Probabilità e Statistica I

Angelo Gilio
Dip. Met. Mod. Mat. - Univ. "La Sapienza"
gilio@dmmm.uniroma1.it

22 settembre 2008

... *Probability doesn't exist* ... (Bruno de Finetti)

Informazioni

Tutoraggio su web:

http://www.dmmm.uniroma1.it/∼gilio/stdinfo/

file: Prog-ProbSt-I.doc
cartelle: esami-AT, esami-Ges, esami-EIT-LT
e-mail: gilio@dmmm.uniroma1.it

Libro adottato:
Primi Passi in Probabilità e Statistica,
R. Scozzafava, Zanichelli, 2001.

Esercizi: *Training Autogeno in Probabilità*, Joseph Toscano, Zanichelli, ed.2000.

Ricevimento:
Lunedì, Mercoledì: 14.00 - 15.30
(Via Scarpa 10, pal. A, 1^0 piano)

Introduzione al CdP

Il calcolo delle probabilità si è sviluppato fra il XV e il XVI secolo, prevalentemente sulla base di studi e considerazioni teoriche riguardanti situazioni e problemi connessi ai giochi d'azzardo.

Si è soliti far risalire l'origini del CdP a certe questioni di scommessa poste dal Cavalier de Méré a Pascal e da questi discusse con Fermat.

Il primo libro sul gioco d'azzardo (*Liber de ludo aleae*) è stato scritto, anche se pubblicato successivamente, agli inizi del 1500 da Gerolamo Cardano (matematico, fisico, medico e astrologo italiano).

Lo sviluppo storico del calcolo delle probabilità è dovuto a grandi scienziati quali Galilei, Bernoulli, Pascal, Fermat, Laplace.

Nel nostro secolo la teoria delle probabilità si è sviluppata in molte direzioni grazie al lavoro di famosi matematici, fra i quali *Kolmogorov* e *B. de Finetti*.

Prop. logiche, eventi, indicatori

L'analisi di situazioni e problemi reali spesso comporta l'esame di fatti e aspetti incerti, che potranno successivamente risultare veri o falsi.

Nei problemi aleatori i fatti incerti sono formalizzati (in modo non ambiguo) mediante *proposizioni logiche* che possono assumere il valore *Vero* oppure *Falso*.

Una proposizione o affermazione logica si indica con il termine di *evento*, che si può definire come un'entità logica a due valori: vero (V) o falso (F).

Un evento definito mediante una proposizione logica che è certamente vera si dice *evento certo* e si indica con il simbolo Ω.

Un evento definito mediante una proposizione logica che è certamente falsa si dice *evento impossibile* e si indica con il simbolo $\emptyset$.

Gli eventi si indicano di solito con le lettere maiuscole: $A\,,\;\; B\,,\;\; \ldots\,,\;\; E\,,\;\; H\,,\;\; \ldots$

Indicatore di E:

$$|E| = \begin{cases} 1 & \text{se } E \text{ è vero,} \\ 0 & \text{se } E \text{ è falso.} \end{cases}$$

Osservazione: $|\Omega| = 1, \quad |\emptyset|$=0.

Relazioni e operazioni logiche tra eventi:

implicazione, uguaglianza, indipendenza logica, incompatibilità;
negazione, unione, intersezione.

Negazione L'evento *contrario* o *negazione* di un evento E è l'evento che è vero quando E è falso ed è falso quando E è vero. L'evento contrario di E si indica con il simbolo E^c.

$$E^c = \begin{cases} \text{vero} & \text{se } E \text{ falso,} \\ \text{falso} & \text{se } E \text{ vero.} \end{cases}$$

Utilizzando gli indicatori si ha: $|E^c| = 1 - |E|$.

Implicazione. Un evento A implica un altro evento B se quando è vero A segue che è vero anche B.

In simboli si scrive $A \Longrightarrow B$ oppure $A \subseteq B$.

$A \Longrightarrow B$ equivale alla diseguaglianza $|A| \leq |B|$.

Uguaglianza. Due eventi A e B si dicono uguali se ognuno dei due implica l'altro, cioè se $A \Longrightarrow B$ e $B \Longrightarrow A$.

Unione. L'*unione* o *somma* (logica) di due eventi A, B è l'evento che è vero quando almeno uno dei due eventi è vero ed è falso quando sia A che B sono falsi. Si indica con $A \vee B$ oppure $A \cup B$.

Osservazioni:

$$A \vee \Omega = \Omega \; ; \quad A \vee \emptyset = A \; ; \quad A \vee A = A \; ; \quad A \vee A^c = \Omega \; .$$

Intersezione

L'*intersezione* (logica) o *prodotto* (logico) di due eventi A, B è l'evento che è vero quando entrambi gli eventi sono veri ed è falso quando almeno uno dei due eventi A, B è falso.

L'evento intersezione di A, B si indica con $A \wedge B$, oppure $A \cap B$, o più semplicemente AB.

Osservazioni:

$$A \wedge \Omega = A \; ; \quad A \wedge \emptyset = \emptyset \; ; \quad A \wedge A = A \; ; \quad A \wedge A^c = \emptyset \; .$$

Incompatibilità

Due eventi A, B si dicono *incompatibili* se non possono essere entrambi veri, cioè se $AB = \emptyset$.

Proprietà degli indicatori:

$$|AB| = |A| \cdot |B| \; ; \quad |A \vee B| = |A| + |B| - |AB| \; ,$$

con $|A \vee B| = |A| + |B|$ nel caso in cui $AB = \emptyset$.

Altre proprietà:

$$AB \implies A \implies A \vee B \; ,$$
$$AB \implies B \implies A \vee B \; .$$

In termini di indicatori:

$$|AB| \;\le\; |A| \;\le\; |A \vee B| \; ; \quad |AB| \;\le\; |B| \;\le\; |A \vee B| \; .$$

Proprietà *distributive* :

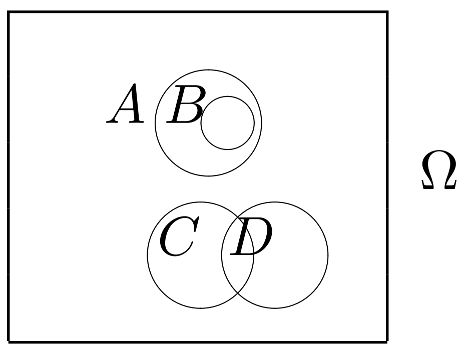
$$(A \vee B) \wedge C = AC \vee BC \; , \quad (A \wedge B) \vee C = (A \vee C) \wedge (B \vee C) \; .$$

Formule di De Morgan :

$$(A \vee B)^c = A^c \wedge B^c \;\; ; \quad (A \wedge B)^c = A^c \vee B^c \; .$$

Diagrammi di Venn.

Consentono una rappresentazione geometrica degli eventi, utile per esaminare le relazioni e operazioni logiche.



Eventi	**Insiemi**
certo	universo
impossibile	vuoto
contrario	complementare
implicazione	inclusione
incompatibili	disgiunti
unione	unione
intersezione	intersezione

Partizione Una famiglia di eventi $\{H_1, H_2, \ldots, H_n\}$ costituisce una partizione di Ω se valgono le seguenti due proprietà :

$$1.\;\; H_i H_j = \emptyset \;, \quad i \neq j \;; \qquad 2.\;\; H_1 \vee H_2 \vee \cdots \vee H_n = \Omega \;.$$

Utilizzando gli indicatori:

$$|H_1| + |H_2| + \cdots + |H_n| = 1 \;. \tag{1}$$

Costituenti.
Osserviamo che $\Omega \wedge \Omega = \Omega$ e che $E \vee E^c = \Omega$, $\forall\, E$.

Sfruttando tali relazioni, considerati degli eventi $A, B, \ldots$, si possono determinare i *casi possibili* o *costituenti*, sviluppando la seguente espressione:

$$\begin{aligned} &(A \vee A^c) \wedge (B \vee B^c) \wedge \cdots = \\ &= (AB \vee AB^c \vee A^cB \vee A^cB^c) \wedge \cdots = \\ &= AB \cdots \vee AB^c \cdots \vee A^cB \cdots \vee A^cB^c \cdots \vee \cdots \ . \end{aligned} \tag{2}$$

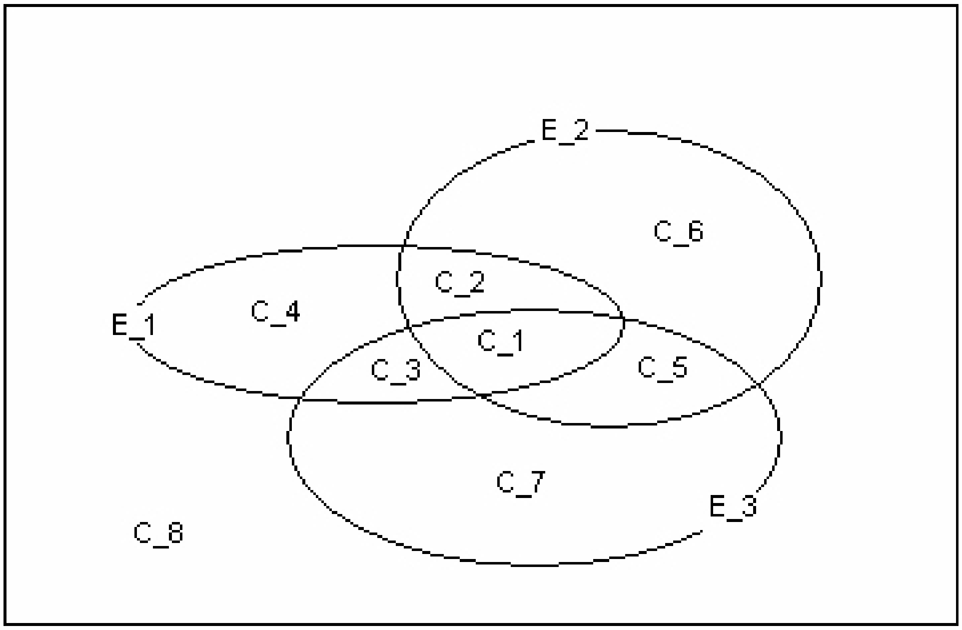
Eliminando le intersezioni impossibili, quelle rimanenti sono i casi possibili relativi agli eventi $A, B, \ldots$. .

In generale, data una famiglia $\mathcal{F}_n = \{E_1, \ldots, E_n\}$, i *casi possibili* o *costituenti*, $C_1, \ldots, C_m$, con $m \leq 2^n$, si ottengono sviluppando l'espressione

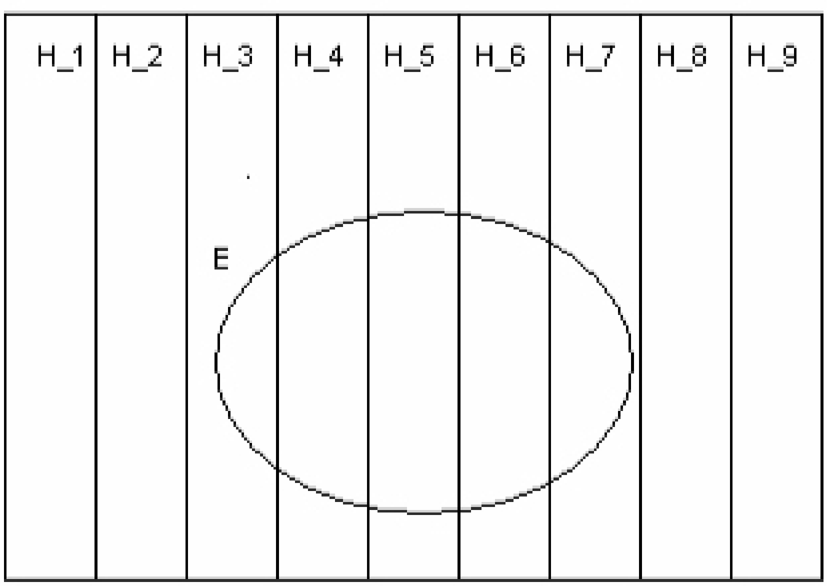
$$(E_1 \vee E_1^c) \wedge (E_2 \vee E_2^c) \wedge \cdots \wedge (E_n \vee E_n^c) = C_1 \vee C_2 \vee \cdots \vee C_m \ .$$

$$C_k = E_1^* E_2^* \cdots E_n^* \,, \quad k = 1, 2, \ldots, m \leq 2^n \,,$$

dove $E_i^* = E_i$, oppure $E_i^* = E_i^c$.



Ω



Ω

Decomposizione di un evento. Dato un evento arbitrario E ed una partizione $\{H, H^c\}$, si ha:

$$E = E \wedge \Omega = E \wedge (H \vee H^c) = EH \vee EH^c \,. \quad (3)$$

Più in generale, data una partizione $\{H_1, H_2, \dots, H_n\}$, si ha:

$$E = EH_1 \vee EH_2 \vee \cdots \vee EH_n \,. \quad (4)$$

In molti casi le formule (3) e (4) sono utili per calcolare la probabilità di E.

Indipendenza logica.

Gli eventi di una famiglia $\mathcal{F} = \{E_1, E_2, \dots, E_n\}$ si dicono logicamente indipendenti se il numero m di costituenti è pari a 2^n.

Esempio 1 Estrazioni con restituzione da un'urna contenente 1 pallina bianca e 1 nera.

Gli eventi

$$E_i = \text{la i-ma pallina estratta è bianca}\ , \quad i = 1, 2, \ldots, 5,$$

sono logicamente indipendenti.

Esempio 2 Estrazioni senza restituzione da un'urna contenente 2 palline bianche e 3 nere. Gli eventi E_1, E_2, E_3 non sono logicamente indipendenti. Ad esempio, se E_1, E_2 sono entrambi veri (cioè le prime due palline estratte sono *bianche*) la terza pallina è certamente *nera* e quindi E_3 è necessariamente falso.

Se invece E_1 è vero ed E_2 è falso, E_3 può essere sia vero che falso, e così via.

L'evento E_5 è *logicamente dipendente* da $E_1, \ldots, E_4$,

nel senso che il suo valore logico è *univocamente* determinato da quello dei primi quattro.

Infatti, assumendo noto il risultato delle prime quattro estrazioni, il risultato della quinta è scontato.

Criterio classico di valutazione della probabilità

In molti problemi aleatori, per ragioni di simmetria o di mancanza di informazioni sul fenomeno studiato, i casi *possibili* sono giudicati **ugualmente possibili**.

In tali situazioni, per valutare il grado di attendibilità di un evento A, è del tutto naturale basarsi sul *numero di casi favorevoli* a ciascuno degli eventi considerati.

Definizione 1 (Classica) Considerato un esperimento aleatorio con m casi possibili, giudicati ugualmente *possibili*, ed un evento E con r casi *favorevoli*, la probabilità $P(E)$ di E è uguale al rapporto $\frac{r}{m}$.

$$P(E) = \frac{\#\ \text{casi favorevoli}}{\#\ \text{casi possibili}}$$

Esempio *(lancio di 2 dadi)*. Indichiamo con X, Y il risultato dei due dadi e con $Z = X + Y$ il totale.

I casi possibili (le coppie (x, y)) sono $6 \times 6 = 36$;

$P(Z=3)=\frac{2}{36}=\frac{1}{18}$,
in quanto le coppie favorevoli sono due : $(1,2),(2,1)$.

$P(X>Y)=\frac{15}{36}=\frac{5}{12}$,
6 coppie favorevoli all'evento $(X=Y)$ e delle restanti 30 quelle favorevoli all'evento $(X>Y)$ sono 15.

Proprietà fondamentali della probabilità

Utilizzando la Definizione (1) si possono dimostrare le seguenti proprietà di base (*assiomi*) della probabilità.

- **P1.** $P(E)\geq 0$, per ogni evento E;
  (il numero r di casi favorevoli è non negativo e quindi $\frac{r}{m}\geq 0$)

- **P2.** $P(\Omega)=1$;
  (per l'evento certo Ω, si ha $r=m$ e quindi $\frac{r}{m}=1$)

- **P3.** se $AB=\emptyset$, allora $P(A\vee B)=P(A)+P(B)$ (*proprietà additiva*).

Dim. di **P3**.
Da $AB = \emptyset$, segue $r_{AB} = 0$ e quindi

$$r_{A \vee B} = r_A + r_B - r_{AB} = r_A + r_B.$$

Pertanto

$$P(A \vee B) = \frac{r_A + r_B}{m} = \frac{r_A}{m} + \frac{r_B}{m} = P(A) + P(B) \; .$$

In particolare, nel caso $B = A^c$, applicando **P2** e **P3** e osservando che $r_{A^c} = m - r_A$ si ottiene

$$P(A^c) = 1 - P(A) \; . \tag{5}$$

Se $AB = \emptyset$, ponendo $C = (A \vee B)^c$, si ha

$$P(C) = 1 - P(A \vee B) = 1 - P(A) - P(B) \; , \tag{6}$$

e quindi per la partizione $\{A, B, C\}$ vale

$$P(A) + P(B) + P(C) = 1 \; . \tag{7}$$

Se $E_1, \ldots, E_n$ sono a due a due incompatibili, si ha:

$$P(E_1 \vee \cdots \vee E_n) \;=\; P(E_1 \vee \cdots \vee E_{n-1}) + P(E_n) = P(E_1 \vee \cdots \vee E_{n-2}) + P(E_{n-1}) + P(E_n) \;= \cdots \;= P(E_1) + P(E_2) + \cdots + P(E_n)\ , \tag{8}$$

Se $E_1, \ldots, E_n$ formano una partizione di Ω si ha:

$$P(E_1) + P(E_2) + \cdots + P(E_n) = 1\ . \tag{9}$$

Proprietà di monotonia. Se $A \subseteq B$, si ha

$$B = A \vee A^c B\,, \quad A \wedge A^c B = \emptyset\,,$$

e da **P1**, **P3** segue

$$P(B) = P(A) + P(A^c B) \geq P(A)\ . \tag{10}$$

Probabilità di $A \vee B$.
Dati due eventi compatibili A e B, si ha

$$A \vee B = A \vee A^c B\,, \quad P(A \vee B) = P(A) + P(A^c B)\,,$$

$$B = AB \vee A^c B\,, \quad P(A^c B) = P(B) - P(AB)\,,$$

e quindi

$$P(A \vee B) = P(A) + P(B) - P(AB)\,. \qquad (11)$$

Iterando la formula (11), per l'unione di tre eventi arbitrari A, B, C si ottiene

$$\begin{aligned} &P(A \vee B \vee C) \; = \cdots = \\ &= P(A) + P(B) + P(C) + \\ &-P(AB) - P(AC) - P(BC) + P(ABC)\;, \end{aligned}$$

equivalente (cfr. *De Morgan*) anche alla formula

$$P(A \vee B \vee C) = 1 - P(A^c B^c C^c). \qquad (12)$$

Paradosso del Cavalier De M$\acute{e}$r$\acute{e}$

Esempio 3 Si effettuano 4 lanci di un dado, A : "esce almeno una volta la faccia 6".

Si effettuano 24 lanci di una coppia di dadi, B : esce almeno una volta la coppia (6,6).

Si racconta che (nel 1654) il Cavalier De M$\acute{e}$r$\acute{e}$ (accanito giocatore di azzardo) valutasse ugualmente probabili A e B sulla base del seguente ragionamento

nel primo esperimento in ognuno dei 4 lanci la faccia 6 ha probabilità $\frac{1}{6}$ e quindi

$$P(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3} \ .$$

Nel secondo esperimento in ognuno dei 24 lanci la coppia (6,6) ha probabilità $\frac{1}{36}$ e quindi

$$P(B) = \frac{1}{36} + \cdots + \frac{1}{36} = \frac{24}{36} = \frac{2}{3} = P(A) \ .$$

Negli esperimenti pratici, però, abbastanza sistematicamente si osservava per uno dei due eventi una frequenza di successo leggermente superiore a quella dell'altro.

Il problema venne sottoposto a Blaise Pascal.
Intuitivamente si può notare che con il ragionamento precedente, se i lanci nel primo esperimento fossero più di 6, ad esempio 7, si avrebbe

$$P(A) = \frac{1}{6} + \cdots + \frac{1}{6} = \frac{7}{6} > 1 \ ,$$

il che è assurdo.

Soluzione. $P(A) = 1 - P(A^c) = 1 - \frac{5^4}{6^4} \simeq 0.51$

$P(B) = 1 - P(B^c) = 1 - \frac{35^{24}}{36^{24}} \simeq 0.49.$

Quindi si ha $P(A) > P(B)$.

Aspetti critici della definizione classica

1) Scelta appropriata dei casi da giudicare ugualmente possibili.

Esempio 4 Un esperimento aleatorio consiste in due lanci di una moneta.

E : *in almeno un lancio esce Testa*.

Casi possibili:

1. C_1: esce Testa al primo lancio (e l'esperimento termina),
2. C_2: esce Croce al primo lancio e Testa al secondo lancio,
3. C_3: esce Croce in entrambi i lanci,

C_1 e C_2 sono favorevoli ad E.

Allora, la probabilità di E è $\frac{2}{3}$?

Non ragionevole!

Non è ragionevole giudicare i tre casi ugualmente possibili.

Infatti, $P(E_1) = \frac{1}{2}$ (se Testa o Croce al primo lancio si giudicano ugualmente possibili).

E_2 ed E_3 sono ugualmente possibili e la loro unione logica coincide con l'evento *Croce al primo lancio*;

In base alla *proprietà additiva*) hanno ciascuno probabilità $\frac{1}{4}$ e quindi una valutazione più adeguata di $P(E)$ è $\frac{3}{4}$.

Tale valutazione è quella che si ottiene direttamente se si considerano come casi possibili (ugualmente possibili) i seguenti quattro, i primi tre dei quali sono quelli favorevoli ad E:

1. esce Testa in entrambi i lanci;
2. esce Testa al primo lancio e Croce al secondo lancio;
3. esce Croce al primo lancio e Testa al secondo lancio;
4. esce Croce in entrambi i lanci.

2) La Definizione 1 non è applicabile sempre.

Esempio 5 Se uno studente sostiene un esame vi sono due casi possibili (lo studente può essere promosso o bocciato). Nessuno, però, si sognerebbe di concludere che la probabilità di essere promosso è pari a $\frac{1}{2}$.

Come si vede già da questo esempio banale, la valutazione della probabilità di uno o più eventi richiede metodi generali e solo in casi particolari ci si può basare sulla Definizione 1.

3) Circolarità.

Il termine *ugualmente possibili* utilizzato nella "Definizione Classica" non può significare altro che *ugualmente probabili* e quindi ... il concetto di probabilità viene definito mediante se stesso.

Impostazione frequentista

Definizione 2 La probabilità di un evento è il limite della frequenza relativa delle prove in cui l'evento si verifica, in una successione di prove identiche, quando il numero delle prove tende all'infinito.

Critiche.

1) Le frequenze non costituiscono una successione numerica data mediante una legge, ma da dei numeri rilevati sperimentalmente. Il concetto di limite utilizzato non è quello rigoroso dell'analisi.

2) Validità. Si deve prendere in considerazione una successione di prove fatte nelle stesse condizioni. Non è sempre possibile.

In conclusione, la definizione classica e la definizione frequentista si possono utilizzare solo come *criteri operativi* di valutazione, utili in certi casi, e non per *definire* la probabilità.

Impostazione soggettiva della probabilità

I precedenti criteri di valutazione possono essere comunque integrati in una impostazione più generale: la teoria *soggettiva*, sviluppata intorno al 1930 dal matematico italiano Bruno de Finetti.

Nell'impostazione soggettiva, che è applicabile in tutte le circostanze, si mantengono rigorosamente distinti gli aspetti *oggettivi* (concernenti gli eventi) da quelli *soggettivi* (concernenti le valutazioni probabilistiche), vale a dire *la logica del certo* dalla *logica del probabile*.

Capita spesso di prendere parte a discussioni accese in cui delle persone esprimono opinioni e valutazioni differenti sulla maggiore o minore attendibilità di fatti incerti in confronto ad altri.

La ragione principale della diversità di valutazioni risiede nel fatto che le persone hanno un'esperienza e uno *stato di informazione* differenti fra di loro.

Esempio 6 Estrazioni con restituzione da un'urna di *composizione incognita* contenente palline bianche e nere. Si vuole valutare la probabilità p di estrarre una pallina bianca alla 1001-ma estrazione.

Se non si conosce il risultato delle precedenti 1000 estrazioni può essere naturale valutare $p = \frac{1}{2}$.

Se si sa che 900 volte la pallina estratta è stata bianca, in mancanza di altre informazioni, si può essere indotti a valutare $p \simeq \frac{9}{10}$.

Il diverso atteggiamento nei due casi è semplicemente dovuto a un diverso *grado di fiducia* nel verificarsi dell'evento considerato.

Questo esempio mette in evidenza che l'informazione di cui si è in possesso (oltre al modo in cui si elabora tale informazione) gioca un ruolo essenziale nella valutazione delle probabilità.

Nell'impostazione soggettiva tale aspetto viene riconosciuto esplicitamente, come si vede dalla definizione seguente.

Definizione 3 Dato un evento E, la probabilità $P(E) = p$ dell'evento E, secondo un dato individuo in un certo stato di informazione, è la *misura numerica* del suo grado di fiducia nel verificarsi di E.

