Probabilità e Statistica

Angelo Gilio

Dip. Met. Mod. Mat. Univ. La Sapienza

gilio@dmmm.uniroma1.it

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

Informazioni

Tutoraggio su web (materiale didattico)

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

https://www.dmmm.uniroma1.it/
(cliccare su personale, gilio, corso di laurea, testi esami, oppure avviso)

Libro adottato: *Incertezza e Probabilità*, Romano Scozzafava, Zanichelli, ed. 2008.

Ricevimento:
lunedì e venerdì: 14.00 - 15.30,
pal. A, primo piano, studio 16.

Esercizi: *Training Autogeno in Probabilità*, Joseph Toscano, Zanichelli, ed. 2008.

Introduzione
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.

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.

I fatti incerti sono formalizzati (in modo non ambiguo) mediante *proposizioni logiche* o *eventi* che possono assumere il valore *Vero* (V) oppure *Falso* (F).

Evento certo, Ω: evento definito mediante una proposizione logica certamente vera.

Evento impossibile, $\emptyset$: evento definito mediante una proposizione logica certamente falsa.

Esempi: $\cdots\cdots$

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

Indicatore

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

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

Relazioni e operazioni logiche

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.

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

Nota: $|E^c| = 1 - |E|$.

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

In simboli: $A \Longrightarrow B$ oppure $A \subseteq B$.
$A \Longrightarrow B$ equivale alla diseguaglianza $|A| \leq |B|$.

Esempi: $\cdots\cdots$

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 $A \cup B$.

Proprietà:
- *associativa* : $(A \vee B) \vee C = A \vee (B \vee C) = A \vee B \vee C$;
- *commutativa* : $A \vee B = B \vee A$.

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.

Proprietà:
- *associativa* : $(A \wedge B) \wedge C = A \wedge (B \wedge C) = A \wedge B \wedge C$.
- *commutativa* : $A \wedge B = B \wedge A$.

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 $AB = \emptyset$.

La corrispondenza tra i valori logici di due eventi A, B e quelli di $A \wedge B$ e $A \vee B$ è riportata nella Tabella 1. Quella tra gli indicatori è riportata nella Tabella 2.

Tabella 1

A	B	AB	$A \vee B$
V	V	V	V
V	F	F	V
F	V	F	V
F	F	F	F

Tabella 2

$\|A\|$	$\|B\|$	$\|AB\|$	$\|A \vee B\|$
1	1	1	1
1	0	0	1
0	1	0	1
0	0	0	0

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:

$$AB \subseteq A \subseteq A \vee B\,, \quad (|AB| \leq |A| \leq |A \vee B|)$$

$$AB \subseteq B \subseteq A \vee B\,, \quad (|AB| \leq |B| \leq |A \vee B|)$$

Proprietà *distributive* :

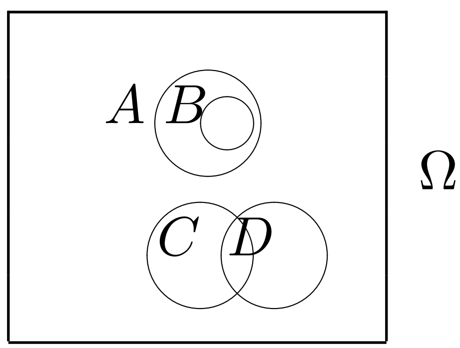
$$(A{\vee}B){\wedge}C = AC{\vee}BC\ , \quad (A{\wedge}B){\vee}C = (A{\vee}C){\wedge}(B{\vee}C)\ .$$

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.



Evento	**Insieme**
certo	universo
impossibile	vuoto
contrario	complementare
implicazione	inclusione
incompatibili	disgiunti
unione	unione
intersezione	intersezione

Partizioni finite dell'evento certo

$\{H_1, H_2, \ldots, H_n\}$ è una partizione di Ω se:

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

Equivalentemente:

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

Esempi: $\cdots\cdots$

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

Considerati degli eventi $A, B, \ldots$, i *costituenti* si ottengono sviluppando l'espressione:

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

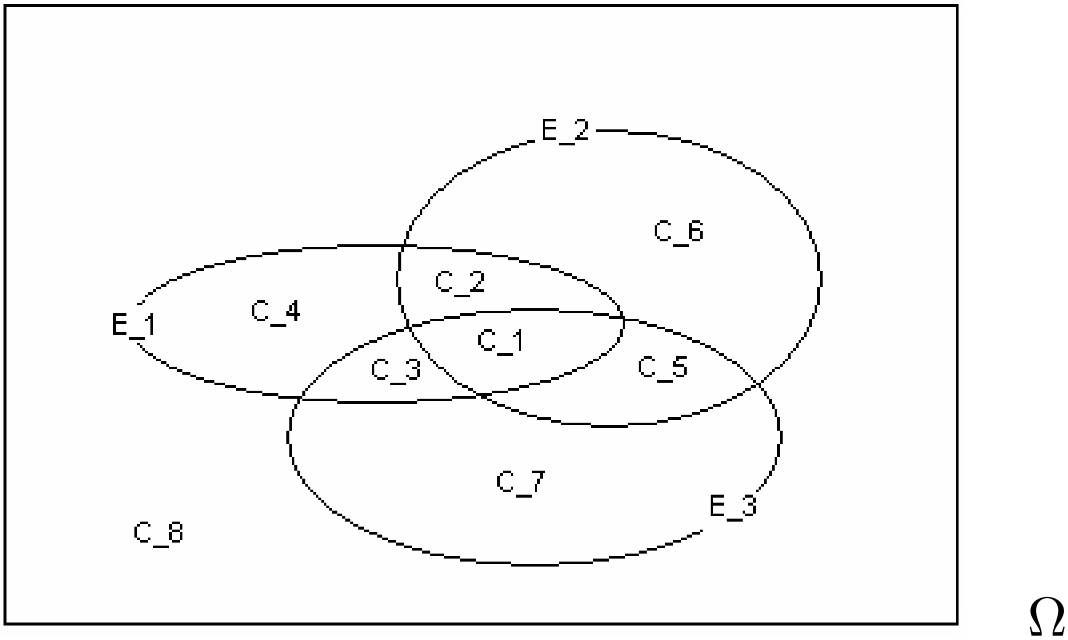
Eliminando le intersezioni impossibili, quelle rimanenti sono i casi possibili relativi agli eventi $A, B, \ldots$. .

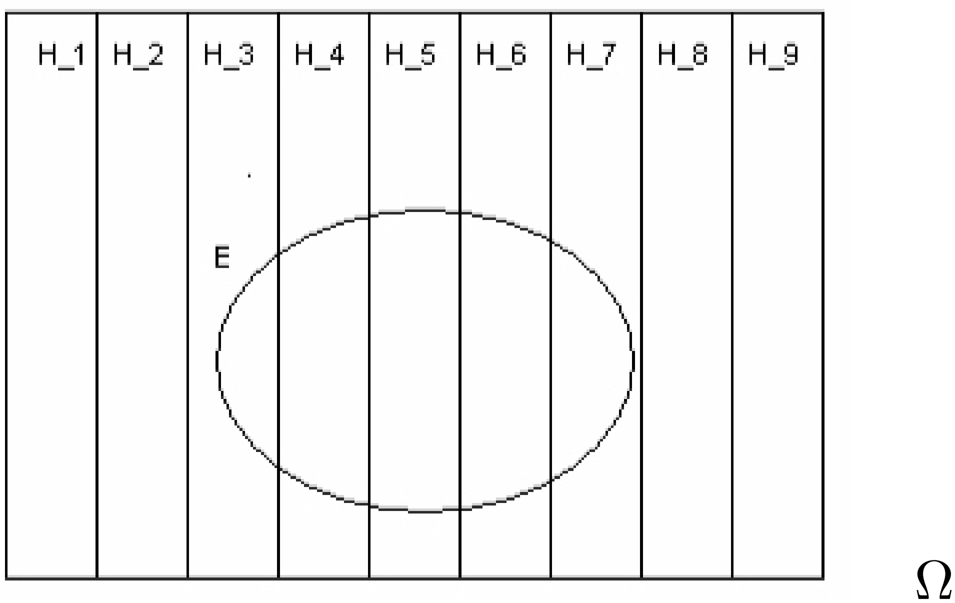
In generale, data una famiglia $\mathcal{F}_n = \{E_1, \ldots, E_n\}$, i *casi possibili* o *costituenti*, $C_1, \ldots, C_m$, con $m \leq 2^n$, si ottengono sviluppando l'espressione

