Analisi Dati con Excel

Giovanni Della Lunga

giovanni.dellalunga@gmail.com

La prima regola di ogni tecnologia è che l'automazione applicata ad un'operazione efficiente ne aumenterà l'efficienza. La seconda è che l'automazione applicata ad un'operazione inefficiente ne aumenterà l'inefficienza.

Bill Gates

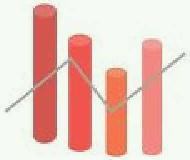
Richiami di Probabilità e Statistica

Fondamenti per l’Analisi e l’Interpretazione dei Dati

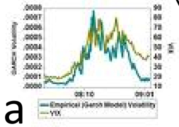
Premessa

Probabilità

Statistica



Econometria



La **probabilità** è un concetto matematico utilizzato per affrontare le condizioni di incertezza.

In probabilità si considera un fenomeno osservabile esclusivamente dal punto di vista della possibilità o meno del suo verificarsi, prescindendo dalla sua natura. Tra due estremi, detti *evento certo* ed *evento impossibile* si collocano eventi più o meno probabili.

La **statistica** è la scienza che ha come fine lo studio quantitativo e qualitativo di un "collettivo".

Occorre distinguere tra il calcolo delle probabilità inteso come modello matematico e la statistica come disciplina. I modelli matematici, infatti, sono concepiti dal matematico come modelli astratti, *modelli*, appunto, mentre è compito dello statistico mettere in relazione questi modelli con i dati.

L'**econometria** può essere definita una branca della statistica che si occupa dell'analisi dei fenomeni economici; in alternativa, può considerarsi un settore dell'economia dedicato alla verifica empirica di modelli formulati in ambito teorico.

Premessa

- La statistica è una disciplina che si occupa della raccolta, analisi, interpretazione, presentazione e organizzazione dei dati.
- Il suo scopo principale è quello di estrarre informazioni significative da dati grezzi per supportare il **processo decisionale in presenza di incertezza**.
- La statistica si divide in due rami principali: la **statistica descrittiva** e la **statistica inferenziale**.

Premessa

Statistica Descrittiva

- La statistica descrittiva si concentra sulla descrizione e sintesi dei dati.
- Utilizza misure come la media, la mediana, la moda, la varianza e la deviazione standard per riassumere le caratteristiche principali di un insieme di dati.
- Viene spesso rappresentata attraverso tabelle, grafici e diagrammi per facilitare la comprensione delle informazioni.

Premessa

Statistica Inferenziale

- La statistica inferenziale, d'altra parte, **utilizza i dati di un campione per fare inferenze e previsioni riguardanti una popolazione più ampia**.
- Questo ramo della statistica si basa sull'analisi dei campioni e l'uso della teoria delle probabilità per trarre conclusioni significative.
- Gli strumenti principali della statistica inferenziale includono le stime puntuali e intervallari, i test di ipotesi e la costruzione di modelli statistici come la regressione.

Premessa

LA STATISTICA
di Trilussa

Sai ched'è la statistica? È na' cosa
che serve pe fà un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che spósa.

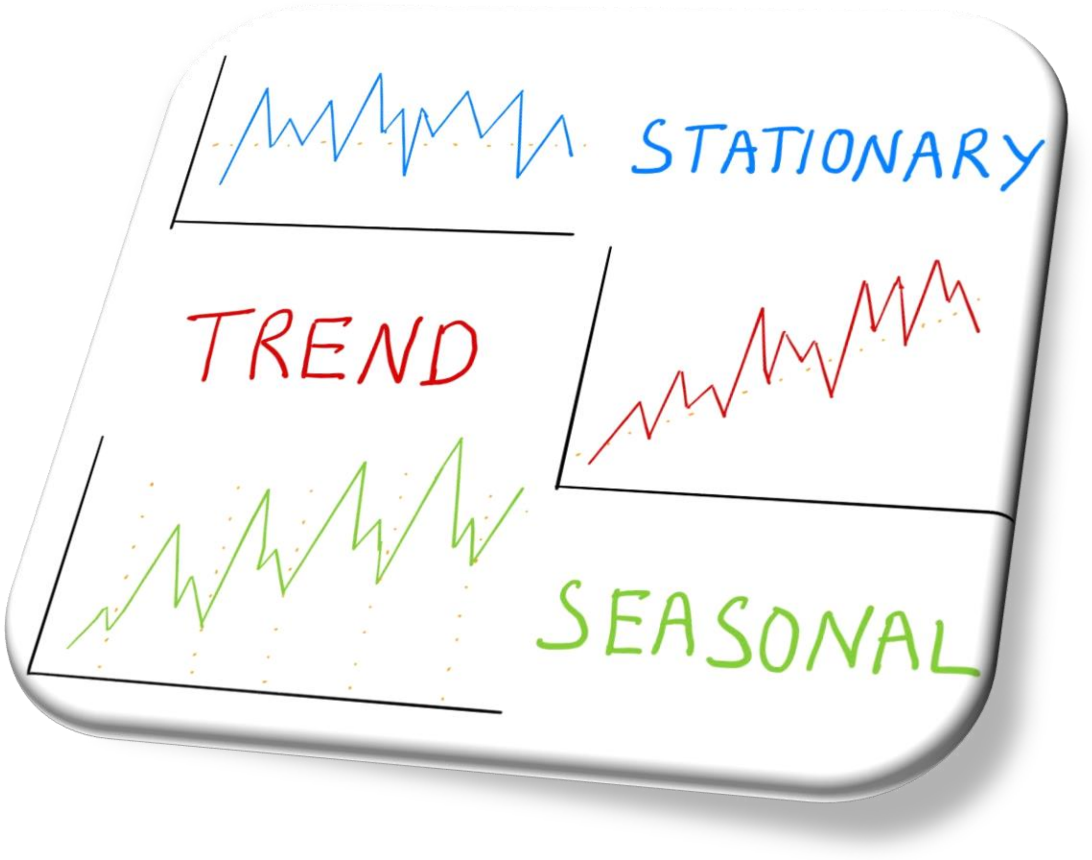
Ma pè me la statistica curiosa
è dove c'entra la percentuale,
pè via che, lì,la media è sempre eguale
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:

e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perch'è c'è un antro che ne magna due.

Premessa

- In statistica, una serie storica (o temporale) esprime la dinamica di un certo fenomeno nel tempo. Le serie storiche vengono studiate sia per interpretare un fenomeno, individuando componenti di trend, di ciclicità, di stagionalità e/o di accidentalità, sia per prevedere il suo andamento futuro.
- In generale, per serie si intende la classificazione di diverse osservazioni di un fenomeno rispetto ad un carattere qualitativo. Se tale carattere è il tempo, la serie viene detta *storica* o *temporale*.
- Il fenomeno osservato, detto *variabile*, può essere osservato in dati istanti di tempo (*variabile di stato*: numero dei dipendenti di un'azienda, quotazione di chiusura di un titolo negoziato in borsa, livello di un tasso di interesse ecc.) o alla fine di periodi di lunghezza definita (*variabili di flusso*: vendite annuali di un'azienda, PIL trimestrale ecc.).



Probabilità

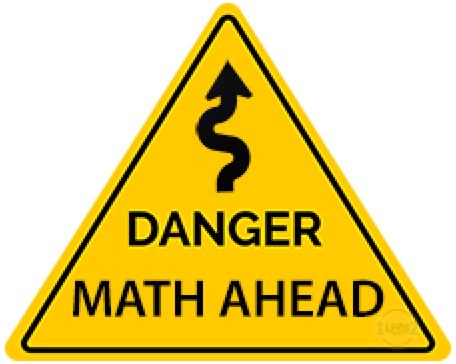
Concetti Fondamentali e Applicazioni

Caso e Probabilità

- Nella vita di tutti i giorni incontriamo continuamente situazioni incerte, fuori delle nostre capacità di controllo e previsione.
- In queste circostanze usiamo parole come caso, incertezza e aleatorietà. La scienza che tratta questi problemi è la **teoria della probabilità**, una branca della matematica nata molto tardi (gli albori sono all'inizio del XVII secolo).
- Nonostante la probabilità sia onnipresente nella scienza e nella tecnologia, la sua conoscenza non è affatto diffusa.
- In genere la probabilità viene presentata come una scienza approssimativa e circondata da un'aura di ingiustificato mistero.

- Ogni tentativo di dare una definizione rigorosa dei concetti probabilistici più elementari si trova di fronte ad un problema; infatti, non solo esistono differenti formalizzazioni e assiomatizzazioni della probabilità ma a queste corrispondono, in generale, molteplici nozioni intuitive di probabilità spesso assai diverse fra loro.

- Al di là delle differenze di carattere formale un elemento comune posseduto da tutte le forme di probabilità riguarda il suo significato intuitivo di valutazione della ***possibilità*** che un dato evento possa accadere o meno.




OGGI MI SONO ISCRITTA A UN NUOVO CORSO
SI CHIAMA "CONOSCERE LA MATEMATICA"
E' DIFFICILE?
PER NIENTE...
SONO ENTRATA E HO DETTO "BUONGIORNO MATEMATICA, MOLTO PIACERE!"
9/30
SCHULZ
www.snoopy.com

PROBABILITA' | Definizioni

- Sia nelle scienze naturali sia in quelle economiche si è soliti assumere che un certo **evento** sia il risultato di un ipotetico *esperimento* intendendo con questo termine l'insieme di tutte "le azioni e le condizioni ambientali che conducono al determinarsi di un fatto".

