con c la v.a. costante uguale a c , abbiamo

$$E(1) = 1 \quad , \quad E(X - E(X)) = 0 .$$

Non è vero in generale che il valor medio del prodotto sia uguale al prodotto dei valori medi, ma una condizione *sufficiente* è che X e Y siano *indipendenti*. In tal caso

$$E(XY) = E(X)E(Y) .$$

Dimostriamo che $E(XY) = E(X)E(Y)$ se X e Y sono indipendenti, supponendo per semplicità che siano v.a. limitate. Poniamo

$$\begin{aligned} B_n^j &= \left\{ \frac{j}{n} \leq X < \frac{j+1}{n} \right\} , \quad C_n^k = \left\{ \frac{k}{n} \leq Y < \frac{k+1}{n} \right\} , \\ A_n^{j,k} &= B_n^j \cap C_n^k = \left\{ \frac{j}{n} \leq X < \frac{j+1}{n} \text{ e } \frac{k}{n} \leq Y < \frac{k+1}{n} \right\} , \\ X_n &= \frac{j}{n} \text{ su } B_n^j , \quad Y_n = \frac{k}{n} \text{ su } C_n^k , \end{aligned}$$

e quindi

$$X_n Y_n = \frac{jk}{n^2} \text{ su } A_n^{j,k} .$$

La v.a. $X_n Y_n$ è discreta e tende uniformemente a XY :

$$|XY - X_n Y_n| \leq |X - X_n| |Y| + |X_n| |Y - Y_n| \leq \frac{1}{n} (\sup |X| + 1 + \sup |Y|) ,$$

per n sufficientemente grande perché si abbia $|X_n| \leq |X| + 1$.

Poiché X e Y sono indipendenti

$$\begin{aligned} P(A_n^{j,k}) &= P(B_n^j)P(C_n^k) \quad \text{e} \\ E(X_n Y_n) &= \sum_{j,k} \frac{jk}{n^2} P(A_n^{j,k}) = \sum_{j,k} \frac{jk}{n^2} P(B_n^j)P(C_n^k) = \\ &= \left(\sum_j \frac{j}{n} P(B_n^j) \right) \left(\sum_k \frac{k}{n} P(C_n^k) \right) = E(X_n)E(Y_n) \end{aligned}$$

(la famiglia di numeri $jk/n^2 P(A_n^{j,k})$ è sommabile) e quindi passando al limite si ottiene il risultato. *q.e.d.*

La nozione di valor medio acquista particolare significato alla luce del

Legge forte dei grandi numeri (Kolmogorov). Siano X_1, X_2, \dots v.a. indipendenti e con la stessa distribuzione della v.a. X . Allora esiste $\mu \in \mathbf{R}$ tale che

$$P\left(\lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mu\right) = 1$$

se e solo se X ammette valor medio $E(X) = \int_{\Omega} X dP$. In tal caso $E(X) = \mu$. Osserviamo che, se due v.a. X e Y hanno la stessa distribuzione F , allora $E(X) = E(Y)$. Infatti, per ogni n, k risulta

$$P\left(\left\{\omega \in \Omega \mid \frac{k}{n} \leq X(\omega) < \frac{k+1}{n}\right\}\right) = P\left(\left\{\omega \in \Omega \mid \frac{k}{n} \leq Y(\omega) < \frac{k+1}{n}\right\}\right)$$

e basta allora riferirsi alla definizione dell'integrale di X e di Y .

3.3 Variabili aleatorie continue

La v.a. si dice *continua* se esiste una funzione reale di variabile reale $f(x)$ integrabile (nel senso di Lebesgue), detta *densità* di X , tale che per ogni intervallo $[a, b[$ si abbia

$$P(a \leq X < b) = \int_a^b f(x) dx .$$

È equivalente chiedere che per ogni $x \in \mathbf{R}$ si abbia

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

Valgono ovviamente le seguenti proprietà

$$f(x) \geq 0 \quad , \quad \int_{-\infty}^{+\infty} f(x) dx = 1 ,$$

$$F \text{ è continua (assolutamente) e } P(a \leq X < b) = F(b) - F(a) .$$

Se B è un qualunque insieme *Boreliano* di \mathbf{R} , risulta

$$P(X \in B) = \int_B f(x) dx .$$

Essendo f integrabile, per l'assoluta continuità dell'integrale, se la misura di B tende a 0 allora $P(X \in B)$ tende a 0. Questo dimostra la continuità di F e permette di osservare che per ogni valore x

$$P(X = x) = 0 .$$

Naturalmente, pur valendo l'additività numerabile, ma non essendo numerabile l'insieme dei punti di un intervallo $[a, b[$ di \mathbf{R} , $P([a, b[)$ può essere diverso da 0 senza alcuna contraddizione.

Se la v.a. X ammette la densità f , risulta

$$E(X) = \int_{\Omega} X dP = \int_{\mathbf{R}} xf(x)dx .$$

Infatti, posto $A_n^k = \{\frac{k}{n} \leq X < \frac{k+1}{n}\}$, si ha

$$\int_{\Omega} X dP = \lim_n \sum_k \frac{k}{n} P(A_n^k) = \lim_n \sum_k \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx ,$$

ma

$$|\frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx - \int_{\frac{k}{n}}^{\frac{k+1}{n}} xf(x)dx| \leq \frac{1}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx ,$$

dunque, sommando su k e usando la disuguaglianza triangolare

$$|\sum_k \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx - \int_{\mathcal{R}} xf(x)dx| \leq \frac{1}{n} \int_{\mathcal{R}} f(x)dx = \frac{1}{n} .$$

Passando al limite per $n \rightarrow \infty$ si ottiene il risultato.

3.4 Funzioni di variabili aleatorie

Sia $H(x)$ una funzione reale di variabile reale *Boreliana*, cioè tale che se $B \subseteq \mathcal{R}$ è un Boreliano di \mathbf{R} , allora $H^{-1}(B)$ è ancora un Boreliano di \mathbf{R} . Perchè H sia Boreliana è sufficiente che per ogni c $H^{-1}(]-\infty, c])$ sia un insieme di Borel. È noto che le funzioni continue sono Boreliane.

Definiamo la v.a. $H(X)$ “funzione della v.a. X mediante la relazione

$$H(X)(\omega) = H(X(\omega)) \text{ ovvero } H(X) = H \circ X .$$

Se X è continua ed f è la sua densità, il valor medio di H si può calcolare nel modo seguente

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

Sia infatti

$$B_n^j = \{\omega | \frac{j}{n} \leq H(X(\omega)) < \frac{j+1}{n}\} = \{\omega | X(\omega) \in I_n^j = H^{-1}([\frac{j}{n}, \frac{j+1}{n}[))\}$$

e dunque

$$P(B_n^j) = \int_{I_n^j} f(x)dx ,$$

allora

$$|\sum_j \frac{j}{n} P(B_n^j) - \sum_j \int_{I_n^j} H(x)f(x)dx| \leq \sum_j \int_{I_n^j} |\frac{j}{n} - H(x)|f(x)dx \leq \frac{1}{n} ,$$

ma la prima serie converge a $\int_{\Omega} H(X)dP$, mentre la seconda, per la σ -additività dell'integrale è uguale a $\int_{\mathbf{R}} H(x)f(x)dx$. Dunque per $n \rightarrow +\infty$ si ha

$$E(H(X)) = \int_{\Omega} H(X)dP = \int_{\mathbf{R}} H(x)f(x)dx .$$

La distribuzione di H è $G(y) = P(H(X) < y)$.

Osservazione. Se due v.a. X e Y hanno la stessa distribuzione F e h è una funzione Boreliana da \mathbf{R} in \mathbf{R} (cioè tale che per ogni Boreliano B si ha che $h^{-1}(B)$ è un Boreliano) allora le v.a. $h(X)$ e $h(Y)$ hanno la stessa distribuzione.

Infatti, se B è il Boreliano $h^{-1}(] - \infty, c[)$, risulta

$$F_{h(X)}(c) = P(\{h(X) < c\}) = P(\{X^{-1}(h^{-1}(] - \infty, c[))\}) = P(\{X^{-1}(B)\}) =$$

$$P(\{Y^{-1}(B)\}) = P(\{Y^{-1}(h^{-1}(] - \infty, c[))\}) = P(\{h(Y) < c\}) = F_{h(Y)}(c) ,$$

perché, avendo X e Y la stessa distribuzione, $P(\{X \in B\}) = P(\{Y \in B\})$, qualsiasi $c \in \mathbf{R}$.

Di conseguenza $E(h(X)) = E(h(Y))$.

La distribuzione G di $H(X)$ può essere eventualmente continua ed ammettere una densità g :

$$G(y) = \int_{-\infty}^y g(t)dt .$$

In tal caso, se si dispone di g , si può calcolare $E(H(X))$ anche mediante la formula

$$E(H(X)) = \int_{\mathbf{R}} g(y)dy .$$

Se ad esempio $H(x) = x^2$ ed f è la densità di X , per $c > 0$ si ha

$$G(c) = P(X^2 < c) = P(-\sqrt{c} < X < +\sqrt{c}) = \int_{-\sqrt{c}}^{+\sqrt{c}} f(x)dx ,$$

e, posto $y = x^2$, quindi $x = \pm\sqrt{y}$, $dy = 2xdx$, $dx = dy/(\pm 2\sqrt{y})$:

$$G(c) = - \int_c^0 \frac{f(-\sqrt{y})}{2\sqrt{y}} dy + \int_0^c \frac{f(\sqrt{y})}{2\sqrt{y}} dy = \int_0^c \frac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}} dy .$$

Dunque la densità di X^2 è

$$g(y) = \begin{cases} 0 & y \leq 0 \\ \frac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}} & y > 0 \end{cases} .$$

Se X e Y sono v.a. reali indipendenti e f e g sono funzioni Boreliane da \mathbf{R} in \mathbf{R} , allora si può dimostrare che $f(X)$ e $g(Y)$ sono indipendenti.

* * *

Indicando, come al solito, con \mathcal{B} i Boreliani di \mathbf{R} , si ha

$$\sigma(f(X)) = \{(f \circ X)^{-1}(B)\}_{B \in \mathcal{B}} = \{X^{-1}(f^{-1}(B))\}_{B \in \mathcal{B}} \subseteq \{X^{-1}(B)\}_{B \in \mathcal{B}} = \sigma(X),$$

perché gli eventi $f^{-1}(B)$ sono alcuni dei (in qualche caso tutti i) Boreliani. Allora $\sigma(f(X))$ e $\sigma(g(Y))$ sono famiglie indipendenti in quanto sottofamiglie rispettivamente di $\sigma(X)$ e $\sigma(Y)$.

& & &

3.5 Varianza

Per variabili aleatorie reali X arbitrarie la varianza $V(X)$ è

$$V(X) = E((X - E(X))^2) = \int_{\Omega} (X - E(X))^2 dP.$$

Se X è continua, con densità f , indicando con μ la media di X , si ha

$$V(X) = \int_{\mathcal{R}} (x - \mu)^2 f(x) dx.$$

La varianza si può calcolare mediante la formula

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

Infatti

$$E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - 2E(X)E(X) + E(X)^2.$$

Per la varianza valgono le proprietà:

$$V(\lambda X) = \lambda^2 V(X);$$

e se le v.a. X, Y sono *indipendenti*

$$V(X + Y) = V(X) + V(Y).$$

Infatti

$$\begin{aligned} E((\lambda X - E(\lambda X))^2) &= E(\lambda^2(X - E(X))^2); \\ E((X + Y - E(X + Y))^2) &= E((X - E(X))^2) + 2E((X - E(X))(Y - E(Y))) + E((Y - E(Y))^2) = \\ &= E((X - E(X))^2) + E((Y - E(Y))^2), \end{aligned}$$

perchè il valor medio centrale, vista l'indipendenza di X, Y , è il prodotto dei valori medi, entrambi nulli.

Affinché la varianza della somma di due v.a. X e Y sia uguale alla somma delle loro varianze è sufficiente (e necessario) che X e Y abbiano covarianza nulla:

$$\text{Cov}(X, Y) \stackrel{\text{def}}{=} E((X - E(X))(Y - E(Y))) = 0 .$$

Ovviamente, se X e Y sono indipendenti, esse hanno covarianza nulla.

Più in generale, se X_1, X_2, \dots, X_n sono a due a due indipendenti, indicando la loro *media aritmetica* con

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n} ,$$

si ha

$$E(M_n) = \frac{1}{n} \sum_k E(X_k) , \quad V(M_n) = \frac{1}{n^2} \sum_k V(X_k) .$$

In particolare, se tutte le v.a. X_k hanno la stessa media μ e varianza σ^2

$$E(M_n) = \mu , \quad V(M_n) = \frac{\sigma^2}{n} .$$

3.6 Disuguaglianza di Chebyshev

Se X è una v.a. reale ≥ 0 e dotata di media $\mu > 0$ finita, per ogni $c > 0$ vale la **disuguaglianza di Markov**

$$P(X \geq c\mu) \leq \frac{1}{c}$$

Infatti

$$\begin{aligned} \mu &= \int_{\{\omega | X(\omega) < c\mu\}} X(\omega) dP(\omega) + \int_{\{\omega | X(\omega) \geq c\mu\}} X(\omega) dP(\omega) \geq \\ &\geq c\mu \int_{\{\omega | X(\omega) \geq c\mu\}} dP(\omega) = c\mu P(X \geq c\mu) . \end{aligned}$$

Se X ammette la densità f si ha naturalmente

$$P(X \geq c\mu) = \int_{c\mu}^{+\infty} f(x) dx ,$$

che si può stimare in modo più preciso disponendo di opportune approssimazioni di f .

La **disuguaglianza di Chebyshev** per una v.a. dotata media μ e di varianza σ^2 finite, afferma che

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2} .$$

Essa si può dedurre immediatamente ponendo $Y = (X - \mu)^2$, per la quale $E(Y) = \sigma^2$, e applicando a Y la disuguaglianza di Markov, perchè

$$P(|X - \mu| \geq c\sigma) = P(Y \geq c^2\sigma^2) .$$

Se X ammette la densità f si ha naturalmente

$$P(|X - \mu| \geq c\sigma) = \int_{-\infty}^{\mu - c\sigma} f(x)dx + \int_{\mu + c\sigma}^{+\infty} f(x)dx .$$

Se ad esempio se $X \sim \mathcal{N}(\mu, \sigma^2)$ (v. sezione 10), introducendo $Y = \sigma^{-1}(X - \mu)$, che è normale standard, si trova

$$P(|X - \mu| \leq \sigma) = P(|Y| \leq 1) = 1 - 2\Phi(-1) = 1 - 2 \int_{-\infty}^{-1} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \sim 0.6826 ,$$

$$P(|X - \mu| \leq 2\sigma) = P(|Y| \leq 2) = 1 - 2\Phi(-2) = 1 - 2 \int_{-\infty}^{-2} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \sim 0.9544 ,$$

$$P(|X - \mu| \leq 3\sigma) = P(|Y| \leq 3) = 1 - 2\Phi(-3) = 1 - 2 \int_{-\infty}^{-3} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \sim 0.9974 .$$

3.7 Funzione generatrice dei momenti

Per una v.a. X si indica con m_X la *funzione generatrice dei momenti*, definita da

$$m_X(t) = E(e^{tX}) = \int_{\Omega} e^{tX} dP ,$$

per tutti i valori di t per i quali l'integrale converge. Se X ammette la densità f , allora

$$m_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx .$$

Supponiamo che esista $t_0 > 0$ tale che l'integrale precedente converga per $\pm t_0$, allora $m_X(t)$ è definita in $[-t_0, t_0]$ ed X ammette momenti $E(X^k)$ di qualsiasi ordine:

$$\int_{-\infty}^{+\infty} |x|^k f(x) dx \leq \frac{k!}{t_0^k} \left(\int_{-\infty}^0 e^{-t_0 x} f(x) dx + \int_0^{+\infty} e^{t_0 x} f(x) dx \right) .$$

Si osservi che

$$\int_{-\infty}^0 e^{-t_0 x} f(x) dx + \int_0^{+\infty} e^{t_0 x} f(x) dx = \int_{-\infty}^{+\infty} e^{|t_0 x|} f(x) dx ,$$

e che $t_0^k |x|^k / k! \leq e^{|t_0 x|}$. Inoltre si ha

$$E(e^{tX}) = \sum_0^{+\infty} \frac{t^k}{k!} E(X^k)$$

per il teorema sulla convergenza dominata:

$$\left| \sum_0^n \frac{(tx)^k}{k!} \right| \leq \sum_0^n \frac{|tx|^k}{k!} \leq e^{|tx|} \leq e^{|t_0 x|} .$$

Dunque, se $m_X(t)$ è definita in un intorno di 0, essa è analitica in t e si ha

$$E(X^k) = \frac{d^k}{dt^k} m_X(t)|_{t=0} .$$

Si verificano immediatamente le seguenti regole di calcolo

1) Se $Y = aX + b$, dove a e b sono costanti

$$m_Y(t) = e^{bt} m_X(at) ,$$

infatti

$$E(e^{(aX+b)t}) = E(e^{(at)X} e^{bt}) .$$

2) Se X e Y sono *indipendenti* si ha

$$m_{X+Y}(t) = m_X(t) m_Y(t) ,$$

infatti

$$E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY})$$

perchè e^{tX} e e^{tY} sono indipendenti come X e Y .

Si può dimostrare che due v.a. hanno la stessa distribuzione se hanno la stessa funzione generatrice dei momenti.

Esempio di applicazione. Se X_1 e X_2 sono v.a. indipendenti con distribuzione di Poisson di parametri λ_1 e λ_2 , allora $X_1 + X_2$ ha funzione generatrice dei momenti

$$e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

e dunque $X_1 + X_2$ è una v.a. di Poisson di parametro $(\lambda_1 + \lambda_2)$.

3.8 Distribuzione uniforme

Si dice che la v.a. X ha una **distribuzione uniforme** sull'intervallo $[a, b]$ se ammette una densità f_X costante su tale intervallo e nulla fuori di esso:

$$f_X(x) = \begin{cases} 0 & x < a \text{ oppure } x > b \\ \frac{1}{b-a} & a \leq x \leq b . \end{cases}$$

Il valore costante deve essere $\frac{1}{b-a}$ perchè l'integrale di f su \mathcal{R} deve valere 1.

