

Probabilità ¹

Un *esperimento (o prova) aleatorio (casuale)* è un esperimento che a priori può avere diversi esiti possibili e il cui risultato effettivo non è prevedibile con certezza.

Esempi:

1. Lancio di due dadi.
2. Scelta a caso di 1 carta da un mazzo di 52 carte.
3. Durata di vita di una sequoia.

In ciascuno di questi esperimenti l'esito non è certo, a priori. Sono però noti i possibili esiti: nel caso del lancio di due dadi, i possibili esiti sono coppie di numeri da 1 a 6, mentre la carta scelta può essere una qualunque delle 52 carte del mazzo, mentre nel rimanente caso i possibili esiti saranno infiniti (il tempo di vita di una sequoia può essere un reale positivo qualunque).

Definizione: Chiamiamo **esito od evento elementare** ogni possibile risultato di un esperimento aleatorio e **spazio campione (o campionario)** l'insieme Ω di tutti i possibili esiti. Nell'esempio 1, $\Omega = \{(p, q) \mid p, q = 1, \dots, 6\}$, nel secondo $\Omega = \{52 \text{ carte}\}$, nel terzo $\Omega = [0, +\infty)$. Lo spazio campione si dice **finito** se ha un numero finito (come in 1,2), **continuo** in caso contrario.

In realtà ci sono eventi di cui misurare la probabilità che non sono eventi elementari: per esempio, potremmo voler sapere qual è la probabilità che in un lancio di 2 dadi i due numeri che escono diano somma pari. L'uscita di una somma pari non è un evento elementare, in quanto ci sono esiti diversi che danno come somma un pari, tutti quelli in cui p e q sono entrambi pari o entrambi dispari. Possiamo allora rappresentare questo evento come l'insieme di tutti gli esiti che concorrono a formare l'evento stesso. Quindi un **evento** A altro non è che un **sottoinsieme di Ω** e diciamo che *l'evento A si è verificato in una prova se l'esito x della prova appartiene ad A* . $A = \Omega$ si chiama **evento certo** e $A = \emptyset$ si chiama **evento impossibile**. L'insieme di tutti gli eventi $A \subseteq \Omega$ è quindi l'insieme dei sottoinsiemi di Ω , che si chiama *insieme delle parti di Ω* ed è indicato con $\mathbb{P}(\Omega)$. Il linguaggio degli insiemi ha in questo caso un'interpretazione particolare, in quanto il complementare $\bar{A} = \Omega \setminus A = \{x \in \Omega, x \notin A\}$ di A rappresenta l'evento opposto ad A , il verificarsi dell'unione $A \cup B$ rappresenta il verificarsi di A o di B , il verificarsi dell'intersezione $A \cap B$ rappresenta il verificarsi contemporaneo dei due eventi, mentre con $A \setminus B$ si rappresenta il verificarsi di A , ma non di B . Inoltre se i due eventi sono disgiunti, cioè $A \cap B = \emptyset$, significa che non possono verificarsi simultaneamente e quindi si dicono **incompatibili (o mutuamente esclusivi)**. Se poi $A \subseteq B$, il verificarsi di A implica il verificarsi di B .

La **probabilità** di un evento è un numero compreso tra 0 ed 1 che esprime il grado di fiducia nel verificarsi di tale evento. Il *Calcolo delle Probabilità* è la disciplina matematica che tratta il calcolo della probabilità di eventi "complessi" conoscendo le probabilità di eventi considerati "semplici" (il problema di come vengono assegnate quest'ultime lo affronteremo dopo). Daremo invece assiomi su come si opera con la probabilità.

Definizione. Sia Ω uno spazio campione. Si chiama (*misura di*) **probabilità** su Ω una funzione $P : \mathbb{P}(\Omega) \rightarrow [0, 1]$ tale che:

- $P(\Omega) = 1$ e $P(\emptyset) = 0$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

La coppia (Ω, P) viene detta **spazio di probabilità**.

¹NOTA. Queste dispense sono in rete a disposizione degli studenti di Scienze Naturali, corso di Statistica, I parte, Prof. Mantovani. Non è autorizzata dall'autore la vendita nè in fotocopia nè in forma elettronica.

Da questi assiomi si deducono altre proprietà:

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cap B) \leq P(A) \leq P(A \cup B)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$.