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

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

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


E_2
C_6
C_2
E_1
C_4
C_1
C_3
C_5
C_7
E_3
C_8
Ω


H_1
H_2
H_3
H_4
H_5
H_6
H_7
H_8
H_9
E
Ω

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

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

Data una partizione $\{H_1, H_2, \ldots, H_n\}$, 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.
Esempi: ....
Da un'urna contenente 5 palline bianche e 3 nere si effettuano 2 estrazioni senza restituzione. Sia A l'evento "la 1^a pallina estratta è bianca" *e B l'evento* "la 2^a pallina estratta è bianca". *Calcolare, in relazione a ciascun evento, il rapporto tra il numero di casi favorevoli e il numero di casi possibili, confrontando i valori ottenuti per A e B.*

Data una famiglia $\mathcal{F}_n = \{E_1, \ldots, E_n\}$, siano $C_1, C_2, \ldots, C_m$ i relativi costituenti. Ogni evento E_i si può scrivere come unione logica dei costituenti ad esso favorevoli.

$$E_i = E_i \wedge \Omega = E_i \wedge (C_1 \vee C_2 \vee \cdots \vee C_m) =$$

$$= E_i C_1 \vee E_i C_2 \vee \cdots \vee E_i C_m =$$

$$= \bigvee_{h: C_h \subseteq E_i} (E_i C_h) \vee \bigvee_{h: C_h \nsubseteq E_i} (E_i C_h) =$$

$$= \bigvee_{h: C_h \subseteq E_i} C_h \, .$$

Esempio 1 Con riferimento a una data partita di calcio tra la *Roma* e l' *Inter*, si considerino gli eventi:

- A : *la Roma a cinque minuti dal termine della gara è in vantaggio sull'Inter*;
- B : *la Roma vince la partita*;
- C : *la Roma a cinque minuti dal termine della gara è in svantaggio sull'Inter.*

A e C sono incompatibili, quindi gli eventi ABC e AB^cC sono impossibili. Pertanto, i costituenti sono i seguenti:

- $C_1 = ABC^c$
  la Roma a cinque minuti dal termine della gara è in vantaggio sull'Inter e vince la partita.

- $C_2 = A^cBC$
  la Roma a cinque minuti dal termine della gara è in svantaggio sull'Inter e vince la partita.

- $C_3 = A^cBC^c$
  a cinque minuti dal termine della gara le squadre sono in parità e la Roma vince la partita.

- $C_4 = AB^cC^c$
  la Roma a cinque minuti dal termine della gara è in vantaggio sull'Inter e non vince la partita.

- $C_5 = A^cB^cC$
  la Roma a cinque minuti dal termine della gara è in svantaggio sull'Inter e non vince la partita.

- $C_6 = A^c B^c C^c$
  a cinque minuti dal termine della gara le squadre sono in parità e la Roma non vince la partita.

A, B, C si possono rappresentare nel modo seguente

$$\begin{aligned} &A = C_1 \vee C_4 = ABC^c \vee AB^cC^c\,, \\ &B = C_1 \vee C_2 \vee C_3 = ABC^c \vee A^cBC \vee A^cBC^c\,, \\ &C = C_2 \vee C_5 = A^cBC \vee A^cB^cC\,. \end{aligned}$$

Indipendenza logica.
Un evento A si dice *logicamente indipendente* da altri eventi $B, C, \ldots$ se, assegnando in tutti i modi possibili il valore logico (vero o falso) agli eventi $B, C, \ldots$, l'evento A rimane incerto, potendo risultare sia vero che falso.

Se A non è logicamente indipendente da $B, C, \ldots$ si possono presentare vari tipi di dipendenza logica.

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

Esempio 2 Estrazioni con restituzione da un'urna contenente 1 pallina bianca e 1 nera.
Gli eventi
E_i = la i-ma pallina estratta è bianca, $i = 1, \ldots, 5$,
sono logicamente indipendenti?

Esempio 3 Estrazioni senza restituzione da un'urna contenente 2 palline bianche e 3 nere.
Gli eventi E_1, E_2, E_3 sono logicamente indipendenti?
Ad esempio, se E_1, E_2 sono entrambi veri (cioè le prime due palline estratte sono *bianche*), cosa si può dire della terza pallina e quindi del valore logico di E_3 ?
Se invece E_1 è vero ed E_2 è falso, cosa si può dire di E_3 ?

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

(Se si suppone noto il risultato delle prime quattro estrazioni, il risultato della quinta è incerto?)

Criterio classico e proprietà 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, è del tutto naturale basarsi sul *numero di casi favorevoli* ad esso.

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

$$P(A^c) = 1 - P(A) \ . \qquad (5)$$

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

$$P(C) = 1 - P(A \vee B) = 1 - P(A) - P(B) \ , \qquad (6)$$

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

$$P(A) + P(B) + P(C) = 1 \ . \qquad (7)$$

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 $\{E_1, \ldots, E_n\}$ è 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 $A \vee B$ in generale.

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 Méré

Esempio 4 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 Méré (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 ... per uno dei due eventi la frequenza di successo leggermente superiore a quella dell'altro.

Il problema venne sottoposto a Blaise Pascal.
... 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 5 Un esperimento aleatorio consiste in due lanci di una moneta.

E : *in almeno un lancio esce Testa.*

Casi possibili:
C_1: esce Testa al primo lancio (e l'esperimento termina);
C_2: esce Croce al primo lancio e Testa al secondo lancio;
C_3: esce Croce in entrambi i lanci.

C_1 e C_2 sono favorevoli ad E.

Allora $\ldots P(E) = \frac{2}{3}$? (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).

... inoltre, l'unione logica di E_2 ed E_3 coincide con l'evento *Croce al primo lancio*, che ha probabilità $\frac{1}{2}$,
... pertanto, $P(E_2) = P(E_3) = \frac{1}{4}$.
... quindi una valutazione più adeguata è $P(E) = \frac{3}{4}$.

Tale valutazione si ottiene direttamente considerando come casi (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 6 Se uno studente sostiene un esame vi sono due casi possibili (lo studente può essere promosso o bocciato) ... nessuno, però, concluderebbe che la probabilità di essere promosso è $\frac{1}{2}$.

... per la valutazione delle probabilità occorrono quindi 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

N prove indipendenti e ripetute nelle stesse condizioni;

$f_N =$ frequenza relativa di "successo" sulle N prove, per un dato evento E;

$$P(E) = lim_{N \to +\infty} f_N \, .$$

Critiche.
1) Le frequenze non costituiscono una successione numerica data mediante una legge, ma sono dei numeri rilevati sperimentalmente. Il concetto di limite utilizzato non è quello rigoroso dell'analisi.
2) (applicabilità) si deve prendere in considerazione una successione di prove fatte nelle stesse condizioni. Non è sempre possibile.
3) (circolarità) il concetto di indipendenza ha un significato probabilistico.

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, ... sono rigorosamente distinti gli aspetti *oggettivi*, concernenti gli eventi, da quelli *soggettivi*, concernenti le valutazioni probabilistiche (*logica del certo* e *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 diversità di valutazioni ... risiede nel fatto che le persone hanno un'esperienza e uno *stato di informazione* differenti fra di loro.

Esempio 7 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 ... è dovuto a un diverso *grado di fiducia* nel verificarsi dell'evento considerato.

... l'informazione di cui si è in possesso e il modo in cui si elabora tale informazione giocano un ruolo essenziale nella valutazione delle probabilità.

... nell'impostazione soggettiva tale aspetto viene riconosciuto esplicitamente ...

Definizione 2 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 comportare a priori una perdita certa.

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

$$\mathcal{G} = S|E| - pS = S(|E| - p) =$$

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

In generale, dati $\mathcal{F}_n = \{E_1, \ldots, E_n\}$ e $\mathcal{P}_n = (p_1, \ldots, p_n)$, con $p_i = P(E_i),\ i = 1, \ldots, n$, si ha

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

con $S_1, \ldots, S_n$ numeri reali arbitrari (non tutti nulli).