Criterio operativo di misura + condizione di coerenza:

Criterio della scommessa
$P(E) = p$ rappresenta *il prezzo che un individuo ritiene equo pagare per ricevere*

$$\begin{array}{lll} 1 & \textit{se si verifica} & E \\ 0 & \textit{se non si verifica} & E. \end{array}$$

Più in generale, se $S \in \mathbb{R}$, $S \neq 0$, l'individuo deve essere disposto a pagare pS per ricevere

$$\begin{array}{ll} S & \textit{se si verifica } E \\ 0 & \textit{altrimenti.} \end{array}$$

Condizione di coerenza:
L'individuo deve essere coerente, cioè le sue valutazioni di probabilità per uno o più eventi non devono essere tali da fargli subire una perdita certa.

Se indichiamo con $\mathcal{G}$ il guadagno aleatorio, si ha

$$\mathcal{G} = S(|E| - p) = \begin{cases} S(1-p)\,, & \text{E } vero \\ -pS\,, & \text{E } falso. \end{cases}$$

In generale, data una famiglia $\mathcal{F}_n = \{E_1, \ldots, E_n\}$ ed un'assegnazione di probabilità $\mathcal{P}_n = (p_1, \ldots, p_n)$ su $\mathcal{F}_n$, con $p_i = P(E_i)$, $i = 1, \ldots, n$, il guadagno aleatorio corrispondente è dato da

$$\mathcal{G} = \sum_{i=1}^{n} S_i(|E_i| - p_i)\ ,$$

dove $S_1, \ldots, S_n$ sono n numeri reali arbitrari (non tutti nulli).

Definizione 4 La valutazione $\mathcal{P}_n$ si dice *coerente* se, per ogni scelta di $S_1, \ldots, S_n$, risulta

$$Min\,\mathcal{G} \;\cdot Max\,\mathcal{G} \leq 0\ .$$

Proprietà fondamentali della probabilità

- **P1.** $P(E) \geq 0$, per ogni evento E;
- **P2.** $P(\Omega) = 1$;
- **P3.** se $AB = \emptyset$, allora $P(A \vee B) = P(A) + P(B)$.

Ricordando il guadagno aleatorio associato alla valutazione $P(E) = p$, cioè

$$\mathcal{G} = \begin{cases} S(1-p)\,, & \text{E } vero \\ -pS\,, & \text{E } falso, \end{cases}$$

la condizione di coerenza $Min\,\mathcal{G} \;\cdot Max\,\mathcal{G} \leq 0$, diventa $S(1-p)(-pS) = -p(1-p)S^2 \leq 0$, che è soddisfatta se e solo se $0 \leq p \leq 1$. (Proprietà **P1**).

In particolare, se $E = \Omega$, si ha

$$Min\,\mathcal{G} = Max\,\mathcal{G} = \mathcal{G} = S(1-p)$$

e la condizione di coerenza richiede che sia $G = 0$, per ogni S. Quindi deve essere uguale a 0, da cui si ottiene $P(\Omega) = 1$. (Proprietà **P2**).

Proprietà additiva (**P3**).
Consideriamo una partizione di Ω, costituita da 3 eventi $\{H_1, H_2, H_3\}$, di probabilità p_1, p_2, p_3.

Il guadagno totale relativo a tre scommesse simultanee sugli eventi H_1, H_2, H_3, con importi S_1, S_2, S_3, è

$$\mathcal{G} = S_1(|H_1| - p_1) + S_2(|H_2| - p_2) + S_3(|H_3| - p_3).$$

Poichè $|H_1| + |H_2| + |H_3| = 1$, per $S_1 = S_2 = S_3 = S$ si ottiene

$$Min\,\mathcal{G} = Max\,\mathcal{G} = \mathcal{G} = \cdots = S[1 - (p_1 + p_2 + p_3)].$$

Per la condizione di coerenza, dev'essere $\mathcal{G} = 0$. Allora, segue $p_1 + p_2 + p_3 = 1$, ovvero

$$P(H_1) + P(H_2) + P(H_3) = 1. \tag{13}$$

Dim. della proprietà additiva:
dati due eventi incompatibili A, B, per la partizione $\{A \vee B, (A \vee B)^c\}$ deve essere soddisfatta la condizione

$$P(A \vee B) + P[(A \vee B)^c] = 1 \; .$$

D'altra parte, per la partizione $\{A, B, (A \vee B)^c\}$ deve valere

$$P(A) + P(B) + P[(A \vee B)^c] = 1 \; ,$$

da cui si ottiene :

$$P(A \vee B) = P(A) + P(B) \; . \tag{14}$$

Le formule (13) e (14), nel caso di una partizione di Ω, $\{H_1, \ldots, H_n\}$, e nel caso di n eventi $E_1, \ldots, E_n$ a due a due incompatibili, diventano rispettivamente

$$P(H_1) + \cdots + P(H_n) = 1 \,, \tag{15}$$

$$P(E_1 \vee \cdots \vee E_n) = P(E_1) + \cdots + P(E_n) \,. \tag{16}$$

Osservazione 1 (*criterio classico di valutazione*)
Se in un dato esperimento aleatorio si hanno m casi possibili $C_1, \ldots, C_m$ giudicati *ugualmente probabili*, poichè

$$P(C_1) + \cdots + P(C_m) = 1 \,,$$

segue

$$P(C_k) = \frac{1}{m} \,, \quad k = 1, \ldots, m.$$

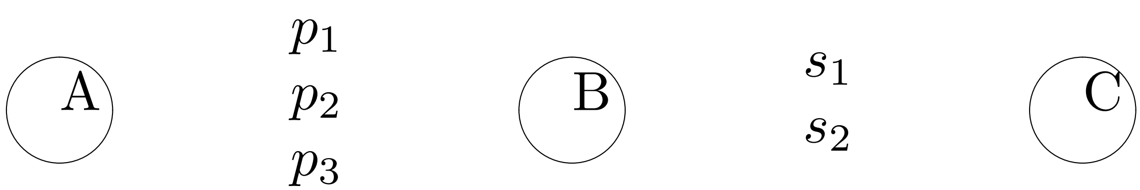
Allora, considerato un evento E con r casi favorevoli, ad esempio $E = C_1 \vee \cdots \vee C_r$, dalla formula (16) si ottiene

$$P(E) = P(C_1) + \cdots + P(C_r) = \frac{r}{m} \,,$$

cioè la probabilità di E è pari al *rapporto tra il numero di casi favorevoli e il numero di casi possibili.*

Richiami di calcolo combinatorio

Esempio 7 Tre località A, B, C sono collegate nel seguente modo : per andare da A a B vi sono 3 percorsi distinti : p_1, p_2, p_3; da B a C vi sono 2 percorsi distinti : s_1, s_2.



I percorsi distinti (per almeno un tratto) che vanno da A a C passando per B non sono 3+2, ma $3 \times 2 = 6$, cioè i seguenti :

$$(p_1, s_1)\ ,\ (p_1, s_2)\ ,\ (p_2, s_1)\ ,\ (p_2, s_2)\ ,\ (p_3, s_1)\ ,\ (p_3, s_2)\ .$$

Il *principio della moltiplicazione* interviene spesso nel calcolo combinatorio.

Nell'Esempio 7 la scelta di un percorso richiede l'esecuzione di una procedura in due passi, con un certo numero di alternative in ogni passo:

1. si sceglie il tratto da A a B (3 alternative);

2. si sceglie il tratto da B a C (2 alternative);

il numero di modi in cui si può svolgere l'intera procedura è pari al prodotto delle alternative in ogni passo $(3 \times 2 = 6)$.

Ogni percorso corrisponde ad una **coppia ordinata** (p_i, s_j) , $i = 1, 2, 3;\ j = 1, 2.$

In generale, dato un insieme S formato da n oggetti $a_1, a_2, \ldots, a_n$, può essere utile calcolare il numero di *disposizioni* o *gruppi ordinati* distinti $(\alpha_1, \ldots, \alpha_r)$, dove $\alpha_i \in S, i = 1, \ldots, r$, che si possono formare utilizzando gli elementi di S.

Due gruppi ordinati si dicono distinti se differiscono per almeno un **elemento** oppure per **l'ordine**.

Per scegliere un gruppo ordinato si esegue una procedura di r passi.

Disposizioni con ripetizione. Le componenti $\alpha_1, \ldots, \alpha_r$ possono essere (in parte o anche tutte) coincidenti. In questo caso si parla di *disposizioni con ripetizione* di classe r di n oggetti. Le alternative in ogni passo sono sempre n; pertanto, in base al principio della moltiplicazione visto nell'Esempio (7), indicando con $D'_{n,r}$ il numero di disposizioni con ripetizione si ha

$$D'_{n,r} = n \times \cdots \times n = n^r \ . \qquad (17)$$

Disposizioni semplici o senza ripetizione. Se $\alpha_i \neq \alpha_j$, per $i \neq j$. In questo caso si parla di *disposizioni semplici o senza ripetizione* (di classe r di n oggetti) e dev'essere ovviamente $r \leq n$. Indicando con $D_{n,r}$ il numero di disposizioni senza ripetizione si ha

$$D_{n,r} = n(n-1) \cdots (n-r+1) \ . \qquad (18)$$

In particolare, per $r = n$ si ha $n - r + 1 = 1$, da cui segue :

$$D_{n,n} = P_n = n \times (n-1) \cdots \times 2 \times 1 = n! \ . \qquad (19)$$

Il simbolo $n!$ si legge *n fattoriale* e rappresenta il prodotto di tutti i numeri da 1 sino a n. Indica il numero di **permutazioni** o *ordinamenti* di n oggetti.

Ad esempio : $3! = 3 \times 2 \times 1 = 6$; $5! = \cdots = 120$.
Per convenzione si pone $0! = 1$. La definizione di $n!$ può esser data in forma ricorsiva:

$$n! = \begin{cases} n \times (n-1)! & \text{se } n \in \mathcal{N} \\ 1 & \text{se } n = 0. \end{cases} \tag{20}$$

Inoltre :

$$D_{n,r} = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}\ . \tag{21}$$

Ad esempio : $D_{5,2} = 5 \times 4 = \frac{5!}{3!}$; $\quad D_{10,4} = 10 \times 9 \times 8 \times 7 = \frac{10!}{6!}$.

Combinazioni

Consideriamo per l'insieme $S = \{a_1, a_2, \ldots, a_n\}$ il calcolo del numero di *gruppi non ordinati* distinti $[\alpha_1, \ldots, \alpha_r]$, dove $\alpha_i \in S, i = 1, \ldots, r$, che si possono formare utilizzando gli elementi di S.

Due gruppi non ordinati si dicono distinti se differiscono per almeno un elemento.

Distinguiamo due casi:

- le componenti $\alpha_1, \ldots, \alpha_r$ possono essere (in parte o anche tutte) coincidenti. In questo caso si parla di *combinazioni con ripetizione* (di classe r di n oggetti).

- $\alpha_i \neq \alpha_j$, se $i \neq j$.

  In questo caso si parla di *combinazioni semplici* (di classe r di n oggetti) e dev'essere ovviamente $r \leq n$. Ogni combinazione semplice rappresenta un sottoinsieme di r oggetti di S e si indica con il simbolo $\{\alpha_1, \ldots, \alpha_r\}$.

Comb. Semplici. Il numero di combinazioni semplici si indica con il simbolo $C_{n,r}$ e rappresenta il numero di sottoinsiemi distinti di r oggetti che si possono formare con gli elementi di S.

Osservando che ogni combinazione semplice dà luogo ad $r!$ disposizioni semplici (distinte per l'ordine), segue:

$$D_{n,r} = r! \times C_{n,r} \ ,$$

e quindi :

$$C_{n,r} = \frac{D_{n,r}}{r!} = \frac{n!}{r!(n-r)!} = \left(\begin{array}{c} n \\ r \end{array} \right) . \qquad (22)$$

Il simbolo $\left(\begin{array}{c} n \\ r \end{array} \right)$ si legge *coefficiente binomiale* n *su* r. Ovviamente essendo:

$$\left(\begin{array}{c} n \\ r \end{array} \right) = \left(\begin{array}{c} n \\ n-r \end{array} \right) = \frac{n!}{r!(n-r)!} \ ,$$

segue che $C_{n,r} = C_{n,n-r}$.

Un'altra formula utile è la seguente :

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r},$$

ovvero: $C_{n,r} = C_{n-1,r-1} + C_{n-1,r}$.

Osserviamo che, volendo costruire un generico sottoinsieme $I \subseteq S$, si deve eseguire una procedura di n passi, con 2 alternative in ogni passo. Infatti, occorre decidere per ciascuno degli elementi $a_1, \ldots, a_n$ se includerlo oppure no in I.

Pertanto, il numero di sottoinsiemi di S, compreso il sottoinsieme vuoto $\emptyset$ e lo stesso S, è dato da

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n ,$$

come segue anche dalla formula del *binomio di Newton* ponendo $a = b = 1$:

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

Combinazioni con ripetizione. Infine, in relazione al numero $C'_{n,r}$ di combinazioni (con ripetizione) di classe r di n oggetti, con un opportuno ragionamento combinatorio si potrebbe verificare che risulta :

$$C'_{n,r} = \binom{r+n-1}{r} = \binom{r+n-1}{n-1} .$$

Coefficiente multinomiale. Dati n interi $r_k \geq 0$, tali che $r_1 + r_2 + \cdots + r_n = r$, si definisce coefficiente multinomiale il seguente:

$$\frac{r!}{r_1! r_2! \cdots r_n!}$$

esso individua il numero di modi in cui r elementi uguali (palline) si possono ripartire in n scatole in modo che la scatola k contenga r_k elementi.

scatola 1	scatola 2	...	scatola n
r_1	r_2	...	r_n

Paradosso di De M$\acute{e}$**r**$\acute{e}$**.** Riprendendo l'Esempio 3, possiamo intanto osservare che lanciando 4 volte un dado i casi possibili sono 6^4 (disposizioni con ripetizione di classe 4 di 6 oggetti). I casi favorevoli all'evento $E_h =$ *la faccia 6 esce esattamente* h *volte* sono dati dal numero

$$\binom{4}{h} 5^{4-h}\,, \quad h = 0, 1, \ldots, 4.$$

Infatti, $\binom{4}{h}$ è il numero di modi distinti in cui la faccia 6 esce in h dei 4 lanci del dado, mentre 5^{4-h} è il numero di modi distinti in cui nei rimanenti $4 - h$ lanci può uscire una delle altre cinque facce diverse da 6. Il prodotto di tali numeri rappresenta i casi favorevoli ad E_h e quindi la probabilità di E_h è data da:

$$P(E_h) \; = \; \frac{\binom{4}{h} 5^{4-h}}{6^4}\,.$$

Pertanto :

$$P(A) \; = \; \sum_{h=1}^{4} \frac{\left(\begin{array}{c} 4 \\ h \end{array} \right) 5^{4-h}}{6^4} \; \simeq 0.51 \; .$$

In modo analogo, si può dimostrare che :

$$P(B) \; = \; \sum_{k=1}^{24} \frac{\left(\begin{array}{c} 24 \\ k \end{array} \right) 35^{24-k}}{36^{24}} \; \simeq 0.49 \; ,$$

e quindi : $P(A) > P(B)$.

Ricordiamo che alla stessa conclusione si può giungere in modo più rapido utilizzando le relazioni

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

Numero Aleatorio

Dati n eventi $E_1, E_2, \ldots, E_n$ ed n numeri reali $x_1, x_2, \ldots, x_n$ si definisce *numero aleatorio semplice* la seguente quantità:

$$X = x_1 \cdot |E_1| + x_2 \cdot |E_2| + \cdots + x_n \cdot |E_n| \qquad (23)$$

Al variare in tutti i modi *possibili*, dei valori 1 o 0 degli indicatori E_i, con $i = 1, \ldots, n$, il numero aleatorio X definisce una funzione reale. Osserviamo che a priori non è conosciuto il valore assunto da X, mentre è noto il suo codominio che è dato da tutte le possibili somme ottenute considerando i costituenti $C_1, \ldots, C_m$ generati dagli eventi $E_1, \ldots, E_n$.

Esempio. Dato un evento E, il suo indicatore è un (particolare) numero aleatorio

$$X = 1 \cdot |E| + 0 \cdot |E^c| = |E|,$$

con valori possibili 0 e 1.

Esempio. Consideriamo il lancio di un dado e definiamo gli eventi

$$E_i = \text{''esce il numero } i\text{''}, \quad i = 1, 2, \ldots, 6.$$

Il risultato aleatorio X del lancio si può rappresentare nel modo seguente

$$X \;=\; 1 \cdot |E_1| + 2 \cdot |E_2| + \cdots + 6 \cdot |E_6| \,.$$

In questo caso $E_1, E_2, \ldots, E_6$ formano una partizione e quindi i coefficienti $x_1, x_2, \ldots, x_6$ rappresentano i possibili valori di X.

Forma Canonica. Dato un numero aleatorio

$$X \;=\; x_1 \cdot |H_1| + x_2 \cdot |H_2| + \cdots + x_n \cdot |H_n| \,,$$

dove $\{H_1, H_2, \ldots, H_n\}$ è una partizione di Ω, il codominio di X è $\{x_1, x_2, \ldots, x_n\}$.

In generale non è così semplice determinare il codominio di un numero aleatorio X.

Esempio Dati tre eventi E_1, E_2, E_3, con le seguenti relazioni logiche

$$E_1E_2 = \emptyset\,, \qquad E_3 \subseteq E_1\,,$$

determiniamo i valori possibili del numero aleatorio

$$X = 2|E_1| - |E_2| + |E_3| = 2\cdot|E_1| + (-1)\cdot|E_2| + 1\cdot|E_3|.$$

Calcoliamo i costituenti generati da E_1, E_2, E_3 (che non formano una partizione) ed i corrispondenti valori di X.

$$C_1 = E_1E_2^cE_3\,,\ \ \chi_1 = 3\,; \qquad C_2 = E_1E_2^cE_3^c\,,\ \ \chi_2 = 2\,;$$

$$C_3 = E_1^cE_2E_3^c\,,\ \ \chi_3 = -1\,; \quad C_4 = E_1^cE_2^cE_3^c\,,\ \ \chi_4 = 0\,.$$

Pertanto l'insieme dei possibili valori (o codominio) di X è $\{-1, 0, 2, 3\}$.

In una situazione generale, con eventi arbitrari e con certe relazioni logiche, per ottenere la forma canonica occorre calcolare i costituenti e i valori possibili di X.

Sia $\{E_1, E_2, \ldots, E_n\}$ una famiglia di eventi e siano $C_1, C_2, \ldots, C_m$ i relativi costituenti. Sia inoltre X il numero aleatorio $X = x_1|E_1| + x_2|E_2| + \cdots + x_n|E_n|$. Se si verifica il costituente C_h, X assume un ben determinato valore χ_h.

$$\begin{array}{ccc} C_1 & \to & \chi_1 \\ C_2 & \to & \chi_2 \\ \vdots & \vdots & \vdots \\ C_m & \to & \chi_m \end{array}$$

Possiamo esprimere il numero aleatorio X in funzione degli indicatori dei costituenti.

$$X = \chi_1 \cdot |C_1| + \chi_2 \cdot |C_2| + \cdots + \chi_m \cdot |C_m| \,.$$

In questo caso i possibili valori di X saranno $\chi_1, \chi_2, \ldots, \chi_m$ (alcuni dei quali potrebbbero anche coincidere). Il legame che esiste tra gli x_i e gli χ_h è il seguente:

$$\chi_h = \sum_{i: C_h \subseteq E_i} x_i \,.$$

Previsione di un N.A.

Siano dati n eventi $E_1, E_2, \ldots, E_n$ e siano

$$p_1 = P(E_1)\,,\, p_2 = P(E_2)\,,\, \ldots\,,\, p_n = P(E_n)$$

le rispettive valutazioni di probabilità. Considerato il numero aleatorio

$$X = x_1|E_1| + x_2|E_2| + \cdots + x_n|E_n|$$

si definisce *Previsione (o Speranza Matematica o Valor Medio)* di X la seguente quantità:

$$\mathbb{P}(X) = p_1x_1 + p_2x_2 + \cdots + p_nx_n \qquad (24)$$

Diamo un'interpretazione in termini di scommessa della previsione di X. Supponiamo che il n.a. X sia una vincita aleatoria, nel senso che

$$\begin{array}{cccc}
\text{se si verifica } E_1 & (|E_1| = 1) & \Rightarrow & \text{si vince } x_1 \\
\text{se si verifica } E_2 & (|E_2| = 1) & \Rightarrow & \text{si vince } x_2 \\
\vdots & \vdots & \vdots & \vdots \\
\text{se si verifica } E_n & (|E_n| = 1) & \Rightarrow & \text{si vince } x_n
\end{array}$$

dove X rappresenta il totale di quello che si riceve.

In termini di scommessa $\mathbb{P}(X)$ indica
la quantità che si è disposti a pagare per ricevere X,
oppure
la quantità che si è disposti a pretendere per pagare X. Infatti osservando che

$$
\begin{array}{llll}
\text{si paga} & p_1x_1 & \text{per ricevere} & \begin{cases} x_1, & \text{se si verif. } E_1, \\ 0, & \text{se si verif. } E_1^c, \end{cases} \\[2ex]
& (\text{ovvero}, & \text{per ricevere} & x_1|E_1|) \\[2ex]
\text{si paga} & p_2x_2 & \text{per ricevere} & x_2|E_2| \\[2ex]
\vdots & \vdots & \vdots & \vdots \\
\text{si paga} & p_nx_n & \text{per ricevere} & x_n|E_n|.
\end{array}
$$

Allora sommando tali termini si può dire che

$$
\begin{array}{lll}
\text{si paga } \sum_{i=}^n p_ix_i & \text{per ricevere} & \sum_{i=1}^n x_i|E_i| \\[2ex]
\text{cioè: si paga} \quad \mathbb{P}(X) & \text{per ricevere} & X.
\end{array}
$$

Esempio. Sia

$$X = |E| = 1|E| + 0|E^c|$$

e sia $p = P(E)$. Si ha

$$\mathbb{P}(X) = \mathbb{P}(|E|) = 1 \cdot p + 0 \cdot (1 - p) = p\,,$$

cioè la previsione dell'indicatore di un evento coincide con la probabilità dell'evento ($\mathbb{P}(|E|) = p$).

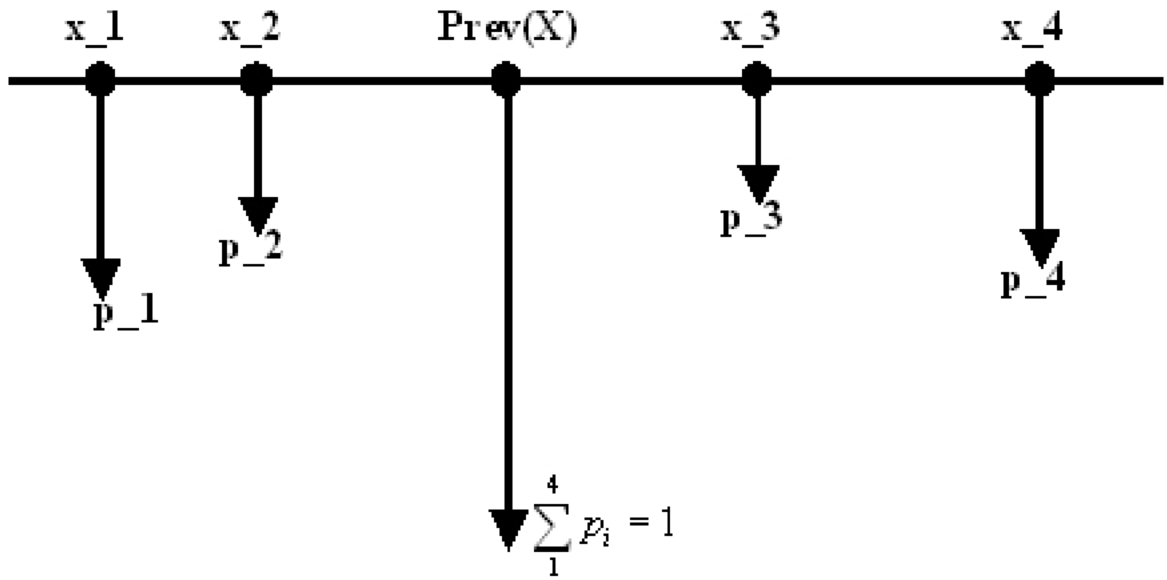
Baricentro. Sia $\{H_1, \ldots, H_n\}$ una *partizione* di Ω. Siano

$$p_1 = P(H_1), p_2 = P(H_2), \ldots, p_n = P(H_n)$$

le rispettive probabilità. Ricordiamo che $\sum_{i=1}^{n} p_i = 1$. La previsione $\mathbb{P}(X)$ del numero

$$X = x_1|H_1| + x_2|H_2| + \cdots + x_n|H_n|$$

si può interpretare come il baricentro di una distribuzione di n masse di pesi p_i e di ascisse x_i con $i = 1, 2, \ldots, n$.



Osservazione. La previsione $\mathbb{P}(X)$ di un n. a. X gode della seguente proprietà

$$\min(X) \leq \mathbb{P}(X) \leq \max(X). \tag{25}$$

Dim. Consideriamo solo il caso in cui

$$X = x_1|H_1| + x_2|H_2| + \cdots + x_n|H_n|$$

con $\{H_1, \ldots, H_n\}$ una *partizione* di Ω.

