# Elementi di probabilità

Alessandra Nardi Gianpaolo Scalia Tomba Caius Gavrila Livio Finos

Dipartimento di Matematica

Università di Roma "Tor Vergata"

8 maggio 2021

# Outline

Il calcolo delle probabilità fa riferimento ad un *prova o esperimento* il cui esito è incerto ma di cui siamo in grado di elencare i possibili risultati $\Omega$.
Un *evento* è una qualsiasi affermazione circa il risultato di un esperimento purché a posteriori si sia in grado di stabilire se tale affermazione si sia verificata o meno.
Formalmente ogni evento è un sottoinsieme di $\Omega$.
Indichiamo con $\mathcal{A}_\Omega$ la classe degli eventi osservabili che sono oggetto di studio
Lanciamo due volte una moneta
$\Omega = \{TT, CC, TC, CT\}$
$\Omega$ costituisce l'evento *certo* poiché uno di questi risultati necessariamente si verificherà .
Come è fatta $\mathcal{A}_\Omega$?

Gli eventi sono assimilabili ad insiemi e quindi su di loro possiamo fare tutte le operazioni definite sugli insiemi

- Negazione o complemento di A
  $A^c = \overline{A} = \{\omega \in \Omega : \omega \ni A\}$
  Il complemento di A contiene tutti gli elementi di $\Omega$ che non si trovano in A. Il complemento di $\Omega$ è l'insieme vuoto $\emptyset$.
- Unione di due eventi
  $A \cup B = \{\omega \in \Omega : \omega \in A \; oppure \; \omega \in B\}$
  L'unione di due eventi si verifica se si verifica almeno uno dei due eventi
- Intersezione di due eventi
  $A \cap B = \{\omega \in \Omega : \omega \in A \; e \; \omega \in B\}$
  L'intersezione di due eventi si verifica solo se si verificano entrambi gli eventi.
- Implicazione
  Se $A \subset B$ allora il verificarsi A implica il verificarsi di B.

- Eventi incompatibili
  A e B sono incompatibili se $A \cap B = \emptyset$ cioè se non hanno elementi comuni. Si tratta di eventi che non possono verificarsi allo stesso tempo. Più eventi sono incompatibili se lo sono a due a due.
- Eventi necessari
  A e B sono necessari se $A \cup B = \Omega$.
  Più eventi sono necessari se la loro unione è $\Omega$

Un insieme di eventi necessari e incompatibili è chiamato *partizione* di $\Omega$.

- Legge di De Morgan
  $(A \cup B)^c = A^c \cap B^c$
- Proprietà associativa
  $(A \cup B) \cup C = A \cup (B \cup C)$
  $(A \cap B) \cap C = A \cap (B \cap C)$
- Proprietà distributiva
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  Da cui
  $A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$

# L'incertezza, la variabilità e la probabilità

Caratteristica non eliminabile degli esperimenti in biomedicina è l'incertezza del risultato finale. Anche immaginando di poter ripetere l'esperimento in condizioni analoghe, l'esito finale sarebbe comunque in qualche misura diverso a motivo di una serie di fattori non controllabili dal ricercatore che concorrono a determinare quella che siamo soliti chiamare variabilità sperimentale (o casuale).
Nel valutare questa variabilità ci serviamo del calcolo delle probabilità che assume per noi l'aspetto di logica dell'incerto.
Mentre il concetto di probabilità è intuitivo, la sua definizione è complessa al punto che esistono diverse scuole di pensiero.

# La probabilità secondo Pascal

Il primo tentativo di dare una definizione formale di probabilità è generalmente attribuito a Blaise Pascal (1623-1662) e documentato dalla sua corrispondenza con Pierre Fermat.
Sotto la pressione di un suo amico, il Cavalier de Méré, accanito giocatore, Pascal definisce "la probabilità di un evento come il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili, *purché questi ultimi siano tutti ugualmente probabili* ".
La definizione è tautologica e limitata ma immediatamente operativa.
Inoltre stabilisce uno stretto legame tra frequenza relativa e probabilità , legame che ritroveremo anche nell'ambito dell'inferenza statistica.

# L'impostazione frequentista

L'impostazione frequentista si fonda sull'idea che in una successione di prove fatte nelle stesse condizioni la frequenza (relativa) di un evento si avvicina alla sua probabilità e che l'approssimazione tende a migliorare con l'aumentare del numero delle prove (legge empirica del caso).
Si stabilisce così uno stretto legame tra frequenze relative e probabilità al punto che quest'ultima viene definita come il limite a cui tende la frequenza relativa dell'evento al tendere ad infinito del numero delle prove.
Se immaginiamo di lanciare una moneta ripetute volte, la probabilità che esca testa sarà pari al limite a cui tende la frequenza relativa delle prove in cui è uscita testa (1/2 soltanto se la moneta non è stata truccata !).

# L'impostazione soggettiva

Se l'evento non è ripetibile?
Ci aiuta l'impostazione soggettiva che definisce la probabilità come il grado di fiducia che una persona (equa e coerente) ha nel verificarsi dell'evento.
Definizione bella, ampia ma non operativa (soprattutto per la difficoltà di trovare una persona equa e coerente).
Aggiungiamo che: la probabilità è il prezzo che un individuo ritiene equo pagare per ricevere 1 se l'evento si verifica (e 0 se non si verifica).
Inoltre: le probabilità degli eventi devono essere attribuite in modo che non sia possibile ottenere con un insieme di scommesse una vincita certa o una perdita certa.