In particolare, se si considera l'intervallo $[0, 1]$, la densità è la funzione caratteristica dell'intervallo

$$f_X(x) = \chi_{[0,1]}(x) .$$

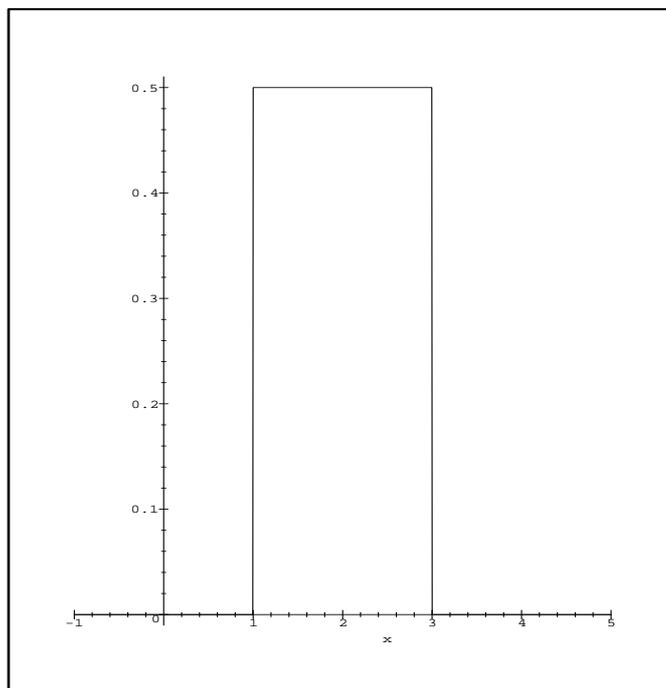


Figura 3.1: Densità uniforme, con $a = 1, b = 3$.

La distribuzione è

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases} .$$

Per ogni k il momento di ordine k è

$$E(X^k) = \int_0^1 x^k dx = \frac{1}{k+1}$$

e quindi il valor medio, la varianza e la deviazione standard sono

$$E(X) = \frac{1}{2}, \quad V(X) = E(X^2) - E(X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}, \quad \sigma_X = \frac{1}{2\sqrt{3}} .$$

La funzione generatrice dei momenti è

$$E(e^{tX}) = \int_0^1 e^{tx} dx = \frac{e^t - 1}{t} = \sum_0^{+\infty} \frac{t^k}{k!(k+1)} .$$

I coefficienti della serie permettono di ritrovare immediatamente tutti i momenti di ordine k .

3.9 Distribuzione di Cauchy

Sia A un punto del piano a distanza $R > 0$ da una retta r : ad esempio, in coordinate cartesiane, sia r l'asse delle ascisse ed $A(0, R)$. Si lanci un proiettile da A verso r in una direzione "casuale", nel senso che la direzione di lancio forma con il semiasse negativo delle ordinate un angolo aleatorio ϕ distribuito uniformemente nell'intervallo $]-\pi/2, +\pi/2[$. Il proiettile colpisce r in un punto di ascissa aleatoria $X = R \cdot \tan(\phi)$. Allora X è funzione di ϕ e

$$\begin{aligned} P(a \leq X < b) &= P(\arctan(a/R) \leq \phi < \arctan(b/R)) = \\ &= \frac{1}{\pi}(\arctan(b/R) - \arctan(a/R)) = \int_a^b \frac{R}{\pi(R^2 + x^2)} dx . \end{aligned}$$

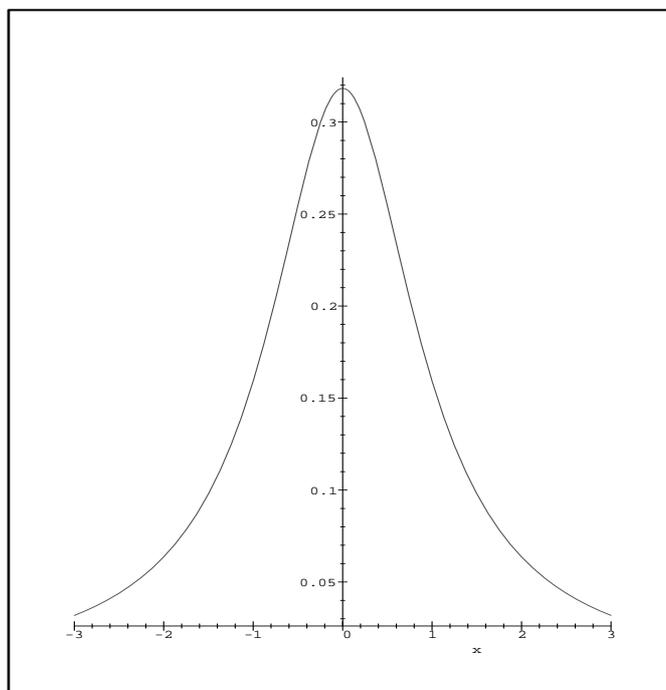


Figura 3.2: Densità di Cauchy, con $R = 1$.

Dunque X ha una distribuzione continua con densità

$$g(x) = \frac{R}{\pi(R^2 + x^2)} .$$

Si dice che X ha **una distribuzione di Cauchy**.

La v.a. X non ammette valor medio, perchè l'integrale

$$\int_{-\infty}^{+\infty} \frac{R|x|}{\pi(R^2 + x^2)} dx$$

diverge a $+\infty$. A maggior ragione X non ammette momenti di ordine superiore.

3.10 Distribuzione normale o di Gauss

Una v.a. si dice normale di parametri μ , σ^2 , dove $\sigma > 0$, se ammette la densità

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

Pertanto la distribuzione è espressa dall'integrale

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dt,$$

che non si può esplicitare in termini di trascendenti elementari.

Per indicare che una v.a. X ha una **distribuzione normale** di parametri μ, σ^2 si scrive

$$X \in \mathcal{N}(\mu, \sigma^2) \text{ oppure } X \sim \mathcal{N}(\mu, \sigma^2).$$

Spesso conviene prendere in considerazione la v.a. ausiliaria *normale standard*

$$Y = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Infatti

$$P(a \leq Y < b) = P(\mu + a\sigma \leq X < \mu + b\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu+a\sigma}^{\mu+b\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

e quindi, posto $y = \frac{x-\mu}{\sigma}$, $dx = \sigma dy$, si ha

$$P(a \leq Y < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} dy.$$

Dunque la **densità normale standard** è

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

e la distribuzione normale standard si indica con

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

L'integrale improprio è ovviamente convergente e tende a 0 per $x \rightarrow -\infty$. Inoltre

$$2\pi\Phi(+\infty)^2 = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = \int_{\mathcal{R}^2} e^{-\frac{(x^2+y^2)}{2}} dx dy =$$

(ricorrendo a coordinate polari)

$$= \int_0^{2\pi} d\theta \int_0^{+\infty} e^{-\frac{\rho^2}{2}} \rho d\rho = 2\pi[-e^{-\frac{\rho^2}{2}}]_0^{+\infty} = 2\pi .$$

Dunque $\Phi(+\infty) = 1$ e Φ è effettivamente una distribuzione di probabilità.

Nel caso della densità normale standard

$$\phi(x) = \phi(-x) , \quad \phi(0) = (2\pi)^{-\frac{1}{2}}$$

è il valore massimo,

$$\phi'(x) = -(2\pi)^{-\frac{1}{2}} x \phi(x) , \quad \phi''(x) = (2\pi)^{-\frac{1}{2}} (x^2 - 1) \phi(x)$$

e il grafico di ϕ ha due flessi simmetrici in ± 1 .

Nel caso generale, la funzione generatrice dei momenti è data dall'integrale

$$m(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx .$$

Riscriviamo l'esponente

$$\begin{aligned} tx - \frac{(x-\mu)^2}{2\sigma^2} &= -\frac{1}{2\sigma^2}(x^2 - 2\mu x - 2\sigma^2 tx + \mu^2) = \\ &= -\frac{1}{2\sigma^2}[(x - (\mu + \sigma^2 t))^2 - \mu^2 - 2\mu\sigma^2 t - \sigma^4 t^2 + \mu^2] . \end{aligned}$$

Allora

$$m(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x - (\mu + \sigma^2 t))^2}{2\sigma^2}} dx$$

ed infine

$$m(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Si ricava immediatamente

$$\begin{aligned} m'(t) &= m(t)(\mu + \sigma^2 t) \\ m''(t) &= m(t)(\mu + \sigma^2 t)^2 + m(t)\sigma^2 . \end{aligned}$$

Pertanto

$$E(X) = \mu , E(X^2) = \mu^2 + \sigma^2 , V(X) = \sigma^2$$

ed i parametri μ e σ sono il valor medio e la deviazione standard di X .

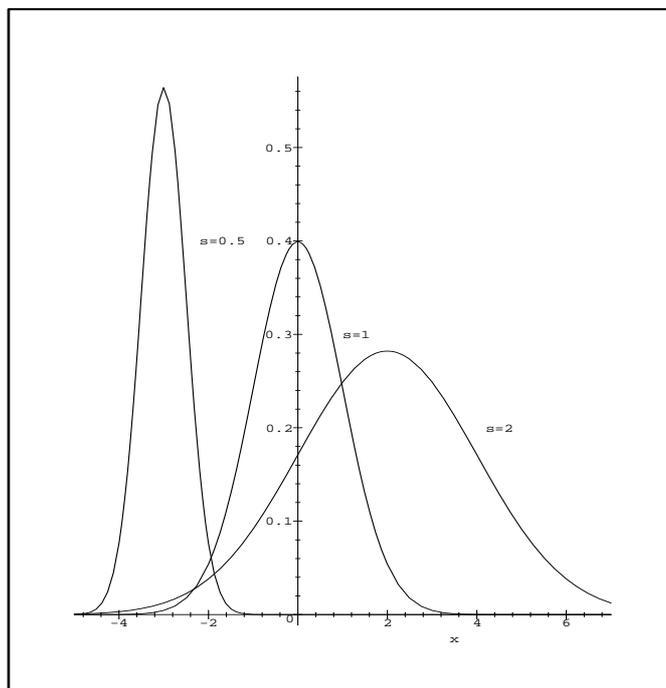


Figura 3.3: Densità normali, per $\mu = -3, 0, 2$ e $s = \sigma = 0.5, 1, 2$.

3.11 Distribuzione di Maxwell

In Teoria cinetica dei gas si ricorre, sotto opportune ipotesi (gas perfetto, stato di equilibrio, temperatura costante) alla *distribuzione di Maxwell*, con densità

$$f(v) = \begin{cases} 4/\sqrt{\pi} v^2 \alpha^{-3} \exp(-v^2/\alpha^2) & v \geq 0 \\ 0 & v < 0 \end{cases} ,$$

dove v è il modulo della velocità \mathbf{v} della generica molecola del gas, $\alpha > 0$ e $\alpha^2 = 2kT/m$, m è la massa della molecola, k la costante di Boltzmann e T la temperatura assoluta.

È facile calcolare i momenti di una v.a. V che segue la distribuzione di Maxwell:

$$E(V^k) = \int_0^{+\infty} v^k f(v) dv ,$$

ricorrendo alla sostituzione

$$v = \alpha y^{1/2} , \quad v^k = \alpha^k y^{k/2} , \quad dv = \alpha dy / 2y^{1/2} .$$

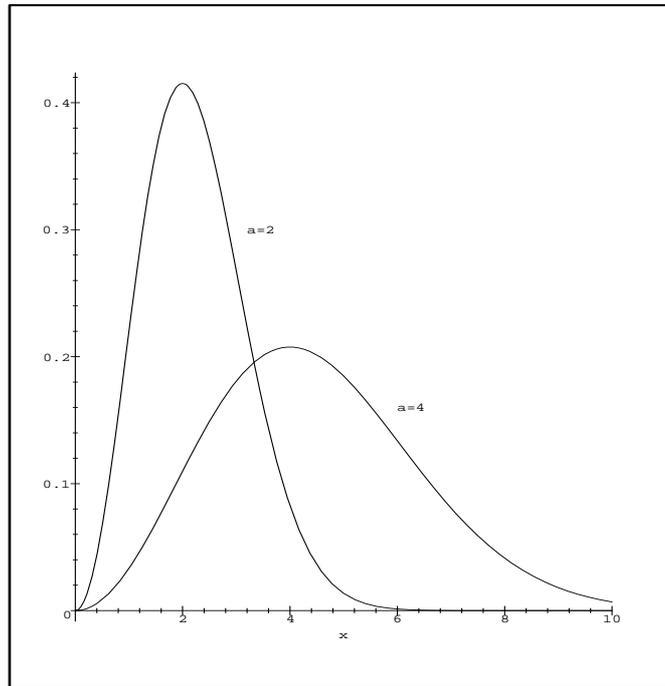


Figura 3.4: Densità di Maxwell, con $a = \alpha = 2, 4$.

Allora si trova

$$E(V^k) = \frac{2\alpha^k}{\sqrt{\pi}} \int_0^{+\infty} y^{(k+1/2)} e^{-y} dy = \frac{2\alpha^k}{\sqrt{\pi}} \Gamma\left(\frac{k+3}{2}\right).$$

Dunque, ricordando che $\Gamma(s+1) = s\Gamma(s)$ e che $\Gamma(1/2) = \sqrt{\pi}$, si ottiene:

$$E(1) = 1, \quad E(V) = \frac{2\alpha}{\sqrt{\pi}}, \quad E(V^2) = \frac{3\alpha^2}{2}, \quad E(V^4) = \frac{15\alpha^4}{4}.$$

L'energia cinetica è definita da $E_c = mV^2/2$. Allora

$$V(V) = \frac{(3\pi - 8)\alpha^2}{2\pi}, \quad V(V^2) = \frac{3}{2}\alpha^2,$$

$$E(E_c) = \frac{m}{2}E(V^2) = \frac{3}{2}kT, \quad V(E_c) = \frac{m^2}{4}\left(\frac{15}{4} - \frac{9}{4}\right)\alpha^4 = \frac{3}{2}(kT)^2.$$

3.12 Distribuzione esponenziale e processi di Poisson

Si consideri una situazione nella quale eventi di un determinato tipo si presentano ad istanti casuali. Indichiamo con $X(s, t)$ il numero di eventi che si presentano nell'intervallo $[s, t]$. Sup-

porremo che siano soddisfatte le condizioni seguenti:

1) La situazione è stazionaria nel senso che il numero e le modalità con le quali gli eventi si presentano dipendono soltanto dall'ampiezza dell'intervallo e non dall'istante iniziale:

$$X(s, t) \sim X(0, t - s) , \quad \text{cioè le due v.a. hanno la stessa distribuzione .}$$

2) La storia fino all'istante s (gli eventi che si sono presentati prima dell'istante s e le modalità con le quali si sono presentati) non influiscono sugli eventi dell'intervallo $[s, t[$. Ad esempio

$$P(X(s, t) = k \mid X(0, s) = j) = P(X(s, t) = k) .$$

Per semplicità scriveremo X_t in luogo di $X(0, t)$.

3) La probabilità che due o più eventi si verificano in un intervallino di tempo Δt è trascurabile rispetto a Δt :

$$P(X(0, \Delta t) > 1) = o(\Delta t) .$$

Se indichiamo con T la v.a. "tempo di attesa perché si verifichi un primo evento, abbiamo:

$$P(T \geq t) = P(X_t = 0) .$$

È facile verificare che la distribuzione di T è

$$F(t) = P(T < t) = 1 - e^{-\lambda t} ,$$

dove λ è un numero positivo, detta **distribuzione esponenziale di parametro λ** . Infatti, essendo T una v.a. non negativa ed essendo indipendenti eventi riguardanti intervalli di tempo disgiunti, la funzione

$$g(t) = P(T \geq t)$$

soddisfa all'equazione funzionale

$$g(t + s) = g(t)g(s) .$$

Infatti si ha

$$\begin{aligned} P(T \geq t + s) &= P(X(0, t + s) = 0) = P(X(0, t) = 0 \wedge X(t, t + s) = 0) = \\ &= P(X_t = 0)P(X_s = 0) = P(T \geq t)P(T \geq s) . \end{aligned}$$

Ma $g(t)$, essendo ovviamente non crescente, deve avere la forma indicata, perchè, posto

$$g(1) = P(T \geq 1) = P(X_1 = 0) = p ,$$

allora deve essere

$$g\left(\frac{1}{n}\right)^n = p .$$

Dunque

$$g\left(\frac{1}{n}\right) = p^{\frac{1}{n}}, \quad g\left(\frac{k}{n}\right) = p^{\frac{k}{n}}.$$

Ma per

$$\frac{k-1}{n} \leq t \leq \frac{k}{n}$$

si ha

$$g\left(\frac{k-1}{n}\right) \geq g(t) \geq g\left(\frac{k}{n}\right)$$

Passando al limite con k ed n tendenti ad infinito in modo che

$$\lim_{n \rightarrow +\infty} \frac{k}{n} = t$$

si ottiene

$$g(t) = p^t.$$

Finalmente $0 \leq p \leq 1$ e, non considerando i casi estremi, di poco interesse, si può porre

$$p = e^{-\lambda},$$

dunque

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

e finalmente

$$F(t) = 1 - g(t) = 1 - e^{-\lambda t}.$$

La densità della distribuzione esponenziale è

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \text{e quindi } F(t) = \int_0^t \lambda e^{-\lambda x} dx.$$

La funzione generatrice dei momenti è

$$m(t) = \int_0^{+\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t} \quad \text{per } t < \lambda.$$

Ovviamente $m(0) = 1$ e

$$E(T) = m'(0) = \frac{\lambda}{(\lambda - t)^2} \Big|_{t=0} = \frac{1}{\lambda}.$$

Inoltre

$$E(T^2) = m''(0) = \frac{2\lambda}{(\lambda - t)^3} \Big|_{t=0} = \frac{2}{\lambda^2}$$

e quindi

$$V(T) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

$$\sigma_T = \frac{1}{\lambda}.$$

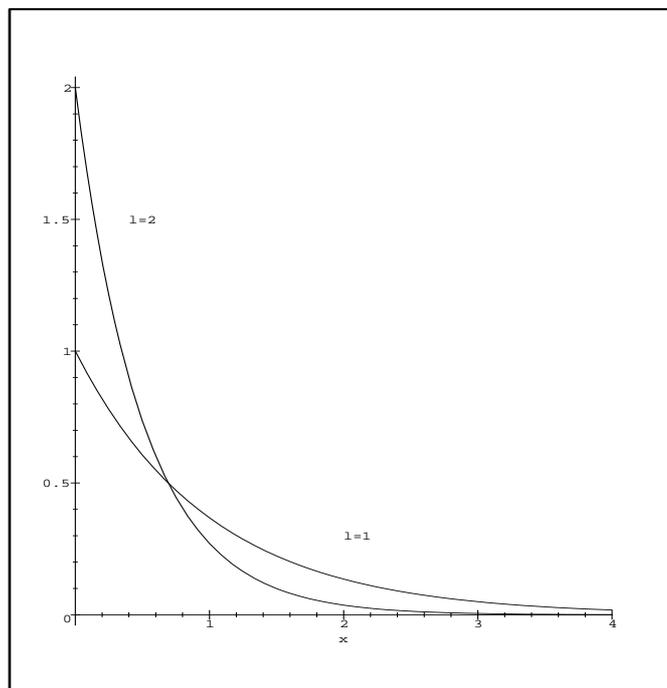


Figura 3.5: Densità esponenziale, con $l = \lambda = 1, 2$.