Come sappiamo, affinchè le probabilità $p_1, p_2, \ldots, p_n$ siano coerenti dev'essere

$$\sum_{i=1}^{n} p_i = 1\,, \quad p_i \geq 0\,, \;\; i = 1, \ldots, n\,.$$

Quindi $\mathbb{P}(X) = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$ è combinazione lineare convessa dei punti (ascisse) x_i. Osserviamo che, per ogni $i = 1, \ldots, n$, si ha

$$\min(X) \;\leq\; x_i \;\leq\; \max(X)\,,$$

e quindi

$$p_i \min(X) \leq p_i x_i \leq p_i \max(X)\,.$$

Allora, sommando per i che va da 1 a n, si ha

$$(\sum_{i}^{n} p_i) \min(X) \leq \sum_{i}^{n} p_i x_i \leq (\sum_{i}^{n} p_i) \max(X)\,,$$

ovvero

$$\min(X) \leq \mathbb{P}(X) \leq \max(X)\,.$$

Combinazione lineare di n.a. Siano X e Y due numeri aleatori

$$\begin{aligned} X &= x_1|E_1| + x_2|E_2| + \cdots + x_n|E_n| \\ Y &= y_1|A_1| + y_2|A_2| + \cdots + y_n|A_m| \end{aligned}$$

e siano a e b due numeri reali arbitrari. La quantità $C = aX + bY$ è il numero aleatorio che si ottiene dalla combinazione lineare degli indicatori degli eventi

$$E_1, \ldots, E_n, A_1, \ldots, A_m,$$

con coefficienti

$$ax_1, \ldots, ax_n, by_1, \ldots, by_m,$$

ovvero

$$\begin{aligned} C = aX + bY = \quad & ax_1|E_1| + ax_2|E_2| + \cdots + ax_n|E_n| + \\ + \quad & by_1|A_1| + by_2|A_2| + \cdots + by_n|A_m| \,. \end{aligned}$$

Linearità della Previsione. Dati due numeri aleatori X e Y e due numeri reali arbitrari a e b si ha

$$\mathbb{P}(aX + bY) = a\mathbb{P}(X) + b\mathbb{P}(Y). \tag{26}$$

Infatti

$$\mathbb{P}(aX + bY) = \mathbb{P}(\textstyle\sum_{i=1}^{n} ax_i|E_i| + \sum_{j=1}^{m} by_j|A_j|) = \\ \textstyle\sum_{i=1}^{n} ax_i p(E_i) + \sum_{j=1}^{m} by_j p(A_j) = a\mathbb{P}(X) + b\mathbb{P}(Y).$$

In particolare $\mathbb{P}(b) = b$. Ciò deriva dal fatto che una costante b si può vedere come il numero aleatorio $Y = b|\Omega|$, quindi $\mathbb{P}(Y) = b \cdot p(\Omega) = b \cdot 1$.

Casi particolari

$$\begin{array}{ll} \mathbb{P}(aX) = a\mathbb{P}(X), & \mathbb{P}(b) = b, \\ \mathbb{P}(aX + b) = a\mathbb{P}(X) + b, & \mathbb{P}(-X) = -\mathbb{P}(X). \end{array}$$

Dati n n.a. $X_1, X_2, \ldots, X_n$, applicando ripetutamente la (26) si ottiene

$$\mathbb{P}(X_1 + X_2 + \cdots + X_n) = \mathbb{P}(X_1) + \mathbb{P}(X_2) + \cdots + \mathbb{P}(X_n).$$

Più in generale si ha

$$\mathbb{P}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \mathbb{P}(X_i)$$

con $a_1, a_2, \ldots, a_n$ n numeri reali arbitrari.

Esempio. Supponiamo di effettuare n estrazioni con restituzione da un urna di composizione nota contenente palline bianche e nere . Indichiamo con

$$E_i = \text{esce pallina bianca alla i-esima estrazione.}$$

Parliamo di successo all'i-esima estrazione nel caso in cui E_i è vero e di insuccesso nel caso contrario. Il numero aleatorio

$$X = |E_1| + |E_2| + \cdots + |E_n|$$

rappresenta il numero di successo nelle n prove. Se supponiamo che gli eventi E_i siano tutti equiprobabili, ovvero

$$P(E_i) = p, \;\; i = 1, 2, \ldots, n$$

allora si ottiene

$$\begin{aligned} \mathbb{P}(X) = \;\; & \mathbb{P}(|E_1| + |E_2| + \cdots + |E_n|) = \\ & \textstyle\sum_{i=1}^{n} \mathbb{P}(|E_i|) = \sum_{i=1}^{n} p(E_i) = np \end{aligned}$$

Se invece considero il numero aleatorio

$$Z = \frac{|E_1| + |E_2| + \cdots + |E_n|}{n}$$

che rappresenta il numero medio di successi, si osserva che $\mathbb{P}(Z) = \mathbb{P}(\frac{X}{n}) = p$ ovvero la previsione della media coincide con la probabilità di successo.

Varianza

Relativamente ad un numero aleatorio X abbiamo visto che la previsione $\mathbb{P}(X)$ si può interpretare come il valore centrale della distribuzione di probabilità di X. In alcuni casi la conoscenza della previsione può essere sufficiente per sintetizzare la distribuzione di probabilità (un analogia la si riscontra con il baricentro).

In generale è troppo restrittivo descrivere delle valutazioni di probabilità solamente attraverso la previsione.

Ad esempio se Enzo mangia 2 polli e Marco non ne mangia, allora la previsione si limita a dire che entrambi mangiano un pollo.

Per tale motivo si introduce il concetto di *varianza*, un indice che valuta la dispersione di X attorno al suo valor medio $\mathbb{P}(X)$.

Definizione. Sia X un numero aleatorio e sia $\mathbb{P}(X) = m$. Si definisce *Varianza* di X e si indica con $Var(X)$ o con σ_X^2 la seguente quantità

$$Var(X) = \mathbb{P}[(X - m)^2] \tag{27}$$

La $Var(X)$ si definisce come la previsione del numero aleatorio $Y = (X - m)^2$. Indicando con $x_1, x_2, \ldots, x_n$ **tutti i possibili valori** di X e con $p_i = P(X = x_i)$ le rispettive probabilità, possiamo rappresentare X nella seguente forma canonica

$$X = x_1|X = x_1| + x_2|X = x_2| + \cdots + x_n|X = x_n|$$

e la previsione $\mathbb{P}(X)$ nella consueta espressione

$$\mathbb{P}(X) = p_1x_1 + p_2x_2 + \cdots + p_nx_n.$$

Il n.a. $Y = (X - m)^2$ si può rappresentare nella forma seguente

$$(x_1 - m)^2|X = x_1| + \cdots + (x_n - m)^2|X = x_n|\,.$$

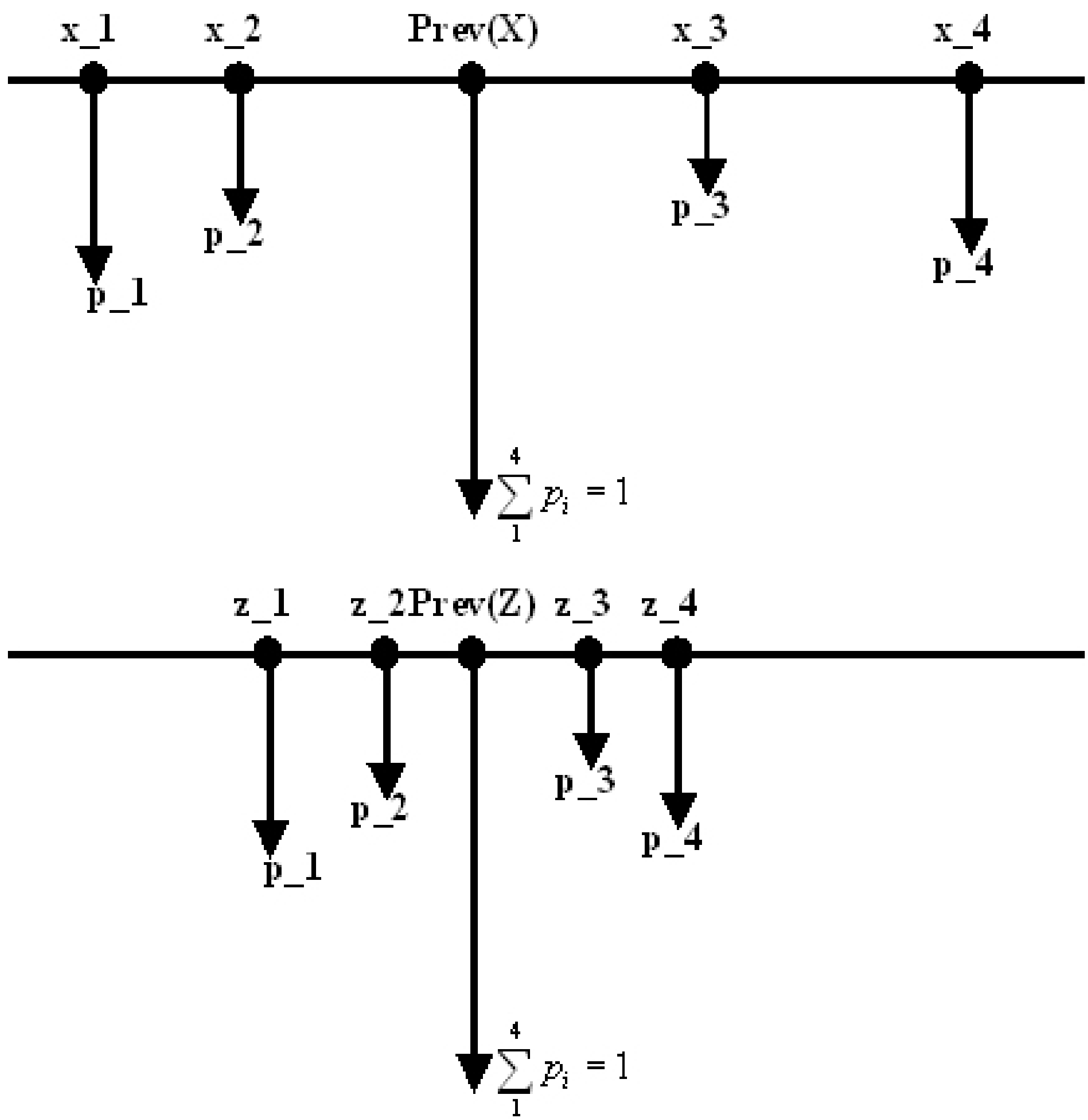
Allora, per la varianza si ottiene

$$Var(X) = \mathbb{P}[(X - m)^2] =$$

$$= p_1(x_1 - m)^2 + p_2(x_2 - m)^2 + \cdots + p_n(x_n - m)^2 =$$

$$= \textstyle\sum_{i=1}^{n} p_i(x_i - m)^2\,.$$

Come si può intuire, la varianza di un n.a. è tanto più grande quanto più la distribuzione di X è *dispersa* attorno al suo valor medio, ovvero quanto più sono probabili i valori di X lontani dal valor medio $\mathbb{P}(X)$.



$$\mathbb{P}(X) = \mathbb{P}(Z), \;\; Var(X) > Var(Z)$$

Dal punto di vista meccanico, dato un sistema di masse, la $Var(X)$ rappresenta il momento d'inerzia rispetto al baricentro e quindi misura la dispersione di massa rispetto al baricentro $\mathbb{P}(X)$.

Scarto quadratico medio (deviazione standard).

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathbb{P}[(X-m)^2]} \tag{28}$$

Esempio. Siano X, Y, Z tre numeri aleatori ripsettivamente con le seguenti distribuzioni:

	-2	-1	1	2
X		$\frac{1}{2}$	$\frac{1}{2}$	
Y	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Z	$\frac{1}{2}$			$\frac{1}{2}$

cioè

$P(X=-1)=P(X=1)=\frac{1}{2}$

$P(X=-2)=P(Y=-1)=P(Y=1)=P(Y=2)=\frac{1}{4}$

$P(Z=-2)=P(Y=2)=\frac{1}{2}$

Le previsioni sono uguali, essendo $\mathbb{P}(X)=\mathbb{P}(Y)=\mathbb{P}(Z)=0$, mentre le varianze sono diverse. Infatti

$$\begin{aligned}
\sigma_X^2 = \quad & \tfrac{1}{2}(-1-0)^2+\tfrac{1}{2}(1-0)^2=1 \\
\sigma_Y^2 = \quad & \tfrac{1}{4}(-2-0)^2+\tfrac{1}{4}(-1-0)^2+ \\
& +\tfrac{1}{4}(1-0)^2+\tfrac{1}{4}(2-0)^2=\tfrac{5}{2} \\
\sigma_Z^2 = \quad & \tfrac{1}{2}(-2-0)^2+\tfrac{1}{2}(2-0)^2=4
\end{aligned}$$

Quindi: $Var(Z)>Var(Y)>Var(X)$.

Esempio. Consideriamo la scelta a caso di una persona da una popolazione statistica costituita da due individui: uno mangia 2 polli (supponiamo, a settimana) e l'altro 0 polli. Sia X il numero (aleatorio) di polli mangiati settimanalmente dalla persona estratta. Si ha

$$X = \begin{cases} 0, & P(X=0) = \frac{1}{2}, \\ 2, & P(X=2) = \frac{1}{2}. \end{cases}$$

Adesso, consideriamo il caso in cui entrambi gli individui mangiano 1 pollo a settimana. Sia Y il numero (aleatorio) di polli mangiati settimanalmente dalla persona estratta. Si ha

$$Y = 1, \;\; P(Y=1) = 1.$$

Allora:

$$\mathbb{P}(X) = \mathbb{P}(Y) = 1\,,$$

mentre:

$$\sigma_X^2 = 1 \; > \; \sigma_Y^2 = 0\,.$$

Varianza di un indicatore. Dato un evento E con $P(E) = p$ e $p(E^c) = 1 - p = q$, essendo $\mathbb{P}(|E|) = p$, si dimostra che $Var(|E|) = pq$. Infatti

$$\sigma^2_{|E|} = \mathbb{P}[(|E| - p)^2] = p(1-p)^2 + q(0-p)^2 = pq\,.$$

Osserviamo che, come mostrato nel grafico sotto, risulta: $Var(|E|) = pq \,\leq\, \frac{1}{4}\,.$

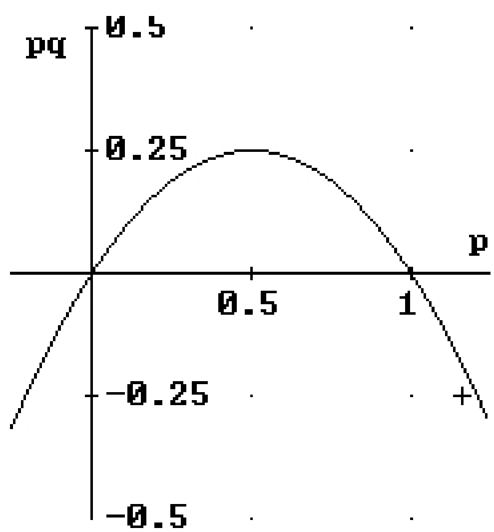


Figura 1: $Var(|E|)$

Proprietà della varianza.

1. $Var(X) \geq 0$;

2. $Var(X + c) = Var(X)$;

3. $Var(aX) = a^2 Var(X)$, per ogni $a \in \mathbb{R}$;

4. $Var(aX + b) = a^2 Var(X)$, per ogni $a, b \in \mathbb{R}$;

5. $Var(-X) = Var(X)$;

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

Ricordando che $m = \mathbb{P}(X)$, dimostriamo la 6.

$$\sigma_X^2 = \mathbb{P}[(X - m)^2] = \mathbb{P}(X^2 - 2mX + m^2) =$$

$$= \mathbb{P}(X^2) - 2m\mathbb{P}(X) + m^2 = \mathbb{P}(X^2) - [\mathbb{P}(X)]^2.$$

Osserviamo inoltre che dalla proprietà 4 per la deviazione standard si ha

$$\sigma_{(aX+b)} = |a|\sigma_X, \text{ per ogni } a, b \in \mathbb{R}.$$

Standardizzazione

Dato un n.a. X con $\mathbb{P}(X) = m$ e deviazione standard σ_X, consideriamo il seguente n.a.

$$Z = \frac{X - m}{\sigma_X}.$$

Utilizzando sia le proprietà della previsione che della varianza si ha

$$\mathbb{P}(Z) = 0, \;\; \sigma_Z^2 = \sigma_Z = 1.$$

Il n.a. $Z = \frac{X-m}{\sigma_X}$ dicesi numero aleatorio **ridotto** o **standardizzato**.

Disuguaglianza di Markov. Dato un n.a. $X \geq 0$, con $\mathbb{P}(X) = m$, e un numero reale $\alpha > m$, si ha

$$P(X \geq \alpha) \leq \frac{m}{\alpha}\,. \tag{29}$$

Disuguaglianza di Cebicev. Dato un n.a. arbitrario X con $\mathbb{P}(X) = m$ e deviazione standard σ si ha

$$P(|X - m| \geq k\sigma) \leq \frac{1}{k^2}\,, \quad \forall k \in \mathbb{R} \;\; (k > 1). \tag{30}$$

Covarianza

Ricordiamo che, dati due numeri aleatori X e Y, si ha $\mathbb{P}(X+Y)$=$\mathbb{P}(X)+\mathbb{P}(Y)$. Posto

$$\mathbb{P}(X)=m_X\,,\quad \mathbb{P}(Y)=m_Y\,,$$

calcoliamo $Var(X+Y)$.

$$\begin{aligned}
&Var(X+Y)=\mathbb{P}\{[(X+Y)-(m_X+m_Y)]^2\}=\\
&=\mathbb{P}\{[(X-m_X)+(Y-m_Y)]^2\}=\\
&=\mathbb{P}[(X-m_X)^2+(Y-m_Y)^2+2(X-m_X)(Y-m_Y)]=\\
&=Var(X)+Var(Y)+2\mathbb{P}[(X-m_X)(Y-m_Y)]\,.
\end{aligned}\tag{31}$$

La quantità $\mathbb{P}[(X-m_X)(Y-m_Y)]$ si definisce **covarianza** di X,Y e si indica con $Cov(X,Y)$, oppure σ_{XY}; quindi

$$Cov(X,Y)=\sigma_{XY}=\mathbb{P}[(X-m_X)(Y-m_Y)]\,. \tag{32}$$

Pertanto: $\sigma^2_{X+Y}=\sigma^2_X+\sigma^2_Y+2\sigma_{XY}\,.$

Osserviamo che

$$\begin{array}{ll} Cov(X,Y) & = \mathbb{P}[(X-m_X)(Y-m_Y)] = \\ & = \mathbb{P}(XY - m_X Y - m_Y X + m_X m_Y) = \\ & = \mathbb{P}(XY) - m_X m_Y - m_Y m_X + m_X m_Y \\ & \Downarrow \\ Cov(X,Y) = & \mathbb{P}(XY) - \mathbb{P}(X)\mathbb{P}(Y). \end{array} \tag{33}$$

In particolare

$$Cov(X,X) = \mathbb{P}(XX) - \mathbb{P}(X)\mathbb{P}(X) = \sigma_X^2.$$

La covarianza di X e Y è una misura della tendenza di X e Y ad associarsi prevalentemente secondo valori

$$\begin{array}{lllll} X & Y & X & Y & \\ 1.grande & grande, & piccolo & piccolo & (\leftrightarrow \sigma_{XY} > 0) \\ 2.grande & piccolo, & piccolo & grande & (\leftrightarrow \sigma_{XY} < 0) \end{array}$$

dove con *grande* indichiamo valori di $X > \mathbb{P}(X)$ e valori di $Y > \mathbb{P}(Y)$ e analogamente per *piccolo* indichiamo valori di $X < \mathbb{P}(X)$ e $Y < \mathbb{P}(Y)$.

Se come tendenza prevale il 1^0 caso si ha correlazione positiva ($\sigma_{XY} > 0$).

Se prevale il 2^0 caso si ha correlazione negativa ($\sigma_{XY} < 0$).

Esempio. Si hanno 2 palline (una bianca e una nera) da ripartire in 2 scatole. Siano
$X =$ "numero di palline nella prima scatola"
$Y =$ "numero di scatole non vuote".
Calcoliamo le rispettive previsioni. Si ha:

$$X \in \{0,1,2\}\,, \quad Y \in \{1,2\}\,.$$

Inoltre, si hanno 4 casi possibili

$$(0,bn) \qquad (bn,0) \qquad\qquad (b,n) \qquad (n,b)$$

$$P(X=0) = P(X=2) = \tfrac{1}{4} \quad P(X=1) = \tfrac{1}{2}$$

$$P(Y=1) = \tfrac{1}{2} \qquad\qquad P(Y=2) = \tfrac{1}{2}$$

con

$$X = 0|X=0| + 1|X=1| + 2|X=2|\,,$$

$$Y = 1|Y=1| + 2|Y=2|\,,$$

e con

$$\mathbb{P}(X) = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1\,,$$

$$\mathbb{P}(Y) = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{2} = \tfrac{3}{2}\,.$$

Come si vede nella tabella sotto, si ha $XY \in \{0, 2\}$.

$Y \setminus X$	0	1	2
1	si $\frac{1}{4}$	no	si $\frac{1}{4}$
2	no	si $\frac{2}{4}$	no

Allora, per la previsione di XY si ottiene

$$\mathbb{P}(XY) = 0 \cdot \frac{1}{4} + 2 \cdot \frac{3}{4} = \frac{3}{2}\,,$$

e utilizzando la (33) risulta

$$Cov(X,Y) = \mathbb{P}(XY) - \mathbb{P}(X)\mathbb{P}(Y) = \frac{3}{2} - \frac{3}{2} = 0.$$

Incorrelazione. Due n.a. X e Y si dicono $incorrelati$ se $Cov(X,Y) = 0$.

Dati due n.a X, Y incorrelati, si ha

$$Var(X+Y) = Var(X) + Var(Y)\,.$$

Proprietà della covarianza.

$$Cov(aX+b, cY+d) = \cdots = ac\,Cov(X,Y)\,.$$

Coefficiente di correlazione
Dati due n. a. X e Y, si definisce coefficiente di correlazione la seguente quantità

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}\,.$$

ρ_{XY} si chiama anche covarianza normalizzata in quanto coincide con la covarianza dei due n. a. ridotti. Infatti

$$Cov(\tfrac{X-m_X}{\sigma_X}, \tfrac{Y-m_Y}{\sigma_Y}) = Cov(\tfrac{X}{\sigma_X} - \tfrac{m_X}{\sigma_X}, \tfrac{Y}{\sigma_Y} - \tfrac{m_Y}{\sigma_Y}) =$$
$$= Cov(\tfrac{X}{\sigma_X}, \tfrac{Y}{\sigma_Y}) = \tfrac{1}{\sigma_X\sigma_Y} Cov(X,Y) = \rho_{XY}.$$

Osservazione. Dati X e Y, consideriamo $X' = aX+b$ e $Y' = cY+d$, con $a,b,c,d \in \mathbb{R}\,, a \neq 0$, $b \neq 0$. Si ha

$$\rho_{X'Y'} = \tfrac{\sigma_{(aX+b)(cY+d)}}{\sigma_{(aX+b)}\sigma_{(cY+d)}} = \tfrac{ac\sigma_{XY}}{|ac|\sigma_X\sigma_Y} = \pm\rho_{XY}.$$

Proprietà.

1. $\rho_{XY} = 0 \Longleftrightarrow \sigma_{XY} = 0 \quad$ (incorrelazione);

2. $-1 \leq \rho_{XY} \leq 1$;

3. $\rho_{XY} = \pm 1 \Longleftrightarrow Y = aX + b$.
   $(a > 0 \Longrightarrow \rho_{XY} = 1\,, \quad a < 0 \Longrightarrow \rho_{XY} = -1)$

La proprietà 3 esprime numericamente il caso in cui tra X e Y esiste una dipendenza lineare. Prima di dimostrare tali proprietà facciamo le seguenti osservazioni:

- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$
- $Var(aX + cY) = a^2Var(X) + c^2Var(Y) + 2acCov(X, Y)$

Dim. della prop. 2. Si ha

$$
\begin{aligned}
Var(\tfrac{X}{\sigma_X}+\tfrac{Y}{\sigma_Y}) = \;& \tfrac{\sigma_X^2}{\sigma_X^2}+\tfrac{\sigma_Y^2}{\sigma_Y^2}+2Cov(\tfrac{X}{\sigma_X},\tfrac{Y}{\sigma_Y}) = \\
& = 2+2\tfrac{Cov(X,Y)}{\sigma_X\sigma_Y} = 2(1+\rho_{XY}) \geq 0\,; \\
Var(\tfrac{X}{\sigma_X}-\tfrac{Y}{\sigma_Y}) = \;& 2(1-\rho_{XY}) \geq 0\,;
\end{aligned}
$$

$\Downarrow$

$-1 \leq \rho_{XY} \leq 1\,.$

Dim. della prop.3. Se $Y = aX + b$, segue

$$\rho_{XY} = \tfrac{Cov(X,aX+b)}{\sigma_X\sigma_{aX+b}} = \tfrac{aCov(X,X)}{|a|\sigma_X\sigma_X} = \pm 1\,.$$

Viceversa, sia $|\rho_{XY}| = 1$ e dimostriamo che esiste una dipendenza lineare tra X e Y.
Se $\rho_{XY} = -1$, segue $Var(\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}) = 0$ cioè il n.a. $\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}$ ha varianza nulla. Allora (con prob. 1) $\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}$ è costante e coincide con il suo valor medio

$\mathbb{P}(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}) = \frac{m_X}{\sigma_X} + \frac{m_Y}{\sigma_Y}$. Quindi

$$\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} \overset{(prob.1)}{=} \frac{m_X}{\sigma_X} + \frac{m_Y}{\sigma_Y}$$
$$\Downarrow$$

$$\begin{array}{ll}
Y = m_Y - \frac{\sigma_Y}{\sigma_X}(X - m_X) & = \underbrace{-\frac{\sigma_Y}{\sigma_X}}X + \underbrace{m_Y + \frac{\sigma_Y}{\sigma_X}m_X} \\
Y = aX + b \text{ con} & a = -\frac{\sigma_Y}{\sigma_X}\ , b = m_Y + am_X \\
\text{oppure} & \\
X = cY + d \text{ con} & c = -\frac{\sigma_X}{\sigma_Y}\ , b = m_X + dm_Y
\end{array}$$

Una dimostrazione analoga si può fare nel caso in cui $\rho_{XY} = 1$ osservando che il n.a. $\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$ ha varianza nulla.