- E' un esperimento la misura di una grandezza fisica, il lancio di un dado o di una moneta, il verificarsi o meno di un particolare stato di natura (es. l'indice MIB30 supera il livello 50.000).

- Si usa il linguaggio della **teoria degli insiemi**: un insieme non vuoto Ω ha come elementi tutti i risultati possibili di un esperimento; l'evento che risulta verificato da un unico risultato (un unico elemento di Ω) viene detto *evento elementare*; altri eventi sono sottoinsiemi di Ω costituiti da più risultati.

Esempio – *consideriamo l'esperimento aleatorio per antonomasia: il lancio di un dado. In questo caso lo spazio campione è formato dall'insieme dei sei numeri che possono risultare dal lancio stesso*

$$\Omega = \{1,2,3,4,5,6\}$$

Vediamo il significato di alcuni elementi dello spazio degli eventi

Ad esempio l'elemento

$$\omega_1 = \{1,2\}$$

corrisponde all'evento "il numero risultante dal lancio è minore o uguale a 2".

Altri elementi sono

$$\omega_2 = \{1,3,5\}$$

vale a dire "il numero risultante è dispari", e

$$\omega_3 = \{2,4,6\}$$

cioè "il numero uscente è pari".

Definizioni

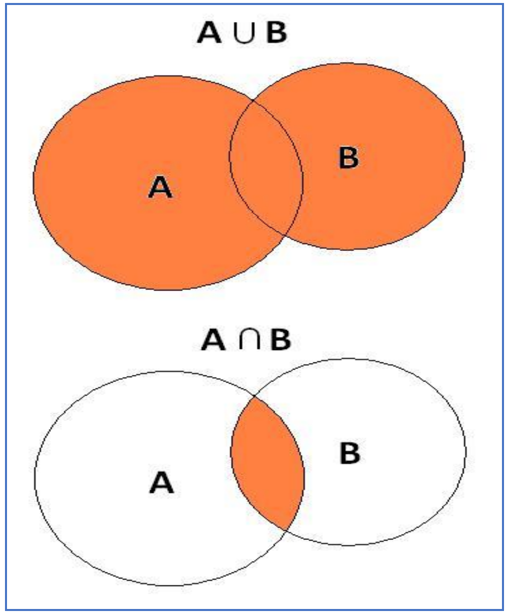
» Gli eventi vengono normalmente indicati con lettere maiuscole.

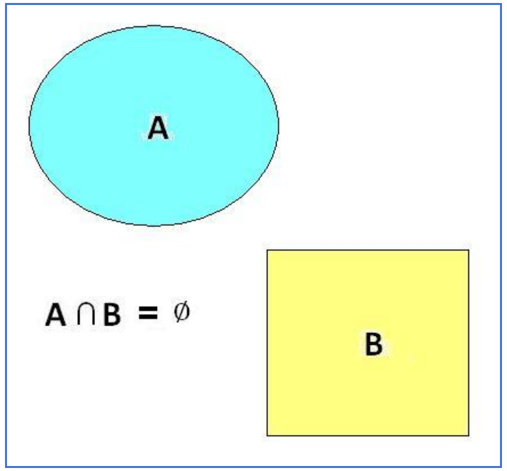
» Dati due eventi A e B, si indica con $A \cup B$ la loro unione, ovvero l'evento costituito dal verificarsi dell'evento A oppure dell'evento B.

» Si indica con $A \cap B$ la loro intersezione, ovvero l'evento costituito dal verificarsi sia dell'evento A che dell'evento B.

» Se $A \cap B = \emptyset$ i due eventi A e B vengono detti *incompatibili* (non possono verificarsi simultaneamente).

» Il complemento di un evento A rispetto a Ω, $\Omega \setminus A$, è detto *negazione* di A e indica il suo non verificarsi (ovvero il verificarsi dell'evento complementare).





PROBABILITA' | Definizione Classica

- Secondo la prima definizione di probabilità, per questo detta *classica*, la probabilità di un evento è ***il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili, purché questi ultimi siano tutti equiprobabili***.
- Questa definizione è spesso attribuita a Pierre Simon Laplace e quindi anche identificata *definizione classica di Laplace*.

- Indicando con Ω l'insieme di casi possibili e con $|\Omega|=n$ la sua cardinalità, con A un evento e con n_A il numero dei casi favorevoli ad A (ad esempio, nel lancio di un dado $\Omega=\{1,2,3,4,5,6\}$, $n = 6$, A = "numero pari", $n_A = 3$), la probabilità di A, indicata con $P(A)$, è pari a:

$$P_A = \frac{n_A}{n} = \frac{3}{6} = \frac{1}{2}$$

» Dalla definizione seguono tre regole:

- la probabilità di un evento è un numero compreso tra 0 e 1;
- la probabilità dell'evento certo è pari a 1; se *A* = "numero compreso tra 1 e 6", $n_A = 6$ e $n_A/n = 1$;
- la probabilità del verificarsi di uno di due eventi incompatibili, ovvero di due eventi che non possono verificarsi simultaneamente, ***è pari alla somma delle probabilità dei due eventi***; se *A* = "numero pari", con *P*(*A*) = 1/2, e *B* = "esce il 3", con *P*(*B*) = 1/6, la probabilità che tirando un dado si ottenga un numero pari oppure un 3 è:

$$P(A \cup B) = \frac{n_{A \cup B}}{n} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

PROBABILITA' | Definizioni

La probabilità $P(A)$ di un evento A è un numero compreso tra zero e uno:

$$0 \leq P(A) \leq 1.$$

Il 100% di probabilità corrisponde ad 1, cioè l'evento accade con certezza, mentre quando l'evento è impossibile la sua probabilità è zero. Conoscendo $P(A)$ si ha anche la probabilità che l'evento A non accada:

$$P(\textit{qualcosa di diverso da } A) = 1 - P(A).$$

Ad esempio se A è l'evento che appaia il numero 3 nel lancio di un dado, poiché $P(A) = 1/6$, allora la probabilità che esca un numero diverso è $1 - 1/6 = 5/6$. Dunque dato un evento A, possiamo calcolare la probabilità anche dell'evento (*qualcosa di diverso da* A), anche detto *evento complementare* ad A.

Analogamente, dati due eventi A e B, possiamo considerare altri eventi, costruiti a partire da questi: possiamo indicare con (A *oppure* B) l'evento che si verifichi o l'uno o l'altro; con (A *e* B) l'evento che si verifichino entrambi.

Se la probabilità che si verifichino entrambi è zero, $P(A \text{ e } B) = 0$, allora i due eventi sono detti *mutuamente incompatibili*, ovvero non possono accadere entrambi e l'evento (A *e* B) è dunque impossibile. Per esempio se consideriamo il lancio di un dado, l'evento che esca 6 e quello che esca 2 sono tra loro incompatibili, perché il lancio ammette un unico risultato. Ci si può convincere che se due eventi sono mutuamente incompatibili, allora

$$P(A \text{ oppure } B) = P(A)+P(B) \text{ (per eventi incompatibili).}$$

Prendendo l'esempio precedente del lancio del dado, in cui A è l'evento che appaia il numero 6 e B è l'evento che appaia il numero 2, allora $P(A \textit{ oppure } B) = 1/3$, poiché sia $P(A)$ che $P(B)$ sono pari ad 1/6 e dunque la loro somma fa 2/6 = 1/3.

Cosa succede però se gli eventi A e B sono *compatibili*, ovvero possono in principio avvenire entrambi? (ad esempio siano l'evento A l'uscita del 2 e l'evento B l'uscita di un numero pari qualsiasi: i due eventi non si escludono a vicenda, perché 2 è un numero pari). In questo caso non è difficile mostrare, facendo ricorso all'interpretazione classica della probabilità (ovvero il rapporto tra il numero dei casi favorevoli al verificarsi di un evento diviso il numero dei casi possibili), che vale la formula più generale:

$$P(A \textit{ oppure } B) = P(A) + P(B) - P(A \textit{ e } B),$$

che si riduce alla formula precedente nel caso di eventi incompatibili – per cui $P(A \textit{ e } B) = 0$.

Definizione Classica

» La definizione classica consente di calcolare effettivamente la probabilità in molte situazioni. Inoltre, **è una definizione operativa** e **fornisce quindi un metodo per il calcolo**.

» Presenta tuttavia diversi aspetti negativi non irrilevanti:

- dal punto di vista formale, è una definizione circolare: *richiede che i casi possiedano tutti la medesima probabilità*, che è però ciò che si vuole definire;
- non definisce la probabilità in caso di eventi non equiprobabili;
- presuppone un numero finito di risultati possibili e di conseguenza non è utilizzabile nel continuo.

Definizione Frequentista

» Per superare tali difficoltà, Richard von Mises propose di definire la probabilità di un evento come *il* ***limite*** *cui tende la frequenza relativa dell'evento al crescere del numero degli esperimenti*:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

» La definizione frequentista si applica ad esperimenti casuali i cui eventi elementari non siano ritenuti ugualmente possibili, ma assume che l'esperimento sia ripetibile più volte, idealmente infinite, sotto le stesse condizioni.