Vediamo ora che nelle ipotesi fatte la v.a. X_t ha una *distribuzione di Poisson* di parametro λt . Infatti, posto $p(t, k) = P(X_t = k)$, si ha

$$p(t, 0) = P(X_t = 0) = P(T \geq t) = e^{-\lambda t} \text{ e per } t \rightarrow 0 \text{ } e^{-\lambda t} = 1 - \lambda t + o(t) ,$$

$$p(dt, 1) = 1 - p(dt, 0) + o(dt)$$

e, per l'indipendenza degli eventi su intervalli disgiunti insieme alla trascurabilità rispetto a dt della probabilità di più eventi in un intervallo di ampiezza dt , si ha

$$\begin{aligned} p(t + dt, k + 1) &= p(t, k + 1)p(dt, 0) + p(t, k)p(dt, 1) + o(dt) = \\ &= p(t, k + 1)(1 - \lambda dt) + p(t, k)\lambda dt + o(dt) . \end{aligned}$$

Dividendo per dt e passando al limite per $dt \rightarrow 0$ si ottiene un sistema di infinite equazioni differenziali lineari del primo ordine, che tuttavia si possono risolvere semplicemente in modo iterativo, conoscendo $p(t, 0) = \exp(-\lambda t)$:

$$\frac{d}{dt}p(t, k + 1) = -\lambda p(t, k + 1) + \lambda p(t, k) ,$$

$$p(0, k + 1) = 0 .$$

È facile verificare per induzione che la soluzione è data dalle formule

$$p(t, k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} .$$

Capitolo 4

Densità congiunta e funzioni di più variabili aleatorie reali

4.1 Densità congiunta (caso discreto)

Siano X e Y v.a. reali discrete (v.a. che possono assumere un numero finito o una infinità numerabile di valori reali x ed y). Per trattare eventi che coinvolgono entrambe le variabili è sufficiente conoscere le probabilità

$$f_{XY}(x, y) = P(X = x \text{ e } Y = y) .$$

La funzione $f_{XY}(x, y)$ si dice *densità congiunta* di X e Y . Ovviamente

$$f_{XY}(x, y) \geq 0 \text{ per ogni coppia } (x, y) \text{ e}$$

$$\sum_x \sum_y f_{XY}(x, y) = 1 .$$

4.2 Densità congiunta (caso continuo)

Siano X e Y v.a. reali. Per trattare eventi che coinvolgono entrambe le variabili è sufficiente conoscere le probabilità

$$P(a \leq X < b \text{ e } c \leq Y < d)$$

per tutti i valori reali a, b, c, d .

Per calcolare tale probabilità è sufficiente conoscere la distribuzione congiunta di X e Y :

$$F_{XY}(x, y) = P(X < x \wedge Y < y) .$$

Infatti

$$[a, b[\times [c, d[= [-\infty, b[\times [-\infty, d[- [-\infty, b[\times [-\infty, c[- ([-\infty, a[\times [-\infty, d[\cup [-\infty, a[\times [-\infty, c[)$$

e quindi

$$P(a \leq X < b \text{ e } c \leq Y < d) = F(b, d) - F(b, c) - F(a, d) + F(a, c) .$$

Diremo che X e Y ammettono una *densità congiunta* se esiste una funzione integrabile $f_{XY}(x, y)$ tale che

$$P(a \leq X < b \text{ e } c \leq Y < d) = \int_a^b \int_c^d f_{XY}(x, y) dx dy .$$

Ovviamente

$$f_{XY}(x, y) \geq 0 , \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1 .$$

La densità congiunta $f_{XY}(x, y)$ è separatamente integrabile rispetto ad x e y (per q.o. y e rispettivamente per q.o. x) e X ed Y ammettono le densità $f_X(x)$ e $f_Y(y)$, dette *densità marginali*, fornite dalle

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

Se X ed Y sono *indipendenti* e ammettono ciascuna densità $f_X(x)$, $f_Y(y)$,

$$P(a \leq X < b \text{ e } c \leq Y < d) = P(a \leq X < b)P(c \leq Y < d) = \int_a^b f_X(x) dx \int_c^d f_Y(y) dy$$

e dunque $f_X f_Y$ è integrabile ed esiste la densità congiunta uguale al prodotto delle densità:

$$f_{XY}(x, y) = f_X(x) f_Y(y) .$$

Viceversa, se vale la formula precedente, allora X e Y sono indipendenti.

In generale, se A è un insieme Boreliano di \mathcal{R}^2 ed E è l'evento

$$E = \{ \omega \in \Omega \mid (X(\omega), Y(\omega)) \in A \} ,$$

la probabilità di E si ottiene nel modo seguente

$$P(E) = \int \int_A f_{XY}(x, y) dx dy .$$

Tale risultato si potrebbe verificare facilmente tenendo conto: i) della possibilità di approssimare A tanto bene quanto si vuole mediante unioni finite di rettangoli disgiunti del tipo $[a, b[\times [c, d[$; ii) del fatto che il risultato è vero per gli eventi corrispondenti a tali rettangoli, vista la definizione stessa di densità congiunta; iii) dell'additività e dell'assoluta continuità dell'integrale.

Si osservi che E è un evento perchè X ed Y sono v.a. e i Boreliani di \mathbf{R}^2 sono la σ -algebra generata ad esempio dai rettangoli $[a, b[\times [c, d[$.

Esempio. Siano X ed Y v.a. indipendenti e uniformi in $I = [0, 1[$. Pertanto f_X e f_Y sono entrambe la funzione caratteristica χ_I di tale intervallo, mentre

$$f_{XY}(x, y) = f_X(x)f_Y(y) = \chi_I(x)\chi_I(y) = \chi_Q(x, y) ,$$

dove $Q = I \times I$ è il quadrato unitario.

Si voglia calcolare la probabilità che la *somma* delle due variabili sia inferiore ad un valore prefissato s . Si trova

$$P(X + Y < s) = \iint_{A=\{(x,y)|x+y<s\}} \chi_Q(x, y) dx dy = \iint_{A \cap Q} dx dy .$$

Al variare di s si ha la distribuzione F_S della v.a. $S = X + Y$ funzione di X ed Y , nel senso usuale che per ogni ω

$$S(\omega) = (X + Y)(\omega) = X(\omega) + Y(\omega) .$$

Dei calcoli elementari forniscono

$$F_S(s) = \begin{cases} 0 & s \leq 0 \\ \frac{s^2}{2} & 0 \leq s < 1 \\ 1 - \frac{(2-s)^2}{2} & 1 \leq s < 2 \\ 1 & 2 \leq s \end{cases} .$$

Dunque S è una v.a. continua con densità f_S lineare a tratti

$$f_S(s) = \begin{cases} 0 & s \leq 0 \\ s & 0 \leq s < 1 \\ (2-s) & 1 \leq s < 2 \\ 0 & 2 \leq s \end{cases} ,$$

derivata della funzione F_S , che è derivabile a tratti.

In modo analogo per esempio

$$P(X + Y \leq s \text{ e } Y < t) = \iint_{\substack{x+y \leq s \\ y < t \\ (x,y) \in Q}} dx dy .$$

Ancora nell'ambito di questo esempio consideriamo il *prodotto* $T = XY$. La sua funzione di distribuzione $F_T(t)$ è nulla per $t < 0$ e uguale a 1 per $t \geq 1$, mentre per $0 \leq t < 1$ si ha

$$\begin{aligned} P(T < t) &= P(XY < t) = \iint_{xy < t} \chi_Q dx dy = \int_0^t dx \int_0^1 dy + \int_t^1 dx \int_0^{\frac{t}{x}} dy = \\ &= t + \int_t^1 \frac{t}{x} dx = t + t[\ln(x)]_t^1 = t(1 - \ln(t)) . \end{aligned}$$

Dunque T è una v.a. continua di densità

$$f_T(s) = \begin{cases} 0 & t \leq 0 \\ -\ln(t) & 0 < t < 1 \\ 0 & 1 \leq t \end{cases} .$$

Si osservi che mentre F_T è continua f_T non è limitata (ma integrabile).

Proponiamo ancora un esempio riguardante v.a. di Cauchy.
Se X e Y sono due variabili di Cauchy indipendenti con densità

$$g_X(x) = \frac{R}{\pi(R^2 + x^2)} \quad \text{e} \quad g_Y(y) = \frac{S}{\pi(S^2 + y^2)} ,$$

anche la loro somma $Z = X + Y$ è una v.a. di Cauchy. Infatti

$$P(Z < t) = \int \int_{x+y < t} \frac{RS}{\pi^2(R^2 + x^2)(S^2 + y^2)} dx dy =$$

con la sostituzione $x = x$, $z = x + y$

$$= \frac{RS}{\pi^2} \int_{-\infty}^t dz \int_{-\infty}^{+\infty} \frac{dx}{(R^2 + x^2)(S^2 + (z-x)^2)} .$$

L'integrale interno si può facilmente calcolare per ogni z fissato, ad esempio con il metodo dei residui, fornendo

$$P(Z < t) = \int_{-\infty}^t \frac{R+S}{\pi((R+S)^2 + z^2)} dz$$

e la densità di Z è

$$g_Z(z) = \frac{R+S}{\pi((R+S)^2 + z^2)} .$$

Se $X_1, X_2, \dots, X_n, \dots$ sono v.a. di Cauchy, indipendenti e di uguale densità $(\pi(1+x^2))^{-1}$, le loro somme

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

sono v.a. di Cauchy con densità $n(\pi(n^2+x^2))^{-1}$ e in questo caso il teorema centrale del limite che studieremo nel Capitolo successivo non vale.

Cambiamento di variabili.

siano X_1, X_2 v.a. continue di densità congiunta $f(x_1, x_2)$ e siano

$$Y_1 = H_1(X_1, X_2) , \quad Y_2 = H_2(X_1, X_2)$$

due loro funzioni, con H_1, H_2 regolari. Brevemente scriviamo $Y = H(X)$. La distribuzione congiunta di Y_1, Y_2 si ottiene mediante l'integrale doppio

$$P(Y_1 < c_1 \wedge Y_2 < c_2) = \int_A f(x_1, x_2) dx_1 dx_2 ,$$

dove

$$A = \{(x_1, x_2) \mid H_1(x_1, x_2) < c_1 \wedge H_2(x_1, x_2) < c_2\} .$$

Se, posto $B = \{(y_1, y_2) \mid y_1 < c_1 \wedge y_2 < c_2\}$, la funzione H è invertibile da B in A con inversa K di classe C^1 , si può effettuare il cambiamento di variabili $x = K(y)$ e dunque

$$P(Y_1 < c_1 \wedge Y_2 < c_2) = \int_B f(K(y)) |\det K'(y)| dy .$$

Se la sostituzione è possibile per tutti i valori di c_1, c_2 , si vede che le nuove v.a. Y_1, Y_2 ammettono la densità congiunta $f(K(y)) |\det K'(y)|$.

4.3 Covarianza e correlazione

Siano X ed Y due v.a. Si definisce la *covarianza* delle due variabili nel modo seguente

$$\begin{aligned} Cov(X, Y) &= E((X - E(X))(Y - E(Y))) = \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) . \end{aligned}$$

Quindi si può calcolare la covarianza mediante la formula

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

Ovviamente se $X = Y$ si ha $V(X) = Cov(X, X)$. Inoltre

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y) .$$

Osservazione. Sia $L^2 = L^2(\Omega, \mathcal{A}, P; \mathbf{R})$ lo spazio delle v.a. che ammettono momento del secondo ordine e quindi varianza. In L^2 la covarianza è una forma bilineare simmetrica positiva e la varianza è la seminorma (il suo annullarsi dice che la v.a. è costante) ad essa associata. Nel sottospazio L_0^2 delle v.a. a media nulla Cov definisce una struttura di spazio di Hilbert.

Se le due variabili sono indipendenti, poichè $E(XY) = E(X)E(Y)$, la covarianza si annulla.

Tuttavia l'annullarsi della covarianza non implica in generale che le due variabili siano indipendenti. Ad esempio se X ha una densità pari: $f(x) = f(-x)$ e ammette momenti fino al terzo ordine, posto $Y = X^2$

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = 0 ,$$

$$E(Y) = E(X^2) = 2 \int_0^{+\infty} x^2 f(x)dx > 0 ,$$

$$Cov(X, Y) = E(X^3) - E(X)E(X^2) = \int_{-\infty}^{+\infty} x^3 f(x)dx = 0 ,$$

ma X ed Y ovviamente non sono indipendenti e Y è funzione di X . Infatti ad esempio in generale (cioè tranne che per v.a. speciali o particolari valori di a) si ha

$$\begin{aligned} P(\{0 \leq X^2 < a^2\} \cap \{-a \leq X < a\}) &= P(\{-a \leq X < a\}) \neq \\ &\neq P(0 \leq X^2 < a^2)P(-a \leq X < a) = P(-a \leq X < a)^2 . \end{aligned}$$

Osservazione. L'evento A è indipendente da sé stesso se e solo se

$$P(A) = P(A \cap A) = P(A)^2 ,$$

dunque se e solo se $P(A) = 1$ oppure $P(A) = 0$.

Una v.a X può essere indipendente da sé stessa solo se per ogni intervallo $[a, b[$ risulta $P(X \in [a, b]) = 0$ o $= 1$.

Se le v.a. X e Y ammettono una densità congiunta $f_{XY}(x, y)$, per la *disuguaglianza di Schwarz*, posto per brevità $\mu = E(X)$ e $\nu = E(Y)$,

$$\begin{aligned} |Cov(X, Y)|^2 &= \left| \int \int (x - \mu)(y - \nu) f_{XY}(x, y) dx dy \right|^2 \leq \\ &\leq \int \int (x - \mu)^2 f_{XY}(x, y) dx dy \int \int (y - \nu)^2 f_{XY}(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx \int_{-\infty}^{+\infty} (y - \nu)^2 f_Y(y) dy . \end{aligned}$$

Dunque

$$|Cov(X, Y)|^2 \leq V(X)V(Y) .$$

La disuguaglianza precedente è vera per v.a. arbitrarie, non essendo affatto necessaria l'esistenza di una densità congiunta, ma solo medie e varianze finite.

Infatti si trova per ogni λ reale

$$\begin{aligned} 0 \leq V(Y - \lambda X) &= E((Y - \lambda X - (\nu - \lambda\mu))^2) = E((\lambda(X - \mu) - (Y - \nu))^2) = \\ &= \lambda^2 V(X) - 2\lambda Cov(X, Y) + V(Y) . \end{aligned}$$

Considerando il discriminante del polinomio di secondo grado in λ si ottiene la disuguaglianza in esame

$$|Cov(X, Y)| \leq \sqrt{V(X)V(Y)}$$

Se le varianze non si annullano, si può introdurre il *coefficiente di correlazione (di Pearson)*, quale normalizzazione della covarianza

$$\rho = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} ,$$

per il quale si ha

$$-1 \leq \rho \leq 1 .$$

Si può interpretare $|\rho|$ come indice della “propensione in media di scarti significativi simultanei delle due v.a. dalla loro media, essendo tali scarti di segno uguale o contrario secondo la positività di ρ .”

Si ha $|\rho| = 1$ se e solo se X ed Y sono “linearmente dipendenti

$$Y = aX + c ,$$

con a e c costanti opportune (c in effetti è una v.a. costante con probabilità 1), perché in tal caso

$$V(Y - \lambda X) = \lambda^2 \sigma_X^2 - 2\lambda \rho \sigma_X \sigma_Y + \sigma_Y^2 \geq 0$$

che si può annullare se e solo se $|\rho| = 1$. In tal caso

$$V(Y - \lambda X) = (\lambda \sigma_X \mp \sigma_Y)^2 ,$$

dove necessariamente deve essere

$$\lambda = a = \pm \frac{\sigma_Y}{\sigma_X} \text{ e } Y - \lambda X = c$$

è una v.a. costante.

Naturalmente si ha in modo simmetrico

$$X = a^*Y + c^* .$$

Coefficiente di regressione γ di Y rispetto a X .

Si può generalizzare il tema precedente della dipendenza “lineare, cercando di scomporre Y in una parte proporzionale a X e in una parte quanto più possibile prossima ad una costante:

$$Y = \lambda X + Z(\lambda) .$$

La decomposizione ottimale si ha per il valore γ di λ che minimizza la varianza della v.a. Z :

$$V(Z(\gamma)) = \min_{\lambda} V(Z(\lambda)) = \min_{\lambda} (\lambda^2 \sigma_X^2 - 2\lambda \rho \sigma_X \sigma_Y + \sigma_Y^2) .$$

Il punto di minimo γ si trova risolvendo l'equazione

$$2\lambda \sigma_X^2 - 2\rho \sigma_X \sigma_Y = 0 .$$

Pertanto

$$\gamma = \rho \frac{\sigma_Y}{\sigma_X} , \quad V(Z(\gamma)) = (1 - \rho^2) \sigma_Y^2 .$$

Il valore minimo $V(Z(\gamma))$ si annulla se e solo se $\rho = \pm 1$, nel qual caso Z è una v.a. costante.

Osserviamo ancora che risulta

$$\text{Cov}(X, Z(\gamma)) = 0 .$$

Infatti, essendo $E(Z) = E(Y) - \gamma E(X)$:

$$\begin{aligned} \text{Cov}(X, Z) &= E((X - E(X))(Z - E(Z))) = \\ &= E((X - E(X))(Y - \gamma X - (E(Y) - \gamma E(X)))) = \\ &= E((X - E(X))(Y - E(Y))) - \gamma E((X - E(X))^2) = \rho \sigma_X \sigma_Y - \rho \frac{\sigma_Y}{\sigma_X} \sigma_X^2 = 0 . \end{aligned}$$

Quindi

$$V(Y) = \gamma^2 V(X) + V(Z) = \gamma^2 \sigma_X^2 + (1 - \rho^2) \sigma_Y^2 = \sigma_Y^2 ,$$

come deve essere.

Osservazione. È equivalente studiare il problema della regressione nello spazio L_0^2 sottraendo preventivamente a X e Y le loro medie. Si trova

$$Y_0 = \gamma X_0 + Z_0 ,$$

dove γX_0 è la proiezione ortogonale (nel senso della forma bilineare Cov) di Y_0 sul sottospazio di dimensione 1 generato da X_0 e ovviamente Z_0 è a media nulla e ortogonale a tale sottospazio.

4.4 La distribuzione Gaussiana bivariata

Consideriamo due v.a. X e Y con densità congiunta

$$f(x, y) = \frac{1}{2\pi} \frac{1}{\sqrt{(1 - \rho^2)\sigma_X\sigma_Y}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left[\frac{(x - \mu_X)^2}{\sigma_X^2} - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X\sigma_Y} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} \right]\right) ,$$

dove si suppone che $\rho^2 < 1$ e $\sigma_X\sigma_Y \neq 0$.