Covarianza di due indicatori. Siano A e B due eventi, si ha

$$\begin{aligned} Cov(|A|,|B|) = \ & \mathbb{P}(|AB|) - \mathbb{P}(|A|)\mathbb{P}(|B|) = \\ & = P(AB) - P(A)P(B)\,. \end{aligned}$$

In particolare dato un evento E di probabilità $P(E) = p$ e ponendo $q = 1 - p$, si ha

$$Cov(|E|,|E^c|) = \ \mathbb{P}(|\emptyset|) - \mathbb{P}(|E|)\mathbb{P}(|E^c|) = -pq\,,$$

ed inoltre essendo $|E^c| = 1 - |E|$ si ha $\rho_{|E||E^c|} = -1$.

Esempio. Estrazioni senza restituzione da un'urna contenente due palline, una bianca e una nera.
A=La prima pallina estratta è bianca;
B=La seconda pallina estratta è bianca.
Si dimostra che: $Cov(|A|, |B|) = -\frac{1}{4}$.
Nello stesso esempio, se si effettuano estrazioni con restituzione, si ha $Cov(|A|, |B|) = 0$.

Esempio. Una pallina bianca e una nera vengono distribuite a caso in due urne U, V. Definiti i n.a.
X = numero di palline bianche in U,
Y = numero di urne non vuote,
si può verificare che $Cov(X, Y) = 0$.

Esempio. Considerate due urne U, V, contenenti ciascuna una pallina bianca e una nera, da U si estrae a caso una pallina e la si inserisce in V. Definiti i n.a.
X="numero di palline bianche in U",
Y="numero di palline bianche in V",
determinare ρ_{XY}.

Matrice delle varianze e delle covarianze. Dati n n.a. $X_1, X_2, \ldots, X_n$ ed m n.a. $Y_1, Y_2, \ldots, Y_m$ consideriamo le seguenti combinazioni lineari:

$$X = \sum_{i=1}^{n} a_i X_i \,, \quad Y = \sum_{j=1}^{m} b_j Y_j \,,$$

con

$$a_i,\, b_j \in \mathbb{R}\,, \quad i = 1, \ldots, n; \; j = 1, \ldots, m.$$

Calcoliamo la covarianza dei n.a. X, Y.

$$\begin{aligned} Cov(X,Y) = \quad & Cov(\textstyle\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \\ & = \cdots = \\ & = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \sigma_{X_i Y_j}. \end{aligned}$$

In particolare

$$Cov(X,X) = Var(X) = \textstyle\sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + 2 \sum_{i<j} a_i a_j \sigma_{X_i X_j}. \tag{34}$$

Indicando con $\mathbf{a}^t = (a_1, a_2, \ldots, a_n)$ il vettore dei coefficienti e con Σ la seguente matrice (delle varianze e covarianze)

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} , \quad \sigma_{ij} = \sigma_{X_i X_j} ,$$

possiamo scrivere la (34) nella seguente forma matriciale

$$Cov(X,X) = Var(X) = \mathbf{a}^t \cdot \Sigma \cdot \mathbf{a} =$$

$$= (a_1, a_2, \ldots, a_n) \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} . \tag{35}$$

Se $X_1, \ldots, X_n$ sono a due a due incorrelati si ha

$$Var(X) = \sum_{i=1}^n a_i^2 \sigma_{X_i}^2 =$$

$$= (a_1, a_2, \ldots, a_n) \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} .$$

Frequenza di successo. Dati n eventi $E_1, E_2, \ldots, E_n$ a 2 a 2 incorrelati e posto $P(E_i) = p_i$, indichiamo con con

$$f_n = \frac{|E_1| + |E_2| + \cdots + |E_n|}{n} \tag{36}$$

la frequenza relativa di successo. Calcoliamo la previsione e la varianza di f_n. Si ha

$$\mathbb{P}(f_n) = \cdots = \tfrac{p_1+p_2+\cdots+p_n}{n}\,,$$

$$Var(f_n) = Var(\tfrac{\sum_{i=1}^n |E_i|}{n}) = \tfrac{1}{n^2}\textstyle\sum_{i=1}^n p_i q_i \leq \tfrac{1}{4n}\,.$$

Al tendere di $n \longrightarrow \infty$ si ha $Var(f_n) \longrightarrow 0$. Ciò significa che la dispersione di f_n attorno al suo valor medio $\frac{\sum_{i=1}^n p_i}{n}$ diventa sempre più piccola al crescere del numero delle prove. Sfruttando la disuguaglianza di Cebicev si perviene alla seguente *legge (debole) dei grandi numeri*

$$\lim_{n \longrightarrow \infty} P(|f_n - \frac{\sum_{i=1}^n p_i}{n}| > \varepsilon) = 0\,.$$

Se in particolare gli E_i sono equiprobabili, con $P(E_i) = p$, si ha: $\lim_{n \longrightarrow \infty} P(|f_n - p| > \varepsilon) = 0$.

Probabilità condizionate

Dati due eventi E ed H, con $H \neq \phi$, si definisce *evento condizionato* il seguente ente logico a tre valori

$$E|H = \begin{cases} \textit{vero} & \text{se } E \text{ ed } H \text{ sono entrambi veri} \\ \textit{falso} & \text{se } E \text{ è falso e } H \text{ è vero} \\ \textit{indeterminato} & \text{se } H \text{ è falso.} \end{cases}$$

L'evento E si chiama *condizionando*, mentre l'evento H si chiama *condizionante*. Gli eventi sono una sottoclasse degli eventi condizionati. Infatti, se $H = \Omega$, segue $E|H = E|\Omega = E$.

La probabilità di $E|H$ (secondo un certo individuo) misura il suo grado di fiducia nel verificarsi di E *assumendo vero* H. Per misurare tale grado di fiducia si può considerare una scommessa (condizionata), che è valida se H risulta vero ed è annullata se H risulta falso. Pertanto, se tale individuo valuta $P(E|H) = p$, significa che egli è disposto a pagare p per ricevere

$$\begin{cases} 1 & \textit{se si verifica} \quad EH \\ 0 & \textit{se si verifica} \quad E^c H \\ p & \textit{se si verifica} \quad H^c. \end{cases} \tag{37}$$

In generale, in tale (ipotetica) scommessa condizionata, l'individuo si impegna a pagare pS, con $S \neq 0$, per ricevere

$$\begin{cases} S & \textit{se si verifica} \quad EH \\ 0 & \textit{se si verifica} \quad E^cH \\ pS & \textit{se si verifica} \quad H^c. \end{cases}$$

Il guadagno aleatorio $\mathcal{G}$ coincide con l'espressione $S|H||E| - pS|H| = S|H|(|E| - p)$ e i suoi possibili valori sono:

$$\begin{cases} S(1-p) & \textit{se si verifica} \quad E \cap H \\ -pS & \textit{se si verifica} \quad E^c \cap H \\ 0 & \textit{se si verifica} \quad H^c. \end{cases}$$

Il valore nullo del guadagno, corrispondente al verificarsi di H^c, esprime il fatto che la scommessa è annullata e tale valore, ai fini della verifica della coerenza, non deve giocare nessun ruolo. Pertanto, per verificare la coerenza della valutazione p, occorre considerare la restrizione, $G|H$, di G ad H, cioè occorre restringersi al caso in cui la scommessa è valida (H vero). Allora la condizione di coerenza diventa

$$Min\, \mathcal{G}|H \ \cdot Max\, \mathcal{G}|H \ \leq 0\,.$$

Osservando che i valori possibili di $\mathcal{G}|H$ sono $S(1-p)$ e $-pS$, segue che p è coerente se e solo se $0 \leq p \leq 1$.

Alcune proprietà.

- $P(H|H) = 1$, $P(\Omega|H) = 1$, $P(\emptyset|H) = 0$,
- Se $H \Longrightarrow E$, allora $P(E|H) = 1$.

Teorema 2 (delle probabilità composte) Dati due eventi E ed H,se le valutazioni di probabilità $P(EH), P(H), P(E|H)$ sono coerenti si ha

$$P(EH) = P(E|H)P(H). \tag{38}$$

Dim. Posto $P(E|H) = p$, la vincita aleatoria (37) corrispondente al pagamento dell'importo p coincide con il n.a.

$$X = 1 \cdot |EH| + 0 \cdot |E^cH| + p \cdot |H^c|\,, \tag{39}$$

che può definirsi come l'*Indicatore* di $E|H$.
D'altra parte, per ricevere X occorre pagare un importo pari alla sua previsione $\mathbb{P}(X)$. Pertanto, operativamente, risulta $\mathbb{P}(X) = p$, cioè

$$1 \cdot P(EH) + 0 \cdot P(E^c H) + p \cdot P(H^c) = p\,,$$

ovvero

$$P(EH) + p[1 - P(H)] = p\,,$$

e quindi

$$P(EH) = P(E|H)P(H)\,.$$

Corollario 1 Se $P(H) > 0$, si ottiene

$$P(E|H) = \frac{P(EH)}{P(H)}\,. \qquad (40)$$

(tale risultato è adottato da molti autori come *definizione* della probabilità di E condizionata ad H).

Nell'impostazione classica il teorema delle probabilità composte si ottiene con il seguente ragionamento.

Siano m i casi possibili, r_H i casi favorevoli ad H ed r_{EH} i casi favorevoli ad EH. Ovviamente, tra i casi favorevoli ad H, quelli favorevoli ad E sono ancora r_{EH}. Allora, assumendo $r_H > 0$, si ha

$$P(H) = \frac{r_H}{m}\,, \quad P(EH) = \frac{r_{EH}}{m}\,, \quad P(E|H) = \frac{r_{EH}}{r_H}\,,$$

pertanto

$$P(E|H) = \frac{r_{EH}}{r_H} = \frac{\frac{r_{EH}}{m}}{\frac{r_H}{m}} = \frac{P(EH)}{P(H)}\,.$$

Corollario 2 Se $H = \Omega$ allora si ha $EH = E$

$$P(E) = P(E|\Omega)P(\Omega) = P(E|\Omega)\,. \qquad (41)$$

Esempio. Estrazioni del lotto. Sia
$X = 1^0$ numero estratto,
$E = (X \leq 45)$ e $H = (X > 30)$. Calcolare $P(E|H)$.

Generalizzazione del teorema delle probabilità composte. Siano $E_1, E_2, \ldots, E_n$ n eventi arbitrari e vediamo come si può esprimere $P(E_1 E_2 \cdots E_n)$.

Iterando la formula relativa al caso di 2 eventi, si ha:

$$\begin{aligned} & P(E_1E_2\cdots E_n) = \\ & = P(E_n|E_1E_2\cdots E_{n-1})P(E_1E_2\cdots E_{n-1}) = \\ & = \cdots = \\ & = P(E_1)P(E_2|E_1)\cdots P(E_n|E_1E_2\cdots E_{n-1}). \end{aligned} \tag{42}$$

Formule analoghe alla precedente si ottengono permutando in tutti i modi possibili l'ordine degli E_i.

Ad esempio, per $n = 3$ si ha

$$\begin{aligned} P(E_1E_2E_3) &= P(E_1)P(E_2|E_1)P(E_3|E_1E_2) = \\ &= P(E_{i_1})P(E_{i_2}|E_{i_1})P(E_{i_3}|E_{i_1}E_{i_2})\,, \end{aligned}$$

dove $\{i_1, i_2, i_3\}$ è una permutazione di $\{1, 2, 3\}$.

Esempio 8 *(Gioco della roulette russa)* In una pistola a 6 colpi viene inserito un solo proiettile e il tamburo viene fatto girare vorticosamente. Quindi 6 prigionieri sono costretti a sottomettersi alla prova della roulette russa. Considerati gli eventi
E_i = *il proiettile esplode all'i-mo colpo*, $i = 1, \ldots, 6$, verificare che tali eventi hanno tutti probabilità $\frac{1}{6}$, cioè che i 6 prigionieri hanno la stessa probabilità di

morire.
Ovviamente, si ha $P(E_1) = \frac{1}{6}$. Inoltre

$$P(E_2|E_1^c) = \frac{1}{5}\,, \quad P(E_3|E_1^c E_2^c) = \frac{1}{4}\,,$$

$$\cdots\cdots\,, \qquad P(E_6|E_1^c \cdots E_5^c) = 1\,.$$

Allora, osservando che

$$E_2 = E_1^c E_2\,, \;\; E_3 = E_1^c E_2^c E_3\,, \;\; \cdots\,, \;\; E_6 = E_1^c \cdots E_5^c E_6\,,$$

applicando la (42) si ha

$$P(E_2) = P(E_1^c) P(E_2|E_1^c) = \frac{5}{6} \times \frac{1}{5} = \frac{1}{6}\,;$$

$$P(E_3) = P(E_1^c) P(E_2^c|E_1^c) P(E_3|E_1^c E_2^c) = \frac{5}{6} \times \frac{4}{5} \times \frac{1}{4} = \frac{1}{6}\,;$$

....

$$P(E_6) = P(E_1^c) P(E_2^c|E_1^c) P(E_3^c|E_1^c E_2^c) \cdots$$

$$\cdots P(E_5^c|E_1^c \cdots E_4^c) P(E_6|E_1^c \cdots E_5^c) =$$

$$= \frac{5}{6} \times \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} \times 1 = \frac{1}{6}\,.$$

Dati A, B, H, con $AB = \emptyset$, $P(H) > 0$, si ha

$$P(A \vee B|H) = \frac{P(AH \vee BH)}{P(H)} = \frac{P(AH)}{P(H)} + \frac{P(BH)}{P(H)} = \\ = P(A|H) + P(B|H)\ , \tag{43}$$

che estende la proprietà additiva al caso di probabilità condizionate.
La formula (43) può essere dimostrata anche nel caso $P(H) = 0$ (introducendo in generale la condizione di coerenza).

Applicando la (43), con $A = E$ e $B = E^c$, si ottiene

$$P(\Omega|H) = P(E \vee E^c|H) = P(E|H) + P(E^c|H) = 1\,,$$

e quindi

$$P(E^c|H) = 1 - P(E|H)\,. \tag{44}$$

In generale, invece, si ha

$$P(E|H^c) \neq 1 - P(E|H)\,.$$

Data una partizione $\{H_1, \ldots, H_n\}$ e un evento qualsiasi E, dalla relazione

$$E = EH_1 \vee \cdots \vee EH_n \,,$$

applicando la proprietà additiva e il teorema delle probabiltà composte si ottiene la seguente *formula di disintegrazione*

$$\begin{aligned} P(E) &= P(EH_1) + \cdots + P(EH_n) = \\ &= P(H_1)P(E|H_1) + \cdots + P(H_n)P(E|H_n). \end{aligned} \tag{45}$$

Esempio 9 *(Problema del condannato)* In un paese orientale un prigioniero è stato condannato a morte da uno sceicco. Prima dell'esecuzione, lo sceicco offre una possibilità di salvezza al condannato, mettendogli a disposizione 2 urne U_1 e U_2, con 2 palline bianche e 2 nere.

Il condannato deve distribuire a suo piacere le palline nelle due urne, con la condizione che nessuna urna rimanga vuota. Verrà poi scelta a caso un'urna da

cui verrà estratta una pallina. Se la pallina estratta sarà bianca il prigioniero sarà graziato.
Qual'è la migliore ripartizione delle palline nelle due urne?
Definiti gli eventi
$B =$ *la pallina estratta è bianca*,
$H_1 =$ *viene utilizzata l'urna* U_1,
$H_2 =$ *viene utilizzata l'urna* U_2,
la decisione migliore per il condannato è quella che rende massima la probabilità di B.

Supponiamo che il condannato inserisca h palline bianche e k palline nere in U_1, con $0 < h+k < 4$, e le rimanenti $4-h-k$ palline in U_2. Con tale strategia si ha

$$P(B) = P(H_1)P(B|H_1) + P(H_2)P(B|H_2) =$$
$$= \frac{1}{2}\left(\frac{h}{h+k} + \frac{2-h}{4-h-k} \; . \right)$$

Sostanzialmente, basta esaminare i seguenti casi (tra parentesi il simbolo $\sim$ indica una decisione equivalente che si può fare a meno di esaminare):

1. $h = 0\,,\ \ k = 1\ (\sim h = 2\,,\ \ k = 1)\,;$

2. $h = 0$, $k = 2$ $(\sim h = 2$, $k = 0)$;

3. $h = 1$, $k = 1$;

4. $h = 1$, $k = 0$ $(\sim h = 1$, $k = 2)$.

- Nel caso 1 si ha $P(B) = \frac{1}{2}(\frac{0}{1} + \frac{2}{3}) = \frac{1}{3}$;
- nel caso 2 si ha $P(B) = \frac{1}{2}(\frac{0}{2} + \frac{2}{2}) = \frac{1}{2}$;
- nel caso 3 si ha $P(B) = \frac{1}{2}(\frac{1}{2} + \frac{1}{2}) = \frac{1}{2}$;
- Nel caso 4 si ha $P(B) = \frac{1}{2}(\frac{1}{1} + \frac{1}{3}) = \frac{2}{3}$.

La decisione migliore pertanto è la n. 4 (una pallina bianca in un'urna e le rimanenti nell'altra).
Il problema si può generalizzare considerando n palline bianche ed n nere, con $n > 2$. Si può ve- rificare che la decisione migliore è sempre quella di mettere una pallina bianca in un'urna e le rimanenti nell'altra. A tale decisione corrisponde per $P(B)$ il valore

(massimo)

$$P(B) = \frac{1}{2}\left(1 + \frac{n-1}{2n-1}\right) \ ,$$

che all'aumentare di n sale verso $\frac{3}{4}$.

Indipendenza Stocastica. Dati E ed H, per le probabilità $P(E)$ e $P(E|H)$ si può avere:

$$P(E|H) > P(E), \ P(E|H) < P(E), \ P(E|H) = P(E).$$

Nel primo caso E è *correlato positivamente* con H, cioè assumendo H vero la probabilità di E aumenta.

Nel secondo caso E è *correlato negativamente* con H.

Nel terzo caso si dice che E è *stocasticamente indipendente* da H : l'ipotesi che H sia vero non fa nè aumentare nè diminuire la probabilità di E. In questo caso si ha

$$P(EH) = P(H)P(E|H) = P(E)P(H) \ , \qquad (46)$$

cioè la probabilità dell'evento intersezione EH coincide con il prodotto delle probabilità di E ed H.
Se $P(E|H) = P(E) > 0$, segue $P(H|E) = P(H)$, cioè se E è stocasticamente indipendente da H, si ha che H è indipendente da E. In tal caso segue anche

$$P(H^c|E) = 1 - P(H|E) = 1 - P(H) = P(H^c) \; .$$

Analoghe considerazioni si possono fare per le probabilità condizionate

$$P(H|E^c) \; , \;\; P(H^c|E^c) \; , \;\; P(E|H^c) \; , \;\; P(E^c|H^c) \; .$$

In generale, data una famiglia arbitraria $\mathcal{F}$ di eventi (tutti di probabilità positiva e minore di 1), gli eventi di $\mathcal{F}$ si dicono **(stocasticamente) indipendenti** se, per ogni sottofamiglia $\{E_1, \ldots, E_n\}$ di $\mathcal{F}$, $n \geq 2$, si ha

$$P(E_1 \cdots E_n) = P(E_1) \cdots P(E_n) \; . \tag{47}$$

Osservazione. Se la condizione (47) vale per un certo intero n (ad esempio, $n = 2$), non è detto che valga per $n + 1$ (nell'esempio fatto, $n + 1 = 3$).

Esempio. Si effettua un'estrazione da un'urna contenente 4 palline, 1 rossa, 1 nera , 1 bianca, 1 gialla. Consideriamo gli eventi
$R=$ "la pallina estratta è rossa",
$N=$ "la pallina estratta è nera",
$B=$ "la pallina estratta è bianca",
$G=$ "la pallina estratta è gialla",
$E = R \vee B$, $H = N \vee B$ $A = G \vee B$.
Si può verificare che la (47) non vale per la terna (E, H, A), mentre vale per ciascuna delle 3 coppie (E, H), (E, A), (H, A).

Teorema di Bayes. Tale teorema mette in evidenza che le probabilità condizionate sono lo strumento teorico per "apprendere dall'esperienza", attraverso l'aggiornamento delle valutazioni di probabilità di una o più *ipotesi*, man mano che lo stato di informazione cambia per effetto di nuovi dati o notizie.

Dati E ed H, per il teorema delle probabilità composte si ha:

$$P(EH) = P(H)P(E|H) = P(E)P(H|E) \ ,$$

da cui, assumendo $P(E) > 0$, si ottiene

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} = \\ = \frac{P(H)P(E|H)}{P(H)P(E|H)+P(H^c)P(E|H^c)} \,. \tag{48}$$

(*teorema o formula di Bayes* nella sua versione più semplice).
Se H rappresenta un'*ipotesi incerta* di probabilità $P(H)$ ed E un fatto osservabile, di probabilità $P(E)$ e di probabilità condizionata ad H pari a $P(E|H)$, la (48) ci mostra come il "grado di fiducia" nei riguardi di H si modifica quando si suppone di aver osservato l'evento E.

Esempio 10 Una persona (Tizio) attende nella stazione ferroviaria di una località A l'arrivo di un amico (Caio) che dovrebbe essere partito dalla stazione di una località B.

Tizio sa che Caio, che è un tipo bizzarro, ha lanciato in aria una moneta, decidendo di partire (ipotesi H) oppure no (ipotesi H^c) a seconda che il risultato del lancio sia stato Testa oppure Croce.

Inoltre, nel caso che il risultato del lancio sia stato Testa (ipotesi H vera), avendo a disposizione 6 treni Tizio ha lanciato in aria un dado, scegliendo il treno corrispondente al risultato ottenuto con il lancio del dado.

Considerati gli eventi
$A_1 =$ *Caio arriva con il primo treno*,
$\dots$
$A_6 =$ *Caio arriva con il sesto treno*,
calcolare la probabilità che Caio sia partito supposto che non sia arrivato con nessuno dei primi cinque treni.
Posto

$$E = A_1^c A_2^c \cdots A_5^c = (A_1 \vee \cdots \vee A_5)^c = A_6 \vee H^c \ ,$$

si tratta di calcolare $P(H|E)$. Nel nostro caso si ha

$$P(H) = P(H^c) = \frac{1}{2}\ , \quad P(A_i|H) = \frac{1}{6}\ , \quad i = 1, \dots, 6.$$

Inoltre, essendo $A_i = A_i H$, si ha per $i = 1, \dots, 6$,

$$P(A_i) = P(A_i H) = P(H)P(A_i|H) = \tfrac{1}{2} \times \tfrac{1}{6} = \tfrac{1}{12}\,,$$

da cui

$$P(E) = P(A_6 \vee H^c) = P(A_6) + P(H^c) = \frac{1}{12} + \frac{1}{2} = \frac{7}{12} .$$

Inoltre

$$P(E|H) = P(A_6 \vee H^c|H) = P(A_6|H) + P(H^c|H) = \\ = \tfrac{1}{6} + 0 = \tfrac{1}{6} .$$

Allora, dalla formula (48) si ottiene

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} = \frac{\frac{1}{2} \times \frac{1}{6}}{\frac{7}{12}} = \frac{1}{7} .$$

Come si vede, supposto che sia rimasta possibile solo l'eventualità che Caio arrivi con l'ultimo treno, la probabilità che Caio sia partito è scesa da $\frac{1}{2}$ a $\frac{1}{7}$.

È da notare che con un ragionamento intuitivo errato si può essere indotti a valutare $P(H|E) = \frac{1}{2}$. Per convincersi che tale valutazione non è "ragionevole", basta considerare il caso in cui i treni anzichè 6 siano 100, oppure 1000, oppure ... e supporre che sia rimasto possibile solo l'arrivo con l'ultimo treno.

Con un ragionamento analogo a quello precedente si può verificare che

$$P(H|A_1^c) = \tfrac{5}{11}\,, \qquad P(H|A_1^c A_2^c) = \tfrac{4}{10}\,,$$

$$P(H|A_1^c A_2^c A_3^c) = \tfrac{3}{9}\,, \quad P(H|A_1^c A_2^c A_3^c A_4^c) = \tfrac{2}{8}\,.$$

Data una partizione $\{H_1, \ldots, H_n\}$ ed un evento E, per ogni fissato $i = 1, \ldots, n$, si ha

$$\begin{aligned} P(H_i|E) = \quad & \tfrac{P(H_i)P(E|H_i)}{P(E)} = \\ & = \tfrac{P(H_i)P(E|H_i)}{P(H_1)P(E|H_1)+\cdots+P(H_n)P(E|H_n)}\,. \end{aligned} \tag{49}$$

(*versione generale del Teorema di Bayes*)

Nelle applicazioni, E rappresenta una "osservazione" in un dato esperimento, mentre $\{H_1, \ldots, H_n\}$ sono le possibili *ipotesi* che "spiegano" l'evento osservato.