# L'impostazione assiomatica (A. Kolmogorov, 1933)

Sia $\Omega$ l'insieme dei possibili eventi "elementari" (se immaginiamo il semplice lancio di un dado bilanciato $\Omega = \{1, 2, 3, 4, 5, 6\}$ )
Indichiamo con $\mathcal{A}_\Omega$ la classe additiva che con contiene tutti i possibili sottoinsiemi di $\Omega$ cioè tutti gli eventi osservabili, semplici e composti.
Questa classe risulta *chiusa* rispetto alle oprazioni di unione e negazione.
La probabilità è una funzione che assegna ad ogni evento (cioè ad ogni insieme in $\mathcal{A}_\Omega$ ) un numero reale e che soddisfa i seguenti assiomi:

1. $P(A) \geq 0$ per ogni elemento $A \in \mathcal{A}_\Omega$
2. $P(\Omega) = 1$ (se l'evento è certo la probabilità è $1$)
3. Se $A$ e $B$ sono incompatibili $P(A \cup B) = P(A) + P(B)$

Questi assiomi costituiscono la base per la cosruzione del calcolo delle probabilità secondo un'impostazione ipotetico deduttiva.

Alcune prime proprietà :

- a $P(\overline{A}) + P(A) = 1$
- b $P(\emptyset) = 0$
  dim: $P(\emptyset) = P(\overline{\Omega}) = 1 - P(\Omega) = 0$
- c $P(A) \leq 1$ per ogni evento $A \in \mathcal{A}_\Omega$
  dim: $P(A) = 1 - P(\overline{A}) \leq 1$ dove $P(\overline{A}) \geq 0$
- d Se $A \subset B$ allora $P(A) \leq P(B)$
  dim: $B = (B \cap A) \cup (B \cap \overline{A}) = A \cup (B \cap \overline{A})$ quindi
  $P(B) = P(A) + P(B \cap \overline{A}) \geq P(A)$ (si noti: $P(B \cap \overline{A}) \geq 0$ e $(B \cap A) \cap (B \cap \overline{A}) = \emptyset$)

# La legge delle probabilità totali

Questa legge estende l'assioma 3 valido per eventi incompatibili al caso di eventi anche compatibili. La presentamo nel caso semplice di 2 eventi ma resta valida anche nel caso di più di 2 eventi.

Se $A \cap B \neq \emptyset$

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

Es:
$A = \{1, 2, 3\}$, $P(A) = 1/2$
$B = \{\text{Dispari}\}$, $P(B) = 1/2$
$A \cap B = \{1, 3\}$, $P(A \cap B) = 1/3$
$P(A \cup B) = 1/2 + 1/2 - 1/3 = 2/3$
per verifica si anche noti che $A \cup B = \{1, 2, 3, 5\}$ quindi
$P(A \cup B) = 4/6 = 2/3$.

# La legge delle probabilità composte

Se $P(B) \neq 0$
si dice probabilità (condizionata) di A dato B, cioè noto che l'evento B si è verificato,

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

Es:
$A = \{1\}$, $P(A) = 1/6$
$B = \{\text{Dispari}\}$, $P(B) = 1/2$
$P(A|B) = 1/3 = \frac{1/6}{1/2}$
(si noti che $A \cap B = A \subset B$ e quindi $P(A \cap B) = P(A)$)
NB: Il condizionamento all'evento $B$ comporta una modifica dello spazio $\Omega$ (e dell'algebra ad esso associata), nel caso in esame: $\Omega_B = \{1, 3, 5\}$ e quindi in $\Omega$ abbiamo $P(A) = 1/3$.

# L'indipendenza

$A$ e $B$ sono eventi indioendenti se si verifica la relazione

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

La probabilità dell'evento $A$ non cambia condizionatamente all'evento $B$. Cioè sapere che l'evento $B$ si è verificato non modifica la probabilità che si verifichi $A$.

Dalla definizione di indipendenza consegue (cioè valgono le seguenti affermazioni *se e solo se* $A$ e $B$ sono indipendenti)

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

Es 1: Lanciamo un dado due volte.
$A = \{6 \text{ al primo lancio di dado}\} =$
$\{6 \text{ al primo lancio di dado e un risultato qualunque al secondo}\} =$
$\{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\},$
$P(A) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$

$B = \{6 \text{ al secondo lancio di dado}\} =$
$\{6 \text{ al secondo lancio di dado e un risultato qualunque al primo}\} =$
$\{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\},$
$P(B) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P((6,6))}{P(B)} = \frac{1/36}{1/6} = 1/6 = P(A)$$

altrettanto vale

$$P(A \cap B) = P((6,6)) = \frac{1}{36} = P(A)P(B) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}$$

Es 2:
$A = \{1, 2\}$, $P(A) = 1/3$
$B = \{\text{Dispari}\}$, $P(B) = 1/2$

$$P(A|B) = \frac{P(1)}{P(B)} = \frac{1/6}{1/2} = 1/3 = P(A)$$

Es 3: Estrazioni senza reinserimento
Un'urna con 5 palline Rosse e 5 palline Verdi.
Estrazioni successive senza reinserire le palline estratte

Determinare:

a $P(\text{pallina rossa alla prima estrazione}) = P(R1)$
sol: $P(R1) = 5/10$

b $P(R1 \cap R2)$
sol: $P(R1 \cap R2) = \frac{5 \cdot 4}{10 \cdot 9} = \frac{2}{9}$ (con la definizione di Pascal).

c $P(R2|R1)$
sol: Con le formule: $P(R2|R1) = \frac{P(R1 \cap R2)}{P(R1)} = \frac{2/9}{1/2} = \frac{4}{9}$
per costruzione logica: "Se la prima estrazione è una pallina rossa, alla seconda estrazione ci sono 4 rosse e 5 verdi", quindi:
$P(R2|R1) = \frac{4 \text{ rosse}}{4 \text{ rosse}+5 \text{ verdi}} = \frac{4}{9}$

d $R2$ e $R1$ sono indipendenti?
per rispondere devo conoscere $P(R1 \cap R2)$, $P(R1)$, $P(R2)$.

e Calcoliamo $P(R2)$, altre sono note.
sol: $P(R2) = P(R2 \cap R1) \cup P(R2 \cap \overline{R1}) =$
$P(R2|R1)P(R1) + P(R2|\overline{R1})P(\overline{R1}) =$
$P(R2|R1)P(R1) + P(R2|V1)P(V1) = \frac{4}{9}\frac{1}{2} + \frac{5}{9}\frac{1}{2} = \frac{1}{2}$.
sol punto [d]: $P(R1 \cap R2)? =?P(R1)P(R2)$, $\frac{4}{9}? =?\frac{1}{2}\frac{1}{2}$ NO! i due eventi sono dipendenti.