Se introduciamo la *matrice di covarianza*

$$\Gamma = \begin{vmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{vmatrix} = \begin{vmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & V(Y) \end{vmatrix}$$

ed il *vettore delle medie* $\mu = (\mu_X, \mu_Y)$, posto $v = (x, y)$, si può scrivere la densità nella forma

$$f(v) = \frac{1}{2\pi} \frac{1}{\sqrt{\det \Gamma}} e^{-\frac{1}{2}(v - \mu, \Gamma^{-1}(v - \mu))} .$$

Si potrebbe facilmente dimostrare che μ_X , μ_Y , σ_X , σ_Y , ρ sono rispettivamente la media di X , la media di Y , la deviazione standard di X , la deviazione standard di Y , il coefficiente di

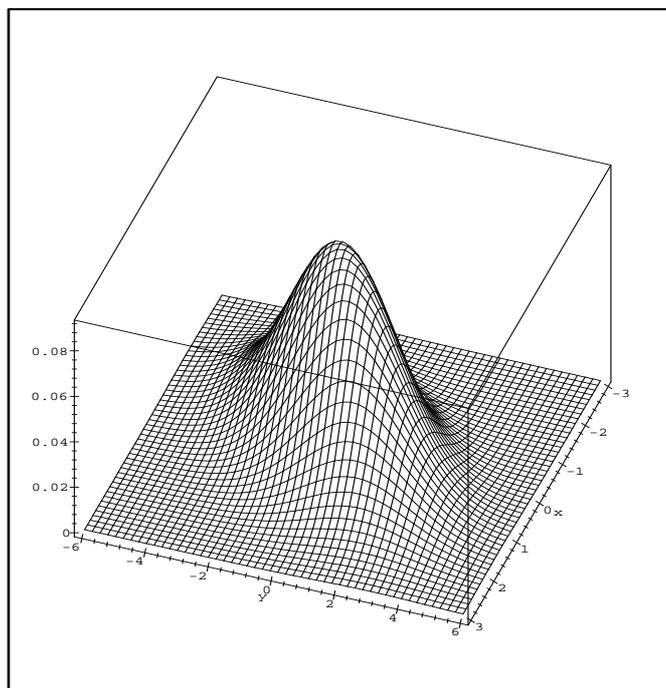


Figura 4.1: Densità Gaussiana bivariata, con $\sigma_X = 1$, $\sigma_Y = 2$, $\rho = 0.5$.

correlazione di X ed Y .

Osserviamo che in questo caso se il coefficiente di correlazione $\rho = 0$, ovvero se $Cov(X, Y) = 0$, allora le variabili X e Y sono indipendenti. Infatti la densità congiunta è in questo caso

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(-\frac{(y - \mu_Y)^2}{2\sigma_Y^2}\right),$$

ovvero il prodotto delle densità di X e Y . (Si ricordi che in generale per v.a. arbitrarie indipendenza implica correlazione nulla ma in caso di correlazione nulla non si ha automaticamente indipendenza.)

Più in generale sia Z una v.a. a valori in \mathcal{R}^m le cui componenti Z_1, Z_2, \dots, Z_m siano v.a. indipendenti normali standard: $Z_j \sim \mathcal{N}(0, 1)$. Sia A una matrice $p \times m$ e μ un vettore di componenti $\mu_1, \mu_2, \dots, \mu_p$ e poniamo

$$\Gamma = AA^*.$$

La v.a. X di componenti X_1, X_2, \dots, X_p data da

$$X = AZ + \mu$$

ha una distribuzione (distribuzione congiunta delle X_j) detta *Gaussiana multivariata* $\mathcal{N}_p(\mu, \Gamma)$. Si può vedere che μ è la media di X , cioè le variabili X_j hanno media μ_j , e Γ è la matrice di covarianza delle X_j :

$$\Gamma_{ij} = E((X_i - \mu_i)(X_j - \mu_j)) .$$

La distribuzione di X non dipende direttamente da A ma solo da Γ , dunque se

$$Y = BZ + \mu \quad \text{con} \quad BB^* = \Gamma$$

allora X e Y sono equidistribuite.

Nel caso in cui Γ sia invertibile, indicando con x il vettore di componenti x_1, x_2, \dots, x_p , la densità congiunta delle X_j è

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \Gamma}} e^{-\frac{1}{2}(x-\mu, \Gamma^{-1}(x-\mu))} .$$

4.5 Variabili normali indipendenti

Se X_1, X_2, \dots, X_n sono v.a. normali indipendenti anche la loro somma segue una distribuzione normale. Precisamente se

$$X_k \sim \mathcal{N}(\mu_k, \sigma_k^2) \quad \text{e} \quad \mu = \sum_k \mu_k, \quad \sigma^2 = \sum_k \sigma_k^2,$$

risulta

$$X = \sum_k X_k \sim \mathcal{N}(\mu, \sigma^2) .$$

Infatti essendo la funzione generatrice dei momenti di X_k

$$m_{X_k}(t) = e^{\mu_k t + \sigma_k^2 \frac{t^2}{2}},$$

per l'indipendenza delle v.a.

$$m_X(t) = \prod_k m_{X_k}(t) = e^{\mu t + \sigma^2 \frac{t^2}{2}},$$

che è la funzione generatrice di una v.a. normale di media μ e varianza σ^2 .

Osserviamo che, anche se le X_k non fossero normali, per l'indipendenza si ha

$$E(X) = \sum_k E(X_k), \quad V(X) = \sum_k V(X_k) .$$

In molti casi praticamente interessanti le v.a. X_k hanno la stessa media μ e varianza σ^2 . In tal caso si ha

$$\begin{aligned} X &= \sum_k X_k \sim \mathcal{N}(n\mu, n\sigma^2) , \\ \bar{X} &= \frac{X}{n} = \frac{\sum_k X_k}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) , \\ \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} &\sim \mathcal{N}(0, 1) . \end{aligned}$$

Dunque la v.a. media \bar{X} ha come media la stessa media μ delle singole X_k , ma varianza tanto minore quante più sono le X_k : σ^2 diviso n .

Questa osservazione è fondamentale per le applicazioni statistiche. Si supponga infatti di voler *stimare* la media incognita μ di una v.a. normale X , ipotizzando invece di conoscerne la deviazione standard σ . Si effettuano n prove indipendenti a ciascuna delle quali è associata una v.a. X_k copia identica di X (cioè X_k e X hanno la stessa distribuzione: $X_k \sim X$). Prefissato un livello α di rischio o equivalentemente un *livello di confidenza* $1 - \alpha$, si osserva il valore della v.a. media \bar{X} e si individua un corrispondente *intervallo di confidenza* $[L_1, L_2]$, dove gli estremi L_1 e L_2 sono v.a., tale che

$$P(L_1 \leq \mu \leq L_2) = 1 - \alpha .$$

Se $Z \sim \mathcal{N}(0, 1)$ e per $z_{\alpha/2}$ si ha

$$\Phi(-z_{\alpha/2}) = \frac{1}{\sqrt{2\pi}} \int_{z_{\alpha/2}}^{+\infty} e^{-x^2/2} dx = \alpha/2 ,$$

allora risulta

$$P(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1 - \alpha ,$$

ovvero

$$P(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha .$$

Dunque l'intervallo aleatorio

$$[\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}]$$

è un intervallo di confidenza per la stima di μ con un livello di confidenza pari a $1 - \alpha$.

Il fatto che la somma di v.a. normali indipendenti sia ancora normale può essere controllato direttamente, senza ricorrere alle funzioni generatrici. Per semplicità consideriamo due v.a. Y_1 ed Y_2 normali standard ed indipendenti. Posto $Y = Y_1 + Y_2$, si ha

$$P(Y < s) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \int \int_{y_1+y_2 < s} e^{-(y_1^2+y_2^2)/2} dy_1 dy_2 .$$

Effettuando il cambiamento di variabili

$$y_1 = t - y_2, \quad y_2 = y_2 \quad \text{con Jacobiano uguale a 1}$$

si trova

$$\begin{aligned} P(Y < s) &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \int_{-\infty}^s dt \int_{-\infty}^{+\infty} dy_2 e^{-[(t-y_2)^2 + y_2^2]/2} = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s dt e^{-t^2/4} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dy_2 e^{-(\sqrt{2}y_2 - \frac{t}{\sqrt{2}})^2/2} = \end{aligned}$$

mediante la sostituzione $\sqrt{2}y_2 - \frac{t}{\sqrt{2}} = z$, con $dy_2 = dz/\sqrt{2}$

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi}\sqrt{2}} \int_{-\infty}^s dt e^{-t^2/4} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz e^{-z^2/2} = \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} \int_{-\infty}^s dt e^{-\frac{1}{2}t^2/2}. \end{aligned}$$

Dunque Y è normale con media 0 e varianza 2.

4.6 Distribuzioni gamma

Si dice che una v.a. X segue una *distribuzione gamma con parametri α e β* se ammette la densità

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1/(\Gamma(\alpha)\beta^\alpha) x^{\alpha-1} e^{-x/\beta} & 0 < x \end{cases},$$

dove $\Gamma(s)$ è la funzione speciale definita da

$$\Gamma(s) = \int_0^{+\infty} z^{s-1} e^{-z} dz, \quad s > 0.$$

Alcune distribuzioni classiche, legate alla distribuzione normale, sono casi particolari della distribuzione gamma.

Ricordiamo alcune proprietà notevoli della funzione speciale Γ .

$$\Gamma(1) = 1,$$

$$\Gamma(s+1) = s\Gamma(s) \text{ e dunque } \Gamma(n+1) = n!,$$

$$\Gamma(1/2) = \sqrt{\pi}.$$

Infatti

$$\int_0^{+\infty} e^{-z} dz = [-e^{-z}]_0^{+\infty} = 1,$$

$$\Gamma(s) = \int_0^{+\infty} z^{s-1} e^{-z} dz = [z^s e^{-z}/s]_0^{+\infty} + \frac{1}{s} \int_0^{+\infty} z^s e^{-z} dz = \frac{1}{s} \Gamma(s+1),$$

sostituendo $z = t^2/2$ e quindi $dz = t dt$ nell'espressione di $\Gamma(s+1)$ si trova

$$\Gamma(s+1) = \int_0^{+\infty} 2^{-s} t^{2s+1} e^{-t^2/2} dt ,$$

dunque per $s = -1/2$

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

Funzione generatrice per la distribuzione gamma .

$$m_X(t) = \int_0^{+\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{+\infty} x^{\alpha-1} e^{-(1/\beta-t)x} dx .$$

Con la sostituzione $z = (1/\beta - t)x$, quindi $x = \beta z/(1 - \beta t)$ e $dx = \beta dz/(1 - \beta t)$, con $t < 1/\beta$:

$$m_X(t) = \frac{1}{\Gamma(\alpha)(1 - \beta t)^\alpha} \int_0^{+\infty} z^{\alpha-1} e^{-z} dz = (1 - \beta t)^{-\alpha} .$$

Derivando si trova

$$m'_X(t) = -\alpha(1 - \beta t)^{-\alpha-1}(-\beta) = \alpha\beta(1 - \beta t)^{-\alpha-1} ,$$

$$m''_X(t) = \alpha(\alpha + 1)\beta^2(1 - \beta t)^{-\alpha-2} .$$

Per $t = 0$ troviamo

$$E(X) = \alpha\beta , \quad E(X^2) = \alpha(\alpha + 1)\beta^2 , \quad V(X) = \alpha\beta^2 .$$

4.7 Distribuzioni χ^2

Siano Y_1, Y_2, \dots, Y_n v.a. normali standard ($\sim \mathcal{N}(0, 1)$) e indipendenti tra loro. Calcoliamo la distribuzione (e la densità) della v.a.

$$\chi^2(n) = Y_1^2 + Y_2^2 + \dots + Y_n^2 .$$

Per l'indipendenza, la densità congiunta delle Y_k è il prodotto delle loro densità $e^{-y_k^2/2}$, dunque la densità congiunta è $e^{-(y_1^2+y_2^2+\dots+y_n^2)/2}$ e quindi

$$P(\chi^2 < u) = (2\pi)^{-n/2} \int \int \dots \int_{y_1^2+y_2^2+\dots+y_n^2 < u} e^{-(y_1^2+y_2^2+\dots+y_n^2)/2} dy_1 dy_2 \dots dy_n .$$

Ricorrendo a coordinate polari ρ, S in \mathcal{R}^n

$$P(\chi^2 < u) = (2\pi)^{-n/2} \int_{S^{n-1}} dS \int_0^{\sqrt{u}} e^{-\rho^2/2} \rho^{n-1} d\rho ,$$

dove S^{n-1} è l'ipersuperficie sferica unitaria in \mathcal{R}^n di misura

$$\int_{S^{n-1}} dS = \frac{2\pi^{n/2}}{\Gamma(n/2)} .$$

Per trovare la densità si può procedere nel modo seguente. Si effettua la sostituzione $\rho^2 = t$ e quindi $\rho = t^{1/2}$, $d\rho = t^{-1/2}dt/2$,

$$\begin{aligned} P(\chi^2 < u) &= (2\pi)^{-n/2} \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^u e^{-t/2} t^{(n-1)/2-1/2} dt/2 = \\ &= \frac{1}{\Gamma(n/2)2^{n/2}} \int_0^u e^{-t/2} t^{n/2-1} dt . \end{aligned}$$

Dunque $\chi^2(n)$ segue una distribuzione gamma con parametri $\alpha = n/2$ e $\beta = 2$. Ad esempio

$$E(\chi^2(n)) = n \quad , \quad V(\chi^2(n)) = 2n .$$

Si dice che $\chi^2(n)$ è un v.a. χ^2 con n gradi di libertà .

Se X_1, X_2, \dots, X_n sono v.a. normali con la stessa distribuzione ($X_k \sim \mathcal{N}(\mu, \sigma^2)$) e indipendenti tra loro e Y_1, Y_2, \dots, Y_n sono le corrispondenti v.a. normali standard ($\sim \mathcal{N}(0, 1)$), indipendenti tra loro)

$$Y_k = \frac{X_k - \mu}{\sigma} ,$$

allora

$$\chi^2 = \sum_k \frac{(X_k - \mu)^2}{\sigma^2} = \sum_k Y_k^2 .$$

ha una distribuzione χ^2 con n gradi di libertà.

Si supponga di voler *stimare* la varianza incognita σ^2 di una v.a. normale X , ipotizzando di conoscerne la media μ . Si effettuano n prove indipendenti a ciascuna delle quali è associata una v.a. X_k copia identica di X . Si osserva quindi il valore della v.a. $Q^2 = \sum_k (X_k - \mu)^2$. Q^2/σ^2 ha una distribuzione $\chi^2(n)$. Prefissato un livello α di rischio o equivalentemente un *livello di confidenza* $1 - \alpha$, si individua un corrispondente *intervallo di confidenza* $[S_1, S_2]$ tale che

$$P(S_1 \leq \sigma^2 \leq S_2) = 1 - \alpha .$$

Per esempio, se f_n è la densità di $\chi^2(n)$, trovato $\chi_{\alpha/2}^2$ tale che

$$\int_0^{\chi_{\alpha/2}^2} f_n(u) du = 1 - \alpha/2, \quad \text{cioè} \quad \int_{\chi_{\alpha/2}^2}^{+\infty} f_n(u) du = \alpha/2$$

e $\chi_{1-\alpha/2}^2$ tale che

$$\int_0^{\chi_{1-\alpha/2}^2} f_n(u) du = \alpha/2, \quad \text{cioè} \quad \int_{\chi_{1-\alpha/2}^2}^{+\infty} f_n(u) du = 1 - \alpha/2 ,$$

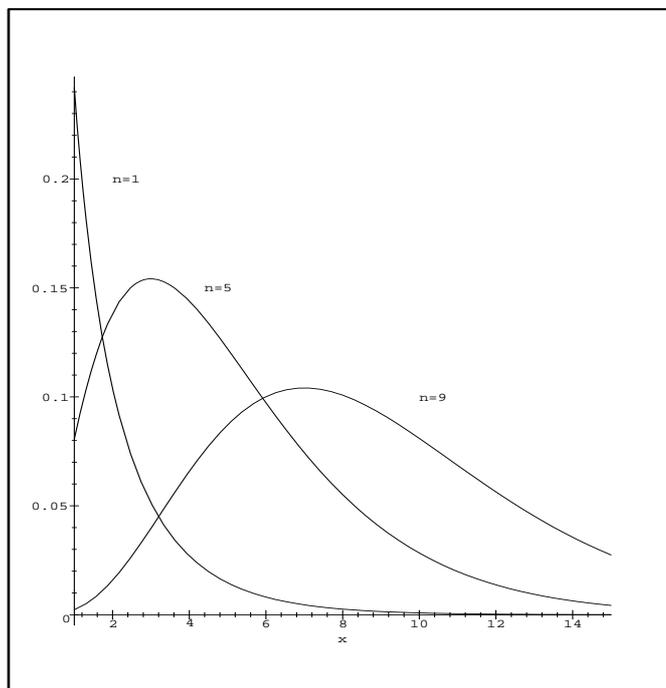


Figura 4.2: Densità $\chi^2(n)$, per $n = 1, 5, 9$.

risulta

$$P(\chi_{1-\alpha/2}^2 \leq Q^2/\sigma^2 \leq \chi_{\alpha/2}^2) = 1 - \alpha .$$

Quindi

$$P\left(\frac{Q^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{Q^2}{\chi_{1-\alpha/2}^2}\right) = 1 - \alpha .$$

L'intervallo aleatorio

$$\left[\frac{Q^2}{\chi_{\alpha/2}^2}, \frac{Q^2}{\chi_{1-\alpha/2}^2}\right]$$

è un intervallo di confidenza per la stima di σ^2 con livello di confidenza $1 - \alpha$.

4.8 Stima simultanea di media e varianza di una variabile normale

Con prove, o misure, indipendenti è anche possibile stimare sia la media che la varianza incognite di una v.a. X , ipotizzando come prima che $X \sim \mathcal{N}(\mu, \sigma^2)$. Se X_k sono le v.a.

associate alle n prove, ciascuna copia identica di X , è usuale considerare gli *stimatori*

$$\bar{X} = \frac{1}{n} \sum_k X_k \text{ e}$$

$$S^2 = \frac{1}{n-1} \sum_k (X_k - \bar{X})^2 .$$

Tali stimatori sono *corretti*, cioè v.a. con media il parametro da stimare:

$$E(\bar{X}) = \mu , \quad V(\bar{X}) = \frac{1}{n} V(X) ,$$

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} E\left(\sum_k (X_k - \mu - (\bar{X} - \mu))^2\right) = \\ &= \frac{1}{n-1} \left(\sum_k E((X_k - \mu)^2) - 2E\left(\sum_k (X_k - \mu)(\bar{X} - \mu)\right) + \sum_k E((\bar{X} - \mu)^2)\right) = \\ &= \frac{1}{n-1} (nV(X) - 2E(n(\bar{X} - \mu)(\bar{X} - \mu)) + n\frac{1}{n}V(X)) = \\ &= \frac{1}{n-1} (nV(X) - 2V(X) + V(X)) = \frac{1}{n-1} (n-1)V(X) = V(X) = \sigma^2 . \end{aligned}$$

S^2 permette di ottenere intervalli di confidenza per σ^2 . Infatti la v.a.

$$C = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) ,$$

cioè segue una distribuzione χ^2 con $n-1$ gradi di libertà.