$P(H_i),\ i = 1, \ldots, n,$ sono le probabilità *iniziali*
$P(H_i|E),\ i = 1, \ldots, n,$ sono le probabilità *finali*
$P(E|H_i),\ i = 1, \ldots, n,$ sono le *verosimiglianze*

La (49) ci mostra come $P(H_i)$ si modifica in $P(H_i|E)$ quando si suppone di aver osservato l'evento E. In questo senso la formula di Bayes rappresenta lo strumento formale per *apprendere dall'esperienza*.

Esempio 11 Uno studente deve rispondere ad un quesito di esame indicando, tra n possibili risposte, quella corretta. Definiamo gli eventi:
$H =$ "lo studente è preparato",
$E =$"lo studente risponde correttamente al quesito".
Indichiamo inoltre con $p = P(H)$ la probabilità iniziale di H e supponiamo che le verosimiglianze siano $P(E|H) = 1$ e $P(E|H^c) = 1/n$. Quanto vale $P(H|E)$? Applicando il teorema di Bayes si ottiene

$$P(H|E) = \cdots = \frac{np}{(n-1)p + 1}\,.$$

Altri possibili aspetti da esaminare sono:

$$P(H|E) > p\,? \quad P(H|E) > P(H^c|E)\,?$$

Test d'ipotesi. Se $P(H|E) > P(H^c|E)$ vuol dire che, assumendo vero E, l'ipotesi H è più probabile dell'ipotesi H^c. Nel caso di n ipotesi, applicando il teorema di Bayes, si può (soltanto) determinare quale delle ipotesi fatte è diventata più probabile rispetto alle altre, mentre non ha senso, in generale, parlare dell'ipotesi più probabile come se fosse (quella) $vera$.

Esempio 12 Un oggetto viene nascosto (da Tizio) in uno fra 7 contenitori, che successivamente vengono divisi in due gruppi di 3 e 4 contenitori rispettivamente.
Un amico (Caio), per cercare l'oggetto, può aprire solo 2 contenitori in uno dei due gruppi a sua scelta.
Conviene scegliere a caso nel gruppo da 3 (strategia A) oppure nel gruppo da 4 (strategia B)?
Supposto di aver aperto 2 contenitori scelti a caso nel gruppo da 3 e di non aver trovato nulla, qual'è la probabilità che l'oggetto stia nel contenitore rimanente?
Si può dimostrare che le due strategie sono equivalenti, cosi come sarebbe indifferente aprire a caso un contenitore del primo gruppo e uno del secondo gruppo.
Infatti, ognuno dei 7 contenitori ha probabilità $\frac{1}{7}$ di contenere l'oggetto nascosto. Allora, definiti gli eventi
$H =$ *l'oggetto è nascosto nel gruppo da 3*,
$H^c =$ *l'oggetto è nascosto nel gruppo da 4*,
$E_1 =$ *Caio apre a caso 2 contenitori nel gruppo da 3 e trova l'oggetto*,

$E_2 =$ *Caio apre a caso 2 contenitori nel gruppo da 4 e trova l'oggetto*, si ha

$$P(H) = \frac{3}{7}\,; \quad P(H^c) = \frac{4}{7}\,.$$

Inoltre

$$P(E_1|H) = \frac{\binom{1}{1}\binom{2}{1}}{\binom{3}{2}} = \tfrac{2}{3}$$

$$P(E_2|H^c) = \frac{\binom{1}{1}\binom{3}{1}}{\binom{4}{2}} = \tfrac{1}{2}\,.$$

Allora, con la strategia A si ha

$$P(E_1) = P(E_1H) = P(H)P(E_1|H) = \frac{3}{7} \times \frac{2}{3} = \frac{2}{7}\,,$$

mentre con la strategia B si ha

$$P(E_2) = P(E_2H^c) = P(H^c)P(E_2|H^c) = \frac{4}{7} \times \frac{1}{2} = \frac{2}{7}\,.$$

Pertanto le due strategie sono equivalenti. Inoltre, si può osservare che aprendo 2 contenitori scelti a caso fra i 7 iniziali la probabilità di trovare l'oggetto è ancora la stessa, essendo data da

$$\frac{\binom{1}{1}\binom{6}{1}}{\binom{7}{2}} = \frac{2}{7}\,.$$

Infine, con riferimento alla seconda domanda, occorre calcolare $P(H|E_1^c)$.
Osservando che $P(E_1^c|H^c) = 1$, dal teorema di Bayes si ottiene

$$P(H|E_1^c) = \frac{P(H)P(E_1^c|H)}{P(E_1^c)} = \frac{P(H)P(E_1^c|H)}{P(H)P(E_1^c|H)+P(H^c)P(E_1^c|H^c)} =$$
$$= \frac{\frac{3}{7}\times\frac{1}{3}}{\frac{3}{7}\times\frac{1}{3}+\frac{4}{7}\times 1} = \frac{1}{5}\,.$$

Ovviamente risulta

$$P(H^c|E_1^c) = \cdots = \frac{4}{5}\,.$$

Teorema 3 Data una famiglia arbitraria $\mathcal{F}$ di eventi (tutti di probabilità positiva e minore di 1), se gli eventi di $\mathcal{F}$ sono (**stocasticamente**) **indipendenti** allora, per ogni sottofamiglia $\{E_1, \ldots, E_n\}$ di $\mathcal{F}$, con $n \geq 2$, e per ogni $1 \leq s \leq n$, si ha

$$\begin{aligned} &P(E_1^c E_2^c \cdots E_s^c E_{s+1} \cdots E_n) = \\ &P(E_1^c)P(E_2^c) \cdots P(E_s^c)P(E_{s+1}) \cdots P(E_n) \ . \end{aligned} \tag{50}$$

Distribuzione Binomiale. Siano dati n eventi $E_1, E_2, \ldots, E_n$ indipendenti ed equiprobabili, di probabilità p. Se E_i risulta vero (risp., falso), si parla convenzionalmente di successo (risp., insuccesso) alla $i - esima$ prova. Poniamo $q = 1 - p$. Indicando con X il numero aleatorio di successi sulle n prove, si ha

$$X = |E_1| + |E_2| + \cdots + |E_n|.$$

I possibili valori di X sono $0, 1, 2, \ldots, n$. Per determinare la distribuzione di probabilità di X, occorre calcolare i valori $P(X = h)$, con $h = 0, 1, \ldots, n$. A tale scopo conviene rappresentare l'evento $(X = h)$

come unione di alcuni dei costituenti della famiglia $\{E_1, E_2, \ldots, E_n\}$. Si ha

$$(X = 0) \; = \; E_1^c E_2^c \cdots E_n^c$$

e quindi

$$P(X = 0) = P(E_1^c E_2^c \cdots E_n^c) = P(E_1^c)P(E_2^c) \cdots P(E_n^c) = \\ = \underbrace{(1-p) \cdot (1-p) \cdots (1-p)}_{n \text{ volte}} = (1-p)^n = q^n \,.$$

In modo analogo si ha

$$(X = n) \; = \; E_1 E_2 \cdots E_n$$

e quindi

$$P(X = n) = P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2) \cdots P(E_n) = \\ = \underbrace{p \cdot p \;\cdots p}_{n \text{ volte}} = p^n \,.$$

In relazione all'evento $(X = 1)$, osserviamo che si ha

$$(X = 1) \; = \\ \underbrace{E_1 E_2^c \cdots E_n^c \vee E_1^c E_2 E_3^c \cdots E_n^c \vee \cdots \vee E_1^c E_2^c \cdots E_{n-1}^c E_n}_{\text{unione di } n \text{ costituenti equiprobabili}}$$

e quindi

$$
\begin{aligned}
&P(X = 1) = \\
&= P(E_1 E_2^c \cdots E_n^c \vee E_1^c E_2 E_3^c \cdots E_n^c \vee \cdots \vee E_1^c E_2^c \cdots E_{n-1}^c E_n) \\
&= P(E_1 E_2^c \cdots E_n^c) + \cdots + P(E_1^c E_2^c \cdots E_{n-1}^c E_n) = \\
&= \underbrace{pq^{n-1} + pq^{n-1} + \cdots pq^{n-1}}_{n\, volte} = npq^{n-1} = \binom{n}{1} p^1 q^{n-1} \,.
\end{aligned}
$$

In generale, fissato $h \in \{0, 1, \ldots n\}$, il numero di costituenti favorevoli all'evento $(X = h)$ è $\binom{n}{h}$ ed ognuno di essi ha probabilità $p^h q^{n-h}$. Pertanto

$$P(X = h) = \binom{n}{h} p^h q^{n-h} \,. \tag{51}$$

Diremo che X ha una *distribuzione binomiale* di parametri n e p, utilizzando il simbolo

$$X \sim \mathbf{B}(n, p) \,.$$

Ovviamente, si ha

$$\sum_{h=0}^{n} P(X = h) = \sum_{h=0}^{n} \binom{n}{h} p^h q^{n-h} = (p + q)^n = 1 \,.$$

In particolare se $p = \frac{1}{2}$, come ad esempio nel caso di n lanci di una moneta ben costruita, si ha

$$P(X = h) = \frac{\binom{n}{h}}{2^n} \quad h = 0, 1, 2, \ldots, n, \qquad (52)$$

e la distribuzione binomiale è simmetrica. Si ha infatti

$$P(X = 0) = P(X = n) = \frac{1}{2^n}\,,$$

$$P(X = 1) = P(X = n - 1) = \frac{n}{2^n}\,,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Se invece $p \neq \frac{1}{2}$ la distribuzione è asimmetrica, come si può vedere dalle seguenti figure:

Figura 2: Binomiale p=1/2, n=10

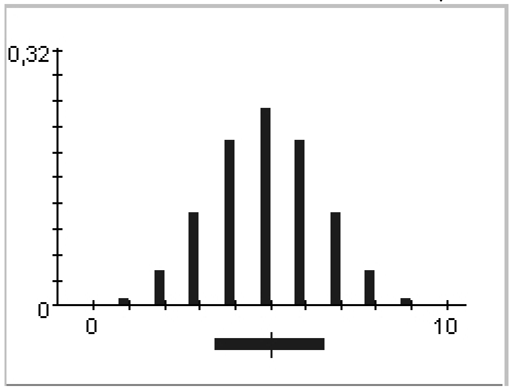
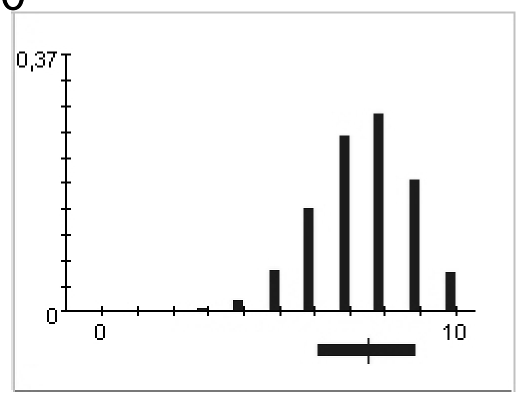
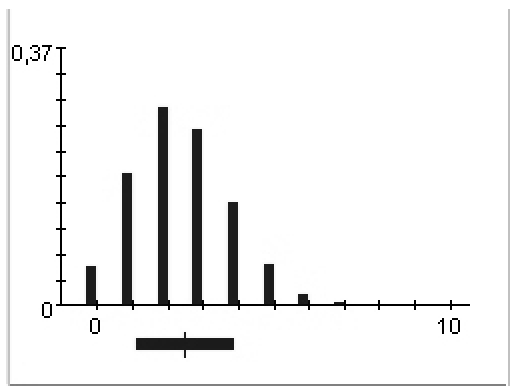


Figura 3: Binomiale p=2/3, n=10 - Binomiale p=1/3, n=10





Per la previsione di X si ha

$$\begin{aligned}\mathbb{P}(X) = \quad & \mathbb{P}(|E_1| + |E_2| + \cdots + |E_n|) = \\ & \sum_{i=1}^{n} \mathbb{P}(|E_i|) = \sum_{i=1}^{n} p(E_i) = np \,.\end{aligned}$$

Per quanto riguarda la varianza, osserviamo che dall'indipendenza degli eventi $E_1, \ldots, E_n$ segue

$$\begin{aligned}& Cov(|E_i|, |E_j|) = P(E_i E_j) - P(E_i)P(E_j) = \\ & = P(E_i)P(E_j) - P(E_i)P(E_j) = 0 \,, \;\; \forall\, i \neq j \,.\end{aligned}$$

Pertanto

$$\begin{aligned}Var(X) = \quad & Var(|E_1| + |E_2| + \cdots + |E_n|) = \\ & \sum_{i=1}^{n} Var(|E_i|) = \sum_{i=1}^{n} pq = npq \,.\end{aligned}$$

In sintesi, se $X \sim \mathbf{B}(n, p)$, allora

$$\mathbb{P}(X) = np \,, \quad Var(X) = npq \,, \quad \sigma_X = \sqrt{npq} \,.$$

Ad esempio:

$$\begin{array}{llll}\mathbf{B}(10, \frac{1}{2}) : & \mathbb{P}(X) = 5 & \sigma_X^2 = 2.50 & \sigma_X = 1.58 \\ \mathbf{B}(10, \frac{3}{4}) : & \mathbb{P}(X) = 7.5 & \sigma_X^2 = 1.88 & \sigma_X = 1.37 \\ \mathbf{B}(10, \frac{1}{4}) : & \mathbb{P}(X) = 2.5 & \sigma_X^2 = 1.88 & \sigma_X = 1.37\end{array}$$

Estrazioni con restituzione da un'urna di composizione nota. Si effettuano n estrazioni con restituzione da un'urna contenente N palline, di cui pN bianche e qN nere ($pN + qN = N$). Indichiamo con E_i l'evento "l'i-esima pallina estratta è bianca", $i = 1, \ldots, n$, e con $X = \sum_{i=1}^{n} |E_i|$ il numero aleatorio di palline bianche estratte nelle n estrazioni.

Gli eventi E_i sono indipendenti ed equiprobabili, quindi $X \sim \mathbf{B}(n, p)$.

Estrazioni senza restituzione da un'urna di composizione nota. Si effettuano n estrazioni senza restituzione da un'urna contenente N palline, di cui pN bianche e qN nere ($pN + qN = N$). Ovviamente deve essere $n \leq N$. Indichiamo con E_i l'evento "l'i-esima pallina estratta è bianca", $i = 1, \ldots, n$, e con $X = \sum_{i=1}^{n} |E_i|$ il numero aleatorio di palline bianche estratte nelle n estrazioni. Valutiamo $P(E_i)$, per $i = 1, \ldots, n$.

$$P(E_1) = \tfrac{pN}{N} = p\,.$$

$P(E_2) = P(E_2 \wedge \Omega) = P(E_2 \wedge (E_1 \vee E_1^c)) =$

$= P(E_2|E_1)P(E_1) + P(E_2|E_1^c)P(E_1^c) =$

$= \frac{pN-1}{N-1}p + \frac{pN}{N-1}q = p\,.$

Con il calcolo combinatorio, calcolando il rapporto fra casi favorevoli e casi possibili, si ha

$$P(E_2) = \frac{(N-1)pN}{N(N-1)} = p\,.$$

In generale, sempre utilizzando il calcolo combinatorio, si ha

$$P(E_i) = \frac{(N-1)(N-2)\cdots(N-i+1)pN}{N(N-1)\cdots(N-i+1)} = p\,.$$

In altro modo, immaginando di fare N estrazioni, le pN palline bianche verranno estratte in uno dei sottoinsiemi di pN prove fra le N e ce ne sono $\binom{N}{pN}$. Quelli favorevoli sono $1 \cdot \binom{N-1}{pN-1}$: 1 corrisponde all'i-ma prova ed $\binom{N-1}{pN-1}$ rappresenta, in relazione alle rimanenti $N-1$

prove, i sottoinsiemi possibili di $pN-1$ prove in cui vengono estratte le rimanenti $pN-1$ palline bianche.

$$P(E_i) = \frac{\overbrace{1}^{prova\ i\text{-}esima} \cdot \binom{N-1}{pN-1}}{\binom{N}{pN}} = p\,.$$

Pertanto gli eventi sono equiprobabili anche nel caso di estrazioni senza restituzione.
Considerando adesso il costituente

$$E_1E_2\cdots E_hE_{h+1}^c\cdots E_n^c\,,$$

si ha

$$P(E_1E_2\cdots E_hE_{h+1}^c\cdots E_n^c) = \frac{D_{pN,h}D_{qN,n-h}}{D_{N,n}} =$$

$$= \frac{pN(pN-1)(pN-2)\cdots(pN-(h-1))qN(qN-1)\cdots(qN-(n-h)+1)}{N(N-1)(N-2)\cdots(N-(h-1))(N-h)\cdots(N-n+1)} =$$

$$= \frac{(pN)!(N-pN)!(N-n)!}{(pN-h)!(N-pN-n+h)!N!} = \frac{\frac{(N-n)!}{(pN-h)!(N-pN-n+h)!}}{\frac{N!}{(pN)!(N-pN)!}} = \frac{\binom{N-n}{pN-h}}{\binom{N}{pN}}\,.$$

Tale risultato dipende solo dagli indici h, n, quindi per ogni scelta degli h successi si ottiene

$$P(E_{i_1} E_{i_2} \cdots E_{i_h} E^c_{i_{h+1}} \cdots E^c_{i_n}) = \frac{\binom{N-n}{pN-h}}{\binom{N}{pN}} \, .$$

Indicando con $A_{h,n}$ il generico costituente

$$E_{i_1} E_{i_2} \cdots E_{i_h} E^c_{i_{h+1}} \cdots E^c_{i_n}$$

favorevole ad $(X = h)$ si ha

$$(X = h) = (\bigvee_{\{i_1, i_2, \ldots, i_n\}} A_{h,n}) \, ,$$

dove l'unione è fatta rispetto a tutti gli $\binom{n}{h}$ costituenti favorevoli ad $(X = h)$. In definitiva si ha

$$P(X = h) = \frac{\binom{n}{h}\binom{N-n}{pN-h}}{\binom{N}{pN}} \, . \qquad (53)$$

Suppondendo di effettuare estrazioni in blocco si giunge ad una formula che si può memorizzare più facilmente

$$P(X = h) = \frac{\binom{pN}{h}\binom{qN}{n-h}}{\binom{N}{n}}\,. \tag{54}$$

La distribuzione di X si chiama distribuzione **Ipergeometrica** e si indica con

$$X \sim \mathbf{H}(N, n, p)\,.$$

Il n.a. X ha il seguente codominio

$$\max\{0, n - qN\} \leq \; X \leq \min\{pN, n\}\,.$$

Esempio. Consideriamo 15 estrazioni senza restituzione da un'urna contenente 12 bianche e 8 nere. Si ha

$$N = 20, \quad pN = 12, \quad qN = 8, \quad n = 15, \quad p = \tfrac{3}{5}.$$

$$\max\{0, n - qN\} = \max\{0, 7\} = 7\,,$$

$$\min\{pN, n\} = \min\{12, 15\} = 12\,,$$

$$X \in \{7, 8, 9, 10, 11, 12\}\,, \quad X \sim \mathbf{H}(20, 15, 3/5)\,.$$

Fissato $h \in \{7, 8, 9, 10, 11, 12\}$, ad esempio $h = 8$, si ha

$$P(X = 8) = \frac{\binom{pN}{h}\binom{qN}{n-h}}{\binom{N}{n}} = \frac{\binom{12}{8}\binom{8}{7}}{\binom{20}{15}}\,.$$

Previsione e Varianza. Essendo gli eventi E_i equiprobabili di probabilità p, come per la binomiale si ha

$$\begin{aligned}\mathbb{P}(X) = \; & \mathbb{P}(|E_1| + |E_2| + \cdots + |E_n|) = \\ & = \sum_{i=1}^n \mathbb{P}(|E_i|) = \sum_{i=1}^n p(E_i) = np\,.\end{aligned}$$

La varianza risulta differente rispetto a quella della binomiale perchè $Cov(|E_i|, |E_j|) \neq 0$. Infatti

$$\begin{aligned}& Cov(|E_i|, |E_j|) = P(E_iE_j) - P(E_i)P(E_j) = \\ & = \tfrac{pN-1}{N-1}p - p^2 = -\tfrac{pq}{N-1}\,.\end{aligned}$$

Quindi

$$Var(X) = Var(|E_1| + |E_2| + \cdots + |E_n|) =$$

$$= \sum_{i=1}^n Var(|E_i|) + 2\sum_{i<j} Cov(|E_i|, |E_j|) =$$

$$= npq + 2\binom{n}{2}\left(-\tfrac{pq}{N-1}\right) = \cdots = npq\left(1 - \tfrac{n-1}{N-1}\right) \leq npq\,.$$

In sintesi se $X \sim \mathbf{H}(N, n, p)$ si ha

$$P(X = h) = \frac{\binom{n}{h}\binom{N-n}{pN-h}}{\binom{N}{pN}} = \frac{\binom{pN}{h}\binom{qN}{n-h}}{\binom{N}{n}}\,,$$

$$\mathbb{P}(X) = np\,, \quad Var(X) = npq(1 - \tfrac{n-1}{N-1})\,.$$

Comportamento Asintotico.

$$\lim_{N\to\infty} P(A_{h,n}) =$$
$$= \lim_{N\to\infty} \tfrac{pN(pN-1)(pN-2)\cdots(pN-(h-1))qN(qN-1)\cdots(qN-(n-h)+1)}{N(N-1)(N-2)\cdots(N-(h-1))(N-h)\cdots(N-n+1)} =$$

$$= \lim_{N\to\infty} \tfrac{pNpN(1-\frac{1}{pN})\cdots pN(1-\frac{(h-1)}{pN})qNqN(1-\frac{1}{qN})\cdots qN(1-\frac{(n-h-1)}{qN})}{N(N-1)(N-2)\cdots(N-(h-1))(N-h)\cdots(N-n+1)} =$$
$$= p^h q^{n-h}\,.$$

Asintoticamente la distribuzione ipergeometrica, fissati i valori n, p, converge alla distribuzione binomiale di parametri n, p. Infatti se la popolazione è molto numerosa (N molto grande rispetto ad n), dal punto di vista numerico è praticamente la stessa cosa fare estrazioni *con* o *senza* restituzione.

Numeri aleatori discreti

Un numero aleatorio X si dice discreto se l'insieme dei suoi possibili valori è finito o infinito numerabile, cioè

$$X \in \{x_1, x_2, \ldots, x_n, \ldots\} \,.$$

Sia $\{p_i\,,\ i \in \mathbb{N}\}$, con $p_i = P(X = x_i)$, la distribuzione di probabilità di X e supponiamo che

$$\textstyle\sum_{i=1}^{\infty} p_i = 1\,, \quad \sum_{i=1}^{\infty} |x_i| p_i < \infty\,.$$

In questo caso la previsione si definisce nel seguente modo

$$\mathbb{P}(X) = \sum_{i=1}^{\infty} p_i x_i\,. \tag{55}$$

La funzione

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i\,, \quad x \in \mathbb{R}\,, \tag{56}$$

si chiama funzione di ripartizione di X.
Osserviamo che $F(x) \in [0,1]\,, \forall\, x \in \mathbb{R}$.

Esempio. La f.d.r di $X = |E|$, con $p = P(E)$, è

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

Basta osservare che per $x < 0$ si ha $(X \leq x) = \emptyset$, mentre per $x \geq 1$ si ha $(X \leq x) = \Omega$. Inoltre, per $0 \leq x < 1$ si ha $(X \leq x) = (X = 0)$. In particolare

$$P(X \leq 0) = \overbrace{P(X = 0)}^{=P(E^c)} + \overbrace{P(X < 0)}^{=0} = P(E^c) = q$$
$$P(X \leq 1) = P(X = 0) + P(X = 1) = q + p = 1$$

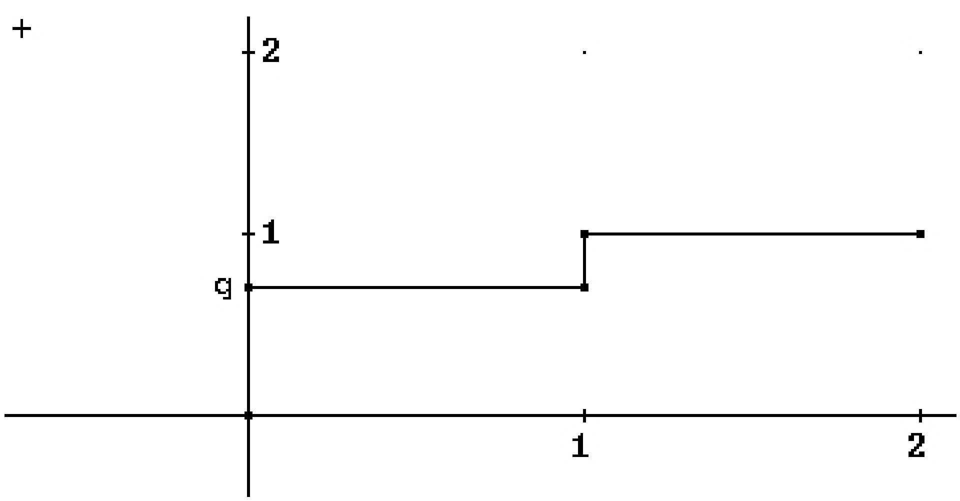


Figura 4: funzione di ripartizione di $|E|$

Distribuzione Geometrica. Data una successione di eventi $E_1, E_2, \ldots, E_n, \ldots$, indipendenti ed equiprobabili, con $P(E_i) = p$, $q = 1 - p$, sia X il n.a. di prove fino al primo successo. Osserviamo che X è un n.a. discreto, infatti

$$X \in \mathbb{N}\,.$$

(In realtà stiamo trascurando il caso in cui $X = \infty$, cioè si ha sempre insuccesso; come vedremo tale evento ha probabilità nulla). Si ha

$$P(X = n) = q^{n-1}p, \ \ n \in \mathbb{N}.$$

Infatti

$$\begin{array}{l} P(X = 1) = P(E_1) = p\,, \\ P(X = 2) = P(E_1^c E_2) = qp\,, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ P(X = n) = P(E_1^c E_2^c \cdots E_{n-1}^c E_n) = pq^{n-1}\,, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{array}$$

Si dice che X ha distribuzione geometrica di parametro p e si indica con

$$X \sim \mathbf{G}(p)\,.$$

La successione $\{p_n\}$, dove $p_n = P(X = n)$, è geometrica di ragione $(1-p)$. Infatti

$$p_1 = p \geq p_2 = p(1-p) \geq p_3 = p(1-p)^2 \geq \cdots ,$$

quindi $(X = 1)$ (successo alla prima prova) è il costituente più probabile.