» Anche tale definizione consente di calcolare la probabilità di molti eventi e da essa si ricavano le stesse tre regole che seguono dalla definizione classica.

» È sufficiente, infatti, sostituire il rapporto tra numero dei casi favorevoli n_A e numero dei casi possibili n con il limite del rapporto per n tendente all'infinito.

Definizione Frequentista

Tuttavia:

1. il "limite" delle frequenze relative non è paragonabile all'analogo concetto matematico; ad esempio, data una successione $\{a_n\}$, si dice che a è il suo limite se per ogni $\varepsilon > 0$ esiste un numero naturale N tale che $|a_n - a| < \varepsilon$ per ogni $n > N$, e, comunque dato ε, è sempre possibile calcolare N; nella definizione frequentista, invece, <u>N non è sempre calcolabile</u>;
2. non tutti gli esperimenti sono ripetibili; ad esempio, ha sicuramente senso chiedersi quale sia la probabilità che vi sia vita su Marte o che tra 2 anni il tasso Euribor a 6 mesi diventi il doppio di quello attuale, <u>ma in casi simili non è possibile immaginare esperimenti ripetibili all'infinito</u>.

Definizione Soggettiva

» **De Finetti** e Savage hanno proposto una definizione di probabilità applicabile ad esperimenti casuali i cui eventi elementari non siano ritenuti ugualmente possibili e che non siano necessariamente ripetibili più volte sotto le stesse condizioni.

La probabilità di un evento è il prezzo che un individuo ritiene equo pagare per ricevere 1 se l'evento si verifica, 0 se l'evento non si verifica.

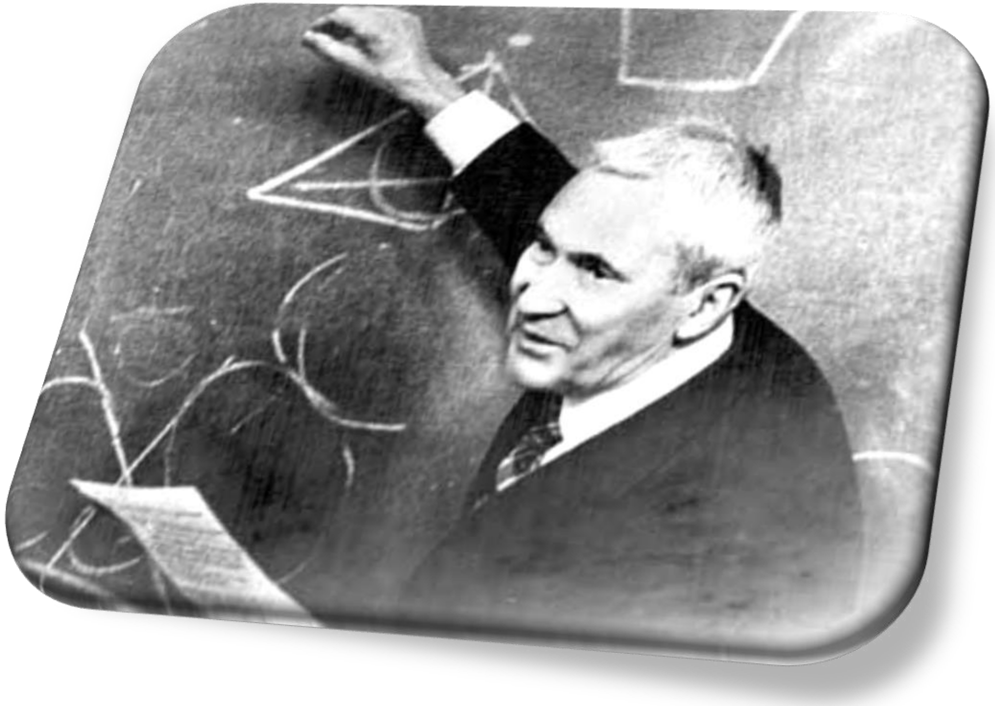
» Al fine di rendere concretamente applicabile la definizione, si aggiunge un criterio di coerenza: *le probabilità degli eventi devono essere attribuite in modo tale che non sia possibile ottenere una vincita o una perdita certa*.

» La definizione soggettiva consente di calcolare la probabilità di eventi anche quando gli eventi elementari non sono equiprobabili e quando l'esperimento non può essere ripetuto.

» Rimane fondata, tuttavia, sull'opinione di singoli individui, che potrebbero presentare diverse ***propensioni al rischio***. Basta pensare che molti sarebbero disposti a giocare 1 euro per vincerne 1000, ma pochi giocherebbero un milione di euro per vincerne un miliardo.

| Definizione Assiomatica

- L'impostazione assiomatica della probabilità venne proposta da **Andrey Nikolaevich Kolmogorov** nel 1933 in *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Concetti fondamentali del calcolo delle probabilità*), sviluppando la ricerca che era ormai cristallizzata sul dibattito fra quanti consideravano la probabilità come limiti di frequenze relative (cfr. impostazione frequentista) e quanti cercavano un fondamento logico della stessa.
- Va notato che la definizione assiomatica **non è una *definizione operativa*** e **non fornisce indicazioni su *come* calcolare la probabilità**. È quindi una definizione utilizzabile sia nell'ambito di un approccio oggettivista che nell'ambito di un approccio soggettivista.
- Il nome deriva dal procedimento per "assiomatizzazione" quindi nell'individuare i concetti primitivi, da questi nell'individuare i postulati da cui poi si passava a definire i teoremi.



- Definiamo *funzione di probabilità* una funzione *P* a valori reali che soddisfa le seguenti proprietà

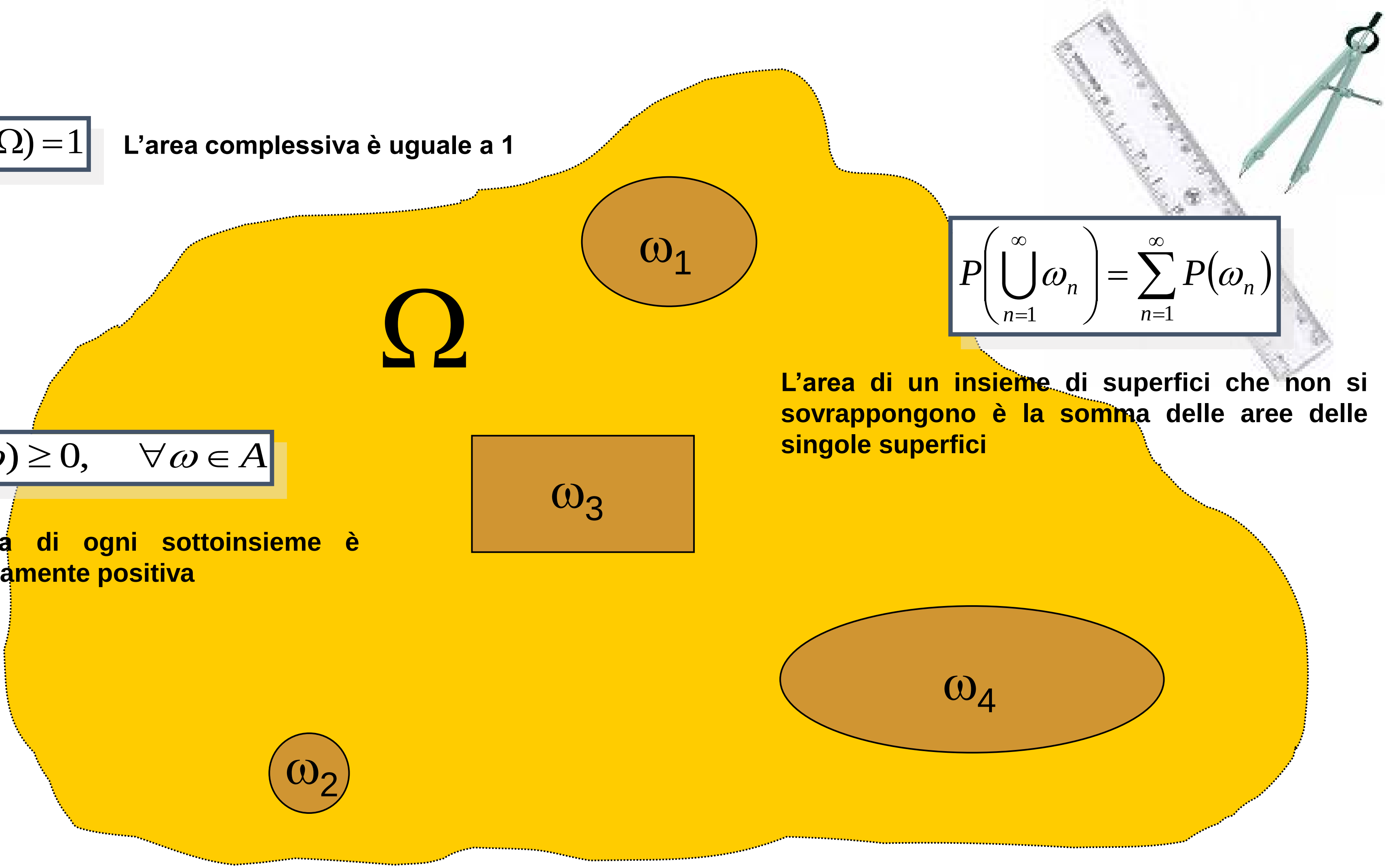
$$\boxed{\begin{aligned} & P(\omega) \geq 0, \qquad \forall \omega \in A \\ & P(\Omega) = 1 \\ & P\left(\bigcup_{n=1}^{\infty} \omega_n\right) = \sum_{n=1}^{\infty} P(\omega_n) \end{aligned}}$$

se gli ω_n sono a due a due disgiunti.