Per dimostrarlo conviene introdurre in primo luogo le variabili Y_k normali standard associate alle X_k :

$$Y_k = \frac{X_k - \mu}{\sigma} , \quad \bar{Y} = \frac{1}{n} \sum_k Y_k = \frac{\bar{X} - \mu}{\sigma} .$$

Allora

$$C = \frac{(n-1)S^2}{\sigma^2} = \sum_k Y_k^2 - n\bar{Y}^2 = \sum_k Y_k^2 - \frac{1}{n} \left(\sum_k Y_k\right)^2 ,$$

come si controlla svolgendo i calcoli seguenti

$$\begin{aligned} C &= \sum_k \frac{(X_k - \bar{X})^2}{\sigma^2} = \sum_k \frac{((X_k - \mu) + (\mu - \bar{X}))^2}{\sigma^2} = \\ &= \sum_k \frac{(X_k - \mu)^2}{\sigma^2} - 2\frac{(\bar{X} - \mu)}{\sigma} \sum_k \frac{(X_k - \mu)}{\sigma} + \sum_k \frac{(\bar{X} - \mu)^2}{\sigma^2} = \end{aligned}$$

$$= \sum_k \frac{(X_k - \mu)^2}{\sigma^2} - 2n \frac{(\bar{X} - \mu)^2}{\sigma^2} + n \frac{(\bar{X} - \mu)^2}{\sigma^2}.$$

La densità congiunta delle Y_k è

$$f(y_1, y_2, \dots, y_n) = (2\pi)^{-n/2} e^{-(\sum_k y_k^2)/2}$$

e per trovare la distribuzione di C si deve allora calcolare

$$P(C < u) = (2\pi)^{-n/2} \int \int \dots \int_{\sum_k y_k^2 - \frac{1}{n} (\sum_k y_k)^2 < u} e^{-(\sum_k y_k^2)/2} dy_1 dy_2 \dots dy_n.$$

Conviene utilizzarla una opportuna trasformazione lineare ortogonale delle variabili, che verrà effettuata per semplicità nel caso $n = 2$.

Poniamo

$$z_1 = \frac{y_1 + y_2}{\sqrt{2}}, \quad z_2 = \frac{-y_1 + y_2}{\sqrt{2}},$$

da cui segue

$$\begin{aligned} z_1^2 + z_2^2 &= y_1^2 + y_2^2 \quad \text{e} \\ y_1^2 + y_2^2 - \frac{(y_1 + y_2)^2}{2} &= z_1^2 + z_2^2 - z_1^2 = z_2^2. \end{aligned}$$

Allora

$$\begin{aligned} P(C < u) &= (2\pi)^{-1} \int_{-\infty}^{+\infty} e^{-z_1^2/2} dz_1 \int_{z_2^2 < u} e^{-z_2^2/2} dz_2 = \\ &= (2\pi)^{-1/2} \int_{-\sqrt{u}}^{+\sqrt{u}} e^{-z_2^2/2} dz_2 = \\ &= \frac{1}{\sqrt{\pi} 2^{1/2}} \int_0^u t^{1/2-1} e^{-t/2} dt. \end{aligned}$$

Dunque in questo caso C ha una distribuzione χ^2 con $2 - 1$ gradi di libertà.

In generale, con un procedimento perfettamente simile, si vedrebbe che $\frac{(n-1)S^2}{\sigma^2}$ ammette la densità $f_{n-1}(t)$ nulla per $t \leq 0$ e

$$f_{n-1}(t) = \frac{1}{\Gamma(\frac{n-1}{2}) 2^{\frac{n-1}{2}}} e^{-t/2} t^{(n-1)/2-1},$$

cioè segue come si era detto una distribuzione χ^2 con $n - 1$ gradi di libertà.

Come nel paragrafo precedente, fissato un livello di confidenza $1 - \alpha$, si trovano $\chi_{\alpha/2}^2 = \chi_{\alpha/2}^2(n - 1)$ tale che

$$\int_{\chi_{\alpha/2}^2(n-1)}^{+\infty} f_{n-1}(u) du = \alpha/2$$

e $\chi_{1-\alpha/2}^2 = \chi_{1-\alpha/2}^2(n-1)$ tale che

$$\int_0^{\chi_{1-\alpha/2}^2(n-1)} f_{n-1}(u) du = \alpha/2 .$$

Allora risulta

$$P(\chi_{1-\alpha/2}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{\alpha/2}^2) = 1 - \alpha .$$

Quindi

$$P\left(\frac{(n-1)S^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}\right) = 1 - \alpha .$$

Osserviamo che le v.a. \bar{Y} e $\chi^2 = (n-1)S^2/\sigma^2$ sono *indipendenti*. Infatti, per esempio nel caso $n=2$, riprendendo i calcoli precedenti, essendo $Z_1/\sqrt{2}$, si ha

$$P(a \leq \bar{Y} < b \text{ e } \chi^2 \leq u) = (2\pi)^{-1/2} \int_{a\sqrt{2}}^{b\sqrt{2}} e^{-z_1^2/2} dz_1 \int_{z_2^2 < u} e^{-z_2^2/2} dz_2 =$$

$$P(a \leq \bar{Y} < b)P(\chi^2 \leq u) .$$

4.9 Distribuzioni *t* di Student

Al fine di determinare intervalli di confidenza per la media μ della v.a. X consideriamo le v.a. indipendenti

$$Y = \sqrt{n} \bar{Y} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ e}$$

$$\chi_{n-1}^2 = \frac{(n-1)S^2}{\sigma^2} .$$

Fomiamo quindi la v.a.

$$T_{n-1} = \frac{Y}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

e calcoliamone la distribuzione.

In generale siano Z e χ_q^2 due v.a. indipendenti, la prima normale standard e la seconda con una distribuzione χ^2 a q gradi di libertà, e si ponga

$$T_q = \frac{Z}{\sqrt{\chi_q^2/q}} .$$

Si dice che T_q ha una *distribuzione *t* di Student con q gradi di libertà*. Essa ha come densità

$$g(t) = \frac{\Gamma((q+1)/2)}{(\pi q)^{1/2} \Gamma(q/2)} \left(1 + \frac{t^2}{q}\right)^{-(q+1)/2} .$$

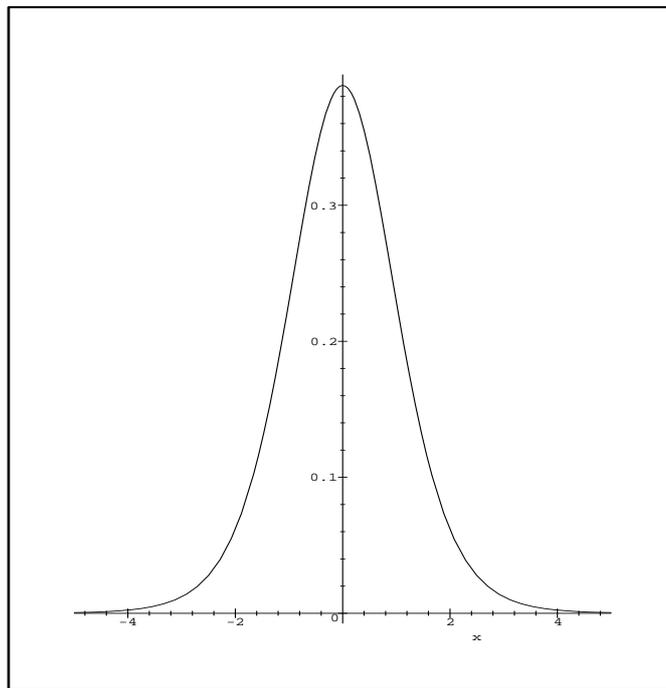


Figura 4.3: Densità di Student, con $q = 9$ gradi di libertà.

Si osservi che la lettera t , nella frase precedente, prima è il nome di una famiglia di distribuzioni, poi serve a indicare una variabile reale.

Il risultato si ottiene ricorrendo alla distribuzione congiunta delle variabili Z e χ_q^2

$$f(z, x) = (2\pi)^{-1/2} e^{-z^2/2} \frac{1}{\Gamma(q/2) 2^{q/2}} x^{q/2-1} e^{-x/2} \text{ per } x > 0$$

e calcolando

$$P(T_q < t) = C \int \int_{z < t\sqrt{x/q}} x^{q/2-1} e^{-x/2} e^{-z^2/2} dx dz ,$$

dove

$$\frac{1}{C} = \sqrt{2\pi} \Gamma(q/2) 2^{q/2} .$$

Allora

$$P(T_q < t) = C \int_0^{+\infty} dx x^{q/2-1} e^{-x/2} \int_{-\infty}^{t\sqrt{x/q}} dz e^{-z^2/2} =$$

con la sostituzione $z = u\sqrt{x}$ nell'integrale interno

$$\begin{aligned} &= C \int_0^{+\infty} dx x^{(q-1)/2} e^{-x/2} \int_{-\infty}^{t/\sqrt{q}} du e^{-u^2 x/2} = \\ &= C \int_{-\infty}^{t/\sqrt{q}} du \int_0^{+\infty} dx x^{(q-1)/2} e^{-(1+u^2)x/2} = \end{aligned}$$

posto $y = (1 + u^2)x/2$

$$\begin{aligned} &= C \int_{-\infty}^{t/\sqrt{q}} du \left(\frac{1+u^2}{2}\right)^{-(1+q)/2} \int_0^{+\infty} dy y^{(q+1)/2-1} e^{-y} = \\ &= C \Gamma\left(\frac{q+1}{2}\right) \int_{-\infty}^{t/\sqrt{q}} du \left(\frac{1+u^2}{2}\right)^{-(1+q)/2} . \end{aligned}$$

Finalmente, ponendo $s = \sqrt{q}u$, $du = q^{-1/2}ds$, $u^2 = s^2/q$:

$$\begin{aligned} P(T_q < t) &= (2\pi)^{-1/2} \Gamma(q/2)^{-1} 2^{-q/2} \Gamma\left(\frac{q+1}{2}\right) q^{-1/2} 2^{(1+q)/2} \int_{-\infty}^t ds \left(1 + \frac{s^2}{q}\right)^{-(1+q)/2} = \\ &= \frac{\Gamma((q+1)/2)}{(\pi q)^{1/2} \Gamma(q/2)} \int_{-\infty}^t ds \left(1 + \frac{s^2}{q}\right)^{-(1+q)/2} \end{aligned}$$

q.e.d..

È bene notare che la densità g è simmetrica

$$g(-t) = g(t) .$$

Tornando alla determinazione di intervalli di confidenze per la media incognita μ , dato un livello di confidenza desiderato $1 - \alpha$, determiniamo $t_{\alpha/2}$ tale che

$$\int_{t_{\alpha/2}}^{+\infty} g(s) ds = \alpha/2$$

e quindi

$$\int_{-\infty}^{-t_{\alpha/2}} g(s) ds = \alpha/2 ,$$

essendo $g = g_{n-1}$ la densità t di Student con $n - 1$ gradi di libertà.

Allora

$$P(-t_{\alpha/2} \leq T_{n-1} = \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}) = 1 - \alpha ,$$

pertanto

$$P(\bar{X} - t_{\alpha/2} S/\sqrt{n} \leq \mu < \bar{X} + t_{\alpha/2} S/\sqrt{n}) = 1 - \alpha .$$

L'intervallo aleatorio (in quanto i suoi estremi sono combinazione delle v.a. \bar{X} e S)

$$[\bar{X} - t_{\alpha/2}S/\sqrt{n}, \bar{X} + t_{\alpha/2}S/\sqrt{n}]$$

è un intervallo di confidenza per la stima della media μ della v.a. normale X , con livello di confidenza $1 - \alpha$.

4.10 Le variabili F e z di Fisher.

Siano χ_p^2 e χ_q^2 due v.a. indipendenti che hanno una distribuzione χ^2 , rispettivamente con p e q gradi di libertà. Consideriamo un loro rapporto normalizzato:

$$F_{p,q} = \frac{\chi_p^2/p}{\chi_q^2/q} = \frac{q}{p} \frac{\chi_p^2}{\chi_q^2},$$

che si dice variabile F con p e q gradi di libertà.

Fisher introdusse una variabile z legata alla precedente dalle relazioni

$$e^{2z} = F_{p,q} \quad , \quad z = \frac{1}{2} \ln F_{p,q} .$$

Queste variabili si incontrano in numerose situazioni, particolarmente in test di *analisi della varianza*.

Vogliamo determinare la distribuzione e la densità di F . È sufficiente considerare la variabile

$$X = \frac{Y}{Z} \quad , \quad \text{dove } Y = \chi_p^2 \quad , \quad Z = \chi_q^2 .$$

La densità congiunta di Y e Z , per la loro indipendenza è (considerando solo valori positivi delle variabili):

$$g(y, z) = C y^{p/2-1} z^{q/2-1} e^{-(y+z)/2} \quad ,$$

dove

$$C = \frac{1}{\Gamma(p/2)2^{p/2}\Gamma(q/2)2^{q/2}} .$$

La distribuzione di X risulta

$$\begin{aligned} P(X < x) &= P(Y < xZ) = \int \int_{z>0, 0<y<xz} g(y, z) dy dz = \\ &= C \int_0^{+\infty} dz z^{q/2-1} \int_0^{xz} dy y^{p/2-1} e^{-(y+z)/2} = \end{aligned}$$

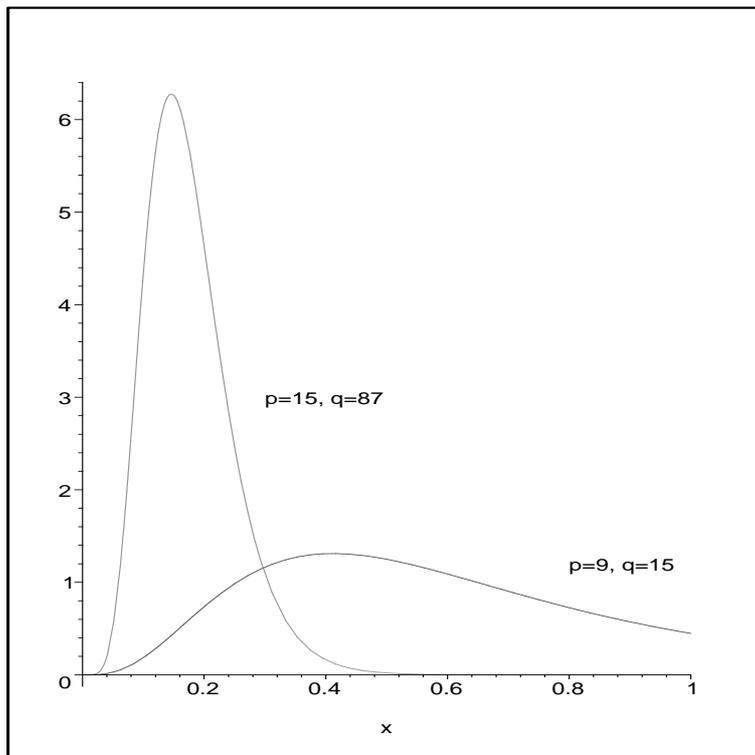


Figura 4.4: Densità di Fisher: v.a. $X_{p,q}$ con parametri $p = 9, q = 15$ e $p = 15, q = 87$.

posto $y = uz$, $dy = zdu$

$$\begin{aligned}
 &= C \int_0^{+\infty} dz z^{q/2} \int_0^x du (uz)^{p/2-1} e^{-(1+u)z/2} = \\
 &= C \int_0^x du u^{p/2-1} \int_0^{+\infty} dz z^{(p+q)/2-1} e^{-(1+u)z/2} =
 \end{aligned}$$

posto $t = (1+u)z/2$, $z = 2t/(1+u)$, $dz = 2dt/(1+u)$

$$\begin{aligned}
 &= C \int_0^x du u^{p/2-1} \left(\frac{2}{1+u}\right)^{(p+q)/2} \int_0^{+\infty} dt t^{(p+q)/2-1} e^{-t} = \\
 &= C \Gamma\left(\frac{p+q}{2}\right) 2^{(p+q)/2} \int_0^x du \frac{u^{p/2-1}}{(1+u)^{(p+q)/2}} .
 \end{aligned}$$

Dunque la densità di $X_{p,q}$ è nulla per $x \leq 0$ e uguale a

$$\frac{\Gamma((p+q)/2)}{\Gamma(p/2)\Gamma(q/2)} \frac{x^{p/2-1}}{(1+x)^{(p+q)/2}} .$$

Non sarebbe difficile calcolare media e varianza di $F_{p,q}$ (Cramér [1], p.242):

$$E(F_{p,q}) = q/(q-2) \quad \text{per } q > 2$$
$$V(F_{p,q}) = \frac{2q^2(p+q-2)}{p(q-2)^2(q-4)} \quad \text{per } q > 4 .$$

Capitolo 5

Funzioni caratteristiche e teorema del limite centrale

5.1 Funzione caratteristica

Sia X una v.a. reale. Si dice *funzione caratteristica* di X la funzione a valori complessi

$$\varphi_X(t) = E(e^{itX}) .$$

La variabile t è reale; per t fissato itX è una v.a. complessa (cioè a valori in \mathcal{C}). (Lo studio di una v.a. complessa Z , volendo, si può ricondurre allo studio di due v.a. reali: la sua parte reale e la sua parte immaginaria. Ad esempio:

$$E(Z) = E(\Re Z + i\Im Z) = E(\Re Z) + iE(\Im Z) .$$

Nel caso discreto, se x_k sono i valori assunti da X e p_k le corrispondenti probabilità

$$\varphi_X(t) = \sum_k e^{itx_k} p_k .$$

Si noti che si tratta di una somma finita o di una serie assolutamente convergente perchè $|e^{itx_k}| = 1$, $p_k \geq 0$ e $\sum_k p_k = 1$.

Nel caso continuo, se $f_X(x)$ è la densità di X :

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx$$

e quindi la funzione caratteristica è data dalla trasformata di Fourier della densità mediante la formula: $\varphi(t) = \mathcal{F}(f)(-t)$.

Osservazione. Assumiamo come definizione della trasformata di Fourier

$$\mathcal{F}(f)(t) = \int_{-\infty}^{+\infty} e^{-itx} f(x) dx ,$$

come ad esempio L.Hörmander. Nella letteratura si trovano varie definizioni, lievemente diverse, che non danno luogo a modifiche delle proprietà fondamentali: invece di $\exp(-itx)$ si può trovare $\exp(itx)$ (I.M.Gelfand-G.E.Šilov), oppure $\exp(-2\pi itx)$ (L.Schwartz, E.M.Stein-G.Weiss), oppure $\exp(-itx)/\sqrt{2\pi}$ (F.Riesz-B.Sz.Nagy, A.Kolmogorov-S.Fomin, K.Yosida).

L'integrale, essendo $|e^{itxk}| = 1$ ed f integrabile, è ben definito.

Nel caso generale di una distribuzione F_X , non continua ovunque né puramente discreta, si dovrebbe usare un integrale di Stieltjes per costruire φ da F .

Proprietà fondamentali delle funzioni caratteristiche .

Con calcoli elementari nel caso discreto, o ricordando i risultati essenziali sulla trasformata di Fourier, si dimostrano le proprietà seguenti, valide anche per v.a. generali:

1) $\varphi(0) = 1$ e, per ogni t , $\varphi(-t) = \overline{\varphi(t)}$ e $|\varphi(t)| \leq 1$.

2) φ è una funzione uniformemente continua e se X ha momenti finiti fino all'ordine k , proprietà che nel caso continuo equivale alla convergenza degli integrali

$$\int_{-\infty}^{+\infty} x^j f(x) dx , \quad j = 0, 1, \dots, k ,$$

allora φ è k volte derivabile con continuità. In tal caso i momenti di X sono forniti dalle formule

$$E(X^j) = \frac{1}{ij} \varphi_X^{(j)}(0) .$$

3) Se X e Y hanno la stessa funzione caratteristica allora X e Y hanno la stessa distribuzione.

