Appunti di probabilità

Angelo Vaccaro

Indice

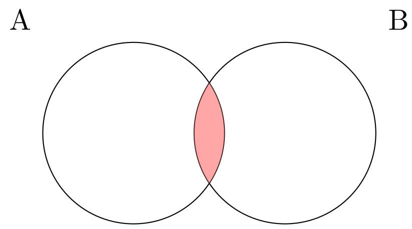
1 Teoria della Probabilità di Base

Un **esperimento aleatorio** è un qualsiasi processo che ha più di un possibile risultato e di cui non possiamo prevedere l'esito con certezza. Chiamiamo **S** lo **spazio campionario**, e è l'insieme di tutti i possibili esiti. I **sottoinsiemi** di S sono detti **eventi**; un evento è un insieme di uno o più esiti che ci interessano.

Operazioni tra Eventi

Intersezione ($A \cap B$)

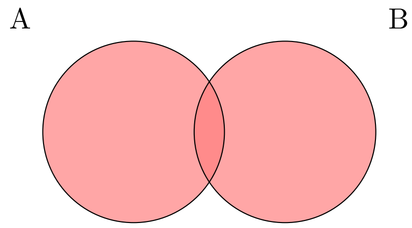
"avviene A" e "avviene B"



$$A \cap B = \{s \in S \mid s \in A \text{ e } s \in B\}$$

Unione ($A \cup B$)

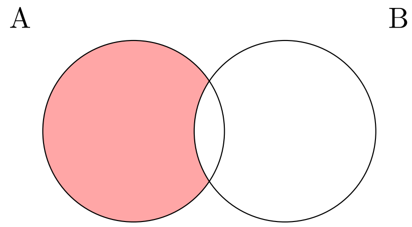
"avviene A" o "avviene B" $\rightarrow$ evento in cui si verifica A o B o entrambi.



$$A \cup B = \{s \in S \mid s \in A \text{ o } s \in B\}$$

Differenza ($A \setminus B$)

"avviene A" ma non "avviene B"



$$A \setminus B = \{s \in S \mid s \in A, s \notin B\}$$

Complementare (A^c)

"non avviene A"

Eventi disgiunti o incompatibili

Due eventi sono disgiunti se non possono accadere contemporaneamente. La loro intersezione è l'insieme vuoto, non ci sono quindi esiti in comune.

- *Esempio:* "esce un numero pari" e "esce il 3" sono eventi disgiunti.

1.1 Calcolo delle probabilità

La probabilità, $P(A)$, è un numero che quantifica la "fiducia" che abbiamo nel verificarsi dell'evento A.

- $P(A) = 1 \rightarrow$ evento A certo (evento sicuro)
- $P(A) = 0 \rightarrow$ evento A impossibile (evento impossibile)

$0 \leq P(A) \leq 1 \rightarrow$ tutti gli altri eventi.

La teoria assegna probabilità agli **insiemi** (eventi), non ai singoli elementi. Nel caso di esiti equiprobabili si arriva alla probabilità del singolo esito, ma il framework è definito sugli eventi.

Non possiamo assegnare valori numeri a caso, dobbiamo rispettare delle regole:

Regola 1.1 (Assiomi di Kolmogorov)**.**

1. $\mathbf{P(S) = 1}$
   La probabilità che si verifichi uno qualsiasi dei possibili risultati è 1 (cioè, è una certezza). Lo spazio campionario contiene tutti gli esiti possibili.

2. $\mathbf{P(\emptyset) = 0}$
   La probabilità dell'evento impossibile è 0.

3. **Se A $\subseteq$ B, allora** $P(B) \geq P(A)$
   Se il verificarsi di A implica il verificarsi di B (tutti gli esiti di A sono anche esiti di B), allora la probabilità di B è maggiore o uguale a quella di A. In pratica se si verifica A, necessariamente si verifica B.

   Da questo assioma derivano altre proprietà importanti:

   - **Probabilità del complementare:**

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

     La probabilità che un evento non accada è 1 meno la probabilità che accada. Questa formula ci facilita alcuni calcoli.

1.2 Regole di Calcolo Fondamentali

1.2.1 Additività per eventi disgiunti

Se 2 eventi non possono accadere insieme (disgiunti), la probabilità che accada o l'uno o l'altro è semplicemente la somma delle probabilità individuali.

Se $A_i \cap A_j = \emptyset$ per ogni $i \neq j$, allora:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Nota: questo vale solo se gli eventi sono a due a due disgiunti.

1.2.2 Regola generale per l'unione

Per due eventi qualsiasi, non necessariamente disgiunti:

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

Togliamo dalla somma la parte ripetuta.

1.3 Probabilità ad Esiti Equiprobabili

Definizione 1.1. Questa è la formula "classica". Funziona solo se tutti gli esiti dello spazio campionario S hanno la stessa identica probabilità di avvenire.

$P(A) = \frac{|A|}{|S|} = \frac{\text{numero di esiti favorevoli}}{\text{numero di esiti possibili}}$

- $|A| \rightarrow$ **cardinalità di A**, n° di esiti che compongono l'evento A.
- $|S| \rightarrow$ **cardinalità di S**, il n° totale di esiti possibili.

1.4 Strumenti di Calcolo Combinatorio

1.4.1 Coefficiente binomiale

Questa formula conta in quanti modi si possono scegliere k oggetti da un insieme di n elementi, **senza** considerare l'ordine e **senza** reinserimento.

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

Esempio 1.1. Devo scegliere 6 carte da un mazzo di 52:

$$\binom{52}{6} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 20,358,520$$

1.4.2 Disposizioni semplici (senza reinserimento)

Questa formula si utilizza invece quando l'**ordine conta**.

$D(n,k) = \frac{n!}{(n-k)!}$

Esempio 1.2. In quanti modi diversi si possono assegnare le 3 posizioni sul podio? (8 atleti)

$$D(8,3) = \frac{8!}{(8-3)!} = \frac{8!}{5!} = 8 \cdot 7 \cdot 6 = 336$$

1.4.3 Disposizioni con ripetizione

Questa è la situazione in cui estrai k oggetti da un insieme di n, l'ordine conta, e puoi estrarre lo stesso oggetto più volte.

$$D'(n,k) = n^k$$

Esempio 1.3. Stiamo scegliendo $k = 3$ cifre da un insieme di $n = 10$ cifre disponibili. L'ordine conta e le cifre si possono ripetere.

$$D'(10,3) = 10^3 = 1000$$

2 Probabilità Condizionata e Teoremi Collegati

2.1 Probabilità condizionata

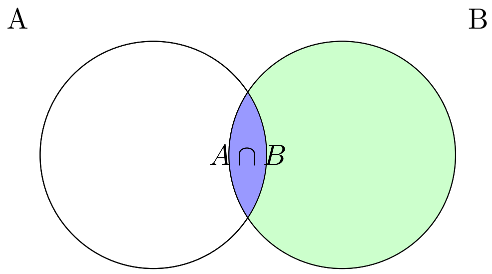
La probabilità condizionata è la probabilità che un evento accada sapendo che un altro evento è già avvenuto.

- **Notazione:** $P(A|B) \rightarrow$ "probabilità di A dato B" o "probabilità di A sapendo che B"

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{con } P(B) > 0$$

$P(A|B)$ non è un nuovo tipo di evento, ma è un nuovo modo di misurare la probabilità di A.

Intuizione geometrica



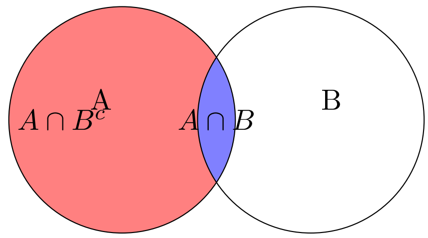
- Quando sappiamo che B è avvenuto, il nostro spazio campionario non è più S, ma si riduce a B.
- Dato che siamo confinati dentro B, l'unico modo perché si verifichi anche A è che l'esito cada nella parte di A che si sovrappone a B, ovvero $A \cap B$.
- La probabilità condizionata $P(A|B)$ è quindi il "peso" dell'area $A \cap B$ rispetto alla nuova area totale 'B'. Ecco perché la formula divide per $P(B)$.