Es 4: il problema del cavaliere de Méré

Nel 1600 in Francia il gioco d'azzardo era di gran moda e le bische clandestine diffusissime. Un gioco estremamente alla moda allora era il seguente: la "casa" scommetteva alla pari con un giocatore che quest'ultimo, lanciando per 4 volte un dado, avrebbe ottenuto almeno una volta il numero 6. Come vedremo successivamente questo gioco è leggermente favorevole alla casa che "in media" vince il $52\%$ delle volte. Un distinto ed intelligente francese dell'epoca, Antoine Gombauld Cavalier de Méré, frequentatore delle case da gioco, conosceva bene quel gioco ed era incuriosito da una sua possibile variante: la casa scommetteva alla pari con un giocatore che quest'ultimo, lanciando per 24 volte una coppia di dadi, avrebbe ottenuto almeno una volta il doppio 6. Anche questo gioco, secondo il Cavalier de Méré, avrebbe dovuto essere leggermente favorevole alla casa per questo motivo: quando si lancia un dado vi sono 6 possibili risultati, quindi la probabilità che esca il 6 sarà $1/6$. Invece la probabilità che esca un doppio 6 sarà di $1/36$ (essendo 36 i risultati distinti del lancio di due dadi), quindi 6 volte più bassa; lanciando la coppia di dadi 6 volte di più $(24 = 6 \cdot 4)$ si dovrebbe controbilanciare l'effetto di considerare un evento meno probabile di un fattore 6 e si dovrebbe avere quindi la stessa probabilità.

Invece non è così, quest'ultimo gioco non è favorevole alla casa, ma al giocatore; ne era consapevole il Cavalier de Méré, non è chiaro se per averlo provato a sue spese o per qualche intuizione teorica. In ogni caso decise di parlarne con un altro brillante francese, Blaise Pascal, che si dilettava di lettere, di filosofia, di teologia, ma anche di matematica. Pascal risolse il problema postogli dal de Méré provando anche che con 25 lanci il gioco sarebbe allora stato favorevole alla casa.

Calcolare la probabilità degli eventi:

$A$= { lanciando 4 volte un dado si ottiene almeno un 6 }

$B$= { lanciando 24 volte 2 dadi si ottiene almeno un doppio 6 }.

L'evento complementare di $A$ è più semplice:
$\overline{A}$ = { lanciando 4 volte un dado non si ottiene alcun 6 }
poi sfrutteremo la proprietà: $P(A) = 1 - P(\overline{A})$.
Indichiamo con $T_1$,$T_2$,$T_3$,$T_4$ gli eventi: nessun 6 al primo lancio, nessun 6 al secondo lancio, . . . , nessun 6 al quarto lancio.
Ora si noti che

- $T_1 \cap T_2 \cap T_3 \cap T_4 = \overline{A}$
- $T_1$,$T_2$,$T_3$,$T_4$ sono indipendenti.

Allora: $P(\overline{A}) = P(T_1 \cap T_2 \cap T_3 \cap T_4) = P(T_1)P(T_2)P(T_3)P(T_4) = (5/6)(5/6)(5/6)(5/6) = (5/6)^4$
Quindi $P(A) = 1 - (5/6)^4 = 0.518$

Analogamente per $P(B)$ indichiamo con $T_1$,$T_2$,...,$T_{24}$ gli eventi: nessun doppio 6 al primo (doppio) lancio, nessun doppio 6 al secondo (doppio) lancio, ..., nessun doppio 6 al ventiquattresimo (doppio) lancio.
$P(T_i) = 35/36$ (per ogni $i = 1, \ldots, 24$)
$P(\overline{B}) = P(T_1 \cap T_2 \cap \ldots \cap T_{24}) = P(T_1)P(T_2)\ldots P(T_{24}) == (35/36)^{24}$
$P(B) = 1 - P(\overline{B}) = 1 - (35/36)^{24} = 0.491$
cosa succede con 25 lanci di due dadi? La situazione si ribalta, e la casa è nuovamente in vantaggio! Verificare per casa: (soluzione P=0.5055)

Es 5: Il problema dei compleanni
Consideriamo 25 persone a una festa. Qual è la probabilità di avere **almeno** 2 compleanni uguali (evento A)?
Essendo eventi indipendenti, la probabilità che i compleanni siano tutti diversi (evento $\overline{A}$):
$P(\overline{A}) = (364/365)(363/365)(362/365)\dots(341/365) = 0.432$
$P(\overline{A}) = 0.432$, $P(A) = 1 - 0.432 = 0.568$
Più preciso:

$$P(\{\text{tutti i compleanni sono diversi}\}) =$$

$$P(\underbrace{\{\text{primi 25 compleanni sono diversi}\}}_{A_{25}}) =$$

$$P(A_{25}) =$$

$$P(\underbrace{\{\text{primi 24 compleanni sono diversi}\}}_{A_{24}} \cap \underbrace{\{\text{il 25-iessimo è diverso dei primi 24}\}}_{B_{25}}) =$$

$$P(A_{24} \cap B_{25}) = P(B_{25}|A_{24}) = 341/365 \cdots P(A_{24}) = \dots$$

NB: Non abbiamo preso in considerazione gli anni bisestili.