Si può dimostrare che, se i punti $x+h$ e $x-h$ ($h > 0$) sono punti di continuità della distribuzione $F(x)$ alla quale corrisponde la funzione caratteristica $\varphi(t)$, allora

$$F(x+h) - F(x-h) = \lim_{T \rightarrow +\infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin(ht)}{t} e^{-itx} \varphi(t) dt .$$

Inoltre, se $|\varphi|$ è integrabile, F è derivabile e la sua derivata f si ottiene mediante la formula

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt .$$

4) Se $Y = bX + a$ allora

$$\varphi_Y(t) = e^{iat} \varphi_X(bt) .$$

5) Se X e Y sono indipendenti

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t) .$$

Esempi .

Se K è una variabile che assume i valori 1 e 0 con probabilità rispettivamente p e $q = 1 - p$:

$$\varphi_K(t) = q + pe^{it} .$$

Quindi se B è una variabile binomiale associata ad n prove indipendenti: $B = K_1 + K_2 \dots + K_n$, dove le K_j sono copie indipendenti di K :

$$\varphi_B(t) = (q + pe^{it})^n .$$

Se $X \sim \mathcal{N}(0, 1)$, allora

$$\varphi_X(t) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} dx = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{-(x-it)^2/2 - t^2/2} dx =$$

posto $z = x - it$

$$= e^{-t^2/2} (2\pi)^{-1/2} \int_{\mathcal{R}-it} e^{-z^2/2} dz = e^{-t^2/2} .$$

Infatti, essendo la funzione $e^{-z^2/2}$ analitica in tutto \mathcal{C} e rapidamente decrescente per $x \rightarrow \pm\infty$, l'integrale sulla retta $\mathcal{R} - it$ risulta uguale a quello sul cammino omotopo \mathcal{R} .

Ricordando la definizione della funzione generatrice dei momenti m_X , si ha la relazione $\varphi_X(t) = m_X(it)$. Si osservi che, essendo e^{it} limitata, φ_X è sempre definita, mentre la convergenza dell'integrale che fornisce m_X dipende dal comportamento della distribuzione.

Nel caso di una v.a. normale $X \sim \mathcal{N}(\mu, \sigma^2)$, poiché $E(e^{zX})$ è definita per ogni numero complesso z :

$$E(e^{zX}) = (2\pi)^{-1/2} \sigma^{-1} \int_{\mathcal{R}} e^{zx} e^{-(x-\mu)^2/(2\sigma^2)} dx = e^{\mu z + \sigma^2 z^2/2} ,$$

che è una funzione analitica di z in tutto il piano complesso. Si ha dunque in particolare

$$\varphi_X(t) = e^{i\mu t + \sigma^2 (it)^2/2} = e^{i\mu t - \sigma^2 t^2/2} .$$

La funzione caratteristica di una v.a. di Cauchy è

$$\varphi_R(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{Re^{itx}}{R^2 + x^2} dx = e^{-R|t|} ,$$

come si può vedere con un utile esercizio di applicazione del metodo dei residui.

5.2 Il teorema di Lévy-Cramér

Riportiamo senza dimostrazione (si veda ad esempio H.Cramér [], p.96, oppure B.V.Gnedenko[]) il seguente risultato fondamentale:

Siano $\varphi_n(t)$ le funzioni caratteristiche delle distribuzioni $F_n(x)$ ed esista per ogni t

$$\varphi(t) = \lim_{n \rightarrow +\infty} \varphi_n(t) .$$

Sia inoltre φ continua in 0. Allora $\varphi(t)$ è la funzione caratteristica di una distribuzione $F(x)$ e per ogni punto x nel quale F è continua si ha

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x) .$$

Una dimostrazione molto interessante e particolarmente istruttiva è proposta da P.Billingsley. Essa tuttavia richiede strumenti più avanzati (convergenza in legge o convergenza debole di misure, teorema di Stone-Weierstrass, famiglie “tight” di distribuzioni e compattezza relativa).

5.3 Comportamento asintotico della distribuzione binomiale

Utilizzando il teorema di Lévy-Cramér possiamo fornire una nuova e semplice dimostrazione del teorema del limite integrale di Laplace-De Moivre.

Teorema. Sia B_n una variabile binomiale associata ad n prove indipendenti, in ciascuna delle quali la probabilità di successo è p . Poniamo

$$\mu = E(B_n) = np , \quad \sigma^2 = V(B_n) = npq , \quad Y_n = \frac{B_n - \mu}{\sigma} .$$

Allora

$$P(a \leq \frac{B_n - np}{\sqrt{npq}} < b) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt .$$

Dimostrazione. La funzione caratteristica di Y_n è

$$\varphi_n(t) = e^{-it\mu/\sigma} (pe^{it/\sigma} + q)^n = e^{-it\mu/\sigma} (p(e^{it/\sigma} - 1) + 1)^n .$$

Allora, poichè

$$e^{it/\sigma} - 1 = \frac{it}{\sigma} - \frac{1}{2} \frac{t^2}{\sigma^2} + O(n^{-3/2}) \quad (\sigma = \sqrt{npq}) ,$$

si trova

$$\begin{aligned} \ln(1 + p(e^{it/\sigma} - 1)) &= \frac{itp}{\sigma} - \frac{1}{2} p \frac{t^2}{\sigma^2} + O(n^{-3/2}) - \frac{1}{2} p^2 \frac{(it)^2}{\sigma^2} + O(n^{-3/2}) = \\ &= \frac{itp}{\sigma} - \frac{1}{2} (p - p^2) \frac{t^2}{\sigma^2} + O(n^{-3/2}) . \end{aligned}$$

Ma

$$\ln \varphi_n(t) = -\frac{itnp}{\sigma} + n \ln(1 + p(e^{it/\sigma} - 1)) =$$

essendo $n(p - p^2) = \sigma^2$

$$= -\frac{itnp}{\sigma} + \frac{itnp}{\sigma} - \frac{1}{2}t^2 + O(n^{-1/2}) \rightarrow -\frac{1}{2}t^2 .$$

Pertanto

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = e^{-\frac{1}{2}t^2} .$$

La funzione limite è continua ed è precisamente la funzione caratteristica di una v.a. normale standard, inoltre la distribuzione normale standard è continua per ogni valore di x , dunque per ogni x si ha quando $n \rightarrow +\infty$

$$P\left(\frac{B_n - np}{\sqrt{npq}} < x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt .$$

Ne segue immediatamente il risultato di De Moivre-Laplace.

5.4 Il teorema del limite centrale

Il risultato precedente concernente le variabili binomiali è un caso particolare del teorema centrale del limite e può essere ampiamente generalizzato. Esistono diversi teoremi, che con ipotesi più o meno forti, conducono alla stessa tesi. Già intravisto da Laplace e Gauss, fù dimostrato nel 1901 da Liapunov; successivi risultati, che fanno uso di metodi diversi, sono associati ai nomi di Lévy, Lindeberg, Khinchin, Feller.

Consideriamo in primo luogo un teorema concernente v.a. equidistribuite, la cui dimostrazione segue il metodo usato per le variabili binomiali.

Teorema. (Lindeberg-Lévy) *Siano X_1, X_2, \dots v.a. mutuamente indipendenti aventi la stessa distribuzione di media μ e varianza σ^2 . Allora la somma, v.a. dipendente da n ,*

$$S_n = X_1 + X_2 + \dots + X_n ,$$

che ha media $E(S_n) = n\mu$ e varianza $V(S_n) = n\sigma^2$, è asintoticamente normale $(n\mu, n\sigma^2)$ nel senso che per ogni a e b si ha

$$\lim_{n \rightarrow +\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt .$$

Dimostrazione. Solo per semplicità sia $\mu = 0$ (ipotesi non restrittiva: altrimenti si ragionerebbe sulle v.a. $Z_k = X_k - \mu$). Sia inoltre $\phi(t)$ la funzione caratteristica della distribuzione comune alle singole X_k , per la quale dunque si ha

$$\phi(t) = 1 - \sigma^2 t^2 / 2 + o(t^2) .$$

Poniamo poi

$$Y_n = \frac{S_n}{\sigma\sqrt{n}} ,$$

che ha come funzione caratteristica

$$\phi_{Y_n}(t) = \phi(t/\sigma\sqrt{n})^n .$$

Allora, per ogni t

$$\phi(t/\sigma\sqrt{n}) = 1 - t^2/2n + o(1/n) ,$$

$$\ln \phi(t/\sigma\sqrt{n}) = -t^2/2n + o(1/n) ,$$

$$\ln \phi_{Y_n}(t) = n \ln \phi(t/\sigma\sqrt{n}) = -t^2/2 + n \cdot o(1/n) \rightarrow -t^2/2 .$$

Si osservi che, essendo t fisso e σ^2 dato, mentre $n \rightarrow +\infty$, scriviamo brevemente $o(1/n)$ invece di $o(t^2/n\sigma^2)$.

Dunque

$$\lim_{n \rightarrow +\infty} \ln \phi_{Y_n}(t) = -t^2/2 .$$

Ricorrendo al teorema di Lévy-Cramér si conclude la dimostrazione. *q.e.d.*

Passando al caso di variabili non equidistribuite enunciamo ora, senza dimostrazione, il

Teorema di Liapunov. *Sia $X_1, X_2, \dots, X_n, \dots$ una successione di v.a.*

- 1) *mutuamente indipendenti,*
- 2) *dotate di media e varianza finite:*

$$E(X_k) = \mu_k \quad , \quad V(X_k) = \sigma_k^2 \quad ,$$

- 3) *per le quali inoltre esista $d > 0$ tale che siano finite le quantità $E(|X_k - \mu_k|^{2+d})$ ed anzi, posto*

$$m_n = \mu_1 + \mu_2 + \dots + \mu_n \quad , \quad s_n^2 = V_n = V(X_1) + V(X_2) + \dots + V(X_n) \quad ,$$

risultati

$$\lim_{n \rightarrow +\infty} \frac{1}{s_n^{2+d}} \sum_{k=1}^n E(|X_k - \mu_k|^{2+d}) = 0 \quad ,$$

Allora la v.a.

$$S_n = X_1 + X_2 + \dots + X_n \quad , \quad \text{per la quale } E(S_n) = m_n \text{ e } V(S_n) = s_n^2 \quad ,$$

è asintoticamente normale (m_n, s_n) :

$$\lim_{n \rightarrow +\infty} P\left(a \leq \frac{S_n - m_n}{s_n} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt \quad .$$

O ancora, per ogni x

$$\lim_{n \rightarrow +\infty} P\left(\frac{\sum_{k=1}^n (X_k - \mu_k)}{\sqrt{\sum_{k=1}^n \sigma_k^2}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad .$$

Si può vedere che il limite è uniforme in x .

Il significato delle varie condizioni imposte per dedurre il teorema del limite centrale è sostanzialmente la trascurabilità asintotica dei singoli addendi rispetto alla loro somma. Per questo il teorema ha un ruolo centrale nel Calcolo delle probabilità, essendo frequenti le situazioni in cui effetti o perturbazioni significative sono la risultante di numerosi effetti o perturbazioni, ciascuna delle quali presa isolatamente sarebbe irrilevante.

A questo proposito, si consideri il seguente semplice esempio: sia K una v.a. che assume i valori $1, 0$ con probabilità p, q e $Z = K - p$. Allora Z ha media nulla e varianza uguale a quella di K , ovvero $\sigma^2 = pq$. Siano poi N_1, N_2, \dots v.a. aleatorie normali di media nulla e varianza σ_k^2 tali che

$$\sum_{k=1}^{+\infty} \sigma_k^2 = \theta^2 < +\infty$$

e supponiamo che tutte le v.a. Z, N_k siano mutuamente indipendenti. Posto

$$R_{n-1} = N_1 + N_2 + \dots + N_{n-1}, \quad S_n = Z + R_{n-1}, \quad \theta_{n-1}^2 = \sum_{k=1}^{n-1} \sigma_k^2$$

risulta

$$E(S_n) = 0, \quad s_n^2 = V(S_n) = \sigma^2 + \theta_{n-1}^2 \rightarrow s^2 = \sigma^2 + \theta^2.$$

Studiamo allora il comportamento asintotico di $Y_n = S_n/s_n$. Siano

$$\varphi_N(t) = e^{-t^2/2}, \quad \varphi_K(t) = pe^{it} + q$$

le funzioni caratteristiche di una normale standard N e di K . Allora

$$\varphi_{N_k}(t) = \varphi_N(\sigma_k t) = e^{-\sigma_k^2 t^2/2}, \quad \varphi_Z(t) = e^{-ipt} \varphi_K(t) = pe^{iqt} + qe^{-ipt},$$

$$\varphi_{Y_n}(t) = \varphi_Z\left(\frac{t}{s_n}\right) \prod_{k=1}^{n-1} \varphi_N\left(\frac{\sigma_k t}{s_n}\right) =$$

$$= (pe^{iqt/s_n} + qe^{-ipt/s_n}) e^{-\theta_{n-1}^2 t^2/2s_n^2} \rightarrow (pe^{iqt/s} + qe^{-ipt/s}) e^{-\theta^2 t^2/2s^2},$$

funzione continua in t e dunque funzione caratteristica di una distribuzione, la quale tuttavia non è una normale standard, ma è la distribuzione della somma $T = U + V$ di due v.a. indipendenti: U , che assume i valori $q/s, -p/s$ con probabilità p, q e V , normale di media nulla e varianza $\beta^2 = \theta^2/s^2$. Dunque

$$P(T < c) = \frac{p}{\sqrt{2\pi}} \int_{-\infty}^{c-q/s} e^{-\frac{v^2}{\beta^2}} dv + \frac{q}{\sqrt{2\pi}} \int_{-\infty}^{c+p/s} e^{-\frac{v^2}{\beta^2}} dv,$$

distribuzione continua, con densità

$$\frac{1}{\sqrt{2\pi}} \left(pe^{-\frac{(x-q/s)^2}{\beta^2}} + qe^{-\frac{(x+p/s)^2}{\beta^2}} \right).$$

In questo caso, ovviamente, la condizione di Liapunov non è soddisfatta:

$$\frac{E(|Z|^{2+d}) + \sum_{k=1}^{n-1} E(|N_k|^{2+d})}{(\sigma^2 + \theta_{n-1}^2)^{(2+d)/2}} \geq \frac{pq^{2+d} + qp^{2+d}}{(\sigma^2 + \theta^2)^{(2+d)/2}} > 0$$

e non può quindi tendere a 0.

Se $X_1, X_2, \dots, X_n, \dots$ sono v.a. di Cauchy, indipendenti e di ugual densità $(\pi(1+x^2))^{-1}$, le loro somme

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

sono v.a. di Cauchy con densità $n(\pi(n^2+x^2))^{-1}$ e in questo caso il teorema del limite centrale non vale.

Capitolo 6

La legge dei grandi numeri

6.1 La legge debole dei grandi numeri

Si dice che per una successione di v.a. X_1, X_2, \dots , dotate di media, vale la **legge debole dei grandi numeri** se

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k)\right| < \varepsilon\right) = 1 .$$

Definizione. La successione X_n di v.a. converge in probabilità alla v.a. X se e solo se

$$\forall \varepsilon > 0 \quad \lim_n P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \varepsilon\}) = 1 .$$

Equivalentemente

$$\forall \varepsilon > 0 \quad \lim_n P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0 .$$

Dunque la legge debole dei grandi numeri si può formulare dicendo che

$$M_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \rightarrow 0$$

in probabilità.

6.2 Convergenza quasi certa, convergenza debole e convergenza in legge

La convergenza in probabilità è una convergenza più debole della convergenza puntuale ($\forall \omega \in \Omega \quad M_n(\omega) \rightarrow 0$), o anche solo della convergenza q.c. ($\lim_n P(\{M_n \neq 0\}) = 0$). Il termine legge

debole deriva dal fatto che debole è la nozione di convergenza coinvolta.

Se X_n e X sono v.a. (reali) in Ω e F_n, F sono le loro distribuzioni, allora

Teorema. *La convergenza in probabilità di X_n ad X implica la convergenza di $F_n(c)$ a $F(c)$ in tutti i punti di continuità della funzione F , o, come si dice, implica la convergenza in distribuzione (ovvero in legge) di X_n ad X . Si scrive*

$$X_n \xrightarrow{D} X .$$

Osservazione. Si può vedere che X_n converge ad X in distribuzione se e solo se per ogni funzione continua e limitata f

$$E(f(X_n)) \rightarrow E(f(X)) .$$

La convergenza delle distribuzioni F_n a F nei punti di continuità di F equivale alla cosiddetta **convergenza debole** delle corrispondenti misure di probabilità μ_n a μ (leggi delle v.a. X_n e X).

* * *

Dimostrazione. Per ogni n ed ε sia

$$A = A_n^\varepsilon = \{ X - \varepsilon < X_n < X + \varepsilon \} .$$

Per l'ipotesi di convergenza in probabilità $P(A) \rightarrow 1$ e $P(A^c) \rightarrow 0$ per $n \rightarrow +\infty$.
Su A

$$X < c - \varepsilon \Rightarrow X + \varepsilon < c \Rightarrow X_n < c ,$$

dunque

$$P(X < c - \varepsilon) = F(c - \varepsilon) \leq P(X_n < c) + P(A^c) = F_n(c) + P(A^c)$$

e passando al limite per $n \rightarrow +\infty$ e poi per $\varepsilon \downarrow 0$ si trova successivamente

$$F(c - \varepsilon) \leq \liminf_n F_n(c) \quad , \quad F^-(c) = F(c) \leq \liminf_n F_n(c) .$$

In modo analogo, su A ,

$$X_n < c \Rightarrow X - \varepsilon < c \Rightarrow X < c + \varepsilon$$

e dunque

$$F_n(c) \leq F(c + \varepsilon) + P(A^c) .$$

e passando al limite per $n \rightarrow +\infty$ e poi per $\varepsilon \downarrow 0$ si trova successivamente

$$\limsup_n F_n(c) \leq F(c + \varepsilon) \quad , \quad \limsup_n F_n(c) \leq F^+(c) .$$

Pertanto nei punti di continuità di F , dove $F^+(c) = F^-(c)$ si ha effettivamente $\lim_n F_n(c) = F(c)$. *q.e.d.*

Al contrario, la convergenza delle distribuzioni non permette di avere alcuna informazione sulla convergenza, anche solo in probabilità, delle v.a. corrispondenti. Si consideri ad esempio la successione di v.a. definite nel seguente modo: sia $\Omega = [0, 1[$ con P uguale alla misura standard di Lebesgue; per ogni intero naturale n dividiamo Ω in 2^n sottointervalli di uguale lunghezza $I_k^n = [(k-1)/2^n, k/2^n[$, $k = 1, 2, \dots, 2^n$ ($P(I_k^n) = 1/2^n$); sia X_n la v.a. che assume il valore 1 sugli intervalli di indice dispari e 0 su quelli di indice pari. Tutte le X_n hanno la stessa distribuzione: $P(\{X_n = 1\}) = 1/2$ e $P(\{X_n = 0\}) = 1/2$, dunque $F_n(x) = 0$ per $x \leq 0$, $F_n(x) = 1/2$ per $0 < x \leq 1$ e $F_n(x) = 1$ per $x > 1$. Tuttavia la successione X_n non converge in probabilità, anzi nessuna sua sottosuccessione può convergere in probabilità: infatti, presi due indici n ed $n+p$, X_n assume un valore costante (0 o 1) su ciascuno dei 2^n intervalli I_k^n , dove invece X_{n+p} alterna con uguale probabilità i due valori sui successivi 2^p sottointervallini di ampiezza $1/2^{n+p}$ nei quali I_k^n è suddiviso. Allora

$$P(\{|X_n - X_{n+p}| = 1\}) = \sum_{k=1}^{2^n} P(\{|X_n - X_{n+p}| = 1\} \cap I_k^n) = \sum_{k=1}^{2^n} \frac{1}{2} \frac{1}{2^n} = \frac{1}{2}.$$

Del resto la distribuzione di una v.a. si occupa del peso probabilistico dei valori assunti dalla v.a., prescindendo dai punti di Ω nei quali tali valori sono assunti. V.a. che assumono valori diversi in ogni punto, possono avere la stessa distribuzione: ad esempio se $X = 0$ su A e $X = 1$ su A^c , mentre $P(A) = P(A^c) = 1/2$, allora $Y = 1 - X$ ha la stessa distribuzione di X , ma è sempre $X(\omega) \neq Y(\omega)$.