2.2 Formula delle Probabilità Totali

Possiamo manipolare la probabilità condizionata per ottenere 2 strumenti potentissimi.

Regola 2.1 (Regola delle catene (o formula inversa))**.** È semplicemente un riarrangiamento della formula iniziale. Serve perché a volte è più facile calcolare $P(A|B)$ e $P(B)$.

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



Regola 2.2 (Formula delle probabilità totali)**.**
Un evento A può essere partizionato dall'evento B e dal suo complementare B^c:

$$A = (A \cap B) \cup (A \cap B^c)$$

Poiché le due parti sono disgiunte, la loro probabilità è la somma delle probabilità:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

Ora, usando la formula moltiplicativa su entrambi i termini, otteniamo la formula delle probabilità totali:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Questa formula ci permette di calcolare la probabilità totale di un evento A "assemblandolo" pezzo per pezzo, considerando tutti i diversi scenari in cui A potrebbe verificarsi. In questo caso, i due scenari sono "il mondo in cui B accade" e "il mondo in cui B non accade (B^c)".

Generalizzazione (Concetto di Partizione)

Invece di dividere il mondo solo in "B accade" e "B non accade", possiamo dividerlo in n scenari $B_1, B_2, \ldots, B_n$. Una **partizione** è un insieme di eventi che sono:

1. **Mutuamente esclusivi:** $B_i \cap B_j = \emptyset$ (non hanno elementi in comune per $i \neq j$)
2. **Esaustivi:** $\bigcup_i B_i = S$ (la loro unione copre tutte le possibilità)

Formula Generale delle Probabilità Totali

Se $\{B_1, B_2, \ldots, B_n\}$ è una partizione, allora la probabilità di un qualsiasi evento A è:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$$

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Stiamo facendo ancora la stessa cosa: calcoliamo la probabilità di A in ogni possibile "mondo" B_i e poi facciamo una media pesata di queste probabilità.

2.3 Teorema di Bayes

La formula delle probabilità totali ci permette di passare dalle **cause** (gli scenari, B_i) agli **effetti** (l'osservazione, A). Il teorema di Bayes fa il percorso inverso: osserviamo un effetto (A) e vogliamo calcolare la probabilità di una specifica **causa** (B_i).

Partiamo da 2 modi di scrivere $P(A \cap B)$:

1. $P(A \cap B) = P(A|B)P(B)$ [definizione di $P(A|B)$]
2. $P(B \cap A) = P(B|A)P(A)$ [definizione di $P(B|A)$]

Poiché l'intersezione è commutativa ($A \cap B = B \cap A$), quindi i termini a destra devono essere uguali. Ora, isolando il termine $P(B|A)$ (la probabilità della causa dato l'effetto), otteniamo la formula:

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

Estensione del Teorema di Bayes (n eventi)

- **Contesto:** Abbiamo una partizione dello spazio campionario $\{B_1, \dots, B_n\}$ (le possibili cause/scenari). Osserviamo un effetto A. Vogliamo sapere qual è la probabilità della i-esima causa, B_i, dato che abbiamo osservato A.

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

- **Numeratore:** È la probabilità che si verifichino sia la causa B_i che l'effetto A, ovvero $P(A \cap B_i)$.
- **Denominatore:** È la probabilità che si verifichi l'effetto A, calcolata sommando su tutte le possibili cause (è la formula delle probabilità totali).
- **In pratica:** La probabilità della causa i dato l'effetto, è il "contributo" della causa i al verificarsi dell'effetto, diviso il "contributo" di tutte le cause messe insieme.

3 Indipendenza degli Eventi

Definizione 3.1. Due eventi A e B sono **indipendenti** se il verificarsi di uno non dà alcuna informazione sul verificarsi dell'altro. Sapere che B è accaduto non cambia in alcun modo la probabilità di A.

Se A e B sono indipendenti, allora $P(A|B) = P(A)$

Regola 3.1 (Formula pratica per l'indipendenza)**.** Sostituendo $P(A|B) = P(A)$ nella formula moltiplicativa $P(A \cap B) = P(A|B)P(B)$, otteniamo la definizione formale di indipendenza. Due eventi A e B sono indipendenti se:

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

Notiamo quindi che la probabilità che 2 eventi indipendenti accadano contemporaneamente è semplicemente il prodotto delle probabilità individuali.

Implicazioni dell'indipendenza

La definizione di indipendenza è coerente con quella di probabilità condizionata.

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

Spiegazione: Questo passaggio conferma la nostra intuizione iniziale. Se 2 eventi sono indipendenti, conoscere l'esito di uno non cambia la probabilità dell'altro. La probabilità condizionata collassa sulla probabilità "base" $P(A)$.

- Si osserva anche che se A e B sono indipendenti, allora lo sono anche A^c con B, A con B^c e A^c con B^c. L'indipendenza si estende a tutte le combinazioni degli eventi e dei loro complementari.

ATTENZIONE: Eventi Disgiunti e Indipendenti sono DIVERSI

- **Disgiunti** $\rightarrow A \cap B = \emptyset$. Se sai che A è accaduto, allora sei certo che B **non** può accadere ($P(B|A) = 0$). Questa è la forma più "forte" di dipendenza.
- **Indipendenti** $\rightarrow$ Sapere che A è accaduto, non cambia per nulla la probabilità di B ($P(B|A) = P(B)$).

L'unica possibilità che 2 eventi possano essere sia disgiunti che indipendenti è se almeno uno dei due ha probabilità zero.

Indipendenza di 3 o più eventi

L'indipendenza può essere estesa a più di 2 eventi, ma la definizione diventa più stringente.
3 eventi A, B, e C sono indipendenti se sono verificate **tutte e quattro** le seguenti condizioni:

1. $P(A \cap B \cap C) = P(A)P(B)P(C)$
2. $P(A \cap B) = P(A)P(B)$
3. $P(A \cap C) = P(A)P(C)$
4. $P(B \cap C) = P(B)P(C)$

Non basta che gli eventi siano indipendenti "a due a due", è necessaria anche la prima condizione, che riguarda l'intersezione di tutti e 3. L'idea si può generalizzare ad n eventi: bisogna verificare che la regola del prodotto valga per tutte le possibili sotto-intersezioni.

4 Definizione delle Variabili Aleatorie

Una variabile aleatoria è una quantità numerica il cui valore dipende dall'esito di un esperimento aleatorio.

Definizione 4.1 (Variabile Aleatoria (v.a.))**.** Una variabile aleatoria (abbreviata con v.a. o r.v. in inglese) non è "variabile" nel senso di incognita algebrica, né "aleatoria" nel senso di caotica. È una **funzione** che associa un numero reale ad ogni esito dello spazio campionario S.

Esempio 4.1.

- **Esperimento:** Lancio di 2 monete. $S = \{TT, CC, TC, CT\}$
- **Variabile aleatoria X:** "Numero di teste"
- **X è una funzione che mappa:**
  - $TT \to 2$
  - $TC \to 1$
  - $CT \to 1$
  - $CC \to 0$
- Invece di parlare dell'evento $\{TC, CT\}$, ora possiamo parlare dell'evento $\{X = 1\}$. Abbiamo trasformato un problema su insiemi in un problema su numeri.

Notazione

- **X (maiuscola):** Rappresenta la v.a. stessa, la "regola" o la "funzione". È l'oggetto concettuale ("il numero totale di successi").
- **x (minuscola):** Rappresenta una specifica realizzazione numerica, un valore che la variabile può assumere (come 2, 5, 10.3, ...).