# Il teorema di Bayes

Dati due eventi $A$ e $B$ con $P(A)$ e $P(B)$ diverse da zero applichiamo la definizione di probabilità condizionata $P(A|B) = P(A \cap B)/P(B)$ da cui
$P(A|B)P(B) = P(A \cap B)$
Possiamo invertire il ruolo dei due eventi ed ipotizzare che sia l'evento A adessere noto $P(B|A) = P(A \cap B)/P(A)$ da cui
$P(B|A)P(A) = P(A \cap B)$
Uguagliando i membri di sinistra delle uguaglianze ricaviamo il Teorema di Bayes nella sua versione più elementare

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

.

# Applicazione del Teorema di Bayes ad un test diagnostico

Un test *diagnostico* di tipo ELISA per anticorpi all'HIV ha i seguenti parametri:
sensitività $= 0.99$
specificità $= 0.99$
Una persona risulta positiva al test. Con quale probabilità la persona è veramente HIV positiva?

Definiamo seguenti eventi:
H+ = persona HIV-positiva,
H- = persona HIV-negativa,
T+ = test positivo,
T- = test negativo
Con l'aiuto di questi eventi definiamo le due caratteristiche del test:
P(T+ | H+) = 0.99 (sensitività)
P(T- | H- ) = 0.99 (specificità)
Attenzione agli eventi complementari
P(T- | H+) = 0.01
P(T+ | H- ) = 0.01
La domanda si può riformulare in termini probabilistici:
$P(H+|T+) = ?$
(qual è la probabilità di essere HIV-positivo quando il test è risultato positivo ?).

Dobbiamo calcolare $P(H^+ \mid T^+)$ conoscendo

- la specificità $P(T^- \mid H^-)$,
- la sensitività $P(T^+ \mid H^+)$ e
- la prevalenza della malattia $P(H^+)$ .

Applicando il **Teorema di Bayes**:

$$P(H^+ \mid T^+) = P(T^+ \mid H^+) \frac{P(H^+)}{P(T^+)}.$$

L'unica quantità non nota è $P(T^+)$. Ma

$$P(T^+) = P(T^+ \cap H^+) + P(T^+ \cap H^-)^1$$

e dalla definizione della **probabilità condizionata**:

$$P(T^+ \cap H^+) = P(T^+ \mid H^+) P(H^+)$$

$$P(T^+ \cap H^-) = P(T^+ \mid H^-) P(H^-) = {}^2 (1 - P(T^- \mid H^-))(1 - P(H^+))$$

[1] Abbiamo decomposto l'evento $T^+$ in due eventi incompatibili ed esaustivi.

[2] Applichiamo due volte la formula dell'evento complementare.

Risulta che

$$P(H^+ \mid T^+) = P(T^+ \mid H^+)\frac{P(H^+)}{P(T^+|H^+)P(H^+)+(1-P(T^-|H^-))(1-P(H^+))}$$

dove tutte le quantità a destra sono conosciute. Calcolando

$$P(H^+ \mid T^+) = 0.99 \cdot \frac{P(H^+)}{0.99 \cdot P(H^+) + 0.01 \cdot (1 - P(H^+))}.$$

Ci serve allora $P(H+)$, la probabilità a priori di essere HIV-positivo.

Una stima attuale è di circa 100.000 infetti in Italia, dunque circa $1/500$, rispetto alla popolazione italiana.
Sostituendo nell'espressione precedente (provare per credere) otteniamo $P(H^+ \mid T^+) = 99/598 = 16.5\%$.

# Cosa non ha funzionato?

Il valore $1/500$ riflette la probabilità che sia infetta una persona scelta **a caso** nella popolazione, ad esempio una situazione di screening.
L'alta percentuale di falsi positivi indica quanto complicato sarebbe uno screening pensato sull'intera popolazione. Tuttavia, nel caso di un soggetto specifico, $P(H^+)$ descrive la sua *probabilità a priori*, cioè prima di effettuare il test, di essere HIV positivo. Se la persona appartiene a un **gruppo a rischio** possiamo ipotizzare che $P(H^+) = 1/100$, da cui $P(H^+ \mid T^+) = 99/198 = 0.50$.

Un test diagnostico di fatto modifica la nostra probabilità a priori di essere malati in una probabilità a posteriori. Nal caso considerato siamo passati da 0.001 a 0.50.
Ne segue che l'utilità del test non è assoluta ma **dipende** in modo critico dalla probabilità a priori dell'individuo.
Se considerassimo un test migliore, arrivando ad una specificità e sensibilità di $0.999\%$ le probabilità calcolate diventerebbero

| | 99% | 99.9% |
|---|---|---|
| Non a rischio | 0.17 | 0.66 |
| A rischio | 0.50 | 0.90 |

# Bayes e COVID-19

Veniamo ai test disponibili per la diagnosi di COVID-19 e consideriamo il test molecolare (PCR based) caratterizzato da un'alta specificità ma anche da una minore sensibilità .
$P(T+ \mid C+) = 0.70 \; P(T- \mid C+) = 0.30$
$P(T- \mid C-) = 0.95 \; P(T- \mid C+) = 0.05$

Consideriamo diversi scenari con diverse probabilità a priori
Ipotizziamo un primo scenario caratterizzato da una bassa probabilità a priori $P(C+) = 0.05$
Avremo
$P(C+ \mid T+) = \frac{(0.70 \cdot 0.05)}{(0.05 \cdot 0.70)+((1-0.95)\cdot(1-0.05))} = 0.42$
$P(C- \mid T+) = 0.58$

$P(C- \mid T-) = \frac{(0.95 \cdot (1-0.05))}{(0.95 \cdot (1-0.05)+(1-0.70)\cdot 0.05)} = 0.98$
$P(C+ \mid T-) = 0.02$