Segnaliamo tuttavia, senza dimostrazione, il seguente notevole risultato:

Teorema (Skorohod). Sia F_n una successione di distribuzioni convergente alla distribuzione F ($F_n(x) \rightarrow F(x)$ in ogni punto di continuità di F). Allora esiste uno spazio di probabilità sul quale sono definite v.a. X_n e X , aventi distribuzione rispettivamente F_n e F , tali che $X_n \rightarrow X$ quasi certamente.

& & &

6.3 Il teorema di Markov, i suoi corollari e il teorema di Khinchin

Teorema di Markov.

Sia $X_1, X_2, \dots, X_n, \dots$ una successione di v.a. dotate di media e varianza, e tali che

$$\lim_{n \rightarrow +\infty} \frac{1}{n^2} V\left(\sum_{k=1}^n X_k\right) = 0.$$

allora per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k)\right| < \varepsilon\right) = 1,$$

cioè per X_1, X_2, \dots vale la legge debole dei grandi numeri.

Dimostrazione. Introduciamo la v.a.

$$Z_n = \frac{1}{n} \sum_{k=1}^n X_k ,$$

risulta

$$E(Z_n) = \frac{1}{n} \sum_{k=1}^n E(X_k) , \quad V(Z_n) = \frac{1}{n^2} V\left(\sum_{k=1}^n X_k\right)$$

e applicando la disuguaglianza di Chebychev (con $\varepsilon > 0$ arbitrario, ponendo $c = \varepsilon/\sqrt{V(Z_n)}$)

$$P(|Z_n - E(Z_n)| < \varepsilon) \geq 1 - V(Z_n)/\varepsilon^2 = 1 - \frac{1}{\varepsilon^2} \frac{1}{n^2} V\left(\sum_{k=1}^n X_k\right)$$

che tende a 1 per $n \rightarrow +\infty$. *q.e.d.*

Se le v.a. X_1, X_2, \dots sono a due a due indipendenti la condizione di Markov diventa

$$\lim_{n \rightarrow +\infty} \frac{1}{n^2} \sum_{k=1}^n V(X_k) = 0 .$$

Si ottiene allora immediatamente il seguente corollario

Teorema di Chebychev. Sia X_1, X_2, \dots una successione di v.a. a due a due indipendenti aventi varianza finita e limitata da una stessa costante:

$$V(X_k) \leq C , \quad k = 1, 2, \dots ,$$

allora per ogni $\varepsilon > 0$ risulta

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k)\right| < \varepsilon\right) = 1 .$$

Infatti in tal caso

$$\frac{1}{n^2} \sum_{k=1}^n V(X_k) \leq \frac{C}{n} \rightarrow 0 .$$

Se inoltre tutte le v.a. hanno la stessa media $E(X_k) = \mu$ allora

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \varepsilon\right) = 1 .$$

Questo risultato è il fondamento teorico della procedura di stima di una grandezza x mediante la *media aritmetica* dei valori ottenuti in n misure indipendenti che, in presenza di perturbazioni

casuali con dispersione finita e limitata, ma in assenza di errori sistematici, forniscono i valori x_1, x_2, \dots, x_n :

$$x \sim \frac{x_1 + x_2 + \dots + x_n}{n} .$$

Infatti ad ogni misura indipendente si può associare una v.a. X_k , supponendo che il suo valor medio $E(X_k)$ sia uguale al valore vero x della grandezza e che i valori da essa assunti siano distribuiti intorno ad x con una dispersione finita, limitata in modo uniforme: $V(X_k) \leq C$.

Un caso particolare del Teorema di Chebychev, con v.a. aventi tutte la stessa media, è il seguente

Teorema di Bernoulli. Si consideri una successione di prove indipendenti, in ciascuna delle quali vi è una probabilità p di successo e $q = 1 - p$ di fallimento. Sia F_n il numero di successi in n prove e $f_n = F_n/n$ la corrispondente frequenza relativa. Allora per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|f_n - p| < \varepsilon) = 1 .$$

Ovvero la frequenza relativa tende, in probabilità, alla probabilità p . Il risultato è immediato osservando che, associando ad ogni prova una v.a. K_j che assume il valore 1 in caso di successo e 0 in caso di fallimento, le K_j sono indipendenti e

$$E(K_j) = p \quad , \quad V(K_j) = pq = p(1 - p) \leq 1/4 .$$

Un poco più generale è il

Teorema di Poisson. Sempre nell'ipotesi precedente di prove indipendenti, ma con probabilità di successo p_j variabile nelle diverse prove, si ha

$$\lim_{n \rightarrow +\infty} P(|f_n - \frac{p_1 + p_2 + \dots + p_n}{n}| < \varepsilon) = 1 .$$

Infatti, come sopra

$$E(K_j) = p_j \quad , \quad V(K_j) = p_j q_j \leq 1/4 .$$

In tutti i risultati precedenti si considerano v.a. con varianza finita. Il seguente teorema evita invece tale ipotesi, considerando però soltanto variabili equidistribuite.

Teorema di Khinchin. Se le v.a. X_1, X_2, \dots sono a due a due indipendenti ed hanno la stessa distribuzione con media finita μ , allora per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|\frac{1}{n} \sum_{k=1}^n X_k - \mu| < \varepsilon) = 1 .$$

* * *

Dimostrazione. Dato $\eta > 0$ ed n intero positivo, per $k = 1, 2, \dots, n$ scomponiamo X_k nel modo seguente:

$$X_k = Y_k + Z_k ,$$

avendo posto

$$\begin{cases} Y_k = X_k & Z_k = 0 & \text{se } |X_k| < n\eta \\ Y_k = 0 & Z_k = X_k & \text{se } |X_k| \geq n\eta \end{cases} .$$

Per ipotesi tutte le v.a. hanno la stessa distribuzione, quella ad esempio di $X = X_1$, con media finita μ . Dunque converge l'integrale

$$a = \int_{\Omega} |X| dP .$$

Le v.a. Y_k hanno media e varianza finite:

$$\mu_n = E(Y_k) = \int_{|X| < n\eta} X dP \rightarrow \mu \text{ per } n \rightarrow +\infty ,$$

$$\sigma_n^2 = V(Y_k) = \int_{|X| < n\eta} X^2 dP - \mu_n^2 \leq n\eta \int_{|X| < n\eta} |X| dP \leq n\eta a .$$

Fissato $\varepsilon > 0$ arbitrario, per $n > \nu$ sarà $|\mu_n - \mu| < \varepsilon$ e per la disuguaglianza di Chebychev

$$P\left(\left|\frac{1}{n} \sum_{k=1}^n Y_k - \mu_n\right| \geq \varepsilon\right) \leq \frac{a\eta}{\varepsilon^2} .$$

Infatti $V = V(1/n \sum_{k=1}^n Y_k) = n/n^2 \sigma_n^2 \leq a\eta$ e, se $\varepsilon^2 = \lambda^2 V$, si ha $\lambda^{-2} = V/\varepsilon^2 \leq a\eta/\varepsilon^2$. Allora, per $n > \nu$ risulta

$$P\left(\left|\frac{1}{n} \sum_{k=1}^n Y_k - \mu\right| \geq 2\varepsilon\right) \leq \frac{a\eta}{\varepsilon^2} .$$

Osserviamo inoltre che, per $n > \nu^*$ opportuno,

$$n\eta P(Z_k \neq 0) = n\eta P(|X| \geq n\eta) \leq \int_{|X| \geq n\eta} |X| dP \leq \eta^2 ,$$

in considerazione della convergenza di $\int_{|X| \geq n\eta} |X| dP$ a $\int_{\Omega} |X| dP$, e dunque $P(Z_k \neq 0) \leq \eta/n$. Allora, essendo $Z_k \geq 0$,

$$P\left(\frac{1}{n} \sum_{k=1}^n Z_k \neq 0\right) = P\left(\sum_{k=1}^n Z_k \neq 0\right) \leq \sum_{k=1}^n P(Z_k \neq 0) \leq \sum_k \eta/n = \eta .$$

Poniamo

$$A_n = \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| \geq 2\varepsilon \right\} , \quad B_n = \left\{ \left| \frac{1}{n} \sum_{k=1}^n Y_k - \mu \right| \geq 2\varepsilon \right\} , \quad C_n = \left\{ \frac{1}{n} \sum_{k=1}^n Z_k \neq 0 \right\} .$$

Ovviamente

$$A_n \subseteq (B_n \cap C_n^c) \cup C_n$$

e dunque

$$P(A_n) \leq P(B_n) + P(C_n) \leq \frac{a\eta}{\varepsilon^2} + \eta .$$

Per l'arbitrarietà di η e ε si ha $\lim_n P(A_n) = 0$ e il teorema è dimostrato. *q.e.d.*

& & &

6.4 La legge forte dei grandi numeri

Si dice che per una successione X_1, X_2, \dots di v.a. dotate di media vale la legge **forte dei grandi numeri** se

$$P(\omega \in \Omega \mid \lim_{n \rightarrow +\infty} \left| \frac{1}{n} \sum_{k=1}^n X_k(\omega) - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| = 0) = 1 .$$

Definizione. Le v.a. X_n convergono *quasi certamente* (q.c.) (a.s. in inglese: “almost surely”), o con probabilità 1, alla v.a. X se e solo se

$$P(\{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}) = 1 ,$$

ovvero

$$P(\{\omega \in \Omega \mid X_n(\omega) \not\rightarrow X(\omega)\}) = 0 .$$

Osservazione. In teoria della misura si usano le espressioni equivalenti “quasi ovunque (q.o.)” e “almost everywhere (a.e.)”.

Dunque vale la legge forte dei grandi numeri se e solo se

$$M_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \rightarrow 0$$

quasi certamente.

Osservazione. Si dimostra che la convergenza quasi certa implica la convergenza in probabilità. Invece la convergenza in probabilità di X_n a X non implica che $X_n \rightarrow X$ q.c. Tuttavia si può sempre trovare una sottosuccessione X_{n_k} che converge q.c. ad X .

* * *

Teorema. Sia X_n convergente q.c. ad X . Allora X_n converge in probabilità ad X .

Dimostrazione. Sia $C \subseteq \Omega$ l'insieme sul quale X_n converge ad X :

$$C = \left\{ \omega \mid \forall k \exists n \forall p \geq n \mid X_n(\omega) - X(\omega) \mid < \frac{1}{k} \right\}.$$

Allora, fissato comunque j , l'insieme

$$A_j = \bigcap_n \bigcup_{p \geq n} \left\{ \mid X_p - X \mid \geq \frac{1}{j} \right\}$$

è contenuto nel complementare di C (se $\omega \in A_j$, per infiniti indici p si ha $\mid X_p(\omega) - X(\omega) \mid \geq 1/j$ e non vi può essere convergenza in ω). Essendo $P(C) = 1$, si ha $P(A_j) = 0$. Ma per la decrescenza dell'unione al crescere di n

$$0 = P(A_j) = \lim_n P(\bigcup_{p \geq n} \left\{ \mid X_p - X \mid \geq \frac{1}{j} \right\}) \geq \lim_n P(\left\{ \mid X_n - X \mid \geq \frac{1}{j} \right\}).$$

Fissato ε , sia j tale che $1/j \leq \varepsilon$, allora

$$\lim_n P(\left\{ \mid X_n - X \mid \geq \varepsilon \right\}) = 0$$

e la convergenza in probabilità è stabilita. *q.e.d.*

Osservazione. Se $X_n \rightarrow X$ in probabilità, non necessariamente $X_n \rightarrow X$ q.c. Basta fornire un controesempio: sia $\Omega = [0, 1[$ con P uguale alla misura di Lebesgue standard su tale intervallo; sia

$$\chi_{n,k} = \chi_{[(k-1)/n, k/n[} \quad n = 1, 2, \dots, \quad k = 1, 2, \dots, n.$$

Ordiniamo queste funzioni caratteristiche formando la successione $X_p = \chi_{n,k}, p = 1 + 2 + \dots + (n-1) + k$. Per ogni $\omega \in [0, 1[$ vi sono infiniti indici p per i quali $X_p(\omega) = 1$ e infiniti per i quali $X_p(\omega) = 0$: quindi la successione X_p non converge in nessun punto. Ma ovviamente per ogni $0 < \varepsilon < 1$

$$P(\left\{ \mid \chi_{n,k} - 0 \mid \geq \varepsilon \right\}) = \frac{1}{n} \rightarrow 0$$

e le v.a. X_p tendono a 0 in probabilità.

Teorema. Sia X_n convergente in probabilità ad X . Allora esiste una sottosuccessione X_{n_k} convergente q.c. ad X .

Dimostrazione. Siano α_n e η_n numeri positivi tali che

$$\lim_{n \rightarrow +\infty} \alpha_n = 0 \quad \text{e} \quad \sum_{n=1}^{+\infty} \eta_n < +\infty.$$

Selezioniamo, in virtù dell'ipotesi di convergenza in probabilità, degli indici $n_1 < n_2 < \dots$ tali che

$$P(\left\{ \mid X_{n_k} - X \mid \geq \alpha_k \right\}) < \eta_k.$$

Siano infine

$$A_j = \bigcup_{k=j}^{+\infty} \left\{ \mid X_{n_k} - X \mid \geq \alpha_k \right\} \quad \text{e} \quad B = \bigcap_{j=1}^{+\infty} A_j.$$

Si ha $A_{j+1} \subseteq A_j$ e, per la continuità di P

$$\sum_{k=j}^{+\infty} \eta_k \geq P(A_j) \rightarrow P(B) .$$

Poiché il resto della serie tende a 0 si ottiene $P(B) = 0$. Ma

$$\forall \omega \in \Omega - B \quad \lim_k X_{n_k}(\omega) = X(\omega) .$$

Infatti se $\omega \notin B$ esiste j tale che $\omega \notin A_j$, cioè per ogni $k \geq j$

$$\omega \notin \{|X_{n_k} - X| \geq \alpha_k\} \quad \text{ovvero} \quad |X_{n_k}(\omega) - X(\omega)| < \alpha_k .$$

Ma $\alpha_k \rightarrow 0$ e il teorema è dimostrato.

Osservazione. Spesso è importante considerare convergenze di v.a. di tipo integrale, principalmente in $L^p, p \geq 1$:

$$X_n \rightarrow X \text{ in } L^p \Leftrightarrow \int_{\Omega} |X_n(\omega) - X(\omega)|^p dP(\omega) \rightarrow 0 .$$

È ben noto in teoria generale dell'integrazione che per misure finite, quali ovviamente sono le misure di probabilità, se $X_n \rightarrow X$ in L^p e $q < p$ allora $X_n \rightarrow X$ in L^q e che $X_n \rightarrow X$ in L^p allora esiste una sottosuccessione X_{n_k} convergente quasi certamente (anche se l'intera successione in generale non converge q.c.)

Inoltre la convergenza in L^p implica la convergenza in probabilità: per la disuguaglianza di Markov si ha

$$P(\{|X_n - X| \geq \varepsilon\}) = P(\{|X_n - X|^p \geq \varepsilon^p\}) \leq \frac{E(|X_n - X|^p)}{\varepsilon^p} \rightarrow 0 .$$

& & &

Prima di riportare alcuni teoremi fondamentali sulla legge dei grandi numeri, premettiamo qualche risultato di grande rilievo ed utilità in numerose questioni.

Lemma di Cantelli. Data una successione di eventi A_n tali che

$$\sum_{n=1}^{+\infty} P(A_n) < +\infty ,$$

posto

$$A = \overline{\lim}_n A_n = \bigcap_n \bigcup_{p>n} A_p = \{A_n \text{ i.o.}\} ,$$

si ha $P(A) = 0$.

Osservazione. i.o. significa *infinitely often* , ovvero per infini valori dell'indice n . Infatti $\omega \in A$

se e solo se $\omega \in A_n$ per infiniti valori di n .

Dimostrazione. Se $E_n = \cup_{p>n} A_p$, allora $E_n \supseteq E_{n+1}$, cioè la successione E_n è generalmente decrescente. Pertanto

$$P(A) = \lim_n P(E_n), \text{ ma } P(E_n) \leq \sum_{p>n} P(A_p) \rightarrow 0 \text{ per } n \rightarrow +\infty . \text{ q.e.d.}$$

Il lemma di Cantelli può essere integrato dal seguente risultato dovuto a Borel:

Lemma. Se gli eventi A_n sono *mutuamente indipendenti* e

$$\sum_n P(A_n) = +\infty ,$$

allora, posto

$$A = \overline{\lim}_n A_n = \cap_n \cup_{p>n} A_p = \{A_n \text{ i.o.} \} ,$$

si ha $P(A) = 1$.

Dimostrazione. Conviene considerare il complementare:

$$P(A^c) = P(\cup_n \cap_{p>n} A_p^c) \leq \sum_n P(\cap_{p>n} A_p^c) = 0 .$$

Perché per ogni n , per l'indipendenza delle A_p , si ha

$$P(\cap_{p>n} A_p^c) = \prod_{p>n} (1 - P(A_p)) = 0 ,$$

in virtù della divergenza della serie $\sum_{p>n} P(A_p)$. *q.e.d.*

Ricordiamo infatti che

$$\prod_n (1 - x_n) = \exp\left(\sum_n \ln(1 - x_n)\right) \rightarrow 0 \text{ se } \sum_n \ln(1 - x_n) \rightarrow -\infty ,$$

come avviene certamente se la serie maggiorante $\sum_n -x_n$ diverge a $-\infty$.

L'insieme dei due risultati (lemma di Cantelli e lemma di Borel) sono spesso presentati congiuntamente come **lemma di Borel-Cantelli**

Molti casi interessanti di eventi per i quali si sa a priori che si verificano o con probabilità 1 o con probabilità 0 sono forniti dalla **legge zero-uno di Kolmogorov**.