Definizione 3 La valutazione $\mathcal{P}_n$ si dice *coerente* se, per ogni $S_1, \ldots, S_n$, 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 $P(A \vee B) = P(A) + P(B)$.

In relazione alla valutazione $P(E) = p$ si ha

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

... la condizione di coerenza diventa ...
$S(1-p)(-pS) = -p(1-p)S^2 \leq 0$,
da cui segue ... $0 \leq p \leq 1$. (Proprietà **P1**).

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

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

e quindi ... $G = 0$, per ogni S, ... da cui si ottiene $P(\Omega) = 1$. (Proprietà **P2**).

Proprietà additiva (**P3**).
Data una partizione di Ω, $\{H_1, \ldots, H_n\}$, con $P(H_1) = p_1, \ldots, P(H_n) = p_n$, si ha

$$\mathcal{G} = S_1(|H_1| - p_1) + \cdots + S_n(|H_n| - p_n).$$

Osservando che $|H_1| + \cdots + |H_n| = 1$ e scegliendo $S_1 = \cdots = S_n = S$ si ottiene

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

e quindi ... dev'essere $\mathcal{G} = 0$.
Allora, segue $p_1 + \cdots + p_n = 1$, ovvero

$$P(H_1) + \cdots + P(H_n) = 1. \tag{13}$$

Dim. della proprietà additiva:

dati A e B incompatibili e considerata la partizione $\{A \vee B, (A \vee B)^c\}$, deve valere

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

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

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

e quindi:

$$P(A \vee B) = P(A) + P(B) \; . \qquad (14)$$

Dati n eventi $E_1, \ldots, E_n$, a due a due incompatibili, si ha

$$P(E_1 \vee \cdots \vee E_n) = P(E_1) + \cdots + P(E_n) \; . \qquad (15)$$

Condizioni sulla probabilità dell'Unione. Dati due eventi A e B e due valutazioni di probabilità $P(A)$ e $P(B)$, consideriamo l'evento unione $A \vee B$. Dalla proprietà di monotonia segue che

$$P(A) \leq P(A \vee B)\,, \quad P(B) \leq P(A \vee B)\,, \tag{16}$$

quindi

$$P(A \vee B) \;\geq\; max\,\{P(A), P(B)\}\,. \tag{17}$$

Osservando, inoltre, che $P(A \vee B) \leq 1$ e che

$$P(A \vee B) = P(A) + P(B) - P(AB) \leq P(A) + P(B)\,,$$

segue

$$P(A \vee B) \;\leq\; min\,\{1, P(A) + P(B)\}\,. \tag{18}$$

La (17) e la (18), assegnati $P(A)$ e $P(B)$, stabiliscono delle limitazioni per $P(A \vee B)$.

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 (15) 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.*

Verifica della coerenza

$\mathcal{F}_n = \{E_1, \ldots, E_n\}$, famiglia di eventi arbitrari, legati da possibili relazioni logiche.
$\mathcal{P}_n = (p_1, \ldots, p_n)$, assegnazione di probabilità su $\mathcal{F}_n$, con $p_i = P(E_i),\ i = 1, \ldots, n$.

$C_1, \ldots, C_m$ ($m \leq 2^n$), costituenti relativi alla famiglia $\mathcal{F}_n$.

Ricordiamo che

$$E_i = \bigvee_{h : C_h \subseteq E_i} C_h. \tag{19}$$

Posto $P(C_h) = \lambda_h$, dev'essere

$$\lambda_h \geq 0\,, \quad h = 1, \ldots, m\,; \quad \sum_{h=1}^{m} \lambda_h = 1\,. \tag{20}$$

Inoltre, dalla (19) segue

$$P(E_i) = \sum_{h : C_h \subseteq E_i} \lambda_h\,. \tag{21}$$

Si consideri il seguente sistema nelle incognite (non negative) $\lambda_1, \ldots, \lambda_m$

$$(\mathcal{S}) : \quad \begin{cases} \sum_{h:C_h \subseteq E_i} \lambda_h = p_i \ , \quad i = 1, \ldots, n; \\ \\ \sum_{h=1}^{m} \lambda_h = 1 \ , \\ \\ \lambda_h \geq 0 \ , \quad h = 1, \ldots, m \ . \end{cases} \tag{22}$$

Si può dimostrare il seguente risultato

Teorema 2 La valutazione di probabilità $\mathcal{P}_n$ è coerente se e solo se il sistema $(\mathcal{S})$ è compatibile.

Propagazione

Sia$\mathcal{F}_n = \{E_1, E_2, \ldots, E_n\}$ una famiglia di eventi e sia $P_n = \{p_1, p_2, \ldots, p_n\}$ una valutazione di probabilità coerente su $\mathcal{F}_n$.

Se consideriamo un ulteriore evento E_{n+1}, sorge il problema di come valutare la probabilità $p_{n+1} = P(E_{n+1})$ in modo tale che la valutazione $P_{n+1} = \{p_1, \ldots, p_n, p_{n+1}\}$ su $\mathcal{F}_{n+1} = \{E_1, \ldots, E_n, E_{n+1}\}$ sia coerente.

... la scelta di p_{n+1} non è arbitraria, ma va fatta in un opportuno intervallo $[p', p''] \subseteq [0, 1]$.

Esempio.
Sia $\mathcal{F}_1 = \{E_1\}$, con $P(E_1) = p_1 \in [0, 1]$. Dato un evento E_2 incompatibile con E_1, con $E_1 \vee E_2 \subset \Omega$, si ponga $P(E_2) = p_2$.
Determinare i valori di p_2 tali che la valutazione (p_1, p_2) su $\mathcal{F}_2 = \{E_1, E_2\}$ sia coerente.

Soluzione: $p_2 \in [0, 1 - p_1]$.
Infatti $P(E_2) \geq 0$, inoltre

$$P(E_1 \vee E_2) \;=\; P(E_1) + P(E_2) \;\leq\; 1\,,$$

e quindi

$$0 \;\leq\; P(E_2) \;\leq 1 - P(E_1)\,.$$

Determinazione dell'intervallo $[p', p'']$.
Siano $C_1, C_2, \ldots, C_m$ i costituenti relativi alla famiglia $\mathcal{F}_n = \{E_1, E_2, \ldots, E_n\}$.
Dato E_{n+1}, distinguiamo 3 classi di costituenti:
(i) costituenti che implicano (cioè favorevoli a) E_{n+1};
(ii) costituenti che non implicano E_{n+1}, ma sono con esso compatibili;
(iii) costituenti incompatibili con E_{n+1}.
Definiamo gli insiemi

$$J_1 = \{h : C_h \wedge E_{n+1} = C_h\}\,,$$

$$J_2 = \{h : \emptyset \subset C_h \wedge E_{n+1} \subset C_h\}\,,$$

$$J_3 = \{h : C_h \wedge E_{n+1} = \emptyset\}\,.$$

Osservando che $\bigvee_{h\in J_3} E_{n+1}C_h = \emptyset$, si ha

$$E_{n+1} = \bigvee_{h=1}^{m} E_{n+1}C_h = \bigvee_{j\in J_1\cup J_2\cup J_3} E_{n+1}C_h =$$

$$= \left(\bigvee_{h\in J_1} E_{n+1}C_h\right) \vee \left(\bigvee_{h\in J_2} E_{n+1}C_h\right) =$$

$$= \left(\bigvee_{h\in J_1} C_h\right) \vee \left(\bigvee_{h\in J_2} E_{n+1}C_h\right) .$$