Ora aumentiamo la probabilità a priori portandola a $P(C+) = 0.50$
Avremo

$P(C+ \mid T+) = \frac{(0.70 \cdot 0.50)}{(0.50 \cdot 0.70) + ((1-0.95) \cdot (1-0.50))} = 0.93$
$P(C- \mid T+) = 0.07$

$P(C- \mid T-) = \frac{(0.95 \cdot (1-0.50))}{(0.95 \cdot (1-0.50) + (1-0.70) \cdot 0.50)} = 0.76$
$P(C+ \mid T-) = 0.24$

Infine consideriamo il caso di una probabilità a priori molto alta
$P(C+) = 0.80$
Avremo

$$P(C+ \mid T+) = \frac{(0.70 \cdot 0.80)}{(0.80 \cdot 0.70) + ((1-0.95) \cdot (1-0.80))} = 0.98$$
$$P(C- \mid T+) = 0.02$$

$$P(C- \mid T-) = \frac{(0.95 \cdot (1-0.80))}{(0.95 \cdot (1-0.80) + (1-0.70) \cdot 0.80)} = 0.44$$
$$P(C+ \mid T-) = 0.56$$

Abbiamo visto che un soggetto ad alto rischio con un risultato negativo del test conserva una probabilità niente affatto trascurabile di essere postivo.

$P(C- \mid T-) = \frac{(0.95 \cdot (1-0.80))}{(0.95 \cdot (1-0.80)+(1-0.70) \cdot 0.80)} = 0.44$

$P(C+ \mid T-) = 0.56$

Immaginiamo che il soggetto ripeta il test ancora con esito negativo: cosa accade in probabilità ?

La probabilià a posteriori del test precedente $P(C+ \mid T-) = 0.56$ diventa adesso la sua probabilità a priori da rivalutare in base al nuovo risultato.

$P(C- \mid T-) = \frac{(0.95 \cdot (1-0.56))}{(0.95 \cdot (1-0.56)+(1-0.70) \cdot 0.56)} = 0.71$

$P(C+ \mid T-) = 0.29$

Ancora non avrà la certezza di essere negativo ma certamente sarà più tranquillo.

# Es 7: La diagnosi di un medico

Sia $E$ un insieme di sintomi che presenta un paziente. Il medico sa che tali sintomi sono associati a *sole* 3 possibili malattie, incompatibili fra di loro, $H_1$, $H_2$ e $H_3$ con probabilità
$P(E|H_1) = 0.90$, $P(E|H_2) = 0.10$ e $P(E|H_3) = 0.30$
Le malattie hanno diverse prevalenze:
$P(H_1) = 0.03$, $P(H_2) = 0.70$ e $P(H_3) = 0.27$
Applichiamo il teorema di Bayes: $P(H|E)$

$$P(H_1|E) = \frac{P(H_1)P(E|H_1)}{P(E)} = \frac{P(H_1)P(E|H_1)}{\sum_{i=1}^{3} P(H_i)P(E|H_i)} = \frac{0.03\ 0.90}{(0.03\ 0.90)+(0.70\ 0.10)+(0.27\ 0.30)} = 0.15169$$

$P(H_2|E) = 0.39326$
$P(H_3|E) = 0.45505$ allora la malattia $H_3$ è più probabile.

# La variabile aleatoria Bernoulliana

Immaginiamo di lanciare una moneta (bilanciata)
La variabile aleatoria che descrive questo elementare esperimento è un'applicazione $X$ che associa valore 1 al risultato testa e valore 0 al risultato croce
Più in generale definiamo come variabile aleatoria un'applicazione che associa al risultato di un esperimento un numero reale
Ad ogni valore del dominio ($\Omega$) resta associata una (e una sola) probabilità di accadimento.
$P(X=0) = P(X^{-1}(0)) = P(Croce) = 1/2$
$P(X=1) = P(X^{-1}(1)) = P(Testa) = 1/2$
dove
$\sum_{i=1}^{c} p(x_i) = 1$, $p(x) \geq 0$

In genere la probabilità di *successo* non è nota (perché altrimenti fare un esperimento ?) e costituisce il parametro incognito di interesse
$P(X = 0) = 1 - \pi$
$P(X = 1) = \pi$
in una formula: $P(X = x) = \pi^x (1 - \pi)^{1-x}$
Quella definita è la famiglia Bernoulliana che contiene distribuzioni diverse, una per ogni valore che $\pi$ può assumere
Una variabile aleatoria si caratterizza definendo l'insieme dei valori che può assumere e la relative probabilità

# Altri esempi di variabili aleatorie

Sia Z il risultato del lancio di un dado a 6 facce.
$\Omega = \{1, 2, 3, 4, 5, 6\}$ (dominio, campo di esistenza),
$P(Z = 1) = 1/6$
$P(Z = 2) = 1/6$
$P(Z = 3) = 1/6$
$P(Z = 4) = 1/6$
$P(Z = 5) = 1/6$
$P(Z = 6) = 1/6$

# Variabili aleatorie e variabili statistiche

Una variabile aleatoria descrive le probabilità dei possibili risultati **osservabili** di un esperimento (per chiarezza possiamo immaginarlo non svolto)

Una variabile statistica sintetizza i risultati osservati

Immaginiamo di aver eseguito $n$ lanci di una moneta, lanci tra loro indipendenti e condotti nelle medesime condizioni, dove la probabilità di successo resta invariata. Sia $n = 1000$, è possibile sintetizzare i risultati **osservati** attraverso una variabile statistica $\tilde{X}$

| Valori osservati | Frequenze assolute $n_i$ | Frequenze relative $f(x_i) = n_i/n$ |
|---|---|---|
| 0 (croce) | 491 | $0.491$ |
| 1 (testa) | 509 | $0.509$ |
| Totale | 1000 | 1 |

Es: lancio di un dado $n = 6000$ volte.

| $x_i$ | $n_i$ | $f(x_i)$ |
|---|---|---|
| 1 | 1000 | 0.1667 |
| 2 | 974 | 0.1623 |
| 3 | 949 | 0.1582 |
| 4 | 1029 | 0.1715 |
| 5 | 1051 | 0.1752 |
| 6 | 997 | 0.1662 |
| Totale | 6000 | 1 |

Prima di svolgere un esperimento i risultati sono **osservabili** e ad essi vengono associate delle probabilità ; quando i risultati dell'esperimento sono stati **osservati** le probabilità vengono sostitute dalle corrispondenti frequenze relative
Torna (e tornerà ) il legame tra probabilità e frequenze relative già stabilito della definizione frequentista della probabilità

Come "si comporta" un variabile aleatoria?