- Osserviamo che una funzione di probabilità così definita è anche una *misura*. La terna $(\Omega, \mathcal{A}, P)$ viene detta *spazio di probabilità*.

PROBABILITA' | Definizione Assiomatica Interpretazione Geometrica

$P(\Omega)=1$ **L'area complessiva è uguale a 1**



$$P\left(\bigcup_{n=1}^{\infty}\omega_n\right)=\sum_{n=1}^{\infty}P(\omega_n)$$

L'area di un insieme di superfici che non si sovrappongono è la somma delle aree delle singole superfici

$P(\omega)\geq 0, \quad \forall \omega \in A$

L'area di ogni sottoinsieme è sicuramente positiva

PROBABILITA' | Probabilità Condizionata

- La comprensione del concetto di probabilità condizionata è fondamentale per sfatare molti luoghi comuni errati
- E' fondamentale per comprendere situazioni che vengono presentate come paradossali, ma tali non sono.

Per dare un'idea della probabilità condizionata consideriamo le seguenti domande:

a) qual è la probabilità P_1 di essere milionari in Italia?

b) qual è la probabilità P_2 di essere milionari se si possiede una casa a Portofino?

Anche se non tutti quelli che hanno una casa a Portofino sono milionari, il buon senso suggerisce che P_2 debba essere molto maggiore di P_1. In casi come questi si parla di eventi correlati: l'evento «possedere una casa a Portofino» è un «indizio» che influenza la probabilità dell'evento «essere milionario». I due eventi sono detti per questo correlati, proprio perché sapere che uno si è verificato cambia la probabilità che si verifichi anche il secondo.

- Per trattare meglio casi come questo, è comodo introdurre la probabilità condizionata, indicata come P(A|B), che esprime la probabilità di osservare l'evento A sapendo che l'evento B è avvenuto.
- Questa quantità è diversa da P(A e B), che è la probabilità che siano accaduti entrambi gli eventi, come da P(A oppure B), la probabilità che accada almeno uno dei due eventi, perché in questi casi l'evento B è un evento possibile ma non certo.
- Nel caso della probabilità condizionata P(A|B), invece, si assume che l'evento B sia accaduto, ovvero che B sia un evento certo.

- Quando si ha motivo di credere che il verificarsi di uno o più eventi influenzano il verificarsi di altri eventi, allora si parlerà di eventi **dipendenti** (condizionati). Così, la probabilità dell'evento A dato che si è già verificato l'evento B (ovvero l'evento B condiziona l'evento A), è:

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

$$P[B] > 0$$

- In tal caso, B diventa il nostro "nuovo" spazio dei campioni; cioè si assume che la prova abbia dato luogo a qualche risultato in B.

Probabilità Condizionata

- Se sappiamo che B è accaduto, quali altri eventi possiamo aver osservato?
- Facciamo un esempio. Se giocando alla roulette il croupier ci annuncia che è uscito il rosso, possiamo essere certi che non sia uscito il numero 20, perché il 20 è un numero nero sul tavolo da gioco.
- Tuttavia, può essere accaduto che sia uscito un numero pari, dato che sul tavolo ci sono esattamente 9 numeri rossi pari.

Probabilità Condizionata

- Dunque, l’essere certi che l’evento [rosso] è accaduto, seleziona tra tutti gli eventi possibili soltanto quelli che non escludono l’accadere di [rosso].
- Gli eventi che non sono esclusi, sono quindi gli eventi compatibili con l’evento [rosso], considerato come certo.
- Tornando dunque al caso generale della probabilità condizionata P(A|B) a rigore è definita solo per eventi A compatibili con B.
- Tuttavia possiamo estendere P(A|B) anche agli eventi A incompatibili con B, cioè tali che l’evento (A e B) è impossibile, semplicemente ponendo, in questo caso, P(A|B) = P(A e B) = 0.

Probabilità Condizionata

Dalla precedente definizione ricaviamo subito che

$$P[A \cap B] = P[A | B]P[B]$$

cioè la probabilità che accadano gli eventi A **e** B è pari al prodotto della probabilità di B per la probabilità di A condizionata a B;

In particolare se l'evento A non è minimamente influenzato dall'accadere di B, cioè se i due eventi sono **indipendenti** avremo

$$P[A | B] = P[A]$$

da cui si deduce la **regola fondamentale per eventi indipendenti**

$$\boxed{P[A \cap B] = P[A]P[B]}$$

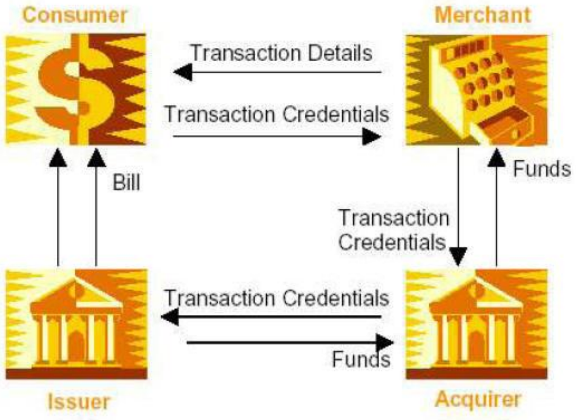
ESEMPIO

» Supponiamo che:
- B rappresenti l'evento "malfunzionamento del sistema informatico" e che la sua probabilità di accadimento sia pari a 0.2;
- A sia l'evento "grave errore durante una transazione".

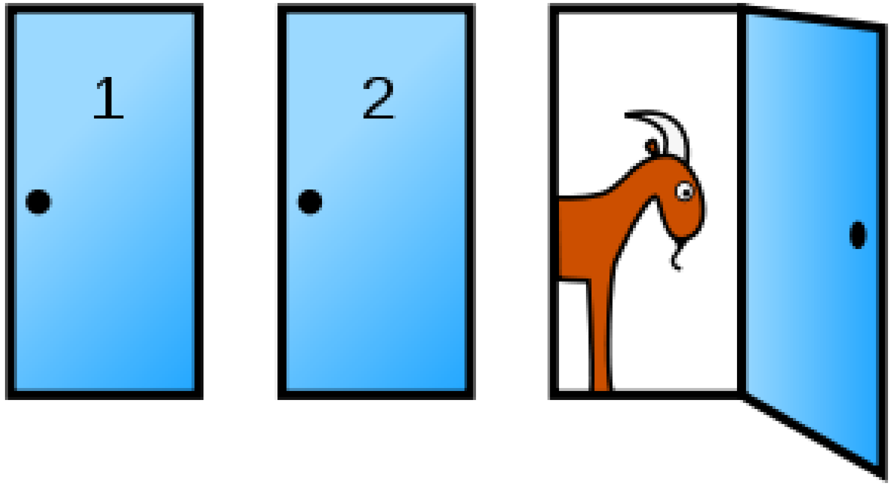
» Se gli aventi A ed I accadono congiuntamente con probabilità 0.05 allora abbiamo