Posto

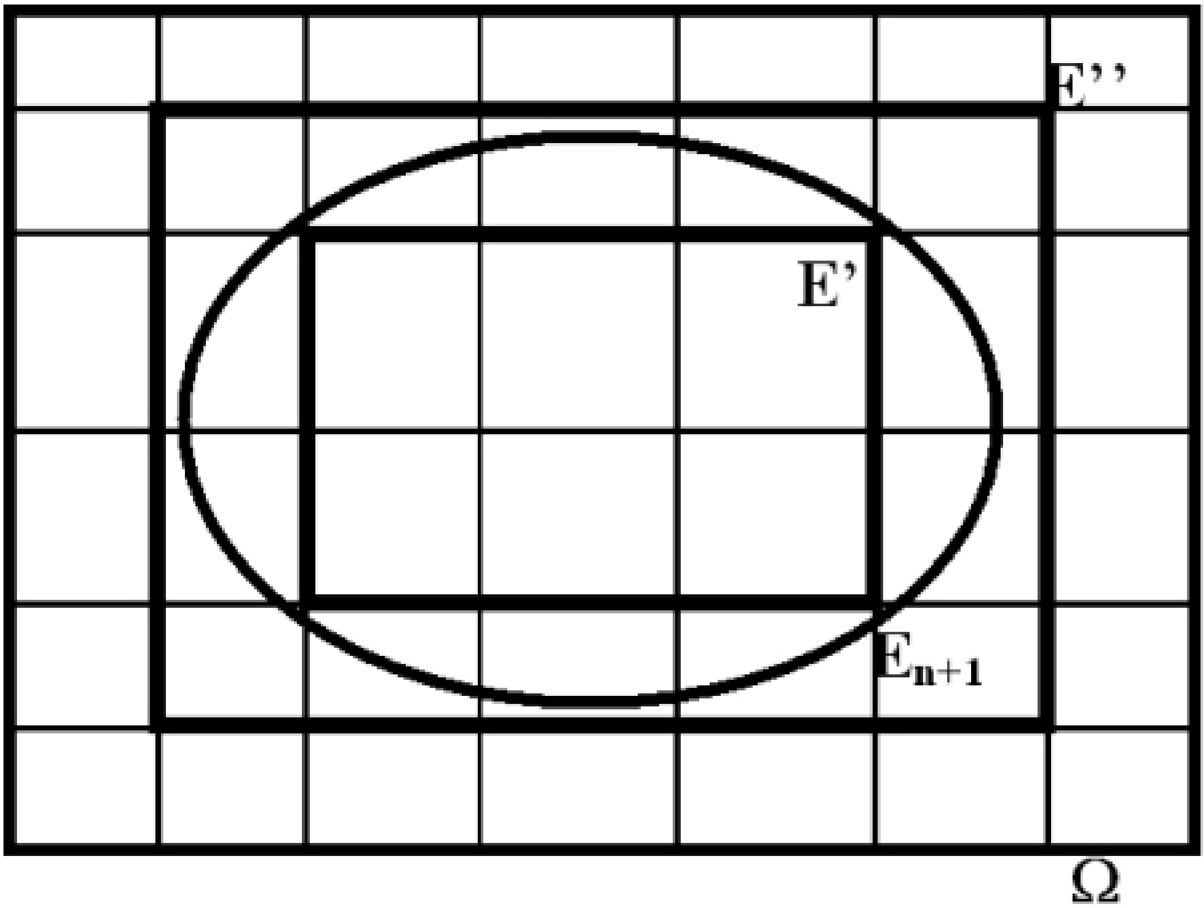
$$E' = \bigvee_{h\in J_1} C_h\,, \quad E'' = \bigvee_{h\in J_1\cup J_2} C_h\,,$$

vale la seguente relazione

$$E' \subseteq E_{n+1} \subseteq E''\,. \tag{23}$$

E' si chiama *massimo evento logicamente dipendente da $E_1, E_2, \ldots, E_n$ contenuto in E_{n+1}*.

E'' si chiama *minimo evento logicamente dipendente da $E_1, E_2, \ldots, E_n$ contenente E_{n+1}*.



Nella rappresentazione grafica i costituenti sono i rettangolini. Dalla relazione (23), supponendo assegnate le probabilità ai costituenti, segue

$$\sum_{h \in J_1} \lambda_h = P(E') \leq P(E_{n+1}) \leq P(E'') = \sum_{h \in J_1 \cup J_2} \lambda_h \, .$$

Avendo supposto coerente la valutazione $\mathcal{P}_n$ su $\mathcal{F}_n$, segue che esistono soluzioni $(\lambda_1, \lambda_2, \dots, \lambda_m)$ del sistema $(\mathcal{S})$ di pag.38. Indicando con S l'insieme delle soluzioni, definiamo le seguenti quantità:

$$p' \,=\, min_{\,(\lambda_1,\lambda_2,\ldots,\lambda_m)\in S}\, P(E')\,,$$

$$p'' \,=\, max_{\,(\lambda_1,\lambda_2,\ldots,\lambda_m)\in S}\, P(E'')\,.$$

... la valutazione $\mathcal{P}_{n+1}$ è coerente se e solo se $p_{n+1} \in [p', p'']$.

Ad esempio, dati due eventi compatibili A, B, con $P(A) = 0.6$, $P(B) = 0.7$, ricordando le (17) e (18), la valutazione di probabilità $(0.6, 0.7, \alpha)$ sulla famiglia $\{A, B, A \vee B\}$ è coerente se e solo se ....

Nell'esempio precedente, posto $P(AB) = \beta$, si può dimostrare che la valutazione $(0.6, 0.7, \beta)$ su $\{A, B, AB\}$ è coerente se e solo se ....

Esempi.

1. Sia $\mathcal{F}_1 = \{E_1\}$, con $P(E_1) = p_1 \in [0,1]$. Dato un evento E_2 incompatibile con E_1, con $E_2 \neq E_1^c$, si ponga $P(E_2) = p_2$.

Per quali valori di p_2 la valutazione (p_1, p_2) su $\mathcal{F}_2 = \{E_1, E_2\}$ è coerente?

Risposta: osservando che $P(E_1 \vee E_2) = p_1 + p_2 \leq 1$, si ha $p_2 \in [0, 1 - p_1]$.

2. Date le probabilità $P(A) = x, P(B) = y$, la valutazione $P(A \vee B) = z$ è un'estensione coerente di (x, y) se e solo se

$$Max\,\{x, y\} \leq z \leq Min\,\{1, x + y\}\,,$$

mentre la valutazione $P(AB) = p$ è un'estensione coerente se e solo se

$$Max\,\{0, x + y - 1\} \leq p \leq Min\,\{x, y\}\,.$$

3. Un sistema S è costituito da 3 moduli M_1, M_2, M_3. Definiti gli eventi $A_i =$ *"il modulo M_i funziona"*, $i = 1, 2, 3$, supponiamo che se M_i funziona allora M_{i+1} funziona, $i = 1, 2$. Posto $P(A_1) = \frac{1}{4}, P(A_3) = \frac{6}{10}$, determinare l'intervallo $[p', p'']$ delle estensioni coerenti $P(A_2) = p$.

Soluzione: Dalla proprietà di monotonia della probabilità segue $P(A_1) \leq P(A_2) \leq P(A_3)$ e quindi $\frac{1}{4} \leq P(A_2) \leq \frac{3}{5}$. Si può verificare che $[p', p''] = [\frac{1}{4}, \frac{3}{5}]$. Infatti, essendo $A_1 \subseteq A_2 \subseteq A_3$, i costituenti sono

$$C_1 = A_1 \wedge A_2 \wedge A_3\,;\ C_2 = A_1^c \wedge A_2 \wedge A_3\,;$$

$$C_3 = A_1^c \wedge A_2^c \wedge A_3\,;\ C_4 = A_1^c \wedge A_2^c \wedge A_3^c.$$

Allora, posto $P(C_h) = \lambda_h$, la coerenza della valutazione

$$P(A_1) = \frac{1}{4}\,,\ P(A_3) = \frac{3}{5}\,,\ P(A_2) = p\,,$$

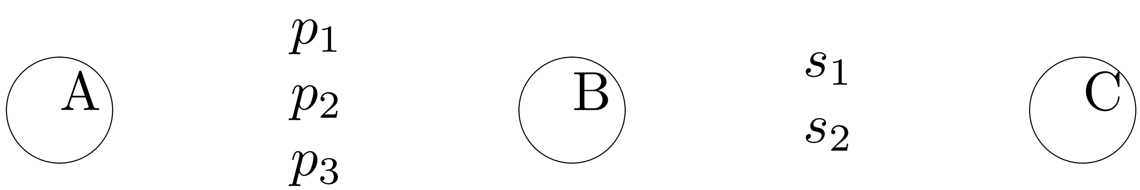
equivale alla risolubilità del sistema

$$\begin{cases} \lambda_1 = \frac{1}{4}\,, \quad \lambda_1 + \lambda_2 = p\,, \quad \lambda_1 + \lambda_2 + \lambda_3 = \frac{3}{5}\,, \\ \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1\,, \quad \lambda_h \geq 0\,,\ \forall\, h\,. \end{cases}$$

Si ha: $p = \frac{1}{4} + \lambda_2 = \frac{3}{5} - \lambda_3\,,\ \lambda_2 + \lambda_3 = \frac{7}{20}$; pertanto, il sistema è risolubile se e solo se $p \in [\frac{1}{4}, \frac{3}{5}]$. In particolare, le soluzioni del sistema sono tutti e soli i vettori $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (\frac{1}{4}, \lambda, \frac{7}{20} - \lambda, \frac{2}{5})$, con $\lambda \in [0, \frac{7}{20}]$.