Inoltre vale la $\sigma-$additività:

$$\sum_{n=1}^{\infty} p_n = \sum_{n=1}^{\infty} p(1-p)^{n-1} =$$

$$= p\sum_{n=1}^{\infty}(1-p)^{n-1} = p\frac{1}{1-(1-p)} = 1 .$$

Dal precedente risultato si ottiene

$$P(X = \infty) = 1 - P(\bigvee_{n=1}^{\infty}(X = n)) = 1 - \sum_{i=1}^{\infty} p_n = 0.$$

Osserviamo che per ogni fissato n si ha

$$(X > n) = E_1^c E_2^c \cdots E_n^c ,$$

quindi

$$P(X > n) = q^n .$$

Proprietà di assenza di memoria. Un numero aleatorio discreto e non negativo X ha una distribuzione geometrica se e solo se vale la seguente proprietà (detta di assenza di memoria)

$$P(X > n + n_0 | X > n_0) = P(X > n), \ \ \forall n, n_0 \in \mathbb{N} \tag{57}$$

il cui significato è: "supposto di non aver avuto successo fino alla prova n_0, la probabilità di non avere successo nelle successive$n$ prove è la stessa di non avere successo nelle prime n prove".

dim. $(\Rightarrow)$ Hp) $X \sim \mathbf{G(p)}$; $\quad Th$) vale la (57).

$$P(X > n + n_0 | X > n_0) = \tfrac{P(X>n+n_0, X>n_0)}{P(X>n_0)} =$$

$$= \tfrac{P(X>n+n_0)}{P(X>n_0)} = \tfrac{q^{n+n_0}}{q^{n_0}} = q^n = P(X > n)\,.$$

$(\Leftarrow)$ Hp) vale la (57); $\quad Th$) $X \sim \mathbf{G(p)}$.

Dalla (57) segue

$$P(X > n_0 + n) = P(X > n) P(X > n_0)\,.$$

Introduciamo la funzione di sopravvivenza definita come

$$S(x) = 1 - F(x) = P(X > x)\,. \qquad (58)$$

Per $n_0 = 1$ si ha

$P(X > n+1) = P(X > 1)P(X > n), \quad \forall n \in \mathbb{N}\,;$
allora
$S(n+1) = S(1)S(n) = S(1)[S(1)S(n-1)] =$
$= \cdots = [S(1)]^{n+1}\,.$
Posto $p = P(X = 1)\,, q = 1 - p = P(X > 1) = S(1),$
segue $S(n+1) = q^{n+1}.$

Osservando che

$$(X > n) \,=\, (X = n+1) \vee (X > n+1)\,,$$

otteniamo

$$P(X = n+1) = P(X > n) - P(X > n+1) =$$

$$= S(n) - S(n+1) = q^n(1-q) = q^n p\,.$$

Previsione e Varianza.

$$\mathbb{P}(X) = \sum_{n=1}^{\infty} npq^{n-1} = p\sum_{n=1}^{\infty} nq^{n-1} = p\sum_{n=1}^{\infty} \frac{d(q^n)}{dq} =$$

$$= p\frac{d}{dq}\sum_{n=0}^{\infty} q^n = p\frac{d}{dq}(\frac{q}{1-q}) = p(\frac{(1-q)+q}{(1-q)^2}) = \frac{p}{p^2} = \frac{1}{p}\,.$$

Con un ragionamento analogo si può provare che

$$\mathbb{P}(X^2) = \sum_{n=1}^{\infty} n^2pq^{n-1} = \frac{1+q}{p^2}\,,$$

$$Var(X) = \mathbb{P}(X^2) - [\mathbb{P}(X)]^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}\,.$$

Distribuzione di Poisson

Si dice che un n.a. X ha una distribuzione di Poisson di parametro λ, con $\lambda \in \mathbb{R}^+$, e si indica

$$X \sim \mathcal{P}(\lambda)\,,$$

se valgono le seguenti condizioni

1) $X \in \mathbb{N}_0 = \{0, 1, 2, 3, \ldots, n, \ldots\}$;

2) $P(X = n) = p_n = \frac{\lambda^n}{n!}e^{-\lambda}, \;\; \forall\, n \in \mathbb{N}_0\,.$

Si ha

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1\,.$$

Nella distribuzione di Poisson la previsione e la varianza coincidono. Infatti

$$\mathbb{P}(X) = \sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} =$$

$$= \sum_{n=0}^{\infty} n\lambda \frac{\lambda^{n-1}}{n!} e^{-\lambda} = \lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \lambda \cdot 1 = \lambda\,,$$

$$\mathbb{P}(X^2) = \sum_{i=0}^{\infty} n^2 p_n = \cdots = \lambda^2 + \lambda\,,$$

$$Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda\,.$$

La distribuzione di Poisson si usa in una grande varietà di applicazioni, ad esempio (sotto certe condizioni) si utilizza come distribuzione di probabilità di un n.a. X che (in un certo *intervallo di tempo*) rappresenta

1. il num. di telefonate che giungono ad un centralino;

2. il num. di clienti che si presentano ad uno sportello;

3. il num. di automobili che transitano ad un casello autostradale;

4. il num. di particelle emesse da un corpo radioattivo;

5. ....

Una delle prime applicazioni si ebbe nell'esercito prussiano, per valutare la probabilità, molto piccola, che un soldato dell'esercito morisse in seguito ad un calcio di cavallo.

Osserviamo che, sotto certe ipotesi, la distribuzione Binomiale si può approssimare con la distribuzione di Poisson.

Consideriamo in proposito l'esperimento di scelta a caso di n punti in un intervallo $[0, a]$. Sia I un sottointervallo di $[0, a]$ di ampiezza t, ad esempio $[0, t]$. Indicando con $X_1, X_2, \ldots, X_n$ le ascisse aleatorie dei punti generati nell'esperimento, definiamo i seguenti eventi

$$E_i = (X_i \in I), \quad i = 1, 2, \ldots, n.$$

Se i punti sono scelti a caso, si può valutare (come si vedrà in seguito con i numeri aleatori continui con distribuzione uniforme)

$$P(E_i) = \frac{t}{a}, \;\; i = 1, 2, \ldots, n,$$

ovvero la probabilità è il rapporto fra le ampiezze degli intervalli. Gli eventi E_i si possono ritenere indipendenti ed equiprobabili. Se definiamo $Y_n = \sum_{i=1}^{n} |E_i|$, ovviamente si ha

$$Y_n \sim \mathbf{B}(n, \frac{t}{a}), \;\; \forall n \in \mathbb{N}.$$

Al variare di n si ha una successione di n.a. $\{Y_n\}$ con distribuzione binomiale.

Supponiamo adesso di far tendere $a \to \infty$ e di mantenere fissa l'ampiezza t di I. In questo modo ad ogni prova la probabilità di successo t/a diventa sempre più piccola e i successi diventano *eventi rari*.

Allo stesso tempo facciamo tendere $n \to \infty$ in modo tale che il rapporto n/a si mantenga costante ($\frac{n}{a} = cost = \alpha$).

Per la previsione di Y_n si ha

$$\mathbb{P}(Y_n) = n\frac{t}{a} = \alpha t = cost = \lambda, \ \ \forall n\,.$$

Si ottiene

$$\lim_{n\to\infty} P(Y_n = h) = \lim_{n\to\infty} \tbinom{n}{h} (\tfrac{t}{a})^h (1 - \tfrac{t}{a})^{n-h} =$$

$$= \lim_{n\to\infty} \tfrac{n(n-1)\cdots(n-h+1)}{h!} (\tfrac{\lambda}{n})^h (1 - \tfrac{\lambda}{n})^{n-h} = \tfrac{\lambda^h}{h!} e^{-\lambda}\,.$$

Pertanto: $\mathbf{B}(n, \frac{\lambda}{n}) \to_{n\to\infty} \mathcal{P}(\lambda)\,,$
cioè, la distribuzione Binomiale (sotto certe condizioni) si può approssimare con la distribuzione di Poisson.

Esempio 13 In una città esiste un parco autobus costituito da 1000 unità. Da considerazioni statistiche si sa che in media il numero di autobus guasti che si osservano ogni mattina è pari a 2. Indichiamo con X il numero aleatorio di autobus guasti in una mattina specifica. Se, per ogni $i = 1, 2, \ldots, 1000$, definiamo l'evento E_i = "l'i-mo autobus è difettoso", possiamo ritenere gli E_i equiprobabili e indipendenti, con $p = P(E_i)$. Si ha $X \sim \mathbf{B}(1000, p)$.
Calcolare p e $P(X > 1)$.

Numeri Aleatori Continui

Sia $\mathcal{A}$ una partizione dell'evento certo Ω, con $card(A) > card(\mathbb{N})$. Si può dimostrare che in questo caso non è possibile attribuire a tutti gli eventi di $\mathcal{A}$ probabilità positiva, ma al più ad un sottoinsieme $\mathcal{B} \subset \mathcal{A}$, con $card(B) = card(\mathbb{N})$.
In queste situazioni si possono presentare numeri aleatori che assumono valori in un insieme con cardinalità del continuo. Un n. a. X di questo tipo può essere visto come una funzione

$$X : \Omega \to \mathbb{R}$$

che ad ogni caso possibile $\omega \in \Omega$ assegna un valore $X(\omega) = x \in \mathbb{R}$.
Ad esempio: X può essere l'istante di tempo in cui si guasta un determinato dispositivo.
In questo genere di applicazioni tipicamente i valori possibili di X sono (per ragioni di carattere matematico) tutti di probabilità nulla, cioè

$$P(X = x) = 0\,, \;\; \forall\, x\,.$$

Distribuzioni (assolutamente) continue. Un numero aleatorio X si dice continuo se

(i) $P(X = x) = 0, \forall x \in \mathbb{R}$ *(non esistono probabilità concentrate, tutti gli eventi hanno probabilità nulla)*;

(ii) esiste una funzione reale $f \geq 0$ (non negativa) integrabile secondo Riemann, tale che per ogni sottoinsieme $A \subseteq R$, misurabile secondo Peano-Jordan, si ha
$$P(A) = P(X \in A) = \int_A f(x)dx \, .$$

La funzione f si chiama *densità di probabilità* del n.a. X.

Osservazioni. Se A è un generico intervallo $[a, b]$ (limitato o non), la (ii) diventa

$$P(X \in [a, b]) = \int_a^b f(x)dx \, ,$$

cioè $P(X \in [a,b])$ coincide con l'area sottesa al diagramma di $f(x)$ nell'intervallo (a,b).
In particolare se $A = (-\infty, +\infty)$ otteniamo

$$P(\Omega) = P(X \in (-\infty, +\infty)) = \int_{-\infty}^{+\infty} f(x)dx = 1. \tag{59}$$

Cioè l'area totale sotto la curva è 1.

Da un punto di vista meccanico la densità di probabilità si può interpretare come la densità di massa con cui una massa unitaria è diffusa sull'asse reale.

Vediamo il legame che sussiste tra la densità di probabilità e la funzione di ripartizione. Data una densità di probabilità $f(x)$, ricordando che

$$F(x) = P(X \leq x) = P(X \in]-\infty, x]),$$

si ha

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

cioè la f.d.r. calcolata in x rappresenta l'area sotto la curva da $-\infty$ a x.

Viceversa data una funzione di ripartizione $F(x)$, se $F(x)$ è derivabile in x, si ha

$$F'(x) = f(x).$$

Quindi una f.d.r. nel caso di distribuzioni continue soddisfa le seguenti proprietà

1. $0 \leq F(x) \leq 1$, $\forall x \in \mathbb{R}$;

2. $F'(x) \geq 0$ ($F(x)$ è non decrescente).

Se abbiamo la f.d.r. $F(x)$ e si vuole calcolare la probabilità di $A = [a, b]$, cioè $P(a \leq X \leq b)$, basta osservare che

$$P(A) = \int_a^b f(x)dx = F(b) - F(a).$$

Sia $g(x)$ una funzione reale di variabile reale integrabile secondo Riemann, per vedere se essa rappresenta una densità di probabilità occorre provare che

1. $g(x) \geq 0\,,\ \ \forall\, x \in \mathbb{R}$;

2. $\int_{-\infty}^{+\infty} g(x)dx = 1.$

Se $g(x) \geq 0$, $\forall\, x \in \mathbb{R}$ e risulta

$$\int_{-\infty}^{+\infty} g(x)dx = k > 0\,,$$

la seguente funzione

$$f(x) = \frac{1}{k} \cdot g(x),\ \forall\, x \in \mathbb{R}\,,$$

ottenuta normalizzando la funzione g, è una densità di probabilità.

Distribuzione Uniforme

Un n.a. continuo X ha una distribuzione uniforme in un intervallo $[a, b]$, in simboli

$$X \sim \mathbf{U}([a, b]) \,,$$

se ha la seguente densità di probabilità

$$f(x) = \begin{cases} k > 0 & \text{se } x \in [a, b], \\ 0 & \text{altrove.} \end{cases}$$

La costante k si determina osservando che

$$\int_{-\infty}^{+\infty} f(x)dx = \int_a^b kdx = k(b-a) = 1 \,,$$

da cui segue

$$k = \frac{1}{b-a} \,.$$

Quindi, se $X \sim \mathbf{U}([a, b])$, si ha

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b], \\ 0 & \text{altrove.} \end{cases}$$

Dato un intervallo $I = (c, d)$ contenuto in $[a, b]$ si prova che la $P(X \in I)$ dipende solo dall'ampiezza dell'intervallo. Infatti se $l = d - c$ è l'ampiezza di I, si ottiene

$$P(X \in I) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a} = \frac{l}{b-a} \cdot$$

Pertanto dati due intervalli I, J contenuti in $[a, b]$, rispettivamente di ampiezza l e λ, si ha

$$P(X \in I) = P(X \in J) \Leftrightarrow l = \lambda.$$

La funzione di ripartizione di $X \sim \mathbf{U}([a, b])$ è definita da

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & \text{se } x < a, \\ \frac{x-a}{b-a}, & \text{se } x \in [a, b], \\ 1, & \text{se } x > b. \end{cases}$$

Esempio 14 Se $X \sim \mathbf{U}([0,1])$, la densità di probabilità è data da

$$f(x) = \begin{cases} 1, & \text{se } x \in [0,1], \\ 0, & \text{altrove,} \end{cases}$$

e la f.d.r. diventa

$$F(x) = \begin{cases} 0, & \text{se } x < 0, \\ x, & \text{se } x \in [0,1], \\ 1, & \text{se } x > 1. \end{cases}$$

Se consideriamo l'intervallo $I = [\frac{1}{2}, \frac{2}{3}]$, è facile verificare che

$$P(I) = P(X \in I) = \tfrac{2}{3} - \tfrac{1}{2} = \tfrac{1}{6}.$$

Se inoltre consideriamo l'intervallo $J = [\frac{1}{2}, \frac{3}{2}]$ si ha

$$P(J) = 1 - \tfrac{1}{2} = \tfrac{1}{2}\,.$$

Previsione. Dato un n.a. continuo X con densità di prob. $f(x)$, la previsione di X si definisce nel modo seguente

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

sotto la condizione

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

cioè che l'integrale sopra esista e sia finito.
Se $X \sim \mathbf{U}([a, b])$ si ha

$$\mathbb{P}(X) = \textstyle\int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \,.$$

Varianza. Dato un n.a. continuo X con densità di prob. $f(x)$, ponendo $\mathbb{P}(X) = m$, la varianza di X è definita da

$$Var(X) = \mathbb{P}[(X-m)^2] = \int_{-\infty}^{+\infty} (x-m)^2 f(x) dx \,,$$

sotto la condizione che l'integrale a secondo membro esista e sia finito.

Deviazione standard. Ricordiamo che lo scarto quadratico medio (o scarto standard o deviazione standard) è dato da: $\sigma_X = \sqrt{Var(X)}\,.$

Se $X \sim \mathbf{U}([a,b])$, si ha

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{+\infty}(x-m)^2 f(x)dx =$$

$$= \int_{-\infty}^{+\infty} x^2 f(x)dx - m^2 = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx - (\frac{a+b}{2})^2 =$$

$$= \cdots = \frac{(b-a)^2}{12}\,,$$

e quindi: $\sigma_X = \frac{b-a}{2\sqrt{3}}\,.$

Proprietà della previsione. Sia X un n.a. continuo ed Y un n.a. definito da

$$Y = h(X)\,,$$

dove h è una funzione definita nel codominio di X. Allora si può provare, sotto l'ipotesi di esistenza di

$\mathbb{P}(Y)$, che risulta

$$\mathbb{P}(Y) = \mathbb{P}[h(X)] = \int_{-\infty}^{+\infty} h(x)f(x)dx.$$

Ad esempio, se $h(X) = X^2$, si ha $\mathbb{P}[h(X)] = \mathbb{P}(X^2)$. Se $h(X) = (X - m)^2$, si ha

$$\mathbb{P}[h(X)] = \mathbb{P}[(X - m)^2] = \sigma_X^2 \,.$$

Linearità della previsione. Sia X un n.a. continuo. Consideriamo $Y = cX + d$, con $c \neq 0$. Allora

$$\mathbb{P}(Y) = \mathbb{P}(cX + d) = \int_{-\infty}^{+\infty}(cx + d)f(x)dx =$$

$$= c\int_{-\infty}^{+\infty} xf(x)dx + d\underbrace{\int_{-\infty}^{+\infty} f(x)dx}_{=1} = c\mathbb{P}(X) + d\,.$$

Osservazione. Le disuguaglianze di Markov e Cebicev valgono anche per n.a. continui. Inoltre

- $\min(X) \leq \mathbb{P}(X) \leq \max(X)$.

Distribuzione Esponenziale.

Un n.a. continuo X con densità di probabilità

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0, \\ 0 & \text{se } x < 0, \end{cases} \quad \lambda \in \mathbb{R}^+ \tag{60}$$

si dice che ha *distribuzione esponenziale* di parametro λ e si indica con $X \sim Exp(\lambda)$.

La distribuzione esponenziale viene utilizzata ad esempio quando X rappresenta

- il tempo di durata di un dispositivo (non soggetto ad usura);

- il tempo di attesa del verificarsi di un certo evento (arrivo di un cliente in una coda, arrivo di una telefonata).

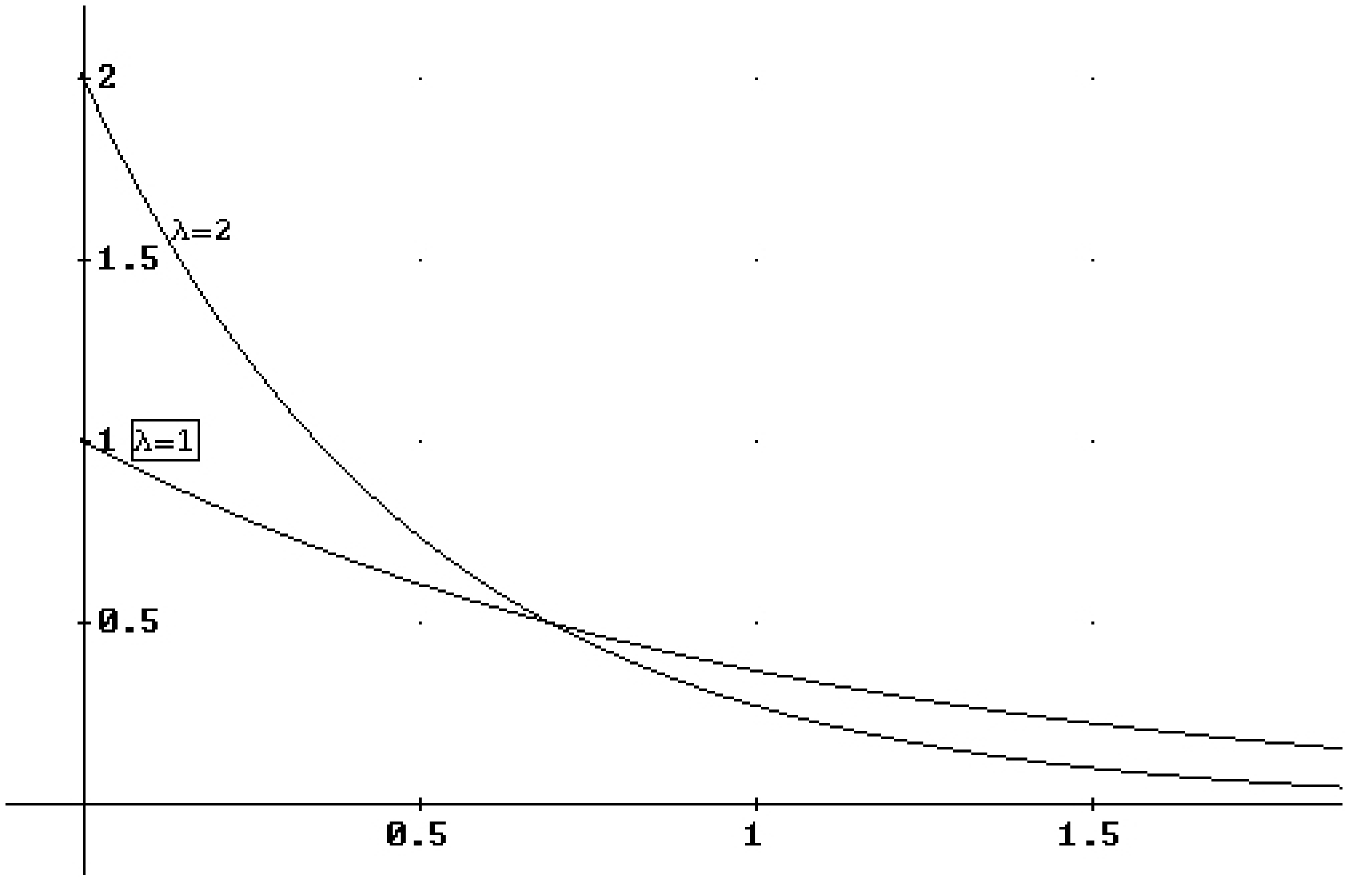


Figura 5: Esponenziale

L'area sotto la curva $y = f(x)$ al crescere del parametro λ si concentra sempre più verso l'origine.
Ricordiamo che l'area totale sotto la curva è uguale a 1. Infatti

$$\int_{-\infty}^{+\infty} f(x) = \int_0^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^{+\infty} = 1 \, .$$

La distribuzione esponenziale è l'analogo nel continuo della distribuzione geometrica. Infatti nel discreto il tempo di attesa può esser visto come il numero di prove necessarie per il verificarsi di un evento (numero di lanci di una moneta fino a quando per la prima volta esce testa).
La funzione di ripartizione è data da

$$F(x) = \begin{cases} \int_0^x \lambda e^{-\lambda t} dt\,, & \text{se } x > 0, \\ 0\,, & \text{se } x \leq 0. \end{cases}$$

Osservando che $\int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}$, si ottiene

$$F(x) = \begin{cases} 1 - e^{-\lambda x}\,, & \text{se } x > 0, \\ 0\,, & \text{se } x \leq 0. \end{cases}$$

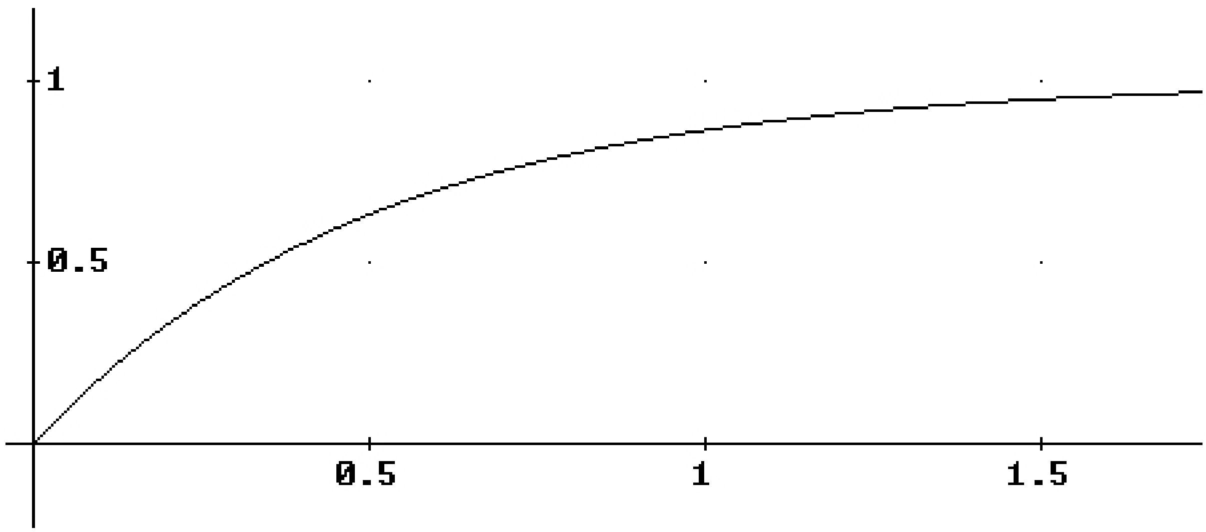


Figura 6: f.d.r. Exp

La funzione $S(x) = 1 - F(x) = P(X > x)$, detta funzione di sopravvivenza, è data da

$$S(x) = \begin{cases} e^{-\lambda x}, & \text{se } x > 0, \\ 1, & \text{se } x \leq 0, \end{cases}$$

La previsione è

$$\mathbb{P}(X) = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = \cdots = \frac{1}{\lambda},$$

mentre

$$\mathbb{P}(X^2) = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \cdots = \frac{2}{\lambda^2}\,.$$

Quindi la varianza e lo scarto sono rispettivamente

$$Var(X) = \mathbb{P}(X^2) - [\mathbb{P}(X)]^2 = \frac{1}{\lambda^2}\,,$$
$$\sigma_X = \frac{1}{\lambda}.$$

Lo scarto quadratico medio coincide con la previsione.