4.1 Variabili Aleatorie Discrete

Definizione 4.2. Una v.a. si dice **discreta** se i valori che può assumere sono in numero finito o al più numerabile. I valori che una v.a. discreta assume sono "separati", "contabili", come i gradini di una scala.

- Possono essere finiti (es. i risultati di un dado: $\{1, 2, 3, 4, 5, 6\}$)
- O infiniti ma enumerabili (es. il n° di email che ricevi in un'ora: $\{0, 1, 2, \dots\}$ all'infinito).

4.1.1 Funzione di massa di probabilità $p(x)$

Definizione 4.3. Per una v.a. discreta possiamo definire una funzione, $p(x)$, che dà la probabilità associata a ciascun singolo valore x che la variabile può assumere. È una sorta di "distribuzione dei pesi" della probabilità totale tra i valori possibili.

$$p(x) = P(X = x)$$

4.2 Proprietà della Funzione di Massa di Probabilità p(x)

1. $p(x) \geq 0$ per ogni x. (Le probabilità non possono essere negative).
2. $\sum p(x) = 1$. (Sommando le probabilità su **tutti** i possibili valori di x, dobbiamo ottenere 1, perché è certo che la v.a. assumerà uno dei suoi valori possibili).

Calcolare la probabilità di un evento A

Se A è un insieme di valori possibili (es. $A = \{2, 4, 6\}$), allora:

$$P(X \in A) = p(2) + p(4) + p(6)$$

Basta sommare le probabilità (i "pesi") dei singoli valori che compongono l'evento.

Esempio 4.2 (Funzione di Massa di Probabilità)**.** Immagina di avere un'urna che contiene 5 palline:

- 3 palline sono rosse e hanno il numero 1 scritto sopra.
- 2 palline sono blu e hanno il numero 2 scritto sopra.

Eseguiamo un esperimento: estraiamo 2 palline dall'urna senza reinserirle.
Definiamo la v.a. X come la **somma** dei numeri sulle 2 palline estratte.

Descriviamo la funzione di massa di probabilità $p(x)$ per la v.a. X.
La nostra v.a. può assumere i valori: $X \in \{2, 3, 4\}$. Calcoliamo ora $p(2)$, $p(3)$, e $p(4)$.

- **Spazio campionario:** Il numero totale di modi per scegliere 2 palline da 5 è $|S| = \binom{5}{2} = \frac{5 \cdot 4}{2} = 10$. Questo calcolo va bene perché non ci interessa l'ordine e non reinseriamo.
- **Calcolo di $p(2) = P(X = 2)$:**
L'unico modo per ottenere somma 2 è estrarre due palline con il numero 1. Ci sono 3 palline rosse (con "1"), quindi il numero di modi per sceglierne 2 è $\binom{3}{2} = 3$.
$$p(2) = \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$$
- **Calcolo di $p(4) = P(X = 4)$:**
L'unico modo per ottenere somma 4 è estrarre due palline con il numero 2. Ci sono 2 palline blu (con "2"), quindi il numero di modi per sceglierne 2 è $\binom{2}{2} = 1$.
$$p(4) = \frac{\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10}$$
- **Calcolo di $p(3) = P(X = 3)$:**
L'unico modo per ottenere somma 3 è estrarre una pallina rossa (con "1") e una blu (con "2"). Il numero di modi è $\binom{3}{1} \cdot \binom{2}{1} = 3 \cdot 2 = 6$.
$$p(3) = \frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}} = \frac{6}{10}$$

Verifica: $p(2) + p(3) + p(4) = \frac{3}{10} + \frac{6}{10} + \frac{1}{10} = \frac{10}{10} = 1$.

5 Variabili Aleatorie Continue

Definizione 5.1. Una v.a. **continua** è una v.a. che può assumere infiniti valori reali all'interno di un intervallo continuo. A differenza delle v.a. discrete, che prendono valori isolati (come 0,1,2,...), una v.a. continua può assumere **qualsiasi** valore reale in un intervallo, ad esempio tra 0 e 1, o tra $-\infty$ e $+\infty$.

Il problema delle v.a. continue

Se X è una v.a. continua (es. l'altezza di una persona a caso), qual è la probabilità che sia *esattamente* 175,0000... cm?
Poiché ci sono infiniti valori possibili, la probabilità di trovarne uno esatto è 0.

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

Questo significa che la funzione di massa $p(x)$ non ha più senso per le continue. Ci serve un nuovo strumento.

5.1 La funzione di densità $f(x)$

Per una v.a. continua, esiste una funzione, $f(x)$, chiamata **densità**, tale che la probabilità che X cada in un insieme (intervallo) A è data dall'area sotto la curva di $f(x)$ in quell'intervallo.

$$P(X \in A) = \int_A f(x)dx$$

Nota: non chiediamo più $P(X = x)$, ma $P(a \leq X \leq b)$.

Proprietà della funzione di densità

1. $f(x) \geq 0$: La densità non può essere negativa (la curva sta sempre sopra l'asse x).
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$: L'area totale sotto l'intera curva deve essere 1. Questo è l'analogo di $\sum p(x) = 1$ per le discrete.

$f(x)$ misura quindi la "densità" di probabilità intorno ad un punto x. Dove $f(x)$ è alta, è più probabile trovare valori di X; dove è bassa, è meno probabile.

Esempio 5.1 (Funzione di densità)**.** Immagina un gioco molto semplice. C'è un timer che parte da 0 e si ferma in un istante di tempo casuale, ma non oltre i 2 secondi. L'istante esatto in cui si ferma, X, può essere qualsiasi numero reale tra 0 e 2. Supponiamo che il meccanismo sia "semplice" e non abbia "preferenze", cioè che la densità di probabilità sia **costante** in questo intervallo [0, 2]. X = istante di tempo in cui il timer si ferma. X è una v.a. continua.
Dobbiamo trovare la sua funzione di densità $f(x)$.

Passaggi per la soluzione

1. **Definire il "supporto":** Il supporto è l'intervallo dove la v.a. può esistere. Qui è $[0,2]$. Al di fuori, $f(x) = 0$. Dobbiamo capire quanto vale $f(x)$ all'interno di $[0,2]$.

2. **Usare la regola dell'area totale:** Abbiamo ipotizzato una densità costante $f(x) = c$ per $x \in [0,2]$. Sappiamo che l'area totale deve essere 1.

$$\int_0^2 f(x)dx = \int_0^2 c\,dx = 1$$

3. **Risolvere l'integrale:**

$$[cx]_0^2 = c(2) - c(0) = 2c = 1 \implies c = \frac{1}{2}$$

4. **Scrivere la funzione completa:**

$$f(x) = \begin{cases} 1/2 & \text{se } 0 \leq x \leq 2 \\ 0 & \text{altrimenti} \end{cases}$$

Questa si chiama **Distribuzione Uniforme** sull'intervallo $[0,2]$.

6 Valore Atteso (o Media)

Ora introduciamo un modo per sintetizzare un'intera distribuzione (sia essa discreta o continua) in un singolo numero che rappresenti il "centro di massa" o il valore "tipico".

6.1 Definizione per v.a. discrete

Definizione 6.1. Il **valore atteso** di una v.a. discreta X, denotato con $E(X)$ o μ, è la somma di tutti i possibili valori di x ponderati per la loro probabilità.

$E(X) = \sum_x x \cdot p(x)$

Il valore atteso è una **media ponderata**. Ogni possibile valore x viene "pesato" per la sua probabilità $p(x)$. I valori più probabili contribuiscono di più alla media.

Esempio 6.1 (Lancio di un dado)**.** Sia X il risultato del lancio. I valori possibili sono $\{1,2,3,4,5,6\}$, ognuno con $p(x) = 1/6$.

$$E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{21}{6} = 3.5$$

Notare che il valore 3.5 non è un valore che il dado può assumere; è la media dei valori se lanciassimo il dado un numero infinito di volte.