Richiami di calcolo combinatorio

Esempio 8 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 8 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 (8), indicando con $D'_{n,r}$ il numero di disposizioni con ripetizione si ha

$$D'_{n,r} = n \times \cdots \times n = n^r \, . \tag{24}$$

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

$$D_{n,r} = n(n-1) \cdots (n-r+1) \, . \tag{25}$$

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

$$D_{n,n} = P_n = n \times (n-1) \cdots \times 2 \times 1 = n! \, . \tag{26}$$

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:

$$n! = \begin{cases} n \times (n-1)! & \text{se } n \in \mathcal{N} \\ 1 & \text{se } n = 0. \end{cases} \tag{27}$$

Inoltre :

$$D_{n,r} = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!} \ . \tag{28}$$

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

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 :

$$C_{n,r} = \frac{D_{n,r}}{r!} = \frac{n!}{r!(n-r)!} = \left(\begin{array}{c} n \\ r \end{array} \right) \ . \qquad (29)$$

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 :

$$C'_{n,r} = \left(\begin{array}{c} r+n-1 \\ r \end{array} \right) = \left(\begin{array}{c} r+n-1 \\ n-1 \end{array} \right) .$$

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$\acute{e}$.

... 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

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

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:

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

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

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

Numeri Aleatori

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:

$$X = x_1 \cdot |E_1| + x_2 \cdot |E_2| + \cdots + x_n \cdot |E_n| \, . \qquad (30)$$

X è un numero ben determinato ma incognito.
I possibili valori di X si ottengono assegnando i valori 1 o 0 agli indicatori degli eventi E_i, $i = 1, \ldots, n$, in tutti i modi *possibili*.
X si può anche considerare come una funzione reale definita sull'insieme dei costituenti $C_1, \ldots, C_m$.
L'insieme dei possibili valori di X costituisce il codominio di tale funzione.

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

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

con valori possibili 0 e 1.

Esempio. (*lancio di un dado*) Definiamo gli eventi

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

Risultato aleatorio del lancio:

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

$E_1, E_2, \dots, E_6$ formano una partizione e quindi $1, 2, \dots, 6$ rappresentano i possibili valori di X.

Forma Canonica.
Dato

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

con $\{H_1, H_2, \dots, H_n\}$ partizione di Ω, il codominio di X è $\{x_1, x_2, \dots, x_n\}$.

In generale non è così semplice determinare il codominio di$X$.

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

$$E_1 E_2 = \emptyset\,, \qquad E_3 \subseteq E_1\,,$$

determiniamo i valori possibili del numero aleatorio

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

Costituenti generati da E_1, E_2, E_3 e corrispondenti valori di X:

$$C_1 = E_1 E_2^c E_3\,, \;\; \chi_1 = 3\,; \qquad C_2 = E_1 E_2^c E_3^c\,, \;\; \chi_2 = 2\,;$$

$$C_3 = E_1^c E_2 E_3^c\,, \;\; \chi_3 = -1\,; \quad C_4 = E_1^c E_2^c E_3^c\,, \;\; \chi_4 = 0\,.$$

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

In generale, per ottenere la forma canonica occorre calcolare i costituenti e i corrispondenti valori 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.
Si consideri inoltre il numero aleatorio

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

Al costituente C_h corrisponde per X un ben determinato valore χ_h.

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

Rappresentazione canonica

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

Legame tra i coefficienti x_i e i valori χ_h:

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

Previsione di un N.A.

Dati n eventi $E_1, E_2, \ldots, E_n$, con

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

si consideri 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:

$$\mathbb{P}(X) = p_1x_1 + p_2x_2 + \cdots + p_nx_n \qquad (31)$$

Interpretazione di $\mathbb{P}(X)$ con il criterio della scommessa.

Supponiamo che il n.a. X sia una vincita aleatoria, nel senso che

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

dove X rappresenta l'importo totale che si riceve.

$\mathbb{P}(X)$ rapresenta *la quantità che si deve pagare (ricevere) per ricevere (pagare)* X. Infatti

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

Allora, sommando tali termini

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

Esempio.

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

con $p = P(E)$. Si ha

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

cioè: $\mathbb{P}(|E|) = p = P(E)$.

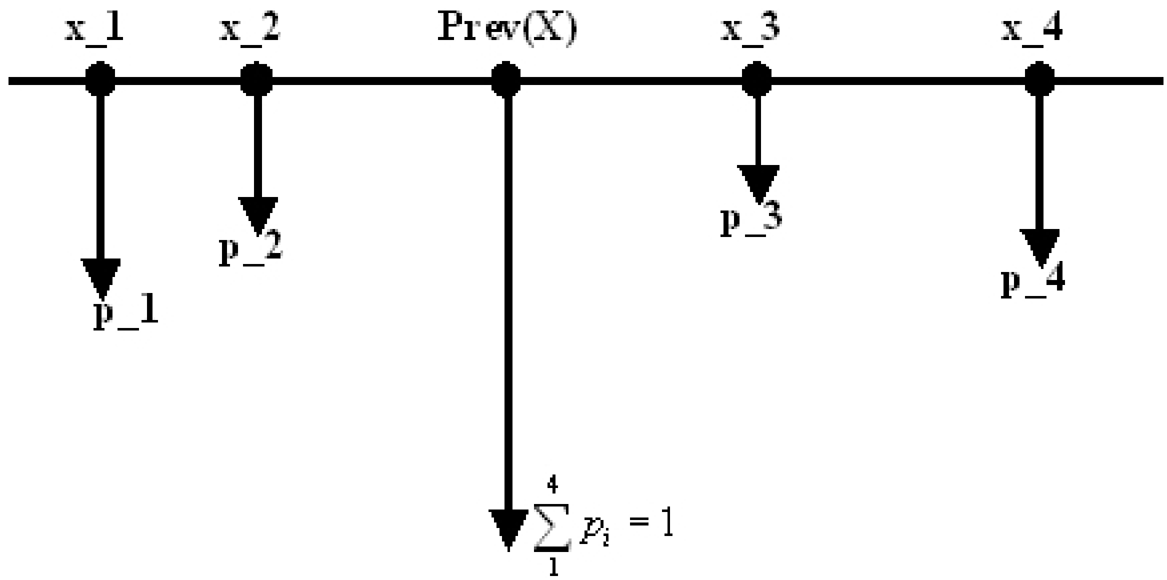
Interpretazione meccanica.
Sia data una *partizione* di Ω, $\{H_1, \ldots, H_n\}$, con

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

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 masse $p_1, \ldots, p_n$, collocate in n punti di ascissa $x_1, \ldots, x_n$.