$$P(A\,|\,B)=\frac{P(A\cap B)}{P(B)}=\frac{0.05}{0.2}=0.25$$

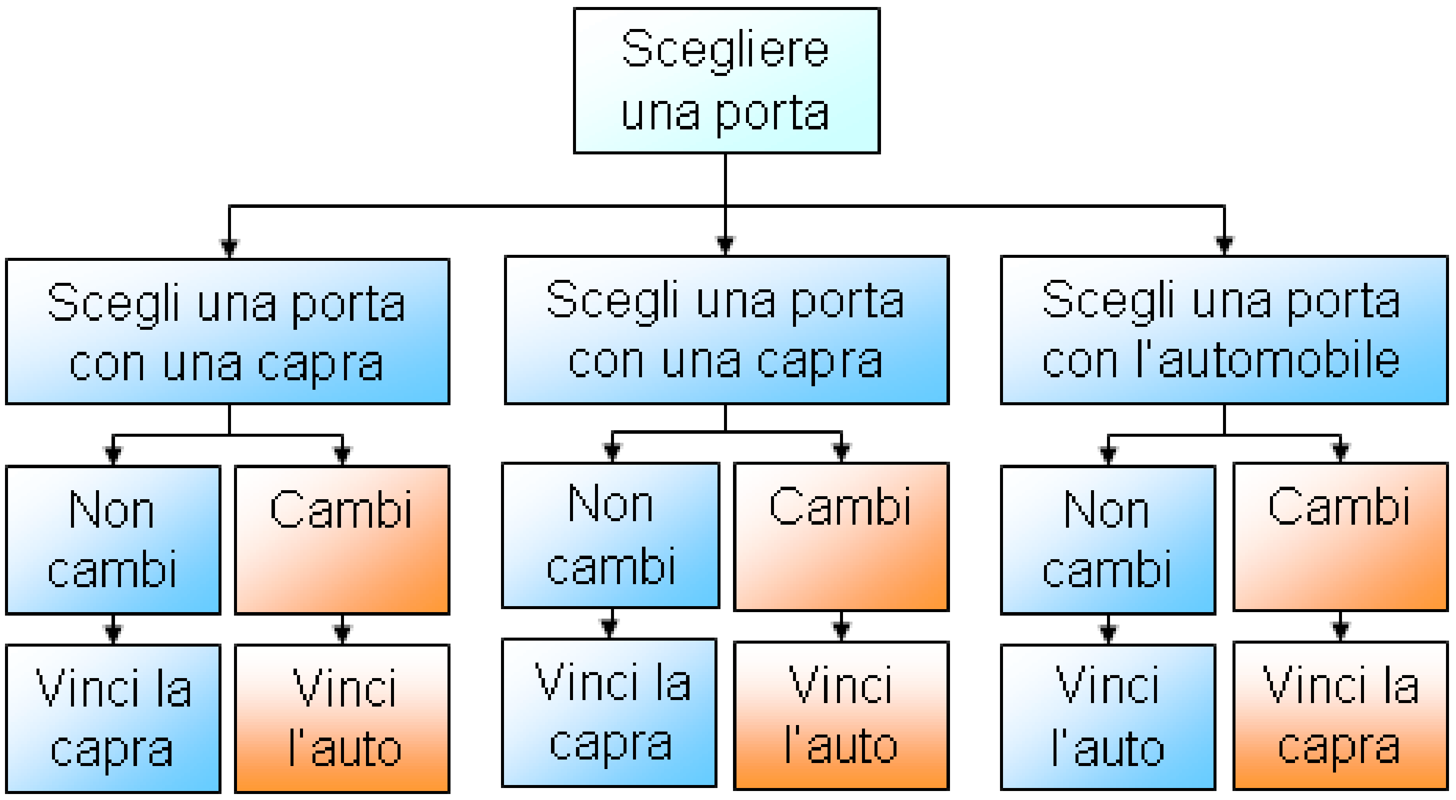
» Questo significa che la probabilità di avere un errore nelle transazioni in corso dato il fatto che si è già verificato un problema con il sistema informatico è pari a 0.25.



Il Problema di Monty Hall



Il Problema di Monty Hall: Soluzione

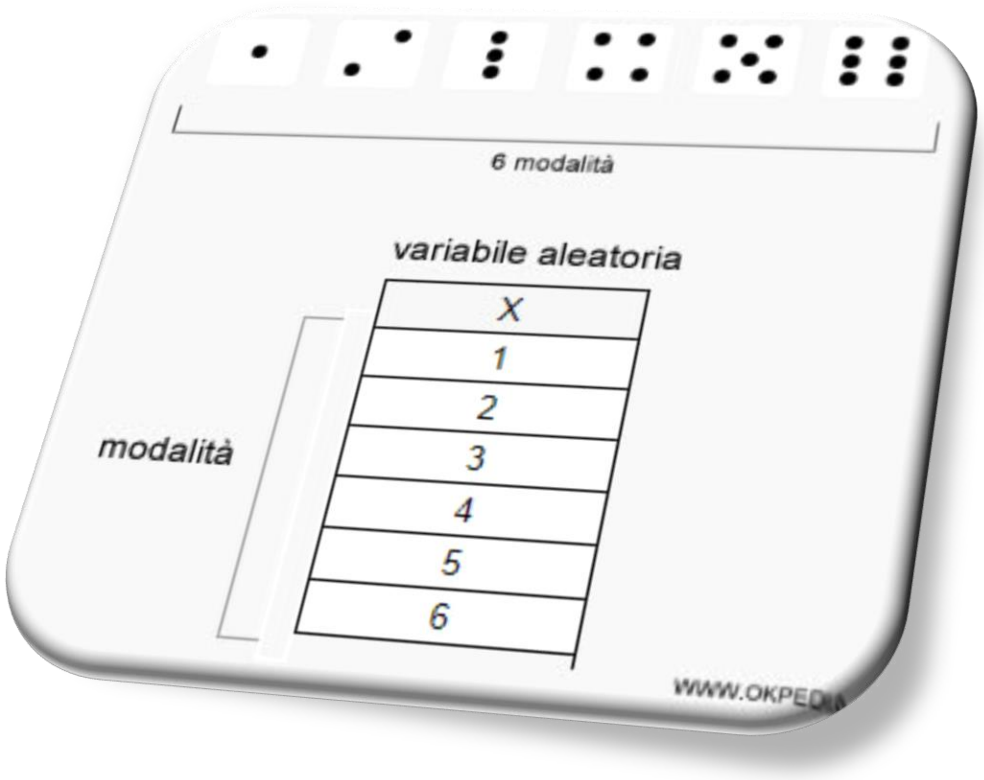


Dopo la scelta del giocatore, il presentatore apre una porta (egli sa dove si trova l'auto) mostrando una capra. Qualsiasi cosa ci sia dietro la scelta iniziale del giocatore, egli cambiando scelta ha il 66,7% di probabilità di vincere l'auto, non cambiandola ne avrebbe il 33,3%.

Variabili Aleatorie

Definizioni e Distribuzioni di Probabilità

- Una variabile aleatoria è una funzione che associa a ogni esito di un esperimento casuale un numero reale.
- In altre parole, una variabile aleatoria mappa gli esiti di un fenomeno aleatorio su valori numerici, permettendo di analizzare quantitativamente la probabilità di diversi risultati.



Esempi di Variabili Aleatorie

» **Lancio di un dado**:

- **Variabile aleatoria X**: rappresenta il numero ottenuto lanciando un dado a sei facce.
- **Possibili valori**: {1,2,3,4,5,6}.

Variabile Aleatoria

» **Esempi di Variabili Aleatorie**

» **Lancio di una moneta**:

- **Variabile aleatoria Y**: rappresenta il numero di teste ottenute in tre lanci di una moneta.
- **Possibili valori**: {0,1,2,3}.

» **Esempi di Variabili Aleatorie**

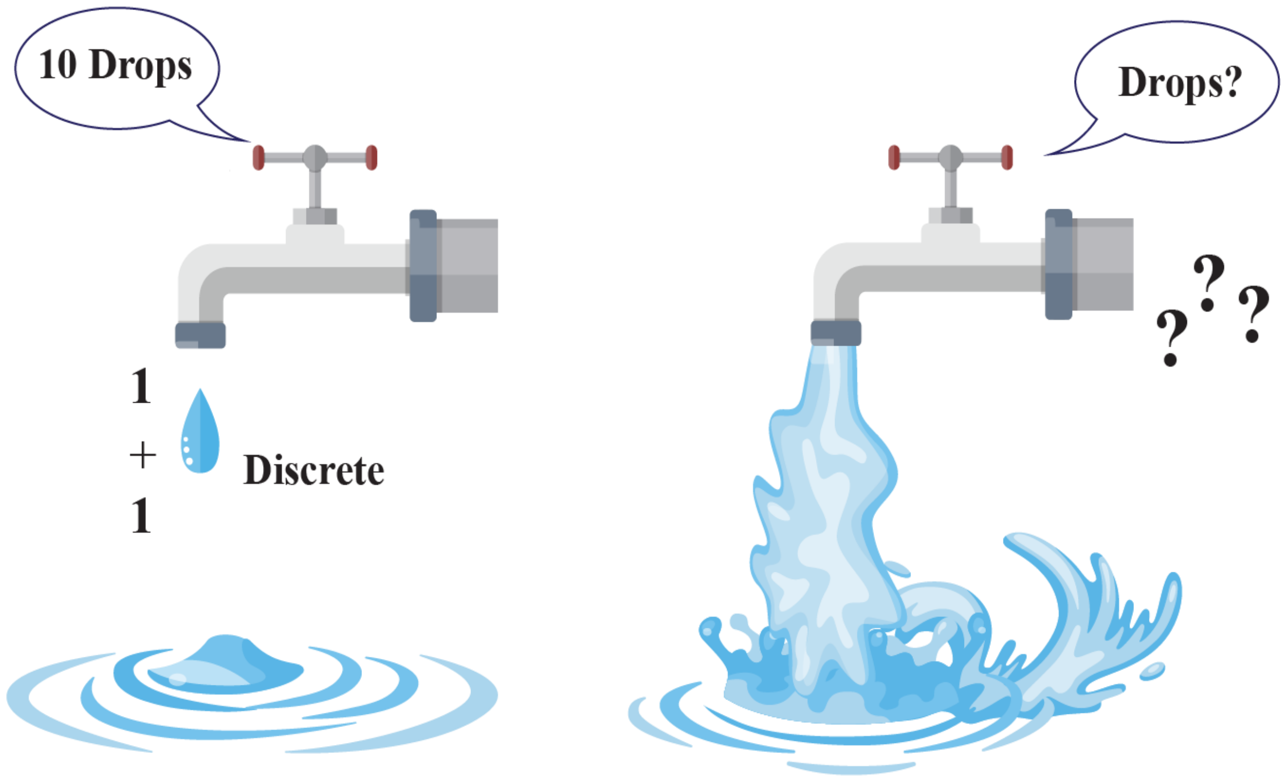
» **Tempo di attesa**:

- **Variabile aleatoria Z**: rappresenta il tempo di attesa (in minuti) per l'arrivo di un autobus a una fermata.
- **Possibili valori**: Un intervallo continuo di numeri reali, ad esempio $[0,\infty)$.