Premettiamo alcune definizioni.

Definizione. Sia $\{X_\gamma\}_{\gamma \in \Gamma}$ una famiglia arbitraria di v.a. Esse si dicono (mutuamente) indipendenti se e solo se ogni sottofamiglia finita $\{X_1, X_2, \dots, X_N\}$ è costituita da v.a. indipendenti, dunque se le σ -algebre $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_N)$ sono indipendenti.

Definizione. Sia $\{X_\gamma\}_{\gamma \in \Gamma}$ una famiglia arbitraria di v.a. Con $\sigma(X_\gamma \mid \gamma \in \Gamma)$ si indica la più piccola σ -algebra generata dalla famiglia di insiemi $\{X_\gamma^{-1}(B) \mid \gamma \in \Gamma, B \in \mathcal{B}\}$.

Definizione. Sia $X_1, X_2, \dots, X_n, \dots$ una successione di v.a. La σ -algebra

$$\tau = \bigcap_{n=1}^{+\infty} \sigma(X_n, X_{n+1}, \dots)$$

si dice σ -algebra di coda della successione.

Teorema zero-uno di Kolmogorov. Se le v.a. X_n sono indipendenti, τ è banale, cioè se $A \in \tau$ allora $P(A) = 0$ oppure $P(A) = 1$.

Dimostrazione. Sia $A \in \tau$. Allora per ogni n si ha $A \in \sigma(X_{n+1}, X_{n+2}, \dots)$. Ma le famiglie $\{X_1, X_2, \dots, X_n\}$ e $\{X_{n+1}, X_{n+2}, \dots\}$ sono indipendenti e quindi per ogni n , A è indipendente da $\{X_1, X_2, \dots, X_n\}$ e dunque è indipendente da $\{X_1, X_2, \dots\}$. Ma $A \in \sigma(X_1, X_2, \dots)$ e dunque A è indipendente da sé stesso: $P(A) = P(A \cap A) = P(A)^2$ e $P(A) = 0$ oppure $P(A) = 1$. *q.e.d.*

Osservazione. Eventi concernenti le v.a.

$$X_* = \liminf_n X_n, \quad X^* = \limsup_n X_n, \quad X = \lim_n X_n,$$

appartengono a τ . Infatti, per ogni n :

$$\sigma(X_*), \sigma(X^*), \sigma(X) \subseteq \sigma(X_n, X_{n+1}, \dots).$$

Ad esempio

$$\left\{ \liminf_n X_n \leq c \right\} = \left\{ \sup_n \inf_{p \geq n} X_p \leq c \right\} = \bigcap_n \bigcup_{p \geq n} \{X_p \leq c\}$$

e l'ultima unione appartiene a $\sigma(X_n, X_{n+1}, \dots)$.

Teorema (criterio di convergenza q.c.). Se, per ogni intero positivo r , si ha

$$\sum_{k=1}^{+\infty} P(|X_k - X| \geq \frac{1}{r}) < +\infty,$$

allora $X_n \rightarrow X$ quasi certamente.

Dimostrazione. Sia C l'insieme dove X_n converge ad X e quindi C^c l'insieme dove X_n non converge ad X . Per $\omega \in C^c$

$$\exists r \forall \nu \exists n > \nu \quad |X_n(\omega) - X(\omega)| \geq \frac{1}{r},$$

ovvero, posto $A_n^r = \{|X_n - X| \geq 1/r\}$,

$$C^c \subseteq \bigcup_r \bigcap_\nu \bigcup_{n > \nu} A_n^r = \bigcup_r A^r, \quad A^r = \overline{\lim}_n A_n^r.$$

(In effetti i due insiemi coincidono).

Per l'ipotesi e il lemma di Cantelli $P(A^r) = 0$. Allora, per la σ -subadditività di P , si ha

$P(C^c) = 0$. *q.e.d.*

Teorema (generalizzazione del precedente criterio di convergenza q.c.). Se, per ogni intero positivo r , esiste una successione di interi $1 = n_1 < n_2 < \dots (\rightarrow +\infty)$ tale che

$$\sum_{k=1}^{+\infty} P\left(\max_{n_k \leq n < n_{k+1}} |X_n - X| \geq \frac{1}{r}\right) < +\infty,$$

allora $X_n \rightarrow X$ quasi certamente.

Dimostrazione. Indicando sempre con C^c l'insieme di non convergenza e ponendo $E_k^r = \{\max_{n_k \leq n < n_{k+1}} |X_n - X| \geq 1/r\}$, per $\omega \in C^c$

$$\exists r \forall n_k \exists n > n_k |X_n(\omega) - X(\omega)| \geq \frac{1}{r}.$$

Dunque, se n è tale che $n_k < n_p \leq n < n_{p+1}$, $\omega \in E_p^r$ e

$$C^c \subseteq \cup_r \cap_k \cup_{p > k} E_p^r = \cup_r E^r, \quad E^r = \overline{\lim}_n E_n^r.$$

Come nel teorema precedente, per l'ipotesi e il lemma di Cantelli $P(E^r) = 0$. Allora, per la σ -subadditività di P , si ha $P(C^c) = 0$. *q.e.d.*

Disuguaglianza di Kolmogorov. Siano $X_1, X_2, \dots, X_k, \dots, X_n$ v.a. mutuamente indipendenti con medie μ_k e varianze σ_k^2 finite. Allora, per ogni $\varepsilon > 0$, posto

$$A_k = \left\{ \left| \sum_{j=1}^k (X_j - \mu_j) \right| < \varepsilon \right\}, \quad E_0 = A_1 \cap A_2 \cap \dots \cap A_n,$$

si ha

$$P(E_0) = P\left(\max_{1 \leq k \leq n} \left| \sum_{j=1}^k (X_j - \mu_j) \right| < \varepsilon\right) \geq 1 - \frac{1}{\varepsilon^2} \sum_{k=1}^n \sigma_k^2.$$

Dimostrazione. Non è restrittivo supporre per semplicità $\mu_k = 0$ (altrimenti si ragionerebbe sulle v.a. $Y_k = X_k - \mu_k$). Poniamo

$$S_k = \sum_{j=1}^k X_j \text{ e dunque } A_k = \{|S_k| < \varepsilon\}.$$

Poniamo inoltre

$$E_k = A_1 \cap A_2 \dots \cap A_{k-1} \cap A_k^c,$$

allora k è il primo indice per il quale $|S_j| \geq \varepsilon$, gli eventi E_k sono disgiunti e $E_0^c = \cup_{k=1}^n E_k$. Dunque

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon\right) = \sum_{k=1}^n P(E_k).$$

Ora

$$V(S_n) = \int_{\Omega} S_n^2 dP = \sum_{k=0}^n \int_{E_k} S_n^2 dP \geq \sum_{k=1}^n \int_{E_k} S_n^2 dP =$$

$$= \sum_{k=1}^n \int_{\Omega} \chi_{E_k} (S_k^2 + 2 \sum_{j>k} S_k X_j + 2 \sum_{i>j>k} X_i X_j + \sum_{j>k} X_j^2) dP \geq \sum_{k=1}^n \varepsilon^2 P(E_k) ,$$

perchè $\sum_{j>k} X_j^2 \geq 0$, $S_k^2 \chi_{E_k} \geq \varepsilon^2$, le v.a. X_j, X_i sono indipendenti da $S_k \chi_{E_k}, \chi_{E_k}$ e $E(X_i) = E(X_j) = 0$. Pertanto

$$P(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} V(S_n) . \quad q.e.d.$$

Siamo ora in grado di dimostrare un risultato fondamentale sulla legge forte dei grandi numeri.

Teorema di Kolmogorov. Se la successione X_1, X_2, \dots di v.a. *mutuamente indipendenti* soddisfa la condizione

$$\sum_{k=1}^{+\infty} \frac{V(X_k)}{k^2} < +\infty ,$$

allora per essa vale la legge forte dei grandi numeri.

* * *

Dimostrazione. Poniamo $\mu_k = E(X_k)$, $\sigma_k^2 = V(X_k)$ e

$$S_n = \sum_{k=1}^n (X_k - \mu_k) , \quad s_n = \frac{1}{n} S_n .$$

In virtù del secondo criterio di convergenza visto sopra, posto per ogni intero q

$$P_q = P(\max_n |s_n| \geq \varepsilon , 2^q \leq n < 2^{q+1}) ,$$

basta far vedere che per ogni $\varepsilon > 0$ si ha

$$\sum_{q=1}^{+\infty} P_q < +\infty .$$

Ma, ricorrendo alla disuguaglianza di Kolmogorov, si ottiene

$$P_q \leq P(\max_n |S_n| \geq 2^q \varepsilon , 2^q \leq n < 2^{q+1}) \leq \frac{1}{(2^q \varepsilon)^2} \sum_{j < 2^{q+1}} \sigma_j^2$$

e dunque

$$\sum_{q=1}^{+\infty} P_q \leq \sum_{q=1}^{+\infty} \frac{1}{(2^q \varepsilon)^2} \sum_{j < 2^{q+1}} \sigma_j^2 =$$

$$= \frac{1}{\varepsilon^2} \sum_{j=1}^{+\infty} \sigma_j^2 \sum_{2^{q+1} > j} 2^{-2q} .$$

Se $2^s \leq j < 2^{s+1}$, allora

$$\sum_{2^{q+1} > j} 2^{-2q} = \sum_{q=s}^{+\infty} = \frac{4}{3} 2^{-2s} \leq \frac{16}{3j^2}$$

e pertanto

$$\sum_{q=1}^{+\infty} P_q \leq \frac{16}{3\varepsilon^2} \sum_{j=1}^{+\infty} \frac{\sigma_j^2}{j^2} < +\infty ,$$

per l'ipotesi del teorema. q.e.d.

& & &

Segue immediatamente come corollario il seguente teorema, già dimostrato nel 1909:

Teorema di Borel. Si consideri uno schema binomiale di prove indipendenti, ciascuna con probabilità p di successo e $q = 1 - p$ di fallimento. Sia $f_n =$ la frequenza relativa dei successi nelle prime n prove. Allora

$$P\left(\lim_{n \rightarrow +\infty} f_n = p\right) = 1 .$$

Infatti la serie

$$\sum_{k=1}^{+\infty} \frac{pq}{k^2}$$

converge.

Ricordiamo infine il seguente risultato, già segnalato precedentemente, anch'esso dovuto a Kolmogorov.

Teorema. Per una successione di v.a. X_k identicamente distribuite, cioè tutte con la distribuzione uguale a quella di una stessa v.a. X ($X_k \sim X$), e mutuamente indipendenti vale la legge forte dei grandi numeri se esse ammettono valor medio finito:

$$\int_{\Omega} |X| dP < +\infty \Rightarrow P\left(\lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \int_{\Omega} X dP\right) = 1 ,$$

e viceversa, se la loro media aritmetica $(X_1 + X_2 + \dots + X_n)/n$ converge quasi certamente ad un limite finito, allora esse ammettono valor medio finito, uguale a tale limite:

$$P\left(\lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mu\right) = 1 , \mu \in \mathbf{R} \Rightarrow \int_{\Omega} |X| dP < +\infty \text{ e } \mu = \int_{\Omega} X dP .$$

* * *

Dimostrazione. Cominciando dalla seconda parte, osserviamo che X è integrabile se (e solo se)

$$\sum_{n=1}^{+\infty} P(\{|X| \geq n\}) \leq +\infty .$$

Infatti in tal caso, essendo

$$\sum_{n=0}^{+\infty} P(\{|X| \geq n\}) = \sum_{n=0}^{+\infty} \sum_{k=n}^{+\infty} P(\{k \leq |X| < k+1\}) = \sum_{k=0}^{+\infty} (k+1)P(\{k \leq |X| < k+1\}) ,$$

la v.a. discreta $F(\omega) = (k+1)$ per $\omega \in \{k \leq |X| < k+1\}$, dominante X , risulta integrabile e basta allora applicare il teorema sulla convergenza dominata. (Inversamente se X è integrabile, allora, essendo $F-1 \leq |X|$, anche F è integrabile e la prima serie converge).

Ma, ponendo $S_n = X_1 + X_2 + \dots + X_n$, si ha

$$\frac{S_n}{n} \rightarrow \mu \text{ q.c.} \Rightarrow \frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \rightarrow 0 \text{ q.c.}$$

e allora, se $A_n = \{|X_n| \geq n\} = \{|X_n/n| \geq 1\}$, può risultare che A_n si verifichi per infiniti indici n solo con probabilità 0:

$$P(\cap_k \cup_{n \geq k} A_n) = 0 .$$

Allora, in considerazione del lemma di Borel-Cantelli e poiché $X_n \sim X$, deve essere

$$\sum_{n=1}^{+\infty} P(\{|X_n| \geq n\}) = \sum_{n=1}^{+\infty} P(\{|X| \geq n\}) < +\infty$$

ed X ammette valor medio finito $E(X)$. Quando avremo dimostrato la prima parte della tesi risulterà, in virtù dell'unicità del limite, che necessariamente $\mu = E(X)$.

Per quanto concerne la prima parte, osserviamo in primo luogo che non è restrittivo supporre $E(X) = 0$, altrimenti ragioneremo sulla successione $X_n - E(X)$. Usiamo poi un metodo di troncamento simile a quello impiegato nella dimostrazione del teorema di Khintchin, ponendo

$$Y_n = \begin{cases} X_n & \text{se } |X_n| < n \\ 0 & \text{se } |X_n| \geq n \end{cases}$$

Allora, essendo le X_n identicamente distribuite (con la distribuzione di X):

$$V(Y_n) \leq E(Y_n^2) \leq \sum_{k=0}^n (k+1)^2 P(\{k \leq |X| < k+1\})$$

e quindi

$$\begin{aligned} \sum_{n=1}^{+\infty} \frac{V(Y_n)}{n^2} &\leq \sum_{n=1}^{+\infty} \frac{1}{n^2} \sum_{k=0}^n (k+1)^2 P(\{k \leq |X| < k+1\}) = \\ &= \sum_{k=0}^{+\infty} (k+1)^2 P(\{k \leq |X| < k+1\}) \sum_{n \geq \max(k,1)} \frac{1}{n^2} \leq 4 \sum_{k=0}^n (k+1) P(\{k \leq |X| < k+1\}) < +\infty , \end{aligned}$$

perché, per $k \geq 1$:

$$\sum_{n \geq k} \frac{1}{n^2} \leq \frac{1}{k^2} + \sum_{n > k} \frac{1}{n(n-1)} = \frac{1}{k^2} + \sum_{n > k} \left(\frac{1}{(n-1)} - \frac{1}{n} \right) = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k}$$

e $2(k+1)^2/k \leq 4(k+1)$. Dunque per la successione Y_n vale la legge forte dei grandi numeri e dunque, posto $\mu_k = E(Y_k)$:

$$P\left(\lim_{n \rightarrow +\infty} \frac{(Y_1 - \mu_1) + (Y_2 - \mu_2) + \dots + (Y_n - \mu_n)}{n} = 0\right) = 1.$$

L'esistenza di $E(X)$ finita e il fatto che X_n ha la stessa distribuzione di X implicano che, indicando con χ_n la funzione caratteristica di $|x| < n$:

$$\begin{aligned} \mu_n &= \int_{|X_n| < n} X_n dP = \int_{\Omega} X_n \chi_n(X_n) dP = \\ &= \int_{\Omega} X \chi_n(X) dP = \int_{|X| < n} X dP \rightarrow 0. \end{aligned}$$

Dunque anche

$$\lim_{n \rightarrow +\infty} \frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} = 0.$$

Sia C l'insieme dove per infiniti valori dell'indice n si ha $X_n \neq Y_n$:

$$C = \bigcap_N \bigcup_{n \geq N} \{ X_n \neq Y_n \}.$$

Allora risulta

$$\begin{aligned} P(C) &= \lim_N P(\bigcup_{n \geq N} \{ X_n \neq Y_n \}), \\ P(\bigcup_{n \geq N} \{ X_n \neq Y_n \}) &\leq \sum_{n \geq N} P(\{ X_n \neq Y_n \}) = \sum_{n \geq N} P(\{|X_n| \geq n\}) \rightarrow 0, \end{aligned}$$

perché X è integrabile e quindi la serie $\sum_n P(\{|X_n| \geq n\})$ converge. Dunque $P(C) = 0$. Ma sul complementare C^c , che ha probabilità 1, si ha definitivamente $X_n = Y_n$ e dunque

$$\lim_{n \rightarrow +\infty} \frac{Y_1 + Y_2 + \dots + Y_n}{n} = \lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \dots + X_n}{n} = 0 = E(X).$$

q.e.d.

& & &

Osservazione. La legge dei grandi numeri, debole o forte, non fornisce alcuna "certezza sul comportamento asintotico delle medie aritmetiche. I risultati indicano soltanto che la *probabilità* di deviazioni significative dal valor medio sono "trascurabili. Nel caso della legge debole si afferma che la *probabilità* di uno scostamento, comunque prefissato, della media aritmetica dal suo valor medio può essere resa arbitrariamente piccola per n sufficientemente grande. Nel caso della legge forte si afferma che la *probabilità* di non convergenza della differenza della media

aritmetica dal suo valor medio è nulla (ma non impossibile).

Anche se questi risultati “suggeriscono di interpretare la probabilità come limite della frequenza, essi non “dimostrano che la probabilità si ottiene come limite (nel senso ordinario) della frequenza.

Bibliografia

- [1] BILLINGSLEY, P.: *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.
- [2] BORKAR, V.S.: *Probability Theory*. Springer-Verlag, New York, 1995.
- [3] CRAMÉR, H.: *Mathematical Methods of Statistics*. Princeton University Press, Princeton, 1971.
- [4] DOOB J.L.: *Measure Theory*. Springer-Verlag, New York 1993.
- [5] GALAMBOS, J.: *Introductory Probability Theory*. Marcel Dekker, New York, 1984.
- [6] GALAMBOS, J.: *Advanced Probability Theory*. Marcel Dekker, New York, 1988.
- [7] GIHMAN, I.I.-SKOROHOD, A.V.: *The Theory of Stochastic Processes I*. Springer-Verlag, Berlin, 1974.
- [8] GNEDENKO, B.: *The Theory of Probability*. Mir Publishers, Moscow, 1973.
- [9] KOLMOGOROV, A.N.: *Grundbegriffe der Wahrscheinlichkeits-Rechnung*. Verlag von Julius Springer, Berlin, 1933.
- [10] ØKSENDAL, B.: *Stochastic Differential Equations*. Springer-Verlag, Berlin, 1989.
- [11] NEGRO, A.: *Teoria della Misura*. Quaderni del Dipartimento di Matematica, Università di Torino, 2001.
- [12] STONE, M.H.: *The Theory of Representations for Boolean Algebras*. Trans. Amer. Math. Soc. 40 (1936), pp. 37-111.
- [13] VAN DER WAERDEN, B.L.: *Mathematical Statistics*. Springer-Verlag, Berlin, 1969.
- [14] VENTSEL, A.D.: *Teoria dei processi stocastici*. Editori Riuniti Edizioni Mir, Roma, 1983.