- Conoscere la caratteristiche di $X$ (ad esempio media e varianza) ci aiuta prevedere i risultati dell'esperimento.

  es: sapere che il lancio di una moneta è descrivibile con una v.a. Bernoulliana $\mathcal{B}e(1/2)$ ci aiuta a prevedere quale sarà la proporzione attesa di teste.

- Osservare i risultati dell'esperimento ci aiuta a capire la v.a. che regola il fenomeno.

  es: sappiamo che la moneta è descrivibile con una v.a. Bernoulliana $\mathcal{B}e(\pi)$, ma condurre l'esperimento ci può aiutare a decidere se la moneta è equilibrata (cioè se la proporzione media di teste e pari a quella delle croci).

# La media di una variabile aleatoria discreta

$X$ è una v.a. discreta se può assumere un numero finito (o al più numerabile) $c$ di valori distinti
Il suo valore atteso (media) si definisce come $E[X] = \sum_{i=1}^{c} x_i \ p(x_i)$
dove $p(x_i) = P(X = x_i)$

Osserviamo l'analogia con la definizione di media per variabili statistiche
$Media[\tilde{X}] = \sum_{i=1}^{c} x_i \ f(x_i)$
Nel caso della v.a. Bernoulliana avremo
$E[X] = 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi$
Eseguito l'esperimento
$Media[\tilde{X}] = 1 \cdot 0.509 + 0 \cdot 0.491 = 0.509$

# La varianza di una variabile aleatoria discreta

$V[X] = E((X - E[X])^2) = \sum_{i=1}^{c}(x_i - E[X])^2 p(x_i)$
in l'analogia con la definizione di varianza per variabili statistiche
$Var[\tilde{X}] = \sum_{i=1}^{c}(x_i - Media[\tilde{X}])^2 f(x_i)$

Nel caso della v.a. Bernoulliana avremo
$V[X] = (1-\pi)^2\,\pi + (0-\pi)^2\,(1-\pi) = \pi\,(1-\pi)$
Eseguito l'esperimento
$Var[\tilde{X}] = (1-0.509)^2\,0.509 + (0-0.509)^2\,0.491 = 0.250$

# La funzione di ripartizione . . .

. . . è la funzione che risponde alla domanda:

"qual' è la probabilità che il risultato sia minore o uguale ad $x$"

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

Proprietà di $F(x)$:

- è non decrescente (se $x < y$ allora $F(x) \leq F(y)$)
- è limitata tra 0 e 1 ( $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$)
- è continua a destra ($\lim_{x \to x_0^+} = F(x_0)$)

# La variabile aleatoria Binomiale

Ritorniamo all'esperimento del lancio di una moneta e immaginiamo di eseguire 3 lanci indipendenti della stessa moneta. Le terne ordinate hanno queste probabilità di essere osservate

| Y | $X_1$ | $X_2$ | $X_3$ | Prob |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | $(1-\pi)^3$ |
| 1 | 1 | 0 | 0 | $(1-\pi)^2\pi$ |
| 1 | 0 | 1 | 0 | $(1-\pi)^2\pi$ |
| 1 | 0 | 0 | 1 | $(1-\pi)^2\pi$ |
| 2 | 1 | 1 | 0 | $(1-\pi)\pi^2$ |
| 2 | 1 | 0 | 1 | $(1-\pi)\pi^2$ |
| 2 | 0 | 1 | 1 | $(1-\pi)\pi^2$ |
| 3 | 1 | 1 | 1 | $\pi^3$ |

La probabilità congiunta della generica terna $(x_1, x_2, x_3)$ è

$$P((X_1, X_2, X_3) = (x_1, x_2, x_3)) =$$
$$\pi^{x_1}(1-\pi)^{1-x_1}\ \pi^{x_2}(1-\pi)^{1-x_2}\ \pi^{x_3}(1-\pi)^{1-x_3}$$

Possiamo essere interessati a conoscere direttamente il valore di Y= Numero di teste (successi) osservate piuttosto che la terna ordinata (l'ordine di apparizione non ci interessa visto che sono eventi indipendenti ottenuti nelle medesime condizioni)

- Y=$\sum_{i=1}^{3} X_i$
- Y può assumere i valori (Dominio di Y ): $\Omega_Y = (0, 1, 2, 3)$.
- Y=k con quale probabilità?

Y=2 con quale probabilità ?
$P(Y=2) = P(X_1 = 1, X_2 = 1, X_3 = 0) + P(X_1 = 1, X_2 = 0, X_3 = 1) + P(X_1 = 0, X_2 = 1, X_3 = 1) = 3\,\pi^2\,(1-\pi)$

Quali leggi della probabilità abbiamo utilizzato? Quali le ipotesi sfruttate?