Tipi di Variabili Aleatorie:

1. **Variabile Aleatoria Discreta**: Assume un numero finito o numerabile di valori distinti.
   1. Esempio: Numero di bambini in una famiglia (0, 1, 2, ...).

2. **Variabile Aleatoria Continua**: Assume un'infinità non numerabile di valori all'interno di un intervallo.
   1. Esempio: L'altezza di una persona, misurata in centimetri.



PROBABILITA’ | Distribuzioni di Probabilità

» La distribuzione di probabilità è una funzione matematica che descrive come le probabilità sono distribuite tra i possibili valori di una variabile aleatoria.

» Esistono vari tipi di distribuzioni di probabilità, a seconda della natura della variabile aleatoria (discreta o continua) e delle proprietà specifiche della distribuzione stessa.

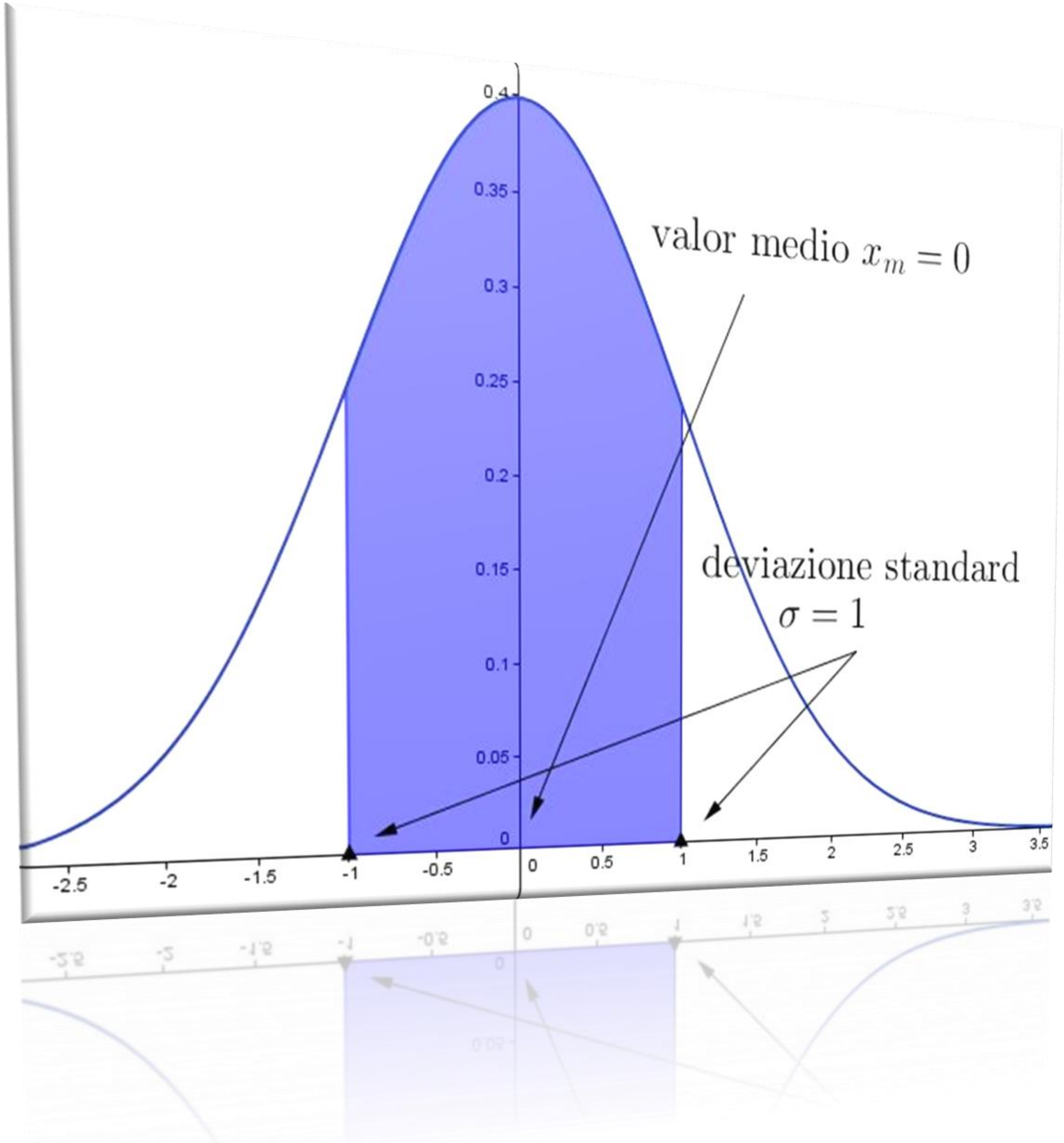
variabile aleatoria | probabilità

modalità

X	P(X)
1	0,17
2	0,17
3	0,17
4	0,17
5	0,17
6	0,17
	1,00

distribuzione di probabilità

WWW.OKPEDIA.IT



Parametri delle Distribuzioni

Ogni distribuzione di probabilità è caratterizzata da uno o più parametri che ne determinano la forma e le proprietà.

- **Media μ**: Indica il valore centrale della distribuzione.
- **Varianza σ^2**: Misura la dispersione dei valori attorno alla media.
- **Skewness**: Indica la simmetria della distribuzione.
- **Curtosi**: Misura la "pesantezza" delle code della distribuzione.

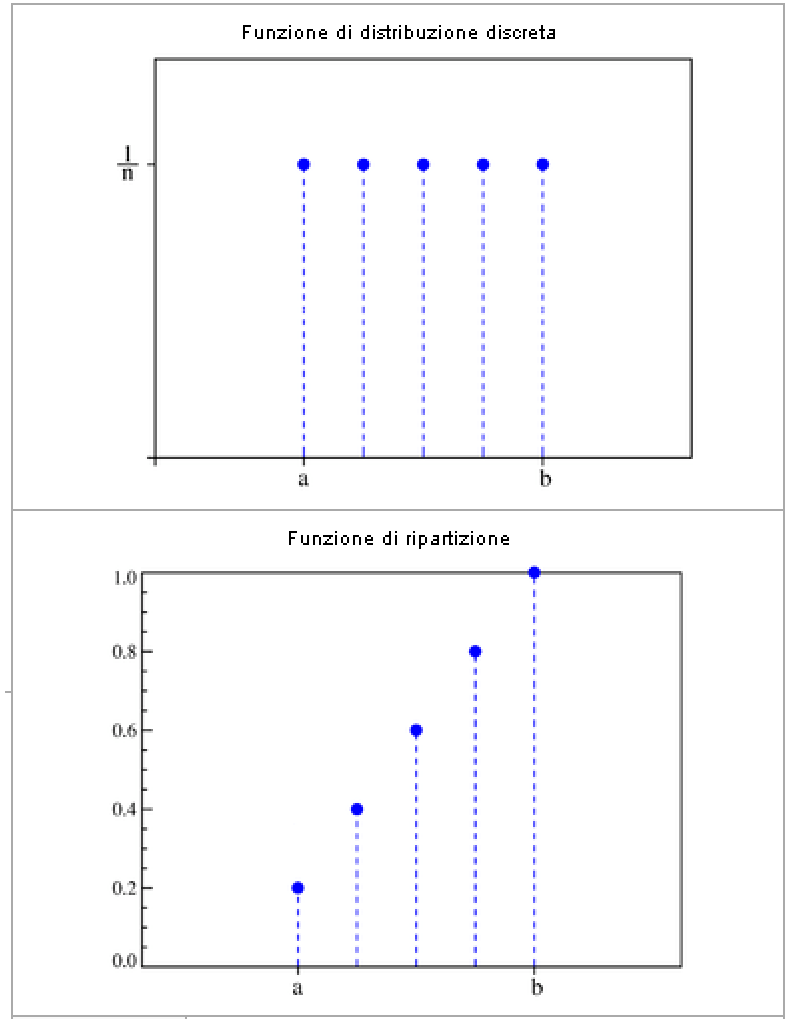
Distribuzione Uniforme

- Sia X una variabile aleatoria che assume valori nel dominio dei numeri naturali 1, 2, ... , n. Diremo che tale variabile ha una distribuzione uniforme se risulta

$$f_X(x) = f_X(x,n) = \begin{cases} 1/n & : \quad x = 1,2,\ldots,n \\ 0 & : \quad altrimenti \end{cases}$$

- In altre parole una distribuzione discreta uniforme è una distribuzione di probabilità discreta che attribuisce la stessa probabilità ad ogni elemento dell'insieme discreto S su cui è definita (in particolare l'insieme dev'essere finito);

- Un esempio di distribuzione discreta uniforme è fornito dal lancio di un dado equilibrato: ognuno dei valori 1, 2, 3, 4, 5 e 6 ha eguale probabilità 1/6 di verificarsi.