6.2 Definizione per v.a. continue

Definizione 6.2. Il concetto è identico: è una media ponderata. Ma siccome abbiamo una distribuzione continua, la somma ($\sum$) viene sostituita dall'integrale ($\int$). Ogni valore x sull'asse reale viene pesato per la sua densità $f(x)$.

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

7 Valore Atteso di una Funzione di una v.a.

E se fossimo interessati non alla media di X, ma alla media di X^2 o $\log(X)$? Chiamiamo $g(X)$ una funzione della variabile aleatoria X.

La regola è semplice: invece di pesare x con la sua probabilità/densità, pesiamo $g(x)$.

Regola 7.1 (LOTUS - Law of the Unconscious Statistician)**.**

- **Discrete:**

  $E[g(X)] = \sum_x g(x) \cdot p(x)$

- **Continue:**

  $E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x)\, dx$

Importanza di LOTUS

La regola per calcolare $E[g(X)]$ è una delle proprietà più potenti del valore atteso. Ci permette di calcolare la media di una variabile trasformata **senza** doverne prima derivare la distribuzione.

Esempio 7.1 (Vincita media di un gioco)**. Scenario di Base:** Riprendiamo il nostro esempio del lancio di un dado.

- Esperimento: Lancio di un dado a 6 facce, non truccato.
- Variabile Aleatoria X: il numero che esce.
- Valori possibili di X: $\{1, 2, 3, 4, 5, 6\}$.
- Funzione di massa $p(x)$: $p(x) = 1/6$ per ciascuno di questi valori.
- Valore medio di X: $E(X) = 3.5$.

Lo Scenario con $g(x)$: Ora, immaginiamo un gioco basato sul lancio di questo dado.

- **Il gioco:** Vinci un ammontare in euro pari al **QUADRATO** del numero uscito.
- Vogliamo calcolare la **vincita media** di questo gioco.
- La vincita non è più X, ma una funzione di X, $g(x) = X^2$.
- Vogliamo calcolare $E[g(X)] = E(X^2)$.

Metodo 1: Lento ma Intuitivo (Creare una nuova v.a. Y = X^2)

1. I valori possibili di Y sono $\{1^2, 2^2, \dots, 6^2\} = \{1, 4, 9, 16, 25, 36\}$.
2. La probabilità di ciascun valore di Y è $p(y) = 1/6$.
3. Calcoliamo $E(Y)$:

$$E(Y) = 1\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + \cdots + 36\left(\frac{1}{6}\right) = \frac{91}{6} \approx 15.17$$

Metodo 2: Veloce e Diretto (Usando LOTUS) Applichiamo direttamente la formula $E[g(X)] = \sum g(x)p(x)$:

$$
\begin{aligned}
E(X^2) &= \sum_{x=1}^{6} x^2 \cdot p(x) \\
&= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\
&= \frac{1+4+9+16+25+36}{6} = \frac{91}{6} \approx 15.17
\end{aligned}
$$

Come vedi, il risultato è identico, ma il calcolo è stato più diretto.

8 Varianza di una Variabile Aleatoria

8.1 Un caso speciale: $E(X^2)$

Da questo arriviamo quindi alla formula per il calcolo di $E(X^2)$, che è un'applicazione diretta di $E[g(X)]$ dove la funzione di trasformazione è semplicemente $g(x) = x^2$.

- **V.A. Discrete:** Per calcolare $E(X^2)$, prendi ogni possibile valore x che la variabile X può assumere, lo elevi al quadrato, lo moltiplichi per la probabilità di quel valore $p(x)$, e sommi i prodotti.

  $$E(X^2) = \sum_x x^2 p(x)$$

- **V.A. Continue:** Il concetto è identico, ma la somma ($\sum$) viene sostituita da un integrale ($\int$).

  $$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx$$

8.2 Proprietà del valore medio

1. **Proprietà di Linearità:**

   $$E(aX + b) = aE(X) + b$$

   Questa è una proprietà importantissima. Se trasformi la tua v.a. X moltiplicandola per una costante a e sommando una costante b, il nuovo valore medio è semplicemente il vecchio valore medio trasformato allo stesso modo.

2. **Additività del valore:**

   $$E(X + Y) = E(X) + E(Y)$$

   Il valore medio della somma di due v.a. (anche se non indipendenti) è la somma dei loro valori medi. Questa proprietà è incredibilmente generale e utile.

3. **Valore atteso della funzione $g(x)$:** Questo punto ribadisce quello che abbiamo discusso: $E[g(X)]$ si calcola come media di $g(x)$ pesata con le probabilità di X.

8.3 La Varianza: $Var(X)$

Il valore medio ci dice il "centro" di una distribuzione, ma non ci dice nulla su quanto i valori siano *sparpagliati* attorno a questo centro.

Definizione 8.1 (Varianza)**. L'idea:** Misurare lo scarto dalla media.

1. Per ogni possibile valore x, calcoliamo lo **scarto dalla media**: $(x - \mu)$, dove $\mu = E(X)$.
2. Questo scarto può essere positivo o negativo. Se facessimo la media degli scarti, $E(X-\mu)$, otterremmo sempre 0. Non è utile.
3. **SOLUZIONE:** Eleviamo gli scarti al quadrato: $(x - \mu)^2$. In questo modo sono tutti non negativi.
4. Ora calcoliamo il valore medio di questi **scarti quadratici**. Questa è la **VARIANZA**.

$$\boxed{Var(X) = E[(X - \mu)^2]}$$

È il valore atteso dello scarto quadratico dalla media.

8.3.1 Formula computazionale della Varianza

$$\begin{aligned}
Var(X) &= E[(X - \mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] && \text{(espandendo il quadrato)} \\
&= E(X^2) - E(2\mu X) + E(\mu^2) && \text{(per linearità della media)} \\
&= E(X^2) - 2\mu E(X) + \mu^2 && \text{(poiché } 2\mu \text{ e } \mu^2 \text{ sono costanti)} \\
&= E(X^2) - 2\mu \cdot \mu + \mu^2 && \text{(poiché } E(X) = \mu\text{)} \\
&= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2
\end{aligned}$$

$$\boxed{Var(X) = E(X^2) - [E(X)]^2}$$

La varianza è la media dei quadrati meno il quadrato della media.

Riassunto del calcolo della Varianza

1. Calcola la media: $\mu = E(X)$.
2. Calcola la media dei quadrati: $E(X^2)$ (usando le formule $\sum x^2 p(x)$ o $\int x^2 f(x)dx$).
3. Sottrai: $Var(X) = E(X^2) - \mu^2$.

Proprietà fondamentale della varianza

- $Var(X) \geq 0$. Poiché è una media di quadrati, la varianza non può mai essere negativa.
- $Var(X) = 0$ se e solo se X non è affatto "aleatoria", ma è una costante.

9 Distribuzioni Congiunte

Finora abbiamo studiato una v.a. alla volta (es. il risultato di un dado). Ma cosa succede se lanciamo 2 dadi e siamo interessati sia al punteggio più alto (X) che a quello più basso (Y)? X e Y sono 2 v.a. definite sullo stesso esperimento e probabilmente sono dipendenti tra loro. Come descriviamo questa relazione?

9.1 Funzione di massa congiunta $p_{X,Y}(x, y)$

Definizione 9.1. Questa funzione dà la probabilità che **contemporaneamente** la v.a. X assuma il valore x **e** la v.a. Y assuma il valore y.

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Mentre $p(x)$ era una lista di probabilità, $p_{X,Y}(x, y)$ è una **tabella** a doppia entrata. Le righe possono rappresentare i valori di X, le colonne i valori di Y, ed ogni cella (x, y) contiene la probabilità che si verifichi quella specifica coppia di valori.