Esempio. Supponiamo che un gene possa presentare due possibili difetti, difetto a e difetto b . Supponiamo che la probabilità che un gene si presenti con a è 0.2, mentre che si presenti con b è 0.3, la probabilità che si presenti con tutti e due è 0.05. Vogliamo calcolare la probabilità che tale gene non presenti alcun difetto, quindi dobbiamo calcolare, chiamato A l'evento "si presenta a " e B l'evento "si presenta b ", la probabilità dell'evento complementare all'unione, cioè $P(\overline{A \cup B}) = 1 - P(A \cup B)$. Quindi basta calcolare $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.3 - 0.05 = 0.45$, da cui $P(\overline{A \cup B}) = 1 - 0.45 = 0.55$.

Adesso vediamo come si possono assegnare le probabilità agli eventi "semplici". Se Ω è uno spazio campione finito, possiamo avere esiti tutti ugualmente probabili (come nel caso dei 2 dadi, se non sono truccati!). In tal caso se gli esiti x_1, x_2, \dots, x_n possibili sono n , ognuno di essi avrà probabilità $\frac{1}{n}$, visto che gli esiti sono eventi tutti mutuamente esclusivi e $1 = P(\Omega) = P(\cup\{x_k\}) = \sum_{k=1}^n P(\{x_k\}) = nP(\{x_k\})$. In un caso come questo, la probabilità di un evento A qualunque dipende solo dal numero $|A|$ dei suoi elementi: $P(A) = \frac{|A|}{n}$, cioè in tal caso la probabilità si ottiene come il rapporto tra il numero dei casi favorevoli (esiti che appartengono ad A) ed il numero dei casi possibili (tutti gli esiti), definizione nota con il nome di *probabilità classica*, che possiamo dare solo quando lo spazio è finito e gli esiti sono equiprobabili (fatto che si prende per dato).

Esempio. Calcoliamo la probabilità che nel lancio di 2 dadi perfetti si ottenga come somma 6 e quella che si ottenga 3. Nel primo caso i casi favorevoli sono gli esiti (p, q) in cui la somma dia 6, cioè $(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)$, che sono 5, nel secondo i favorevoli sono $(1, 2), (2, 1)$, quindi 2. Essendo i casi possibili 36, otteniamo che $P(\text{esce } 6) = \frac{5}{36}$, mentre $P(\text{esce } 3) = \frac{2}{36} = \frac{1}{18}$. (EX. Calcolare la probabilità che esca un numero minore di 4). Notiamo che se prendiamo come esiti possibili non la coppia di numeri dei due dadi, ma la loro somma, otteniamo uno spazio campione finito (gli esiti possibili sono gli 11 numeri dal 2 al 12) con esiti non tutti equiprobabili. Ci sono altri modi per assegnare la probabilità agli eventi semplici, tra cui quello detto *criterio frequentista* che si basa sulla frequenza relativa, nel senso che si usa l'esperienza del passato (frequenza relativa dedotta da statistiche) per definire la probabilità di un evento futuro. In quest'ottica la probabilità di un evento può essere (idealmente) definito con:

$$P(A) = \lim_{n \rightarrow +\infty} \frac{k}{n},$$

dove k rappresenta numero di successi (casi in cui l'evento si è verificato) ed n rappresenta il numero di prove eseguite. Operativamente questo limite (impossibile da eseguire in pratica, visto che si dovrebbe ripetere l'esperimento infinite volte) viene inteso come il rapporto $\frac{k}{n}$, con n molto grande, che dà quindi una stima statistica della probabilità dell'evento (questo è il metodo che usano, per esempio, le assicurazioni prima di stipulare una polizza). Ci sono altri modi per definire la probabilità, ma, (tranne in casi particolari, che non tratteremo), le varie concezioni di probabilità godono tutte degli assiomi enunciati.

Per calcolare la probabilità secondo lo schema della probabilità classica, dobbiamo saper contare i numeri di casi favorevoli e quelli possibili. Quindi abbiamo bisogno di metodi per farlo e questi ci sono forniti dal Calcolo combinatorio.