Proprietà di Assenza di memoria. Un numero aleatorio continuo e non negativo X ha distribuzione esponenziale se e solo se vale la seguente proprietà (detta di assenza di memoria)

$$P(X > x_0 + x | X > x_0) = P(X > x), \ \ \forall x_0, x \in \mathbb{R}_0^+. \tag{61}$$

Se X rappresenta il tempo (aleatorio) fino al guasto di un dispositivo, la proprietà di assenza di memoria ha il seguente significato: supposto che il dispositivo non si guasti sino al tempo x_0, la probabilità che non si guasti per un ulteriore tempo x è la stessa che il dispositivo non si guasti nell'intervallo $[0, x]$.

Tale proprietà è valida per le apparecchiature che, durante il loro funzionamento, non sono soggette ad usura (o, più realisticamente, quando l'usura è trascurabile).

dim.($\Rightarrow$) *Hp)* $X \sim Exp(\lambda)$; *Th)* vale la (61).

$$P(X > x_0 + x | X > x_0) = \frac{P(X > x_0 + x, X > x_0)}{P(X > x_0)} =$$

$$= \frac{P(X > x_0 + x)}{P(X > x_0)} = \frac{S(x_0 + x)}{S(x_0)} = \frac{e^{-\lambda(x_0 + x)}}{e^{-\lambda x_0}} = e^{-\lambda x} =$$
$$= S(x) = P(X > x) \,.$$

($\Leftarrow$)*Hp)* vale la (61); *Th)* $X \sim Exp(\lambda)$.
Da quanto visto nella precedente dimostrazione la proprietà di assenza di memoria si può scrivere anche come:

$$\frac{S(x_0 + x)}{S(x_0)} = S(x) \,,$$

cioè

$$S(x + x_0) = S(x) S(x_0) \,.$$

Essendo la funzione di sopravvivenza definita come $1 - F(x)$, con $F(x)$ crescente, allora $S(x)$ è positiva e decrescente e quindi

$$S(x) > 0, \;\; S'(x) < 0, \;\; \forall x \in R.$$

Osserviamo che

$$\frac{S'(x+x_0)}{S(x+x_0)} = \frac{S(x_0)S'(x)}{S(x_0)S(x)} = \frac{S'(x)}{S(x)} = -\lambda, \quad \lambda > 0,$$

quindi

$$\begin{array}{l} D[ln(S(x))] = \frac{S'(x)}{S(x)} = -\lambda \Rightarrow \\ ln(S(x)) = -\lambda x + k \, , \end{array}$$

allora

$$S(x) = e^{-\lambda x} e^{k} \, .$$

Essendo X un n.a. non negativo, si ha $S(0) = 1$, per cui $e^k = 1$. Allora

$$S(x) = e^{-\lambda x} \, ,$$

ovvero $X \sim Exp(\lambda)$.

Distribuzione normale standard

Un n.a. continuo X, con densità di probabilità

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}, \tag{62}$$

si dice che ha *distribuzione normale standard* (di parametri 0,1) e si indica con $X \sim N_{0,1} = N$. La densità $f(x)$ si indica con $N(x)$, mentre la funzione di ripartizione $F(x)$ si indica con $\Phi(x)$. Di tale funzione non è possibile dare un'espressione, ma si possono cercare soltanto alcuni valori riportati su apposite tavole.

Alcune proprietà:

1. il diagramma della densità ha un andamento a forma di campana (con il massimo nell'origine e due flessi in $x = -1$, $x = 1$) ed è simmetrico rispetto all'asse y, cioè $N(x)$ è una funzione pari ($N(-x) = N(x)$);

2. dalla simmetria di $N(x)$, per ogni $x \in \mathbb{R}$ si ha

$\Phi(-x) = 1 - \Phi(x)$, e quindi

$$P(|X| \leq x) = P(-x \leq X \leq x) = \int_{-x}^{x} N(t)dt =$$

$$= \Phi(x) - \Phi(-x) = 2\Phi(x) - 1\,;$$

$$P(|X| > x) = 1 - P(|X| \leq x) = 2[1 - \Phi(x)]\,;$$

3. in particolare

$$\Phi(1) \simeq 0.8413\,, \quad \Phi(2) \simeq 0.9772\,, \quad \Phi(3) \simeq 0.9987\,,$$

e quindi

$$P(|X| \leq 1) = 2\Phi(1) - 1 \simeq 0.6826\,;$$

$$P(|X| \leq 2) = 2\Phi(2) - 1 \simeq 0.9544\,;$$

$$P(|X| \leq 3) = 2\Phi(3) - 1 \simeq 0.9974\,.$$

Si può verificare che

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} xN(x)dx = \cdots = 0\,,$$

$$\mathbb{P}(X^2) = \int_{-\infty}^{+\infty} x^2 N(x) dx = \cdots = 1 \,,$$

e quindi:

$$Var(X) = \mathbb{P}(X^2) = 1 \,.$$

In generale, si dice che X ha una distribuzione normale di parametri m, σ, con $m \in \mathbb{R}, \sigma > 0$, se la densità di X ha la seguente forma:

$$f(x) = N_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \,, \quad x \in \mathbb{R} \,. \qquad (63)$$

In simboli, si scrive: $X \sim N_{m,\sigma}$. La funzione di ripartizione si indica con $\Phi_{m,\sigma}(x)$.

Il diagramma della densità ha un andamento a forma di campana (con il massimo in $x = m$ e due flessi in $x = m - \sigma$, $x = m + \sigma$) ed è simmetrico rispetto alla retta $x = m$.
Se $Y = aX + b$, con $X \sim N_{m,\sigma}$, $a > 0$, indicando con G la funzione di ripartizione di Y e g la sua densità, si

ha:

$$G(y) = P(Y \leq y) = P(X \leq \frac{y-b}{a}) = \Phi_{m,\sigma}(\frac{y-b}{a}),$$

e quindi

$$g(y) = G'(y) = \Phi_{m,\sigma}(\frac{y-b}{a}) \cdot \frac{1}{a} = \cdots = N_{m_Y,\sigma_Y}(y),$$

dove

$$m_Y = am + b, \quad \sigma_Y = a\sigma. \tag{64}$$

Si può dimostrare che, se $a < 0$, risulta $Y \sim N_{m_Y,\sigma_Y}$, con

$$m_Y = am + b, \quad \sigma_Y = -a\sigma. \tag{65}$$

In altri termini, se dal n.a. X, con distribuzione normale, si passa al n.a. $Y = aX + b$, con $a \neq 0$, la distribuzione rimane di tipo normale, con i parametri che cambiano come indicato nella (64), oppure (65). In particolare, se $Y = \frac{X-m}{\sigma}$, si ha

$$m_Y = 0, \quad \sigma_Y = 1, \tag{66}$$

cioè la distribuzione diventa normale standard. Allora, tenendo conto che, se $Y = \frac{X-m}{\sigma}$, si ha $\mathbb{P}(Y) = 0$, $\sigma_Y = 1$, e che $X = \sigma Y + m$, si ottiene

$$\mathbb{P}(X) = \mathbb{P}(\sigma Y + m) = m, \;\; \sigma_X^2 = Var(\sigma Y + m) = \sigma^2 \,.$$

Pertanto i parametri m, σ sono rispettivamente la previsione e lo scarto quadratico medio. Lo stesso risultato si può ottenere con calcoli diretti, verificando che

$$\mathbb{P}(X) = \int_{-\infty}^{+\infty} x N_{m,\sigma}(x) dx = \cdots = m \,,$$

$$Var(X) = \int_{-\infty}^{+\infty} (x-m)^2 N_{m,\sigma}(x) dx = \cdots = \sigma^2 \,.$$

Se $X \sim N_{m,\sigma}$, osservando che

$$(X \leq x) \Longleftrightarrow (\frac{X-m}{\sigma} \leq \frac{x-m}{\sigma}) \,,$$

e che

$$\frac{X-m}{\sigma} \; \sim \; N \,,$$

si ottiene

$$\Phi_{m,\sigma}(x) = P(X \leq x) = P(\tfrac{X-m}{\sigma} \leq \tfrac{x-m}{\sigma}) = \Phi(\tfrac{x-m}{\sigma})\,.$$

Inoltre, per ogni $k > 0$, si ha

$$P(|X - m| \leq k\sigma) = P(m - k\sigma \leq X \leq m + k\sigma) =$$

$$= \Phi_{m,\sigma}(m + k\sigma) - \Phi_{m,\sigma}(m - k\sigma) = \Phi(k) - \Phi(-k) =$$

$$= 2\Phi(k) - 1\,.$$

Come mostrano le formule precedenti, utilizzando le tavole della distribuzione normale standard è possibile calcolare i valori di una distribuzione normale con parametri m, σ arbitrari.

Esercizio. Un sistema S è costituito da due dispositivi A e B in parallelo funzionanti simultaneamente (e quindi S funziona finchè almeno uno dei due dispositivi funziona). Siano X e Y i tempi aleatori di durata di A e B, rispettivamente, e supponiamo che le loro densità siano:

$$f_1(x) = e^{-x}\,,\ x \geq 0\,,\ \ f_2(y) = 2e^{-2y}\,,\ y \geq 0\,,$$

con $f_1(x) = f_2(y) = 0$ per $x < 0$, $y < 0$. Si supponga inoltre che valga la condizione

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)\,,\ \ \forall\,(x, y)\,.$$

(i) calcolare la densità di probabilità $g(z)$ del tempo aleatorio Z fino al guasto di S.
(ii) è possibilite calcolare, con i dati del problema, la probabilità che A si guasti prima di B?

Soluzione.
(i) ricordando che, per ogni $\lambda > 0$, $x \geq 0$, si ha

$$\int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}\,;$$

inoltre, osservando che $Z = max\ \{X, Y\}$, segue $(Z \leq z) = (X \leq z, Y \leq z)$ e quindi

$$P(Z \leq z) = P(X \leq z, Y \leq z) = P(X \leq z)P(Y \leq z) =$$

$$= (1 - e^{-z})(1 - e^{-2z}) = 1 - e^{-z} - e^{-2z} + e^{-3z}\,.$$

Allora, derivando rispetto a z, si ottiene

$$g(z) = e^{-z} + 2e^{-2z} - 3e^{-3z}\,.$$

Nota: la densità $g(z)$ *è una combinazione lineare, con coefficienti* $1, 1, -1$*, di tre densità esponenziali di parametri* $1, 2, 3$.

(ii) per quanto riguarda la seconda domanda, la risposta è si e no. Si ha

$$p = P(X < Y) = P[(X, Y) \in \{(x, y) : x < y\}] = \cdots = \frac{1}{3}\,.$$

I dettagli saranno esaminati ampliando il discorso al caso di distribuzioni multiple (per vettori aleatori).

Vettori Aleatori

In molti esperimenti aleatori, indicando con Ω l'insieme dei possibili risultati, al generico risultato dell'esperimento, $\omega \in \Omega$, sono associati n numeri reali $x_1, \ldots, x_n$, con $n \geq 2$, che costituiscono i valori di n numeri aleatori $X_1, \ldots, X_n$. Tali n. a. sono le componenti di un vettore aleatorio $X = (X_1, \ldots, X_n)$, che può essere visto come una funzione definita su Ω a valori in $\mathbb{R}^n$, cioè

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longrightarrow x = X(\omega)\,. \end{aligned}$$

Due casi importanti da considerare sono i v. a. discreti e v.a. continui.

Vettori aleatori discreti. Un vettore aleatorio $X = (X_1, X_2, \ldots, X_n)$ si dice discreto se esiste un insieme finito o numerabile $C \subset \mathbb{R}^n$ tale che

- $P(X = x) > 0\,, \;\; \forall x \in C,$

- $P(X = x) = 0\,, \;\; \forall x \notin C,$

dove, ponendo $x = (x_1, \ldots, x_n)$, l'evento $(X = x)$ rappresenta l'evento

$$(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)\,.$$

Vettori aleatori continui. Un vettore aleatorio $X = (X_1, X_2, \ldots, X_n)$ si dice continuo se

- $P(X = x) = 0\,, \;\; \forall x \in \mathbb{R}^n,$
- $\exists f : \mathbb{R}^n \to R$ tale che

  (i) $f(x) \geq 0,\; \forall x \in \mathbb{R}^n;$

  (ii) $\forall\, A \subseteq \mathbb{R}^n$, misurabile secondo Peano - Jordan, si ha

  $P(X \in A) = P(A) = \int_A f(x)dx =$

  $= \int \cdots \int_A f(x_1, \ldots, x_n) dx_1 \cdots dx_n\,.$

La funzione $f(x)$ si chiama densità di probabilità congiunta del v. a. X.

Proprietà di normalizzazione:

$\int_{\mathbb{R}^n} f(x)dx =$

$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n)dx_1 \cdots dx_n = 1\,.$

La funzione di ripartizione congiunta di $X = (X_1, \ldots, X_n)$ è definita nel seguente modo:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)\,.$$

Nel caso continuo si ha:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)dt_1 \cdots dt_n\,.$$

Analizziamo in dettaglio il caso discreto, con $n = 2$. Per semplicità di notazione, indichiamo con (X, Y) il v.a. (X_1, X_2). Va osservato che tale notazione può essere utilizzata anche nel caso in cui $n > 2$, indicando con X e Y due sottovettori del vettore aleatorio $(X_1, \ldots, X_n)$.

Distribuzioni marginali. Sia

$$X \in C_x\,,\;\; Y \in C_y\,,\;\; (X,Y) \in C \subseteq C_x \times C_y\,.$$

Ricordiamo che C è l'insieme (al più numerabile) dei punti di $\mathbb{R}^2$ che hanno probabilità positiva. Quindi per ogni coppia $(x_h, y_k) \in C$ si ha

$$P(X = x_h, Y = y_k) = p_{x_h, y_k} > 0\,.$$

Fissato un punto $x_h \in C_x$, osservando che

$$\Omega = \bigvee_{y_k \in C_y} (Y = y_k)\,,$$

possiamo decomporre l'evento $(X = x_h)$ nel seguente modo

$$(X = x_h) = (X = x_h) \wedge \Omega =$$

$$= (X = x_h) \wedge [\textstyle\bigvee_{y_k \in C_y}(Y = y_k)] =$$

$$= \textstyle\bigvee_{y_k \in C_y}(X = x_h, Y = y_k)\,.$$

Quindi, $\forall\, x_h \in C_x$, si ha

$$P(X = x_h) = p_{x_h} = \sum_{y_k \in C_y} P(X = x_h, Y = y_k) =$$

$$= \sum_{y_k} p_{x_h,y_k}\,;$$ (distribuzione marginale di X).

In modo analogo si ottiene

$$P(Y = y_k) = p_{y_k} = \sum_{x_h \in C_x} P(X = x_h, Y = y_k) =$$

$$= \sum_{x_h} p_{x_h,y_k}\,;$$ (distribuzione marginale di Y).

Distribuzioni marginali condizionate.

$$p_{x_h|y_k} = P(X = x_h | Y = y_k) =$$
$$= \frac{P(X{=}x_h, Y{=}y_k)}{P(Y{=}y_k)} = \frac{\overbrace{p_{x_h,y_k}}^{congiunta}}{\underbrace{p_{y_k}}_{marginale}}\,.$$

La distribuzione $\{p_{x_h|y_k}\,,\ x_h \in C_X\}$ si chiama distribuzione marginale di X condizionata al valore fissato y_k di Y.

In maniera analoga, la distribuzione $\{p_{y_k|x_h}\,,\ y_k \in C_Y\}$ si chiama distribuzione marginale di Y condizionata al valore fissato x_h di X, ovvero

$$p_{y_k|x_h} = P(Y = y_k|X = x_h) =$$
$$= \frac{P(X=x_h,Y=y_k)}{P(X=x_h)} = \frac{\overbrace{p_{x_h,y_k}}^{congiunta}}{\underbrace{p_{x_h}}_{marginale}}\,.$$

Dalle ultime relazioni, per il teorema delle probabilità composte, si ottiene:

$$P(X = x_h, Y = y_k) = P(Y = y_k|X = x_h)P(X = x_h) =$$

$$= P(X = x_h|Y = y_k)P(Y = y_k)\,.$$

ovvero

$$p_{x_h,y_k} = p_{x_h|y_k} \cdot p_{y_k} = p_{y_k|x_h} p_{x_h}.$$

Osserviamo che, in generale, risulta

$$p_{x_h,y_k} \neq p_{x_h} \cdot p_{y_k}\,.$$

Indipendenza stocastica. I numeri aleatori X, Y si dicono stocasticamente indipendenti (in breve, indipendenti) se, $\forall\,(x_h, y_k)$, vale

$$P(X = x_h, Y = y_k) = P(X = x_h) \cdot P(Y = y_k)\,,$$

ovvero la distribuzione congiunta è data dal prodotto delle marginali

$$p_{x_h, y_k} = p_{x_h} \cdot p_{y_k},\ \forall\,(x_h, y_k).$$

Quindi, se X, Y sono indipendenti, le distribuzioni condizionate coincidono con le marginali

$$p_{x_h | y_k} = p_{x_h}, \quad p_{y_k | x_h} = p_{y_k}\,.$$

Esempio. Si lancia due volte un dado, definendo

$X =$ risultato del primo lancio;

$Y =$ risultato del secondo lancio.

Ovviamente, X, Y sono indipendenti e quindi, per ogni coppia $(m, n) \in \{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}$, si ha

$$P(X = m, Y = n) = P(X = m)P(Y = n) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

$Y \setminus X$	1	2	3	4	5	6	$P(Y = n)$
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$P(X = m)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

Esempio. Due estrazioni senza restituzione da un'urna contenente cinque palline numerate da 1 a 5. I numeri aleatori

$$X = \text{risultato della prima estrazione,}$$
$$Y = \text{risultato della seconda estrazione,}$$

non sono indipendenti. Infatti, ad esempio

$$P(X = 2, Y = 1) = \tfrac{1}{20}\,,$$

$$P(X = 2)P(Y = 1) = \tfrac{1}{5} \cdot \tfrac{1}{5} = \tfrac{1}{25} \neq \tfrac{1}{20}\,.$$

Esempio. Siano X, Y due n. a. indipendenti con distribuzione di Poisson, rispettivamente di parametri λ_1 e λ_2, ovvero

$$X \sim \mathcal{P}(\lambda_1), \quad Y \sim \mathcal{P}(\lambda_2).$$

Calcoliamo la distribuzione di probabilità del n. a.

$Z = X + Y$. Osserviamo che, fissato $n \in \mathbb{N}$, si ha

$$P(Z = n) = P[\bigvee_{i=0}^{n}(X = i, Y = n - i)] =$$

$$= \sum_{i=0}^{n} P(X = i, Y = n - i) =$$

$$= \sum_{i=0}^{n} e^{-\lambda_1} \frac{\lambda_1^i}{i!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n-i}}{(n-i)!} = \cdots =$$

$$= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!} .$$

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

Inoltre, si può verificare che

$$X|(Z = n) \sim \mathcal{B}(n, \tfrac{\lambda_1}{\lambda_1+\lambda_2}) ,$$

$$Y|(Z = n) \sim \mathcal{B}(n, \tfrac{\lambda_2}{\lambda_1+\lambda_2}) .$$

Teorema. Se X ed Y sono indipendenti, si ha: $Cov(X, Y) = 0$.

Dim.: Supponiamo che, $\forall\,(x_h, y_k) \in C$, sia

$$P(X = x_h, Y = y_k) = P(X = x_h)P(Y = y_k) .$$

Allora, segue

$$\mathbb{P}(XY) = \sum_{x_h} \sum_{y_k} x_h y_k p_{x_h, y_k} = \sum_{x_h} \sum_{y_k} x_h y_k p_{x_h} p_{y_k} =$$

$$= (\sum_{x_h} x_h p_{x_h})(\sum_{y_k} y_k p_{y_k}) = \mathbb{P}(X)\mathbb{P}(Y)\,,$$

e quindi: $Cov(X, Y) = 0$.
Osserviamo che il viceversa non vale, come mostra il seguente controesempio.

Esempio. Si consideri il seguente vettore aleatorio (X, Y), con la distribuzione congiunta riportata nella tabella:

$Y \setminus X$	-1	0	1
-1	a	/	a
0	/	b	/
1	a	/	a

Si ha

$$C = \{(-1, -1),\, (-1, 1),\, (0, 0),\, (1, -1),\, (1, 1)\}\,,$$

con $P(X=0,Y=0)=b$ e $P(X=x,Y=y)=a$ negli altri casi. Ovviamente, deve essere:

$$4a+b=1\,,\;\; a\geq 0\,,\; b\geq 0\,.$$

Come si può verificare, si ha

$$X\in\{-1,0,1\}\,,\;\; Y\in\{-1,0,1\}\,,\;\; XY\in\{-1,0,1\}\,,$$

con

$$P(X=-1)=P(Y=-1)=P(XY=-1)=2a\,,$$

$$P(X=0)=P(Y=0)=P(XY=0)=b\,,$$

$$P(X=1)=P(Y=1)=P(XY=1)=2a\,.$$

Pertanto X,Y ed XY hanno la stessa distribuzione di probabilità. Inoltre

$$\mathbb{P}(X)=\mathbb{P}(Y)=\mathbb{P}(XY)=0\,,$$

e quindi $Cov(X, Y) = 0$, ovvero X, Y sono incorrelati. Però X ed Y non sono indipendenti, in quanto risulta

$$P(X = x, Y = y) \neq P(X = x)P(Y = y)\,,$$

ad esempio:

$$P(X = 0, Y = 0) = b \neq P(X = 0)P(Y = 0) = b{\cdot}b = b^2\,.$$

Vettori aleatori continui: distribuzioni marginali e condizionate.

Dato un vettore aleatorio continuo $(X_1, \ldots, X_n)$, sia $f(x_1, \ldots, x_n)$ la sua densità congiunta. Le densità marginali $f_1(x_1), \ldots, f_n(x_n)$ dei n. a. $X_1, \ldots, X_n$ sono date dalle seguenti formule:

$$f_i(x_i) =$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n\,,$$

$$i = 1, \ldots, n\,.$$

Ovvero, per calcolare $f_i(x_i)$ si integra $f(x_1,\ldots,x_n)$ rispetto alle variabili $x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n$.
Consideriamo in particolare il caso $n=2$, indicando con (X,Y) il v. a. (X_1,X_2) e con $f(x,y)$ la densità congiunta. Si ha

$$f_1(x)=\int_{-\infty}^{+\infty} f(x,y)dy\,,\quad f_2(y)=\int_{-\infty}^{+\infty} f(x,y)dx\,.$$

Le densità condizionate (di $Y|x$ ed $X|y$), per fissati valori x,y ed assumendo $f_1(x)>0$, $f_2(y)>0$, sono definite nel seguente modo

$$f_2(y|x)=\frac{f(x,y)}{f_1(x)}\,,\qquad f_1(x|y)=\frac{f(x,y)}{f_2(y)}\,.$$

Pertanto

$$f(x,y)=f_1(x)f_2(y|x)=f_2(y)f_1(x|y)\,.$$

Se risulta

$$f(x,y)=f_1(x)f_2(y)\,,\ \ \forall\,(x,y)$$

i n. a. si dicono stocasticamente indipendenti e in questo caso si ha

$$f_2(y|x) = f_2(y)\,,\ \forall\, y\,; \quad f_1(x|y) = f_1(x)\,,\ \forall\, x\,,$$

cioè le densità condizionate coincidono con le densità marginali.
Osserviamo che la relazione di indipendenza tra X e Y può essere definita anche richiedendo che valga

$$F(x,y) = F_1(x)F_2(y)\,,\ \ \forall\,(x,y)\,,$$

cioè

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)\,,\ \ \forall\,(x,y)\,.$$

Come già visto nel caso discreto, si può dimostrare che, se X e Y sono indipendenti, segue che sono incorrelati, mentre il viceversa non vale. Infatti, assumendo

$f(x,y) = f_1(x)f_2(y)\,,\ \ \forall\,(x,y)$, si ottiene

$$\mathbb{P}(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = \cdots$$

$$= (\int_{-\infty}^{+\infty} xf_1(x)dx)(\int_{-\infty}^{+\infty} yf_2(y)dy) = \mathbb{P}(X)\mathbb{P}(Y)\,,$$

e quindi $Cov(X,Y) = 0$. Per mostrare attraverso un controesempio che il viceversa non vale, introduciamo la distribuzione uniforme su un insieme $A \subset \mathbb{R}^2$, limitato e misurabile.
Si dice che (X,Y) ha distribuzione uniforme su A, in simboli

$$(X,Y)\ \sim\ U(A)\,,$$

se la densità congiunta assume un valore costante $k > 0$ su A ed è nulla altrove. Imponendo la condizione

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1\,,$$

ovvero

$$\int\int_A f(x,y)dxdy = 1\,,$$

si ottiene $k = \frac{1}{\mu(A)}$, dove $\mu(A)$ è l'area di A.