Osservazione. La previsione gode della seguente proprietà

$$\min(X) \leq \mathbb{P}(X) \leq \max(X). \tag{32}$$

Dim. Consideriamo solo il caso

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

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

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

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

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

Combinazioni lineari di n.a. Siano dati due numeri aleatori

$$X = x_1|E_1| + x_2|E_2| + \cdots + x_n|E_n|$$
$$Y = y_1|A_1| + y_2|A_2| + \cdots + y_n|A_m|$$

e siano a e b due numeri reali arbitrari. La quantità $Z = aX + bY$ 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

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

Linearità della Previsione. Dati X, Y, a, b si ha

$$\mathbb{P}(aX + bY) = a\mathbb{P}(X) + b\mathbb{P}(Y). \tag{33}$$

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 $\mathbb{P}(b) = b$. Infatti, b si può interpretare come il numero aleatorio $Y = b|\Omega|$ e quindi

$$\mathbb{P}(Y) = b \cdot P(\Omega) = b \cdot 1 = b\,.$$

Casi particolari:

$$\mathbb{P}(aX) = a\mathbb{P}(X), \qquad \mathbb{P}(b) = b,$$

$$\mathbb{P}(aX + b) = a\mathbb{P}(X) + b, \quad \mathbb{P}(-X) = -\mathbb{P}(X).$$

Dati n n.a. $X_1, X_2, \ldots, X_n$, dalla (33) si ottiene

$$\begin{aligned}&\mathbb{P}(X_1 + \cdots + X_n) = \mathbb{P}(X_1 + \cdots + X_{n-1}) + \mathbb{P}(X_n) = \\ &= \cdots = \mathbb{P}(X_1) + \mathbb{P}(X_2) + \cdots + \mathbb{P}(X_n).\end{aligned}$$

Più in generale, dati n numeri reali arbitrari $a_1, a_2, \ldots, a_n$, si ha

$$\mathbb{P}(\sum_{i=1}^{n} a_i X_i) = \mathbb{P}(\sum_{i=1}^{n-1} a_i X_i) + a_n \mathbb{P}(X_n) = \cdots = \sum_{i=1}^{n} a_i \mathbb{P}(X_i)\,.$$

Esempio. (n *estrazioni con restituzione da un urna di composizione nota, contenente palline bianche e nere*) Definiamo

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

Per convenzione, diciamo che si ha un "successo" (oppure "insuccesso") nell'i-esima prova se E_i è vero (oppure falso). Il numero aleatorio

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

rappresenta il "numero di successi su n prove".
Se assumiamo gli eventi E_i equiprobabili, con

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

si ottiene

$$\mathbb{P}(X) = \mathbb{P}(|E_1| + |E_2| + \cdots + |E_n|) = \cdots = np\,.$$

Per il numero aleatorio

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

che rappresenta il numero medio di successi o frequenza relativa di successo, si ha $\mathbb{P}(Z) = \mathbb{P}(\frac{X}{n}) = p$.
(*la previsione della frequenza relativa coincide con la probabilità di successo in ogni prova*)

Probabilità dell'unione di n *eventi: principio di inclusione-esclusione.*

Dati n eventi $E_1, \ldots, E_n$ e supposto che siano state assegnate le probabilità

$$P(Ei_1 Ei_2 \cdots Ei_h)\,,\ \forall \{i_1, i_2, \ldots, i_h\} \subseteq \{1, 2, \ldots, n\},$$

con $1 \leq h \leq n$, calcoliamo la probabilità dell'evento $A = E_1 \vee \cdots \vee E_n$. Si ha

$$A^c = E_1^c E_2^c \cdots E_n^c\,, \quad P(A) = 1 - P(A^c) = 1 - \mathbb{P}(|A^c|)\,;$$

$$|A^c| = |E_1^c| \cdots |E_n^c| = (1 - |E_1|) \cdots (1 - |E_n|) =$$

$$= 1 - \sum_i |E_i| + \sum_{\{i,j\}} |E_i E_j| - \sum_{\{i,j,k\}} |E_i E_j E_k| + \cdots$$

$$\cdots + (-1)^n |E_1 E_2 \cdots E_n|\,.$$

Allora, posto

$$s_1 = \sum_i \mathbb{P}(|E_i|) = \sum_i P(E_i)\,,$$

$$s_2 = \sum_{\{i,j\}} \mathbb{P}(|E_i E_j|) = \sum_{\{i,j\}} P(E_i E_j)\,,$$

$$s_3 = \sum_{\{i,j,k\}} \mathbb{P}(|E_i E_j E_k|) = \sum_{\{i,j,k\}} P(E_i E_j E_k)\,,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$s_n = \mathbb{P}(|E_1 E_2 \cdots E_n|) = P(E_1 E_2 \cdots E_n)\,,$$

dalla linearità della previsione segue

$$\mathbb{P}(|A^c|) = P(A^c) = 1 - s_1 + s_2 - s_3 + \cdots + (-1)^n s_n =$$

$$= \sum_{h=0}^n (-1)^h s_h\,,$$

dove per convenzione $s_0 = 1$, e quindi

$$P(A) = \sum_{h=1}^n (-1)^{h+1} s_h = \sum_i P(E_i) - \sum_{\{i,j\}} P(E_i E_j) +$$

$$+ \sum_{\{i,j,k\}} P(E_i E_j E_k) - \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n)\,.$$

Problema delle concordanze.
Supponiamo di dover inserire n lettere in n buste, ognuna con un proprio indirizzo. Se le lettere vengono inserite a caso nelle buste, qual'è la probabilità p che vi siano zero concordanze, cioè che nessuna lettera capiti nella propria busta?
Definito l'evento
$A =$ *"almeno una lettera capita nella sua busta"*,
si ha:

$$P(E_i) = \frac{1}{n}\,,\ P(E_iE_j) = \frac{1}{n(n-1)} = \frac{(n-2)!}{n!}\,,$$

$$P(E_iE_jE_k) = \frac{1}{n(n-1)(n-2)} = \frac{(n-3)!}{n!}\,,\ \cdots$$

$\cdots\ ,\ P(E_1E_2\cdots E_n) = \frac{1}{n!}$, da cui segue

$$s_1 = \sum_i P(E_i) = \sum_i \frac{1}{n} = 1\,,$$

$$s_2 = \sum_{\{i,j\}} P(E_iE_j) = \sum_{\{i,j\}} \frac{(n-2)!}{n!} = \binom{n}{2} \cdot \frac{(n-2)!}{n!} = \frac{1}{2}\,,$$

$$s_3 = \sum_{\{i,j,k\}} P(E_iE_jE_k) = \sum_{\{i,j,k\}} \frac{(n-3)!}{n!} =$$

$$= \binom{n}{3} \cdot \frac{(n-3)!}{n!} = \frac{1}{3!}\,,\ \cdots\cdots\ ,\ s_n = P(E_1E_2\cdots E_n) = \frac{1}{n!}\,.$$

Allora, per n abbastanza grande (in questo caso basta $n > 7$) si ottiene

$$p = P(A^c) = \sum_{h=0}^{n}(-1)^h s_h = \sum_{h=0}^{n} \frac{(-1)^h}{h!} \simeq$$

$$\simeq \sum_{h=0}^{+\infty} \frac{(-1)^h}{h!} = e^{-1}\,.$$

Pertanto, la probabilità di zero concordanze è all'incirca uguale all'inverso del numero di Nepero ($e^{-1} \simeq 0.367879$).

Se indichiamo con X il numero aleatorio di concordanze, si ha

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

ed essendo $P(E_i) = \frac{1}{n}$, segue: $\mathbb{P}(X) = 1$.

Nota: $X \in \{0, 1, \ldots, n-2, n\}$, con

$$P(X = k) = \cdots = \frac{1}{k!} \sum_{h=0}^{n-k} \frac{(-1)^h}{h!}\,.$$

Varianza.

La previsione $\mathbb{P}(X)$ si può interpretare come il valore centrale della distribuzione di probabilità di X e in alcuni casi la sua conoscenza è sufficiente per sintetizzare la distribuzione di probabilità (come accade per il baricentro nel caso di un sistema di masse).

In generale, assegnare solo la previsione non è sufficiente. Un altro aspetto importante è quello di descrivere la dispersione della distribuzione di probabilità.

Esempio: Enzo mangia 2 polli e Marco non ne mangia nessuno . . . . "in media" ne mangiano uno a testa!

Per misurare la dispersione della distribuzione di probabilità di X (attorno al suo valor medio $\mathbb{P}(X)$), si introduce il concetto di *varianza*.