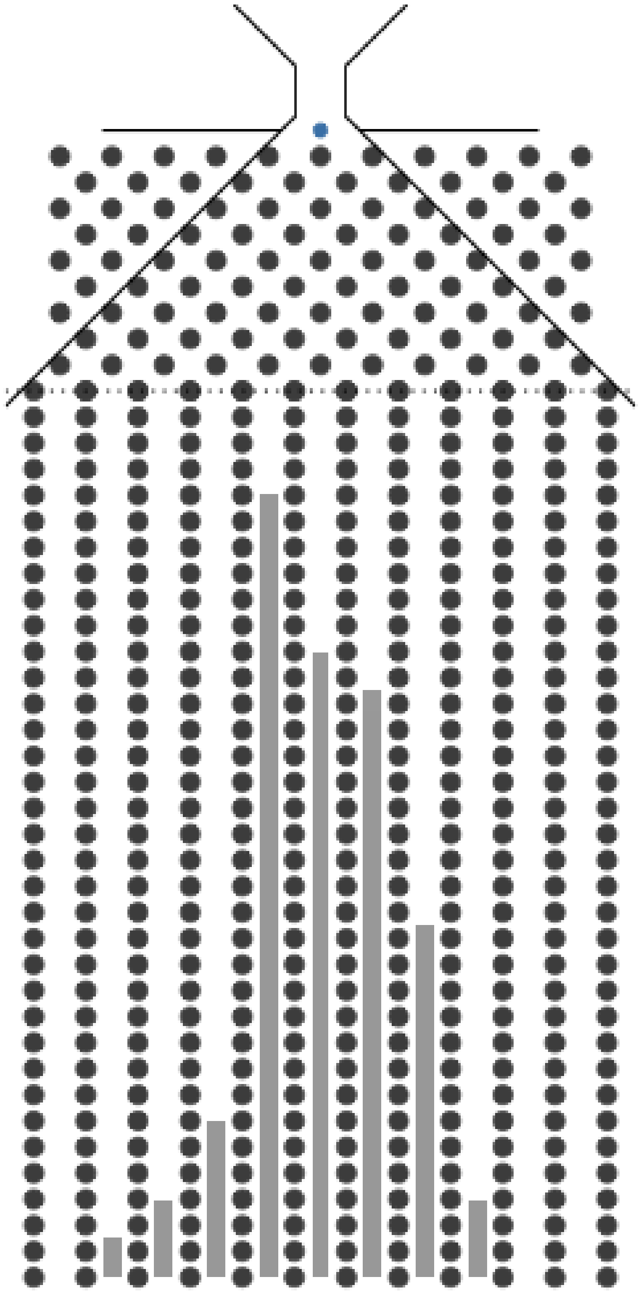


Distribuzione Binomiale:

» La distribuzione (finita, discreta) binomiale si origina dall'osservazione ripetuta (n volte) di una *prova di Bernoulli,* caratterizzata da due esiti che chiameremo "successo" e "insuccesso" con probabilità p e (1-p) rispettivamente.

» La probabilità di successo p non si altera ad ogni successiva osservazione, che viene quindi definita come indipendente;

» il conteggio di successi in n sequenze di osservazioni determina la *variabile aleatoria binomiale*.



Distribuzione Binomiale

» Per il calcolo della **distribuzione binomiale** si usa la funzione **DISTRIB.BINOM**

Sintassi

DISTRIB.BINOM(num_successi;prove;probabilità_s;cumulativo)

- Num_successi numero di successi nelle prove effettuate.
- Prove numero di prove indipendenti effettuate.
- Probabilità_s probabilità di successo in ciascuna prova.
- Cumulativo valore logico che determina il tipo di funzione calcolata.

Se il valore cumulativo è VERO, DISTRIB.BINOM restituirà la funzione di ripartizione, ossia la probabilità di ottenere un numero di successi minore o uguale al valore num_successi. Se il valore cumulativo è FALSO, verrà restituita la distribuzione di probabilità, ossia la probabilità di ottenere un numero di successi uguale al valore num_successi.

Distribuzione Binomiale

Esempio 41.1

Si effettuano 20 lanci di un dado; il successo sia di ottenere il numero tre.

1 Calcolare la probabilità di ottenere 2 volte il numero tre.

prove	20
probabilità_s	0,1667

Nella finestra della funzione DISTRIB.BINOM per Cumulativo scegliere FALSO

P(X=2)	0,1982

=DISTRIB.BINOM(2;D41;D42;FALSO)

Esempio 41.2

Si effettuano 20 lanci di un dado; il successo sia di ottenere tre.

2 Calcolare la probabilità di ottenere al massimo 2 volte il numero tre.

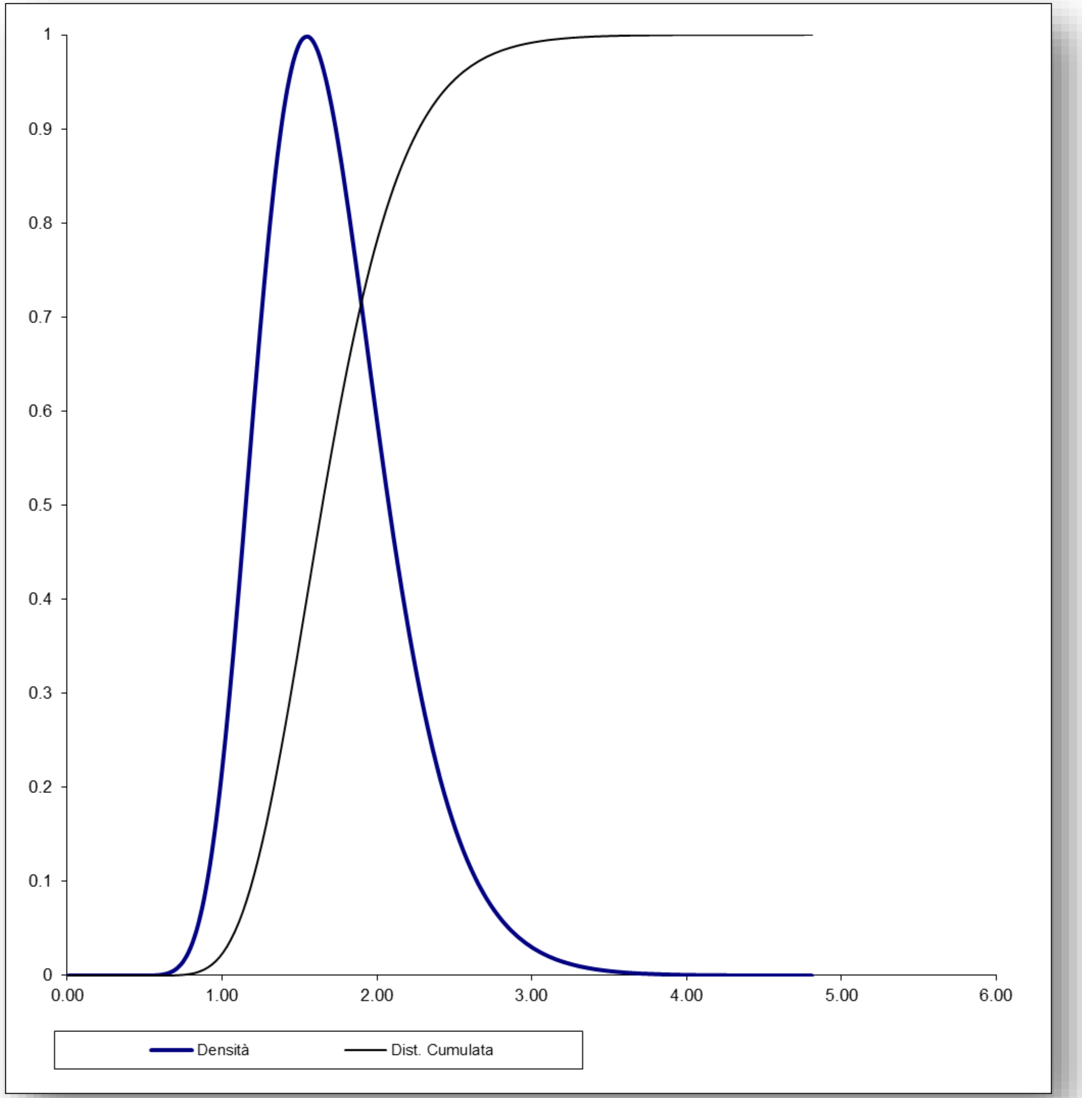
Nella finestra della funzione per Cumulativo scegliere VERO

P(X<=2)	0,3287

=DISTRIB.BINOM(2;D41;D42;VERO)