9.2 Distribuzioni marginali

Abbiamo la tabella congiunta $p_{X,Y}(x, y)$, che contiene tutte le informazioni sulla coppia (X, Y). Come possiamo recuperare da essa la distribuzione della sola X (o della sola Y), ignorando l'altra variabile?

Definizione 9.2 (Funzione di massa marginale)**.** Per trovare la probabilità totale che X assuma un certo valore x, dobbiamo considerare tutti i modi in cui questo può accadere, indipendentemente da cosa faccia Y. Questo significa sommare tutte le probabilità lungo la riga (o colonna) corrispondente a quel valore x nella nostra tabella congiunta.

- **Marginale di X (somma lungo le colonne per ogni riga):**

  $$p_X(x) = P(X = x) = \sum_y p_{X,Y}(x, y)$$

- **Marginale di Y (somma lungo le righe per ogni colonna):**

  $$p_Y(y) = P(Y = y) = \sum_x p_{X,Y}(x, y)$$

Esempio 9.1 (Lancio di due dadi)**.** Sia X il punteggio più basso e Y il punteggio più alto. Vogliamo calcolare $p_X(1) = P(X = 1)$. Questo può accadere in 11 modi: $(1, 1), (1, 2), \dots, (1, 6)$ e $(2, 1), \dots, (6, 1)$. La probabilità di ogni singola coppia (x, y) con dadi standard è $1/36$.

$$\begin{aligned} p_X(1) &= p_{X,Y}(1, 1) + p_{X,Y}(1, 2) + \dots + p_{X,Y}(1, 6) \\ &= \frac{1}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} + \frac{2}{36} = \frac{11}{36} \end{aligned}$$

9.3 Indipendenza tra Variabili Aleatorie

Questo è il caso speciale che ci permette di ricostruire la congiunta dalle marginali.

Definizione 9.3. Due v.a. X e Y sono **indipendenti** se conoscere il valore di una non dà alcuna informazione sul valore dell'altra.
In pratica, X e Y sono indipendenti se per ogni possibile coppia (x, y), la probabilità congiunta è il prodotto delle probabilità marginali.

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

Questa è l'estensione diretta della definizione di indipendenza degli eventi.

9.4 Valore atteso di una funzione congiunta $E[g(X,Y)]$

Questo estende l'idea di $E[g(X)]$ a 2 variabili. Se abbiamo una funzione g che dipende sia da X che da Y, il suo valore atteso si calcola:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y)$$

In pratica: per ogni cella (x,y) della tabella, calcoli $g(x,y)$, lo moltiplichi per la probabilità di quella cella $p_{X,Y}(x,y)$, e poi sommi i risultati di tutte le celle.

9.4.1 Conseguenze dirette dell'indipendenza

Se X e Y sono indipendenti, allora:

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

Questo è un teorema fondamentale. Se 2 v.a. sono indipendenti, la media del loro prodotto è semplicemente il prodotto delle loro medie.

Attenzione

Il contrario **non** è vero in generale! Se $E(XY) = E(X)E(Y)$, non è detto che X e Y siano indipendenti. Questa condizione è meno forte dell'indipendenza.

9.5 Covarianza: $Cov(X, Y)$

Vogliamo un indice che ci dica se, quando X è sopra la sua media, anche Y tende ad essere sopra la sua media (**relazione positiva**) o se tende ad essere sotto la sua media (**relazione negativa**).

Definizione 9.4 (Covarianza)**.**

1. Prendiamo gli scarti di X dalla sua media: $(X - E(X))$.
2. Prendiamo gli scarti di Y dalla sua media: $(Y - E(Y))$.
3. Moltiplichiamo questi scarti: $(X - E(X))(Y - E(Y))$.
   - Se X e Y sono entrambi sopra la media, il prodotto è $(+) \cdot (+) = (+)$.
   - Se X e Y sono entrambi sotto la media, il prodotto è $(-) \cdot (-) = (+)$.
   - Se uno è sopra e l'altro è sotto, il prodotto è $(+) \cdot (-) = (-)$.
4. Calcoliamo la media di questi prodotti. Questa è la **COVARIANZA**.

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Formula computazionale

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

La covarianza è la media del prodotto meno il prodotto delle medie.

Interpretazione del segno della Covarianza

- **$Cov(X, Y) > 0$:** C'è una relazione lineare **positiva**. In media, a valori alti di X corrispondono valori alti di Y.
- **$Cov(X, Y) < 0$:** C'è una relazione lineare **negativa**. In media, a valori alti di X corrispondono valori bassi di Y.
- **$Cov(X, Y) = 0$:** Non c'è relazione **lineare**. Le variabili sono **scorrelate**.

10 Correlazione

10.1 Proprietà importanti della Covarianza

1. $Cov(X, X) = Var(X)$
   Se calcoliamo la covarianza di una v.a. con se stessa, otteniamo la varianza. Questo perché:

$$Cov(X, X) = E(X \cdot X) - E(X)E(X) = E(X^2) - [E(X)]^2 = Var(X)$$

2. **Se X e Y sono indipendenti, allora** $Cov(X, Y) = 0$
   Se X e Y sono indipendenti, sappiamo che $E(XY) = E(X)E(Y)$. Sostituendo nella formula computazionale della covarianza:

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

3. **Variabili Indipendenti $\implies$ Variabili Scorrelate**
   L'implicazione opposta non è sempre vera.

$$\text{Ind} \implies \text{Scorr (Cov=0)} \quad \text{(Sempre Vero)}$$

$$\text{Scorr (Cov=0)} \implies \text{Ind} \quad \text{(Falso in generale)}$$

   Questo perché la covarianza misura solo la forza della relazione **lineare**. Due v.a. possono avere una relazione forte ma non-lineare ed avere comunque $Cov = 0$. L'indipendenza è un concetto molto più forte, che implica l'assenza di qualsiasi tipo di relazione.

10.2 Normalizzazione della covarianza: la Correlazione

La covarianza ha un problema: dipende dall'unità di misura delle variabili. Il suo valore numerico non è facilmente interpretabile.

- **In sintesi:** La covarianza ci dà il **segno** (la direzione) della relazione lineare, ma il suo **valore assoluto** non ci dice nulla sulla **forza** di tale relazione in modo standardizzato.

Per risolvere questo problema, "normalizziamo" la covarianza, cioè la dividiamo per un fattore che tiene conto della dispersione delle singole variabili.

Definizione 10.1 (Coefficiente di Correlazione)**.**

$$Corr(X, Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Il denominatore è il prodotto delle deviazioni standard di X e Y. (La deviazione standard, $\sigma_X = \sqrt{Var(X)}$, è semplicemente la radice quadrata della varianza e ha la stessa unità di misura della variabile originale).

Proprietà del Coefficiente di Correlazione

1. È un numero puro, **senza unità di misura**. Poiché dividiamo un'unità di misura (es. $kg \cdot m$) per un'altra identica, le unità si cancellano. Questo lo rende un indice universale.

2. È sempre compreso tra -1 e +1: $-1 \leq Corr(X, Y) \leq +1$.

   - $Corr(X, Y) \approx +1$: Forte correlazione lineare **positiva**. I punti su un grafico a dispersione tendono a disporsi quasi perfettamente su una retta crescente.
   - $Corr(X, Y) \approx -1$: Forte correlazione lineare **negativa**. I punti tendono a disporsi su una retta decrescente.

- $Corr(X,Y) \approx 0$: Relazione lineare debole o assente. (Questo non esclude relazioni non-lineari!).

3. Vale esattamente +1 o -1 solo se c'è una relazione lineare perfetta.

   - $Corr(X,Y) = 1$ se e solo se $Y = aX + b$ con $a > 0$.
   - $Corr(X,Y) = -1$ se e solo se $Y = aX + b$ con $a < 0$.