Definizione. Dato un numero aleatorio X, con $\mathbb{P}(X) = m$, si definisce *Varianza* di X, indicata con $Var(X)$ oppure σ_X^2, la seguente quantità

$$Var(X) = \mathbb{P}[(X - m)^2]\,. \tag{34}$$

Indicando con $x_1, x_2, \ldots, x_n$ **i possibili valori** di X, si ha

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

Posto $P(X = x_i) = p_i$, segue

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

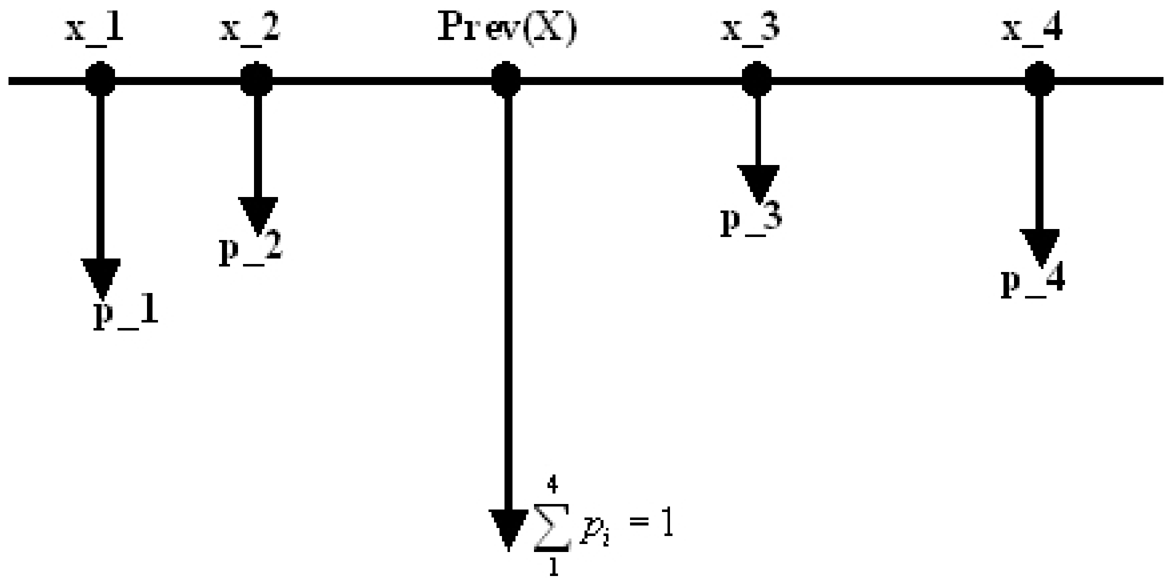
Posto $Y = (X - m)^2$, si ha

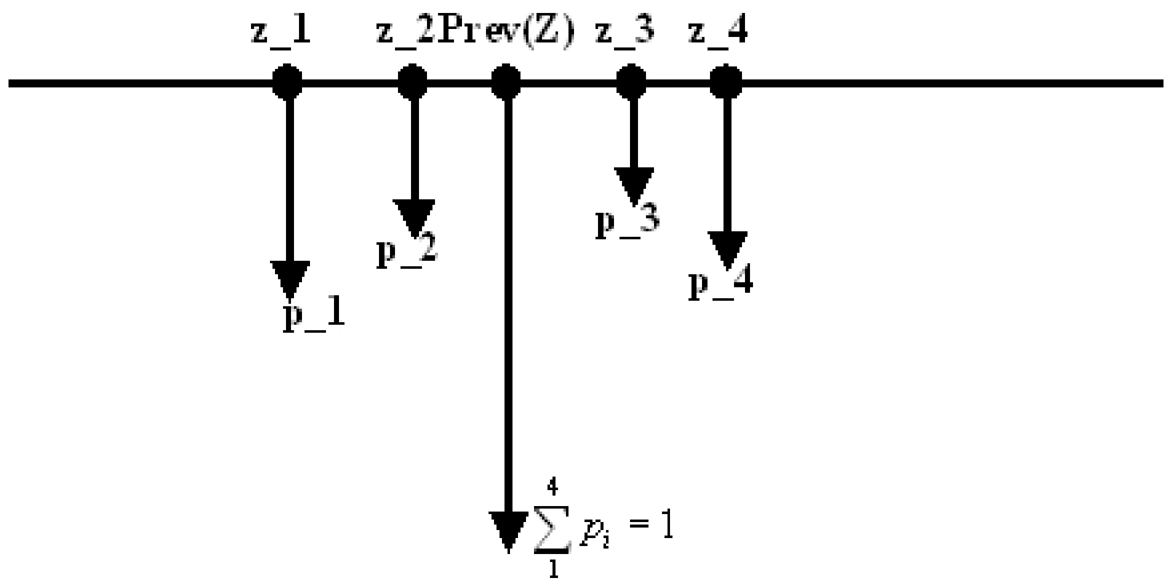
$$Y = (x_1 - m)^2|X = x_1| + \cdots + (x_n - m)^2|X = x_n|\,.$$

Allora,

$$\begin{aligned}
&Var(X) = \mathbb{P}[(X - m)^2] = \mathbb{P}(Y) = \\
&= p_1(x_1 - m)^2 + p_2(x_2 - m)^2 + \cdots + p_n(x_n - m)^2 = \\
&= \textstyle\sum_{i=1}^{n} p_i(x_i - m)^2\,.
\end{aligned}$$

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





Osservazione:

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

Dal punto di vista meccanico, la $Var(X)$ coincide con il momento d'inerzia rispetto al baricentro (com'è noto, tale momento d'inerzia misura la dispersione di massa rispetto al baricentro).

Scarto quadratico medio (deviazione standard).

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathbb{P}[(X-m)^2]}\,. \qquad (35)$$

Esempio. Siano dati tre numeri aleatori X, Y, Z, 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

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

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

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

Si ha $\mathbb{P}(X) = \mathbb{P}(Y) = \mathbb{P}(Z) = 0$, mentre per le varianze risulta

$$\sigma_X^2 = \tfrac{1}{2}(-1-0)^2 + \tfrac{1}{2}(1-0)^2 = 1\,,$$
$$\sigma_Y^2 = \cdots = \tfrac{5}{2}\,, \qquad \sigma_Z^2 = \cdots = 4\,.$$

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

Esempio. ... popolazione statistica costituita da due individui: uno mangia 2 polli (supponiamo, a settimana) e l'altro 0 polli.
X = numero (aleatorio) di polli mangiati ogni settimana dalla persona estratta.

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

... altra popolazione statistica, in cui entrambi gli individui mangiano 1 pollo a settimana.
Y = numero (aleatorio) di polli mangiati settimanalmente dalla persona estratta.

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

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 ha $Var(|E|) = pq$. Infatti

$$\sigma_{|E|}^2 = \mathbb{P}[(|E| - p)^2] = \cdots = pq\,.$$

Nota: $Var(|E|) = pq \; \leq \; \frac{1}{4}\,.$

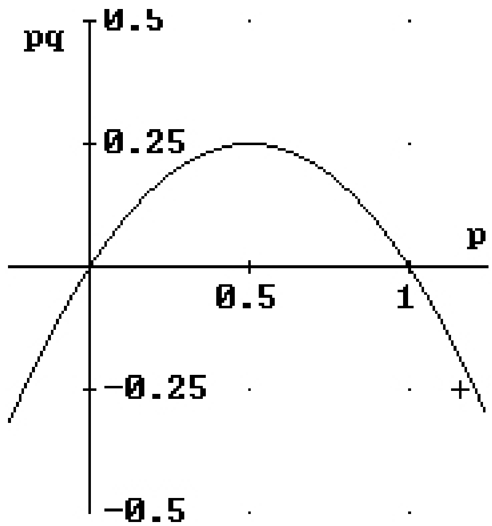


Figura 1: $Var(|E|)$

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

Dim. della 6 ($m = \mathbb{P}(X)$):

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

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

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

$$P(X \geq \alpha) \leq \frac{m}{\alpha}\,. \tag{36}$$

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

$$P(|X - m| \geq k\sigma) \leq \frac{1}{k^2}\,, \quad \forall k \in \mathbb{R} \;\; (k > 1). \tag{37}$$

Covarianza Ricordiamo che, dati due numeri aleatori X e Y, con $\mathbb{P}(X) = m_X$, $\mathbb{P}(Y) = m_Y$, si ha

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

Calcoliamo $Var(X + Y)$.

$$\begin{aligned} & Var(X + Y) = \mathbb{P}\{[(X + Y) - (m_X + m_Y)]^2\} = \\ & = \cdots\cdots\cdots\cdots = \\ & = Var(X) + Var(Y) + 2\mathbb{P}[(X - m_X)(Y - m_Y)] \, . \end{aligned} \tag{38}$$

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

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y) \, . \tag{39}$$

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

Osserviamo che

$$
\begin{aligned}
Cov(X, Y) &= \mathbb{P}[(X - m_X)(Y - m_Y)] = \\
&= \cdots = \mathbb{P}(XY) - \mathbb{P}(X)\mathbb{P}(Y)\,.
\end{aligned} \tag{40}
$$

In particolare

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

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

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

dove X *grande* e Y *grande* significano $X > \mathbb{P}(X)$ e $Y > \mathbb{P}(Y)$ e, analogamente, X *piccolo* e Y *piccolo* significano $X < \mathbb{P}(X)$ 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. n palline vengono distribuite a caso in 2 scatole. Ogni pallina ha 2 alternative e quindi ci sono 2^n ripartizioni possibili, ognuna di probabilità $\frac{1}{2^n}$. Siano
$X =$ "numero di palline nella prima scatola"
$Y =$ "numero di scatole non vuote".
Si ha:

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

Per ogni fissato h, si ha

$$P(X = h) = P(X = n - h) = \frac{\binom{n}{h}}{2^n}.$$

Inoltre

$$P(Y = 1) = P(X = 0) + P(X = n) = \tfrac{2}{2^n} = \tfrac{1}{2^{n-1}},$$

$$P(Y = 2) = 1 - \tfrac{1}{2^{n-1}}.$$

Indicando con E_i l'evento *"la i-ma pallina capita nella prima scatola"*, $i = 1, \dots, n$, si ha $P(E_i) = \frac{1}{2}$, $\forall\, i$.