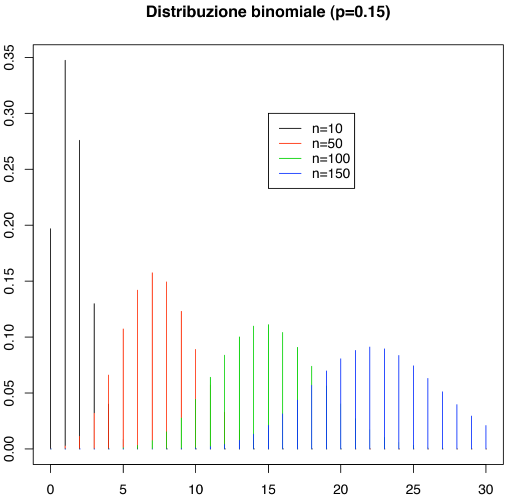
In generale

$$P(Y=k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

Come calcolare $\binom{n}{k}$ (combinazioni di $n$ su $k$), il numero di possibili terne non ordinate con k teste n-k croci?

- devo assegnare $n$ posizioni (l'ordine di lancio) a $k$ teste e $n-k$ croci.
  prima testa : ho n posizioni possibili.
  seconda testa : ho n-1 posizioni possibili.
  ...
  k-esima testa: ho n-k+1 posizioni possibili.
  prima croce: ho n-k posizioni possibili.
  (n-k)-esima croce: ho 1 sola posizione possibile.
  quindi ho $n!$ possibili sequenze di $k$ teste e $n-k$ croci
- ora però devo considerare che le $k$ teste sono **indistinguibili** e che esistono
  - $k!$ possibili ordinamenti indistinti delle teste.
  - $(n-k)!$ possibili ordinamenti indistinti delle croci.

  (quindi la medesima sequenza ordinata di 0 e 1 è ottenibile con $k!$ ordinamenti solo apparentemente diversi delle teste e $(n-k)!$ ordinamenti possibili delle croci
- quindi in tutto ho ci sono $\binom{n}{k} = \frac{n!}{k!\ (n-k)!}$ modi per ottenere $k$ teste con $n$ lanci




Cosa accade se $n \to \infty$?

# Dal discreto al continuo: un passaggio delicato

Cosa accade se il risultato del nostro esperimento è una variabile continua $(\Omega = \mathbb{R})$?
Non è più possibile assegnare una probabilità ad ogni singolo valore osservabile (sono davvero troppi!)
Passiamo dal punto all'intervallo (per quanto piccolo conterrà sempre un numero infinito di punti) e definiamo

$$P(x \leq X < x + dx) = \int_x^{x+dx} f(t)dt \approx f(x)dx$$

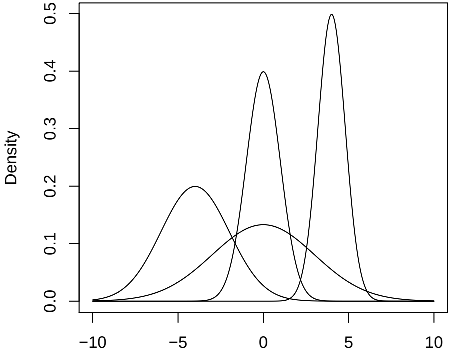
La funzione $f(x)$ prende il nome di densità e distribuisce la massa di probabilità unitaria su $\Omega$
La probabilità di osservare valori compresi in un dato intervallo è l'area corrispondente sotto questa curva
Qualcosa di simile accade anche per le variabili statistiche continue: la distribuzione viene calcolata per classi e rappresentata attraverso l'istogramma (che sotto alcune condizioni è una stima della densità )

Tra le variabili aleatorie continue ha un ruolo importante la famiglia Normale

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

$\Omega_X = \mathbb{R}$
$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]$
Caratteristiche del modello sono:

- l'esistenza di un valore *centrale* $\mu$, valori in un suo intorno sono i più probabili
- la simmetria rispetto a tale valore centrale
- una probabilità che decresce (secondo $\sigma$) man mano che ci si allontana da $\mu$




Ferme restando queste caratteristiche, il modello contiene infinite possibilità ...

La variabile aleatoria normale con media nulla è spesso chiamata a descrivere in probabilità l'effetto del caso o *errore casuale*
Numerose sono state le conferme sperimentali a questa ipotesi e il teorema centrale del limite (che riportiamo nella versione di Liapunov) le ha dato sostegno teorico
Siano $X_1, X_2 \dots$ variabili aleatorie indipendenti con media $E[X_n] = \mu_n$, varianza $V[X_n] = \sigma_n^2$ e momento terzo assoluto rispetto alla media $\alpha_n^3 = E|X_n - \mu_n|^3$ finiti.
Sia inoltre $S_n = X_1 + X_2 + \cdots + X_n$ e, per $n \to \infty$, sia $\frac{(\alpha_1^3 + \cdots + \alpha_n^3)^{1/3}}{(\sigma_1^2 + \cdots + \sigma_n^2)^{1/2}} \to 0$
Allora la variabile aleatoria $U_n = \frac{(S_n - E[S_n])}{\sqrt{V[S_n]}}$ tende in distribuzione, per $n \to \infty$, alla Normale standardizzata (media 0 e varianza unitaria)

Ora il caso viene spesso considerato intuitivamente come il sommarsi degli effetti di numerosi fattori, singolarmente poco rilevanti
In altre parole come la somma $S_n$ di un gran numero di variabili aleatorie nessuna delle quali ha, singolarmente, un peso rilevante.
Diventa allora semplice ammettere che siano rispettate le condizioni del teorema centrale e che l'effetto del caso si manifesti come una legge normale.
Meno semplice è comprendere per quali ulteriori fenomeni questo modello possa essere ipotizzato e quindi il senso reale dell'affermazione
*assumiamo un modello normale*
Poincarè osservava criticamente che tutti giuravano sulla distribuzione normale perché gli sperimentatori assumevano che essa era dimostrata matematicamente, e i matematici la ritenevano acquisita sperimentalmente . . .