11 Funzione di Ripartizione (CDF)

Definizione 11.1 (Cumulative Distribution Function - CDF)**.** La funzione di ripartizione $F_X(x)$, calcolata in un punto x, ci dà la probabilità **accumulata** fino a quel punto.

$$F_X(x) = P(X \leq x)$$

Se scegli un punto x sull'asse numerico, $F_X(x)$ ti dice qual è la probabilità totale che la tua v.a. X assuma un valore minore o uguale ad x. È una probabilità "cumulativa", che parte da 0 e arriva fino a 1.

11.1 Caso discreto

$$F_X(x) = \sum_{t \leq x} p(t)$$

Il grafico di $F_X(x)$ per una v.a. discreta è una **funzione a gradini**, che fa un salto in corrispondenza di ogni valore che la v.a. può assumere. L'altezza del salto in x_i è esattamente $p(x_i)$.

11.2 Caso continuo

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

Il grafico di $F_X(x)$ per una v.a. continua è una funzione **continua** e non decrescente. Parte da 0 (a $-\infty$) e sale fino ad 1 (a $+\infty$).

Relazione inversa (importantissima)

Se la funzione di ripartizione $F_X(x)$ è l'integrale della densità $f(x)$, allora la densità $f(x)$ è la **derivata** della funzione di ripartizione.

$$f(x) = F'_X(x) = \frac{d}{dx}F_X(x)$$

12 Proprietà della Varianza

Regola 12.1.

$$Var(aX + b) = a^2 Var(X)$$

- La **costante additiva b sparisce!** Se aggiungi una costante b a tutti i valori, stai solo "traslando" l'intera distribuzione. La sua forma e la sua dispersione non cambiano. La varianza, che misura la dispersione, rimane la stessa.
- La **costante moltiplicativa a esce al quadrato!** La varianza misura uno scarto quadratico. Se scali la variabile di un fattore a, lo scarto dalla media viene scalato di a, ma lo scarto quadratico viene scalato di a^2.

12.1 Varianza di una somma di v.a.

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

La varianza della somma **non** è semplicemente la somma delle varianze. C'è un termine in più che dipende da come le 2 variabili si muovono insieme.

- **Il termine $2Cov(X,Y)$:**
  - Se X e Y sono **positivamente correlate** ($Cov > 0$), quando X è alta anche Y tende ad essere alta. I loro movimenti si sommano, aumentando la dispersione complessiva. La covarianza aggiunge variabilità.
  - Se X e Y sono **negativamente correlate** ($Cov < 0$), quando X è alta, Y tende ad essere bassa. I loro movimenti si "compensano" a vicenda. La loro somma $X+Y$ tenderà ad essere più stabile e meno dispersa. La covarianza negativa riduce la variabilità totale.

Caso speciale (importantissimo): X e Y scorrelate

Se X e Y sono scorrelate, allora $Cov(X,Y) = 0$.

Se $Cov(X,Y) = 0$, allora $Var(X+Y) = Var(X) + Var(Y)$

Solo in questo caso speciale (che include anche il caso in cui X e Y sono indipendenti), la varianza della somma è la somma delle varianze.

13 Classi Notevoli di Variabili Aleatorie

13.1 Variabile Aleatoria di Bernoulli

È il modello matematico per un **singolo** esperimento che ha solo 2 possibili esiti. Per convenzione li chiamiamo **successo (1)** e **insuccesso (0)**.

- **Parametro:** Ha un solo parametro, p, che è la probabilità di successo. Si scrive $X \sim Bernoulli(p)$.
- **Funzione di massa $p(x)$:**
  - $P(X = 1) = p$
  - $P(X = 0) = 1 - p$
- **Valore atteso $E(X)$:**
$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$
  La media di una Bernoulli è semplicemente la sua probabilità di successo.
- **Varianza $Var(X)$:**
  - Prima calcoliamo $E(X^2)$: $E(X^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$.
  - Poi uso la formula $Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p)$.

  $$\boxed{E(X) = p, \quad Var(X) = p(1 - p)}$$

Modella il lancio di una moneta, la risposta ad una domanda sì/no, ...

13.2 Variabile Aleatoria Binomiale

È il modello matematico per il **numero di successi X in n prove di Bernoulli indipendenti ed identiche**. Stiamo sommando n variabili di Bernoulli indipendenti.

- **Connessione:** $X = X_1 + X_2 + \cdots + X_n$, dove ogni X_i è una $Bernoulli(p)$.
- **Parametri:** Ha 2 parametri: n (numero di prove) e p (probabilità di successo in ogni prova). Si scrive $X \sim Bin(n, p)$.
- **Funzione di massa $p(k) = P(X = k)$:**

  $$\boxed{P(X = k) = \tbinom{n}{k} p^k (1 - p)^{n-k}}$$

- **Valore atteso $E(X)$:** Invece di usare la definizione con la sommatoria, usiamo la proprietà della somma.
$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = \underbrace{p + \cdots + p}_{n \text{ volte}} = np$$

  $$\boxed{E(X) = np}$$

- **Varianza $Var(X)$:** Usiamo la formula per la varianza di una somma. Poiché le prove di Bernoulli sono indipendenti, le variabili X_i sono indipendenti, e quindi anche scorrelate ($Cov(X_i, X_j) = 0$).

  $$Var(X) = Var(X_1+\cdots+X_n) = Var(X_1)+\cdots+Var(X_n) = \underbrace{p(1-p) + \cdots + p(1-p)}_{n \text{ volte}} = np(1-p)$$

  $$\boxed{Var(X) = np(1-p)}$$

13.3 Variabile Aleatoria di Poisson

La distribuzione di Poisson è il modello per contare il **numero di eventi che accadono in un intervallo di tempo (o spazio) fissato**, quando questi eventi sono rari e accadono indipendentemente l'uno dall'altro.

- **Valori possibili:** X può assumere i valori $\{0, 1, 2, 3, \dots\}$. È una v.a. discreta con un'infinità numerabile di valori.
- **Parametro:** Ha un solo parametro, λ (lambda), che rappresenta il **numero medio** di eventi nell'intervallo. $\lambda = E(X)$.
- **Funzione di massa $p(k) = P(X = k)$:**

  $$\boxed{P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}}$$

- **Relazione con la binomiale:** La Poisson può essere vista come un'approssimazione della Binomiale quando n è molto grande e p è molto piccolo. In questo caso, $\lambda \approx np$.
- **Proprietà notevole:** Per una Poisson, **media e varianza sono uguali**.

  $$\boxed{E(X) = \lambda, \quad Var(X) = \lambda}$$

14 Altre Classi Notevoli di Variabili Aleatorie

14.1 Variabile Aleatoria Uniforme (Continua)

È il modello per un esito che ha la stessa probabilità di cadere in qualsiasi punto di un intervallo $[a, b]$. Non ci sono "zone preferite".

- **Esempio:** L'istante di tempo casuale in cui si ferma un timer tra 0 e 2 secondi.
- **Parametri:** I 2 estremi dell'intervallo, a e b. Si scrive $X \sim U(a, b)$.
- **Funzione di densità $f(x)$:** Deve essere costante all'interno dell'intervallo $[a, b]$ e 0 altrove. L'area totale sotto la curva (che è un rettangolo) deve essere 1.

  $$\text{Area} = \text{Base} \times \text{Altezza} = (b-a) \cdot c = 1 \implies c = \frac{1}{b-a}$$

  $$\boxed{f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \le x \le b \\ 0 & \text{altrimenti} \end{cases}}$$

- **Valore atteso $E(X)$:**

$$E(X) = \frac{a+b}{2}$$

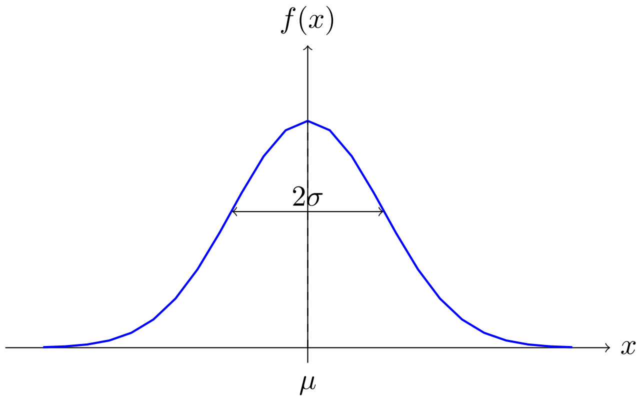
- **Varianza $Var(X)$:**

$$Var(X) = \frac{(b-a)^2}{12}$$

La varianza dipende solo dall'ampiezza dell'intervallo $(b - a)$. Più l'intervallo è largo, più la distribuzione è dispersa, e maggiore è la varianza.