Inoltre

$$X = \sum_{i=1}^{n} |E_i| = \sum_{h=0}^{n} h \cdot |X = h| \,,$$

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

e quindi

$$\mathbb{P}(X) = \sum_{i=1}^{n} \mathbb{P}(|E_i|) = \frac{n}{2} \,.$$

In alternativa:

$$\mathbb{P}(X) = \sum_{h=0}^{n} h \frac{\binom{n}{h}}{2^n} = \frac{n}{2} \,.$$

$$\mathbb{P}(Y) = \frac{1}{2^{n-1}} + 2(1 - \frac{1}{2^{n-1}}) = 2 - \frac{1}{2^{n-1}} \,.$$

Quindi

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

Valori possibili per (X, Y):

$$(0,1)\,,\ (1,2)\,,\ \cdots\,,\ (n-1,2)\,,\ (n,1)\,.$$

Valori possibili per XY:

$$0\,,\ 2\,,\ 4\,,\ \cdots\,,\ 2(n-1)\,,\ n\,.$$

Quindi, la previsione di XY è uguale a

$$0 \cdot \tfrac{1}{2^n} + 2 \cdot \frac{\binom{n}{1}}{2^n} + \cdots + 2(n-1) \cdot \frac{\binom{n}{n-1}}{2^n} + n \cdot \tfrac{1}{2^n} =$$

$$= 2 \cdot \textstyle\sum_{h=0}^{n} h \cdot \dfrac{\binom{n}{h}}{2^n} - 2n \cdot \dfrac{\binom{n}{n}}{2^n} + n \cdot \tfrac{1}{2^n} =$$

$$= 2 \cdot \tfrac{n}{2} - n \cdot \tfrac{1}{2^n} = n - \tfrac{n}{2^n} = \mathbb{P}(X)\mathbb{P}(Y)\,.$$

Pertanto $Cov(X, Y) = 0$.

Esempio. . . . esempio precedente con 2 palline da ripartire a caso in n scatole.
Verificare che i valori possibili di X sono 0, 1, 2, con rispettive probabilità

$$\frac{(n-1)^2}{n^2}, \quad \frac{2(n-1)}{n^2}, \quad \frac{1}{n^2};$$

che i valori possibili di Y sono 1, 2, con

$$P(Y=1) = \frac{n}{n^2} = \frac{1}{n}, \quad P(Y=2) = 1 - \frac{1}{n}.$$

Quindi

$$\mathbb{P}(X) = \frac{2}{n}, \quad \mathbb{P}(Y) = 2 - \frac{1}{n}, \quad \mathbb{P}(X)\mathbb{P}(Y) = \frac{4}{n} - \frac{2}{n^2}.$$

Inoltre, i valori possibili per (X, Y) sono:

$$(0,1), \quad (0,2), \quad (1,2), \quad (2,1),$$

con rispettive probabilità

$$\frac{n-1}{n^2}, \quad \frac{(n-1)(n-2)}{n^2}, \quad \frac{2(n-1)}{n^2}, \quad \frac{1}{n^2}.$$

Infine, i valori possibili per XY sono 0, 2, con

$$P(XY = 0) = \frac{(n-1)^2}{n^2}\,, \qquad P(XY = 2) = \frac{2n-1}{n^2}\,,$$

e quindi

$$\mathbb{P}(XY) = \frac{4}{n} - \frac{2}{n^2} = \mathbb{P}(X)\mathbb{P}(Y)\,.$$

Pertanto $Cov(X, Y) = 0$.

Incorrelazione.
X ed Y si dicono *incorrelati* se $Cov(X, Y) = 0$.

Se X, Y sono 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

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

ρ_{XY} si chiama anche covarianza normalizzata in quanto

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

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

Osservazione. Posto

$$X' = aX+b\,,\;\; Y' = cY+d\,,\;\; a,b,c,d \in \mathbb{R}\,, a \neq 0,\; c \neq 0\,,$$

si ha

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

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 corrisponde al caso di dipendenza lineare tra X e Y. Per dimostrare tali proprietà, osserviamo che:

- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$,

- $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$,

- $Var(aX + cY) =$

  $= a^2Var(X) + c^2Var(Y) + 2acCov(X,Y)$.

(Dim. della 2.) Si ha

$$Var(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}) = \frac{\sigma_X^2}{\sigma_X^2} + \frac{\sigma_Y^2}{\sigma_Y^2} + 2Cov(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}) =$$

$$= 2 + 2\frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 2(1 + \rho_{XY}) \geq 0\,;$$

$$Var(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) = 2(1 - \rho_{XY}) \geq 0\,;$$

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

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

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

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

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

ovvero

$$Y = m_Y - \tfrac{\sigma_Y}{\sigma_X}(X - m_X) \quad = \underbrace{-\frac{\sigma_Y}{\sigma_X}}X + \underbrace{m_Y + \frac{\sigma_Y}{\sigma_X}m_X}$$

$Y = aX + b$ con $\qquad a = -\frac{\sigma_Y}{\sigma_X}\,, b = m_Y - am_X$

oppure

$X = cY + d$ con $\qquad c = -\frac{\sigma_X}{\sigma_Y}\,, d = m_X - cm_Y$

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. Dati due eventi A e B, si ha

$$Cov(|A|, |B|) = \mathbb{P}(|AB|) - \mathbb{P}(|A|)\mathbb{P}(|B|) =$$

$$= P(AB) - P(A)P(B)\,.$$

In particolare, dato un evento E, con $P(E) = p$, posto $q = 1 - p$ ed osservando che $|E^c| = 1 - |E|$, si ha

$$Cov(|E|, |E^c|) = -Var(|E|) = -pq\,, \quad \rho_{|E||E^c|} = -1\,.$$

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: $Cov(|A|, |B|) = -\frac{1}{4}$.
. . . effettuando 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-covarianze.
Dati $n + m$ numeri aleatori

$$X_1,\, X_2,\, \ldots,\, X_n,\, Y_1,\, Y_2,\, \ldots,\, Y_m,$$

consideriamo le combinazioni lineari:

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

con

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

Posto $Cov(X_i, Y_j) = \sigma_{X_i Y_j}$, si ha

$$Cov(X, Y) = Cov(\textstyle\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) =$$

$$= \cdots = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \sigma_{X_i Y_j}.$$

In particolare

$$Cov(X, X) = Var(X) = \textstyle\sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + 2 \sum_{i<j} a_i a_j \sigma_{X_i X_j}. \tag{41}$$

Indicando con $\mathbf{a}^t = (a_1, a_2, \ldots, a_n)$ il vettore dei coefficienti e con Σ la seguente matrice (delle varianze-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 (41) nella seguente forma matriciale

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

$$= (a_1, a_2, \ldots, a_n) \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{42}$$

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

$$f_n = \frac{|E_1| + |E_2| + \cdots + |E_n|}{n} \tag{43}$$

la frequenza relativa di successo. Si ha

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

$$Var(f_n) = Var\left(\frac{\sum_{i=1}^{n} |E_i|}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} p_i q_i \le \frac{1}{4n} .$$

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 eventi E_i sono equiprobabili, con $P(E_i) = p$, si ha:

$$\lim_{n \longrightarrow \infty} P(|f_n - p| > \varepsilon) = 0.$$