Esempio. Supponiamo che $(X, Y) \sim U(C)$, dove C è il cerchio di raggio 1 e centro nell'origine. Allora

$$f(x, y) = \frac{1}{\pi}, \quad (x, y) \in C,$$

con $f(x, y) = 0$ altrove. Si dimostra che

$$f_1(x) = \frac{2}{\pi}\sqrt{1 - x^2}, \quad x \in [-1, 1],$$

con $f_1(x) = 0$ altrove. Inoltre

$$f_2(y) = \frac{2}{\pi}\sqrt{1 - y^2}, \quad y \in [-1, 1],$$

con $f_2(y) = 0$ altrove. Allora $\mathbb{P}(X) = \mathbb{P}(Y) = 0$. Inoltre

$$\mathbb{P}(XY) = \int\int_C xyf(x, y)dxdy = \cdots = 0,$$

pertanto X e Y sono incorrelati. D'altra parte $f(x, y) \neq f_1(x)f_2(y)$, pertanto X e Y non sono indipendenti.

Somme di numeri aleatori.
Dato un vettore aleatorio continuo (X, Y), con densità $f(x, y)$, sia

$$Z = X + Y \,, \quad G(z) = P(Z \le z) \,, \; z \in \mathbb{R} \,.$$

Si ha

$$G(z) = P(X + Y \le z) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x, y) dy =$$

$$= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z} f(x, t - x) dt =$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f(x, t - x) dx \right) dt = \int_{-\infty}^{z} g(t) dt \,,$$

dove g è la densità di Z data da

$$g(z) = G'(z) = \int_{-\infty}^{+\infty} f(x, z - x) dx \,.$$

Caso notevole: se X e Y sono indipendenti, ovvero $f(x,y) = f_1(x)f_2(y)$, si ottiene

$$g(z) = \int_{-\infty}^{+\infty} f_1(x)f_2(z-x)dx\,,$$

che rappresenta l'*integrale di convoluzione* di f_1, f_2 e si indica con il simbolo: $g = f_1 * f_2$ (si noti che: $f_1 * f_2 = f_2 * f_1$).

Esempi:

(i) X, Y indipendenti e ugualmente distribuiti, con densità esponenziale di parametro λ.

$$f_1(t) = f_2(t) = \lambda e^{-\lambda t} = 0\,,\ t \geq 0\,;$$

con $f_1(t) = f_2(t) = 0\,,\ t < 0$. Per $z < 0$ si ha

$$g(z) = \int_{-\infty}^{+\infty} f_1(x)f_2(z-x)dx = \int_{-\infty}^{+\infty} 0\,dx = 0\,;$$

Per $z \geq 0$ si ha

$g(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z-x) dx =$

$= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx = \lambda^2 e^{-\lambda z} \int_0^z\, dx = \lambda^2 z e^{-\lambda z}\,.$

(*distribuzione Gamma, di parametri* c, λ, *con* $c = 2$)

(ii) $f_1 = N_{m_1,\sigma_1} = N_1\,,\ f_2 = N_{m_2,\sigma_2} = N_2$; si ha

$$g(z) = \int_{-\infty}^{+\infty} N_1(x) N_2(z-x) dx = \cdots = N_3(z)\,,$$

con

$$N_3 = N_{m_3,\sigma_3}\,, \quad m_3 = m_1 + m_2\,, \quad \sigma_3^2 = \sigma_1^2 + \sigma_2^2\,.$$

Pertanto, dalla convoluzione di due distribuzioni normali si ottiene una distribuzione ancora normale.

(iii) $X \sim U([0,a])\,,\ Y \sim U([0,a])$; si ha

$$g = U([0,a]) * U([0,a]) = T\,,$$

dove

$$T(z) = \begin{cases} \frac{z}{a^2}, & 0 \le z \le a, \\ \\ \frac{2a-z}{a^2}, & a < z \le 2a, \\ \\ 0, & altrove. \end{cases}$$

Teorema limite centrale.

Data una successione di numeri aleatori $X_1, \ldots, X_n, \ldots$, indipendenti ed ugualmente di- stribuiti, con $\mathbb{P}(X_i) = m$, Var$(X_i) = \sigma^2$, si consideri la successione delle medie aritmetiche

$$Y_1 = X_1\,,\ Y_2 = \frac{X_1 + X_2}{2}\,,\ Y_n = \frac{X_1 + \cdots + X_n}{n}\,,\ \ldots\,,$$

e quella delle medie aritmetiche ridotte $Z_1, \ldots, Z_n$. Ovviamente $\mathbb{P}(Y_n) = m$, Var$(Y_n) = \frac{\sigma^2}{n}$ e quindi $Z_n = \frac{Y_n - m}{\sigma/\sqrt{n}}$. Indicando con F_i la funzione di ripartizione di Z_i, la successione $F_1, \ldots, F_n, \ldots$ converge alla funzione di ripartizione (di una distribuzione normale standard) Φ, ovvero si ha

$$\lim_{n\to+\infty} F_n(z) = \lim_{n\to+\infty} P(Z_n \leq z) =$$

$$= \Phi(z) = \int_{-\infty}^{z} N(t)dt\,, \quad \forall\, z \in \mathbb{R}\,.$$

Rette di regressione

Dato un vettore aleatorio (X, Y), cerchiamo una retta di equazione $y = a + bx$ che meglio si adatti alla distribuzione di probabilità congiunta di (X, Y), ovvero che risulti più vicina possibile a tale distribuzione. Da un certo punto di vista, si potrebbe pensare di voler stimare Y mediante una funzione lineare $a + bX$, con i coefficienti a, b da determinare sulla base di un opportuno criterio. Un criterio ben noto in statistica è il metodo dei minimi quadrati che consiste nel cercare i valori a, b che rendono minima la previsione del numero aleatorio $(Y - a - bX)^2$. La retta che si ottiene si chiama retta di regressione di Y su X. Considerando il caso continuo e ponendo $\mathbb{P}[(Y-a-bX)^2]=g(a,b)$, se la densità congiunta è $f(x,y)$, si ha (applicando la linearità della previsione)

$$g(a,b)=\int\int_{\mathbb{R}^2}(y-a-bx)^2 f(x,y)dxdy=$$

$$\mathbb{P}(Y^2)+a^2+b^2\mathbb{P}(X^2)-2a\mathbb{P}(Y)-2b\mathbb{P}(XY)+2ab\mathbb{P}(X).$$

Uguagliando a zero le derivate parziali di $g(a,b)$ rispetto ad a,b (indicando con $m_1,m_2,\sigma_1,\sigma_2$ le previsioni e gli scarti standard di X e Y, e con ρ il coefficiente di correlazione) si ha

$$\begin{cases} \frac{\partial g}{\partial a}=2a-2m_2+2bm_1=0\,, \\ \\ \frac{\partial g}{\partial b}=2b(m_1^2+\sigma_1^2)-2(m_1m_2+\rho\sigma_1\sigma_2)+2am_1=0\,. \end{cases}$$

Ricavando a dalla prima equazione ($a=m_2-bm_1$) e risolvendo rispetto a b la seconda, si ottiene

$$a=m_2-\rho\frac{\sigma_2}{\sigma_1}m_1\,,\quad b=\rho\frac{\sigma_2}{\sigma_1}\,.$$

Pertanto, l'equazione della retta di regressione di Y su X è data da: $y = m_2 + \rho\frac{\sigma_2}{\sigma_1}(x - m_1)$.

Simmetricamente, l'equazione della retta di regressione di X su Y è: $x = m_1 + \rho\frac{\sigma_1}{\sigma_2}(y - m_2)$, che si può scrivere: $y = m_2 + \frac{1}{\rho} \cdot \frac{\sigma_2}{\sigma_1}(x - m_1)$.

Le equazioni delle due rette di regressione si possono anche scrivere, rispettivamente,

$$\frac{y - m_2}{\sigma_2} = \rho \cdot \frac{x - m_1}{\sigma_1}; \qquad \frac{x - m_1}{\sigma_1} = \rho \cdot \frac{y - m_2}{\sigma_2}.$$

Le due rette si incontrano nel punto di coordinate (m_1, m_2) e, nel caso $\rho = 0$, sono perpendicolari e di equazioni: $y = m_2\,, \;\; x = m_1\,.$
Se $|\rho| = 1$, le due rette coincidono ed hanno equazione

$$y = m_2 \pm \frac{\sigma_2}{\sigma_1}(x - m_1)\,,$$

($+$ oppure $-$, a seconda che sia $\rho = 1$, oppure $\rho = -1$).

Inferenza statistica

In molte applicazioni statistiche si studiano popolazioni in cui una o più caratteristiche numeriche sono incognite. Tali caratteristiche costituiscono quindi un numero (o un vettore) aleatorio, che possiamo indicare con Θ, a cui viene assegnata (sulla base dell'informazione iniziale) una distribuzione iniziale, in particolare una densità iniziale $\beta(\theta)$ nel caso continuo. Per ridurre l'incertezza sul parametro Θ si procede all'osservazione di un vettore (a priori aleatorio) $\mathbf{X} = (X_1, \ldots, X_n)$ estratto dalla popolazione in oggetto. Fare inferenza su Θ significa applicare il procedimento bayesiano, che consiste nel determinare la distribuzione finale di Θ condizionata al vettore osservato $\mathbf{x} = (x_1, \ldots, x_n)$, ovvero $\beta(\theta|\mathbf{x})$. Se, riferendoci al caso continuo, indichiamo con $c(\mathbf{x}, \theta)$ la densità congiunta del vettore aleatorio $(\mathbf{X}, \Theta)$ e con $\alpha(\mathbf{x}|\theta) = \alpha(x_1, \ldots, x_n|\theta)$ la densità di $\mathbf{X}$ condizionata a un fissato valore θ di Θ, si ha

$$c(\mathbf{x}, \theta) = \beta(\theta)\alpha(\mathbf{x}|\theta) = \alpha(\mathbf{x})\beta(\theta|\mathbf{x}) \,.$$

Quindi (*teorema di Bayes per vettori aleatori*)

$$\beta(\theta|\mathbf{x}) = \frac{c(\mathbf{x},\theta)}{\alpha(\mathbf{x})} = \frac{\beta(\theta)\alpha(\mathbf{x}|\theta)}{\int_\Theta c(\mathbf{x},\theta)d\theta} =$$

$$= \frac{\beta(\theta)\alpha(x_1,\ldots,x_n|\theta)}{\int_\Theta \beta(\theta)\alpha(x_1,\ldots,x_n|\theta)d\theta} = k(\mathbf{x})\beta(\theta)\alpha(\mathbf{x}|\theta)\,,$$

con

$$k(\mathbf{x}) = \frac{1}{\int_\Theta \beta(\theta)\alpha(\mathbf{x}|\theta)d\theta}\,.$$

Una situazione tipica nelle applicazioni statistiche è quella in cui, per il vettore delle osservazioni (o misure) $\mathbf{X} = (X_1, \ldots, X_n)$, valgono le seguenti proprietà:
(i) indicando con $f_i(x_i)$ la densità di X_i, si ha $f_1(\cdot|\theta) = \cdots = f_n(\cdot|\theta) = f(\cdot|\theta)$, ovvero per ogni fissato θ, i numeri aleatori $X_1, \ldots, X_n$ sono ugualmente distribuiti condizionatamente a θ;

(ii) per ogni fissato θ, i numeri aleatori $X_1, \ldots, X_n$ sono stocasticamente indipendenti condizionatamente a θ.

Da (i) e (ii) segue

$$\alpha(x_1, \ldots, x_n|\theta) = f_1(x_1|\theta) \cdots f_n(x_n|\theta) =$$

$$= f(x_1|\theta) \cdots f(x_n|\theta)\,,$$

e quindi

$$\beta(\theta|x_1, \ldots, x_n) = k(x_1, \ldots, x_n)\beta(\theta) f(x_1|\theta) \cdots f(x_n|\theta)\,,$$

con

$$k(x_1, \ldots, x_n) = \frac{1}{\int_\Theta \beta(\theta) f(x_1|\theta) \cdots f(x_n|\theta) d\theta}\,.$$

Quando valgono (i) e (ii) il vettore aleatorio $\mathbf{X} = (X_1, \ldots, X_n)$ si dice un campione casuale.

La distribuzione finale di Θ (condizionata al vettore delle osservazioni $\mathbf{x}$) può essere utilizzata per determinare un insieme A, ad esempio un intervallo $[\theta_1, \theta_2]$ (possibilmente di lunghezza minima), tale che per un opportuno valore α risulti $P(\Theta \in A \,|\, \mathbf{x}) = \alpha$. In particolare, se vale $P(\theta_1 \leq \Theta \leq \theta_2 \,|\, \mathbf{x}) = \alpha$ l'intervallo

$[\theta_1, \theta_2]$ si dice un *intervallo di confidenza al* $100\alpha\%$.

Applicazioni:
(a) *campionamento da una popolazione normale con parametro* m *incognito*.
Indichiamo con Θ il parametro incognito m di una distribuzione normale e supponiamo che sia $\beta(\theta) = N_{m_0,\sigma_0}(\theta)$. Supponiamo inoltre che, per $i = 1, \ldots, n$, sia

$$f_i(x_i|\theta) = f(x_i|\theta) = N_{\theta,\sigma}(x_i)\,.$$

Ad esempio, Θ potrebbe rappresentare la lunghezza incognita di una sbarra, a cui viene assegnata come densità iniziale (sulla base dell'informazione a priori) una distribuzione normale di parametri m_0, σ_0. Tale sbarra viene misurata n volte in modo da ottenere un vettore $\mathbf{x} = (x_1, \ldots, x_n)$ che costituisce il valore osservato di un campione casuale $\mathbf{X} = (X_1, \ldots, X_n)$, per le cui componenti si assume (in base a certe ipotesi sullo strumento di misura) che sia $X_i|\theta \sim N_{\theta,\sigma}$.

Si ha quindi

$$\beta(\theta|x_1,\ldots,x_n) = k(x_1,\ldots,x_n)\beta(\theta)f(x_1|\theta)\cdots f(x_n|\theta) =$$

$$= k(x_1,\ldots,x_n)N_{m_0,\sigma_0}(\theta)N_{\theta,\sigma}(x_1)\cdots N_{\theta,\sigma}(x_n) =$$

$$= \cdots = N_{m_n,\sigma_n}(\theta)\,,$$

dove, indicando con $\bar{x}$ la media aritmetica dei valori $x_1,\ldots,x_n$, è

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\,,$$

$$m_n = \frac{\frac{1}{\sigma_0^2}\cdot m_o + \frac{n}{\sigma^2}\cdot\bar{x}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} = \lambda m_0 + (1-\lambda)\bar{x}\,,$$

dove

$$\lambda = \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\,, \quad 1-\lambda = \frac{\frac{n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\,\cdot$$

Come si vede, m_n è una media ponderata dei valori $m_0, \bar{x}$ con rispettivi pesi $\lambda, 1-\lambda$. Inoltre, per n grande si ha $\sigma_n << \sigma_0$, ovvero la distribuzione finale di Θ è molto più concentrata di quella iniziale (infatti σ_n

tende a 0 per $n \to +\infty$).
In particolare, per $n = 1$, si ha

$$\beta(\theta|x_1) = k(x_1)\beta(\theta)f(x_1|\theta) =$$

$$= k(x_1)N_{m_0,\sigma_0}(\theta)N_{\theta,\sigma}(x_1) =$$

$$= \cdots = N_{m_1,\sigma_1}(\theta)\,,$$

con

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}\,, \qquad m_1 = \frac{\frac{1}{\sigma_0^2}\cdot m_o + \frac{1}{\sigma^2}\cdot x_1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}\,.$$

(b) *Campionamento da una popolazione binomiale con parametro p incognito.*

Si tratta del caso in cui una proporzione incognita p di individui di una data popolazione possiede una certa caratteristica. Indicando con Θ il parametro incognito p, si ha intanto $\Theta \in [0,1]$. L'obiettivo è quello di fare inferenza su Θ estraendo n individui dalla popolazione ed osservando quanti di essi possiedono la data caratteristica. Possiamo considerare un campione casuale $\mathbf{X} = (X_1, \ldots, X_n)$, dove $(X_i = 1)$ significa che l'i-mo individuo osservato possiede la data caratteristica, mentre $(X_i = 0)$ significa che l'i-mo individuo osservato non possiede tale caratteristica. Pertanto

$$P(X_i = 1 \,|\, \theta) = f_i(1 \,|\, \theta) = \theta \,,$$

$$P(X_i = 0 \,|\, \theta) = f_i(0 \,|\, \theta) = 1 - \theta \,,$$

e quindi, osservando che $f_i(x_i \,|\, \theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, si ha

$$\alpha(x_1, \ldots, x_n|\theta) = f_1(x_1|\theta) \cdots f_n(x_n|\theta) =$$

$$= f(x_1|\theta) \cdots f(x_n|\theta) = \cdots = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} \,.$$

Se, come distribuzione iniziale di Θ si sceglie una distribuzione beta di parametri r_0, s_0, si può verificare che la distribuzione finale è ancora di tipo beta e, posto

$\sum_i x_i = h$, risulta

$$\beta(\theta|x_1,\ldots,x_n) = k(x_1,\ldots,x_n)B_{r_0,s_0}(\theta)\theta^h(1-\theta)^{n-h} =$$

$$= \cdots = B_{r_n,s_n}(\theta)\,, \quad (r_n = r_0 + h\,,\;\; s_n = s_0 + n - h)\,.$$

(c) *Campionamento da una popolazione esponenziale con parametro λ incognito.*
Supponiamo che la durata aleatoria fino al guasto di un certo tipo di dispositivi abbia una distribuzione esponenziale con parametro incognito, che indichiamo con Θ (anzichè λ).
Sia $(X_1,\ldots,X_n)$ il vettore aleatorio costituito dall'osservazione delle durate fino al guasto di n di tali dispositivi. Si ha

$$X_i\,|\,\theta \,\sim\, f = Exp(\theta) = G_{1,\theta}\,,\;\; i = 1,\ldots,n\,;$$

$$f(x\,|\,\theta) = \theta e^{-\theta x}\,,\;\; x \geq 0\,; \quad f(x\,|\,\theta) = 0\,,\;\; x < 0\,;$$

$$\alpha(x_1,\ldots,x_n|\theta) = f(x_1|\theta)\cdots f(x_n|\theta) =$$

$$= \theta e^{-\theta x_1}\cdots\theta e^{-\theta x_n} = \theta^n e^{-\theta\sum_i x_i}\,;$$

$$\beta(\theta|x_1,\ldots,x_n) = k(x_1,\ldots,x_n)\beta(\theta)\theta^n e^{-\theta\sum_i x_i}\,.$$

Se per Θ si assume una distribuzione iniziale di tipo Gamma, ad esempio $\beta(\theta) = G_{c_0,\lambda_0}(\theta)$, avendo osservato un vettore $\mathbf{x} = (x_1,\ldots,x_n)$ si ha

$$\beta(\theta|\mathbf{x}) = k(\mathbf{x})\frac{\lambda_0^{c_0}}{\Gamma(c_0)}\theta^{c_0-1}e^{-\lambda_0\theta}\theta^n e^{-\theta\sum_i x_i} =$$

$$= k(\mathbf{x})\frac{\lambda_0^{c_0}}{\Gamma(c_0)}\theta^{c_0+n-1}e^{-\theta(\lambda_0+\sum_i x_i)} = G_{c_n,\lambda_n}(\theta)\,,$$

con $c_n = c_0 + n\,,\quad \lambda_n = \lambda_0 + \sum_i x_i$.
Pertanto la distribuzione finale di Θ è ancora di tipo Gamma e si ha

$$\mathbb{P}(\Theta\,|\,x_1,\ldots,x_n) = \frac{c_n}{\lambda_n} = \frac{c_0+n}{\lambda_0+\sum_i x_i}\,;$$

$$Var(\Theta\,|\,x_1,\ldots,x_n) = \frac{c_n}{\lambda_n^2} = \frac{c_0+n}{(\lambda_0+\sum_i x_i)^2}\,.$$

(d) *Campionamento da una popolazione di Poisson con parametro λ incognito.*
Supponiamo che il numero aleatorio di "arrivi" in un certo fenomeno abbia una distribuzione di Poisson con parametro incognito Θ.
Gli "arrivi" potrebbero, ad esempio, essere le telefonate che giungono ad un centralino in un fissato intervallo di tempo durante il giorno.
Sia $(X_1, \ldots, X_n)$ il vettore aleatorio costituito dall'osservazione del numero di telefonate che arrivano a tale centralino nel fissato intervallo di tempo in n giorni distinti. Si ha

$$X_i \,|\, \theta \sim f = \mathcal{P}(\theta)\,, \quad i = 1, \ldots, n\,;$$

$$f(x \,|\, \theta) = \tfrac{\theta^x}{x!} e^{-\theta}\,, \quad x = 0, 1, \ldots, k, \ldots\,.$$

$$\alpha(x_1, \ldots, x_n | \theta) = f(x_1|\theta) \cdots f(x_n|\theta) =$$

$$= \tfrac{\theta^{x_1}}{x_1!} e^{-\theta} \cdots \tfrac{\theta^{x_n}}{x_n!} e^{-\theta} = \tfrac{\theta^{\sum_i x_i}}{x_1! \cdots x_n!} e^{-n\theta}\,;$$

$$\beta(\theta | x_1, \ldots, x_n) = \tfrac{k(x_1, \ldots, x_n)}{x_1! \cdots x_n!} \beta(\theta) \theta^{\sum_i x_i} e^{-n\theta}\,.$$

Se per Θ si assume una distribuzione iniziale (di tipo Gamma) $\beta(\theta) = G_{c_0,\lambda_0}(\theta)$, condizionatamente ad un vettore osservato $\mathbf{x} = (x_1, \ldots, x_n)$ si ha

$$\beta(\theta|\mathbf{x}) = \tfrac{k(\mathbf{x})}{x_1!\cdots x_n!}\tfrac{\lambda_0^{c_0}}{\Gamma(c_0)}\,\theta^{c_0-1}e^{-\lambda_0\theta}\theta^{\sum_i x_i}e^{-n\theta} =$$

$$= \tfrac{k(\mathbf{x})}{x_1!\cdots x_n!}\tfrac{\lambda_0^{c_0}}{\Gamma(c_0)}\,\theta^{c_0+\sum_i x_i-1}e^{-\theta(\lambda_0+n)} = G_{c_n,\lambda_n}(\theta)\,,$$

con $c_n = c_0 + \sum_i x_i\,, \quad \lambda_n = \lambda_0 + n$.

Pertanto la distribuzione finale di Θ è ancora di tipo Gamma e si ha

$$\mathbb{P}(\Theta\,|\,x_1,\ldots,x_n) = \frac{c_n}{\lambda_n} = \frac{c_0 + \sum_i x_i}{\lambda_0 + n}\,;$$

$$Var(\Theta\,|\,x_1,\ldots,x_n) = \frac{c_n}{\lambda_n^2} = \frac{c_0 + \sum_i x_i}{(\lambda_0 + n)^2}\,.$$

Esercizi di Calcolo delle probabilità

1. Un numero aleatorio X ha distribuzione normale con parametri m, σ. Sia $Y = X + c$, $Z = aX + bY$. Calcolare $Cov(Y, Z)$ e il valore z_0 tale che $P(Z \geq z_0) = \Phi(-1)$.

2. Da una stanza S_1, in cui ci sono 4 uomini e 4 donne, escono a caso tre persone che entrano in una stanza S_2, in cui ci sono 2 uomini e 2 donne. Successivamente, da S_2 esce a caso una persona. Calcolare la probabilità che tale persona sia una donna.

3. Sia T il triangolo di vertici i punti $(0,0)$, $(1,1)$, $(1,0)$. Stabilire se un vettore aleatorio (X,Y) può avere una densità congiunta del tipo $f(x,y) = x + by$ per $(x,y) \in T$, con $f(x,y) = 0$ altrove, dove b è un opportuno valore da calcolare.

4. Una porta è chiusa con due serrature. In una scatola ci sono 8 chiavi: 2 capaci di aprire la prima serratura, 2 capaci di aprire la seconda e 4 che non aprono nessuna serratura. Si prendono a caso 2 chiavi. Siano definiti gli eventi:
   $A_i =$ la i-ma chiave estratta apre la prima serratura, $i = 1,2$;
   $B_j =$ la j-ma chiave estratta apre la seconda serratura, $j = 1,2$;
   $C_k =$ la k-ma chiave estratta non apre nessuna serratura, $k = 1,2$.
   Calcolare la probabilità che si riesca ad aprire la porta con le 2 chiavi prese a caso.