# Probabilità e inferenza

La logica inferenziale si basa sull'idea (un pò strana) di osservare un sottoinsieme della popolazione di interesse ed arrivare poi a conclusioni che possano essere generalizzate all'intera popolazione.

Ovviamente perchè questo abbia senso devono essere rispettate condizioni precise circa il modo in cui tale sottoinsieme viene scelto.

Altrettanto ovviamente il risultato finale riferito all'intera popolazione sarà affetto da un errore che deve essere quantificato.

La definizione della popolazione è un elemento essenziale dell'inferenza e richiede rigore. Pensiamo a quanta attenzione è dedicata alla definizione dei criteri di inclusione ed esclusione nei protocolli sperimentali.

Circa la selezione del campione, il principio fondamentale è che tutte le unità della popolazione abbiano la stessa probabilità di entrare a far parte del nostro campione (campionamento casuale semplice).

Il nostro obiettivo è ottenere un campione che sia *rappresentativo* della popolazione.
Ma cosa vuol dire rappresentativo? E perché ci affidiamo al caso?

La parola *rappresentativo*, spesso abusata, traduce l'idea che il nostro campione sia eterogeneo così come la popolazione da cui proviene. Se ad esempio nella nostra popolazione la frequenza di donne è $0.6$, vorremmo ritrovare la stessa proporzione nel nostro campione. Allo stesso modo vorremmo un distribuzione per età simile alla popolazione e il discorso si estende a tutte le carateristiche di interesse nel nostro studio (e anche a quelle che non lo sono).
Resta da capire perchè il caso dovrebbe aiutarci e a quali condizioni.

Consideriamo l'i-esimo individuo estratto casualmente dalla nostra popolazione e indichiamo con $X_i$ la presenza di una data mutazione con $X_i = 0$ se assente e $X_i = 1$ se presente. Sia $F$ la frequenza relativa con cui la mutazione compare nella popolazione.
Indichiamo poi con $\pi$ la porbabilità che l'individuo estratto abbia la mutazione. Avremo

$$P(X_i = 1) = \pi$$
$$P(X_i = 0) = 1 - \pi$$

Il campionamento casuale crea uno stretto legame tra $\pi$ ed $F$, pur restando queste quantità concettualmente diverse.

Per chiarire la natura di questo legame ricordiamo la definizione classica di probabilità come rapporto tra casi favorevoli e casi possibili, a condizione che tutti i casi osservabili siano ugualmente probabili. Ora la condizione di equiprobabilità è assicurata proprio dal campionamento casuale. Ne segue che la probabilità che l'individuo estratto presenti la mutazione coinciderà con il rapporto tra il numero di individui nella popolazione che presentano la mutazione (casi favorevoli) e la numerosità totale della popolazione stessa (casi possibili). Formalmente avremo

$$\pi = F$$

.

Questo risultato può essere ovviamente esteso a tutti gli individui che estrarremo e che costituiranno il nostro campione.

Il legame tra campione e popolazione è tuttavia solo una garanzia *in probabilità* .
Quello che vorremmo è che la frequenza degli individui del nostro campione che presentano la mutazione (che indichiamo con $f$) fosse uguale a quella nella popolazione.
(In una moneta bilanciata la probabilità che esca testa è $1/2$ ma se la lanciamo 6 volte è nessuno scommetterebbe sull'evento (escono 3 teste e 3 croci)).
Ci viene in aiuto la legge dei grandi numeri che ci garantisce che la frequenza relativa tenderà alla probbailità $\pi$ al tendere ad infinito della numerisità del nostro campione.

$$p \rightarrow_{n \rightarrow +\infty} \pi = F$$

L'uguaglianza si attiene soltanto al limite quindi non riusciremo mai ad avere esattamente la stessa frequenza della popolazione ma potremo avere valori molto simili se il campionamento casuale sarà accompagnato da un'alta numerosità .

# Randomizziamo

Immaginiamo ora di voler confrontare l'effetto di due trattamenti alternativi la cui efficacia potrebbe essere ridotta dalla presenza della mutazione.
Ipotizziamo di aver già estratto un campione casuale di volontari: adesso il punto critico è come decidiamo quale dei due trattamenti assegnare a ciascuno.
La logica della randomizzazione è quella di affidare ancora un volta questa decisione al caso. Più esattamente, se indichiamo con $T_i$ il trattamento che assegneremo all'individuo selezionato, sarà

$$P(T_i = 1) = 1/2$$
$$P(T_i = 2) = 1/2$$

La scelta casuale del trattamento esclude che la decisione possa essere legata alle caratteristiche del paziente.
In particolare garantiamo che il trattamento assegnato e la presenza della mutazione restino due eventi indipendenti.
Formalmente

$$P(X_i = 1|T_i = 1) = P(X_i = 1) \; P(X_i = 1|T_i = 2) = P(X_i = 1)$$

Ancora una volta si tratta di una garanzia in probabilità . Servirà una elevata numerosità in entrambi i sottocampioni per garantire che la frequenza degli individui che presentano la mutazione sia simile tra coloro a cui è stato assegnato il trattemento 1 coloro a cui è stato assegnato il trattamento 2.

Vale la pena sottolineare che quanto descritto per la presenza di una mutazione può essere applicato a qualsiasi caratteristica che sia di interesse o meno.
Campionamento casuale, randomizzazione ed elevata numerità ci garantiscono di ottenere due campioni che siano *rappresentativi della popolazione* e *comparabili tra loro*.
Non possiamo eliminare l'eterogenità dei soggetti ma possiamo fare in modo che sia simile nei due sottocampioni in modo che l'unico elemento che li differenzia si ail trattamento assegnato.
In questo modo, se nel gruppo che ha seguito il trattamento 1 si osserverà una maggiore efficacia, questa sarà attrubuibile all'effetto del trattamento.
L'accento adesso è tutto sulla comparabiltà ma possiamo rinunciare alla rappresentatività ?