14.2 Variabile Aleatoria Normale (o Gaussiana)

È la distribuzione di probabilità più importante in assoluto. Moltissimi fenomeni naturali (altezze, pesi, errori di misurazioni) seguono questa distribuzione. Il grafico è la famosa forma a **campana**.



- **Parametri:** La distribuzione Normale è definita da 2 parametri:
  1. **μ (mu), la media:** Questo parametro determina il **centro** della campana. È il valore atteso $E(X)$.
  2. **σ^2 (sigma quadro), la varianza:** Questo parametro determina la **forma** della campana.
     - Un σ^2 grande significa molta variabilità, quindi la campana sarà larga e schiacciata.
     - Un σ^2 piccolo significa poca variabilità, quindi la campana sarà stretta e appuntita.
     - σ (senza il 2) è la deviazione standard.
- **Notazione:** Se una v.a. Normale ha media μ e varianza σ^2 scriviamo:

$$X \sim N(\mu, \sigma^2)$$

- **Funzione di densità $f(x)$:**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Produce la curva della campana. Il grafico è simmetrico rispetto alla sua media μ.

14.2.1 La Normale Standard (Z)

Esiste una Normale Standard (Z) che funge da riferimento per tutte le altre.

- È una normale con media 0 e varianza 1.

$$Z \sim N(0, 1)$$

È importante perché qualsiasi calcolo di probabilità per una qualsiasi variabile Normale $N(\mu, \sigma^2)$ può essere ricondotto ad un calcolo sulla Normale Standard $N(0, 1)$. I valori di probabilità per la Z sono tabulati in apposite tavole statistiche.

14.2.2 Standardizzazione: da X a Z

Se abbiamo una v.a. $X \sim N(\mu, \sigma^2)$, come la trasformiamo in una $Z \sim N(0, 1)$?

- **La formula inversa è:** $X = \sigma Z + \mu$. Se prendi una media Z (media 0, var 1), la "scali" di σ e la "trasli" di μ, ottieni una X (media μ, var σ^2).
- **La Standardizzazione è:**

$$Z = \frac{X-\mu}{\sigma}$$

1. **Sottrazione della media** $(X - \mu)$**:** Calcoliamo lo scarto di X dalla sua media. Il risultato è una nuova variabile con media 0.
2. **Divisione per la deviazione standard** $\frac{(X-\mu)}{\sigma}$**:** Dividiamo questo scarto per la deviazione standard. Questo processo si chiama "riscalare" e fa sì che la nuova variabile abbia varianza (e deviazione standard) pari ad 1.

Il risultato di questa trasformazione, z, viene chiamato **z-score**. Uno z-score di, es., $z = 1.5$ significa che il valore originale x si trova a 1.5 deviazioni standard sopra la media.

14.3 Il Calcolo Pratico di Probabilità con la Normale

Il problema: Vogliamo calcolare $P(a \leq X \leq b)$, dove $X \sim N(\mu, \sigma^2)$. Non possiamo risolvere l'integrale della densità Normale a mano.

La strategia in 3 passi

1. **Standardizzare l'intervallo:** Applichiamo la trasformazione $z = (x - \mu)/\sigma$ a tutti e tre i membri della disuguaglianza.

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right)$$

   Questo diventa: $P(z_a \leq Z \leq z_b)$, dove $z_a = (a - \mu)/\sigma$ e $z_b = (b - \mu)/\sigma$.

2. **Usare la Funzione di Ripartizione della Z (Φ):** La funzione di ripartizione della Normale Standard Z è così importante che ha un suo simbolo speciale: la lettera greca maiuscola Φ (Phi).

$$\Phi(z) = P(Z \leq z)$$

   I valori di $\Phi(z)$ si trovano nelle tavole statistiche o si calcolano con un software. Per trovare la probabilità di un intervallo, usiamo la proprietà:

   $$P(z_a \leq Z \leq z_b) = \Phi(z_b) - \Phi(z_a)$$

3. **Cercare i valori sulle tavole e calcolare.**

Esempio 14.1 (Altezza della popolazione)**. Setup:** L'altezza X è $N(175, 25)$. Quindi $\mu = 175$ e $\sigma^2 = 25$, il che significa $\sigma = \sqrt{25} = 5$.
Domanda: Qual è la frazione di popolazione più alta di 180cm? Vogliamo calcolare $P(X \geq 180)$.

Svolgimento:

1. **Standardizziamo:**

$$P(X \geq 180) = P\left(\frac{X - 175}{5} \geq \frac{180 - 175}{5}\right) = P(Z \geq 1)$$

2. **Usiamo Φ e la regola del complementare:** Le tavole ci danno $\Phi(z) = P(Z \leq z)$ (l'area a sinistra). Noi vogliamo l'area a destra.

$$P(Z \geq 1) = 1 - P(Z < 1) = 1 - \Phi(1)$$

3. **Cerchiamo il valore sulle tavole:** Andando a cercare sulle tavole della Normale Standard, si trova che $\Phi(1) \approx 0.8413$.

4. **Calcolo finale:**

$$P(X \geq 180) = 1 - 0.8413 = 0.1587$$

   Circa il 15.87% della popolazione è più alto di 180cm.

Esempio 14.2 (Problema Inverso)**.** Questo è il "gioco al contrario". Non ci viene dato un valore x per calcolare una probabilità, ma ci viene data una probabilità per trovare il valore x corrispondente.

Domanda: Trovare l'altezza x tale che il 5% della popolazione sia più bassa. Vogliamo trovare x tale che $P(X \leq x) = 0.05$.

Svolgimento:

1. **Standardizziamo:**
$$P(X \leq x) = P\left(Z \leq \frac{x - 175}{5}\right) = 0.05$$
Questo significa che: $\Phi\left(\frac{x-175}{5}\right) = 0.05$.

2. **Usiamo le tavole al contrario (Tavole dei Quantili):** Dobbiamo trovare quale z-score ha un'area alla sua sinistra pari a 0.05. Questo valore si chiama **quantile**. Cercando 0.05 all'interno della tabella (o usando la simmetria), si trova che lo z-score cercato è $z \approx -1.645$.

3. **Risolviamo per x:**
$$\frac{x - 175}{5} = -1.645$$
$$x - 175 = 5 \cdot (-1.645) = -8.225$$
$$x = 175 - 8.225 = 166.775$$
Il 5% della popolazione ha un'altezza inferiore a circa 166.8 cm.

Come usare la Tabella per Z Positivi

- La **colonna di sinistra** ti dà il valore di z fino al primo decimale.
- La **riga in alto** ti dà il secondo decimale.
- Il valore all'incrocio tra riga e colonna è $\Phi(z)$.

Esempio: Per trovare $\Phi(1.23)$:

1. Vai alla riga 1.2.
2. Spostati a destra fino alla colonna 0.03.
3. Il valore che trovi è 0.8907. Quindi $\Phi(1.23) = 0.8907$.

Come trovare i valori per Z NEGATIVI?