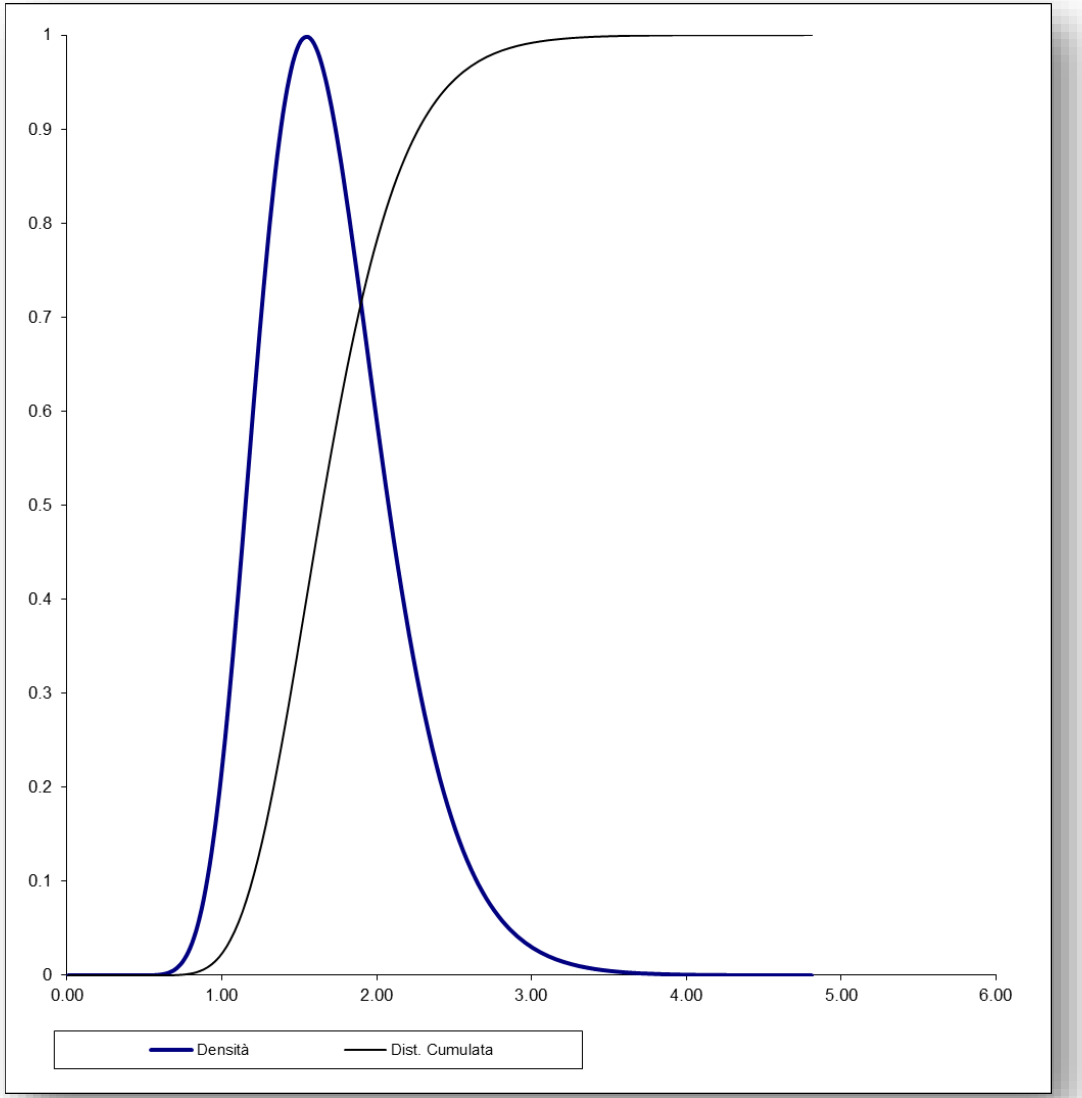
Funzione di Distribuzione di Probabilità (CDF)

- **Definizione**: La funzione di distribuzione cumulativa (CDF, Cumulative Distribution Function) di una variabile casuale X rappresenta la probabilità che X assuma un valore minore o uguale a un certo valore x.
- **Caratteristiche**:
  - **Monotona non decrescente**: La CDF aumenta o rimane costante al crescere di x.
  - **Valori**: La CDF varia tra 0 e 1.
  - **Interpretazione**: Fornisce la probabilità accumulata fino al valore x.
  - **Continuità**: Può essere continua o avere salti (discontinua) se X è una variabile casuale discreta.



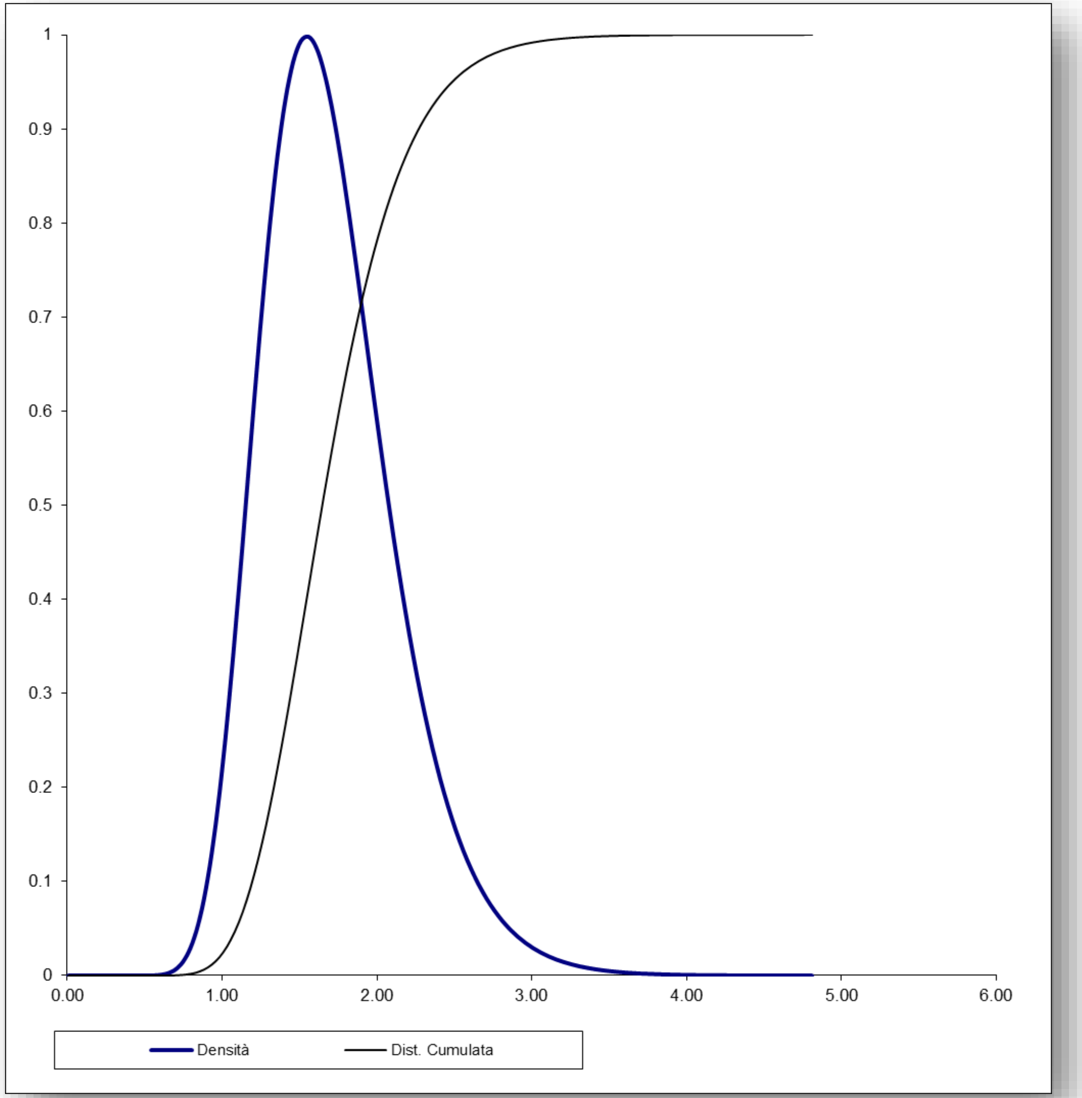
Densità di Probabilità (PDF)

- **Definizione**: La funzione di densità di probabilità (PDF, Probability Density Function) di una variabile casuale continua X descrive la densità della probabilità per ogni valore specifico di x.
- **Caratteristiche**:
    - **Non negativa**
    - **Area totale sotto la curva** =1.
    - **Non è una probabilità diretta**: non rappresenta la probabilità che X sia esattamente uguale a x, ma piuttosto la densità di probabilità attorno a x.



Esempi

- **CDF**:
  - Se X è il risultato di un lancio di dado, la CDF al valore 3 è la probabilità che il risultato sia 1, 2 o 3.
- **PDF**:
  - Se X è una variabile casuale che rappresenta l'altezza in una popolazione, la PDF al valore 170 cm descrive quanto comune è un'altezza di 170 cm, non la probabilità di esattamente 170 cm.



Riepilogo

- **CDF (Funzione di Distribuzione Cumulativa)**:
  - Rappresenta la probabilità cumulata fino a un certo valore.
  - Monotona non decrescente.
  - Valori tra 0 e 1.
- **PDF (Funzione di Densità di Probabilità)**:
  - Rappresenta la densità di probabilità per un valore specifico.
  - Non negativa.
  - L'integrale su tutto il dominio è pari a 1.
  - Derivata della CDF per variabili continue.

Conoscere la differenza tra CDF e PDF è cruciale per l'analisi e l'interpretazione dei dati probabilistici e statistici!