La tabella non li riporta perché non è necessario, grazie alla simmetria della curva a campana. Si usa la seguente regola:

$$\Phi(-z) = 1 - \Phi(z)$$

Esempio: Per trovare $\Phi(-1.23)$:

1. Trova $\Phi(1.23)$ dalla tabella: $\Phi(1.23) = 0.8907$.
2. Calcola $1 - 0.8907 = 0.1093$.
3. Quindi $\Phi(-1.23) = 0.1093$.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
+3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
+3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
+3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
+3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997
+4	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998

15 Statistiche Campionarie

- **Il contesto:** Immagina di voler conoscere l'altezza media di tutti gli italiani. È impossibile misurarli tutti. Cosa facciamo? Estraiamo un **campione** (es. 1000 persone) e misuriamo la loro altezza.
- **Modellizzazione (i.i.d.):** Modelliamo le osservazioni del nostro campione (n dati) come n variabili aleatorie $X_1, X_2, \ldots, X_n$. Facciamo un'assunzione fondamentale: che queste v.a. siano **i.i.d.** (indipendenti e identicamente distribuite).
  - **Indipendenti:** La scelta di una persona nel campione non influenza la scelta di un'altra.

– **Identicamente distribuite:** Tutte le X_i seguono la stessa distribuzione di probabilità (quella della popolazione generale da cui stiamo campionando). Hanno tutte la stessa media μ e la stessa varianza σ^2 (che però non conosciamo).

Definizione 15.1 (Statistica Campionaria)**.** "Ogni combinazione di tali variabili è detta statistica campionaria."
Spiegazione: Una statistica campionaria è una funzione calcolata a partire dai dati del campione. È essa stessa una variabile aleatoria, perché se estraessimo un campione diverso, otterremmo un valore diverso.

15.1 La Media Campionaria $\bar{X}$

Definizione 15.2. È la statistica campionaria più importante, usata per stimare la vera media della popolazione μ.

$$\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Proprietà della Media Campionaria

- **Valore Atteso di $\bar{X}$:** $E(\bar{X}) = \mu$
  Il valore atteso della media campionaria è uguale alla vera media della popolazione. Questo significa che la media campionaria è uno **stimatore corretto** (o non distorto). In media, non sovrastima né sottostima la vera media.
- **Varianza di $\bar{X}$:** $Var(\bar{X}) = \sigma^2/n$
  Questa è una proprietà FONDAMENTALE: all'aumentare della dimensione del campione n, la varianza della media campionaria **diminuisce**. Le medie calcolate su campioni più grandi sono più "stabili", "precise" e meno disperse attorno alla vera media μ.

15.2 I Due Teoremi Fondamentali della Statistica

Regola 15.1 (Legge dei Grandi Numeri (LLN))**. Traduzione in parole semplici:** Man mano che la dimensione del campione n diventa infinitamente grande, la probabilità che la media campionaria $\bar{X}$ sia "vicina quanto si vuole" alla vera media μ tende a 1 (diventa una certezza).
Significato pratico: Se prendi un campione abbastanza grande, la media che calcoli sarà una stima molto, molto buona della vera media.

Regola 15.2 (Teorema del Limite Centrale (TLC))**.** La LLN ci dice che la media campionaria converge a μ. Il TLC ci dice **come** si distribuisce durante questa convergenza.
Il risultato magico: Qualunque sia la distribuzione di partenza delle X_i, se n è abbastanza grande (di solito $n \geq 30$), la distribuzione della media campionaria $\bar{X}$ sarà **approssimativamente Normale**.

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Perché è così potente? Ci permette di usare la distribuzione Normale per fare calcoli e inferenza sulla media di una popolazione, anche se non sappiamo assolutamente nulla sulla forma della distribuzione di quella popolazione. Questa è la spina dorsale della statistica inferenziale.

16 Approssimazione Normale della Binomiale

16.1 Il Problema: Calcoli Complessi con la Binomiale

Esempio 16.1. Lanciamo una moneta 100 volte. Qual è la probabilità che il numero di teste X sia tra 40 e 70 (inclusi)?

Identificazione del modello: Il numero di successi (teste) in $n = 100$ prove di Bernoulli indipendenti con $p = 0.5$ è una v.a. Binomiale $X \sim B(100, 0.5)$.
Il calcolo esatto: Per trovare la risposta, dovremmo sommare le probabilità per ogni singolo valore:

$$P(40 \leq X \leq 70) = \sum_{k=40}^{70} P(X = k) = \sum_{k=40}^{70} \binom{100}{k} (0.5)^k (0.5)^{100-k}$$

Sarebbe un incubo computazionale da fare a mano.

16.2 La Soluzione: il Teorema del Limite Centrale

Il TLC ci offre una via d'uscita elegante.

- **L'idea:** Poiché la Binomiale X è una somma di n v.a. i.i.d. (le singole Bernoulli), se n è abbastanza grande, la sua distribuzione può essere approssimata da una Normale.
- **Verifica della condizione:** La regola pratica per una "buona" approssimazione è $np \geq 5$ e $n(1-p) \geq 5$. Nel nostro caso: $100 \cdot 0.5 = 50$, che è maggiore di 5. L'approssimazione sarà eccellente.
- **Identificazione della Normale approssimante:** La X Binomiale sarà approssimata da una Y Normale con la stessa media e varianza.
  - Media: $\mu = np = 100 \cdot 0.5 = 50$.
  - Varianza: $\sigma^2 = np(1-p) = 100 \cdot 0.5 \cdot 0.5 = 25$.

  Quindi, $X \sim B(100, 0.5)$ viene approssimata da $Y \sim N(50, 25)$.

16.3 Il Dettaglio Cruciale: La Correzione di Continuità

- **Il problema:** Stiamo approssimando una distribuzione a gradini (la Binomiale, che vive solo sugli interi 40, 41, 42...) con una curva liscia (la Normale, che vive su tutti i numeri reali).
- **Visualizzazione:** Immagina la distribuzione Binomiale come un istogramma, con barre centrate su 40, 41, etc. La barra per $X = 40$ in realtà copre l'intervallo da 39.5 a 40.5. La barra per $X = 70$ copre da 69.5 a 70.5.
- **La logica della correzione:** La probabilità $P(40 \leq X \leq 70)$ è la somma delle aree delle barre da 40 a 70. Per approssimare bene quest'area con la curva Normale, dobbiamo integrare sull'intervallo che copre **tutte le barre intere**, dal loro inizio alla loro fine.
- **La regola:** "si applica una correzione di continuità, cioè una modifica dell'intervallo di integrazione".

  La richiesta discreta $P(40 \leq X \leq 70)$ diventa una richiesta continua $P(39.5 \leq Y \leq 70.5)$.

Calcoli Finali dell'Esempio

Abbiamo trasformato il problema in $P(39.5 \leq Y \leq 70.5)$ dove $Y \sim N(50, 25)$.

1. **Standardizzare l'intervallo ($\mu = 50, \sigma = 5$):**
   - Estremo inferiore: $z_1 = (39.5 - 50)/5 = -2.1$.
   - Estremo superiore: $z_2 = (70.5 - 50)/5 = 4.1$.

   La nostra probabilità diventa $P(-2.1 \leq Z \leq 4.1)$.

2. **Usare la Φ:**
   $$P(-2.1 \leq Z \leq 4.1) = \Phi(4.1) - \Phi(-2.1)$$
   Usando la simmetria $\Phi(-2.1) = 1 - \Phi(2.1)$:
   $$\Phi(4.1) - (1 - \Phi(2.1)) = \Phi(4.1) + \Phi(2.1) - 1$$

3. **Cercare i valori sulle tavole e calcolare:** $\Phi(4.1) \approx 1$ e $\Phi(2.1) \approx 0.9821$.
   $$P \approx 1 + 0.9821 - 1 = 0.9821$$

Conclusione: C'è circa il 98.2% di probabilità di ottenere un numero di teste compreso tra 40 e 70.