Cominciamo con illustrare lo *schema delle scelte successive*: supponiamo di dover comprare un'auto ed in una concessionaria vendono 5 modelli di auto, ognuno dei quali è disponibile in 3 motorizzazioni ed in 6 colori: tra quante auto possiamo scegliere? Ovviamente dobbiamo operare delle scelte successive: prima scegliamo il modello, (5 possibilità) poi la motorizzazione (3) ed infine il colore (6): ad ogni passo il numero di opzioni viene moltiplicato per il numero di opzioni del passo precedente, quindi in totale abbiamo $5 \times 3 \times 6 = 90$ auto. Allora il **principio del prodotto delle possibilità** dice che, se ogni elemento di un insieme è individuabile mediante k scelte successive, la prima tra r_1 possibilità, la seconda tra r_2, \dots , l'ultima tra r_k , l'insieme ha $r_1 \times r_2 \times \dots \times r_k$ elementi.

Definizione: Una **permutazione** di n oggetti è un allineamento degli n oggetti. Due permutazioni degli stessi oggetti differiscono solo per l'ordine. Il numero totale di tali permutazioni è dato da $P_n = n! = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1$.

Un anagramma di una parola composta da lettere distinte rappresenta un esempio tipico di permutazione: Milano ha 6 lettere distinte e quindi $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 = 720$ anagrammi (non tutti significativi).

Pensiamo ora di avere sempre n oggetti (distinti) e volerne allineare k ($k \leq n$). Avremo quindi a che fare con una **disposizione di n oggetti in k posti**, dove due tali disposizioni differiscono o per almeno un elemento diverso o per l'ordine. Il numero totale di tali disposizioni è dato da

$$D_{n,k} = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - (k - 1)) = \frac{n!}{(n - k)!}.$$

Quanti numeri di 3 cifre diverse si formano con le cifre 1, 2, 3, 4, 5? È una disposizione (in quanto conta l'ordine) $D_{5,3} = 5 \cdot 4 \cdot 3 = 60$.

Quanti numeri di 3 cifre diverse si formano con le cifre 0, 2, 3, 4, 5?

Una **disposizione con ripetizione di n oggetti in k posti** è un allineamento di k oggetti tra n oggetti ripetibili. Due tali disposizioni differiscono o per almeno un elemento diverso o per l'ordine o per il numero di ripetizioni ($33322 \neq 33222$). il numero totale è dato da $D_{n,k}^* = n^k$.

Quanti numeri di 3 cifre si formano con le cifre 1, 2, 3, 4, 5? $D_{5,3}^* = 5^3 = 125$.

Quanti numeri di 3 cifre si formano con le cifre 0, 2, 3, 4, 5?

Qual è la probabilità che tra di voi ci siano due persone che compiono gli anni nello stesso giorno? Detto k il numero di studenti presenti, consideriamo l'evento A : "Su k , almeno due hanno lo stesso giorno di nascita" ed il suo complementare \bar{A} : "Su k , tutti hanno giorni di nascita diversi". Calcoliamo $P(\bar{A})$ (più facile). Casi possibili: 365^k . Casi favorevoli: $365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - k + 1)$.

Quindi

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - k + 1)}{365^k}$$

che vale, per diversi k :

k	10	20	23	30	40	50	60	70	366
$P(A)$	0.117	0.411	0.507	0.706	0.891	0.970	0.994	0.999	1

Se dobbiamo scegliere k oggetti su n dati senza che interessi l'ordine, parliamo di **combinazione di n oggetti di classe k** ($0 \leq k \leq n$), che coincide con un qualunque sottoinsieme di k elementi dell'insieme dato di n oggetti.

Il numero totale di tali combinazioni è dato da

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{(n-k)!k!} \stackrel{def}{=} \binom{n}{k} \text{ (coefficiente binomiale)}$$

Ex. Un allenatore di basket ha una panchina con 9 elementi. Quante squadre può mandare in campo? $C_{9,5} = \binom{9}{5} = \frac{9!}{4!5!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5!}{4!5!} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2} = 3 \cdot 7 \cdot 6 = 126$.

Ex. Si gioca a poker con un mazzo di 32 carte, dal 7 all'asso per ogni seme. Quante mani, cioè un insieme di 5 carte prese dalle 32 date, contengono un poker? Quante un full (tris+coppia)? Qual è la probabilità di avere un poker servito? Qual è la probabilità di avere un full servito?

Probabilità condizionata

Se la probabilità misura il grado di aspettative che abbiamo nel realizzarsi di un evento, è chiaro che la probabilità dello stesso evento può cambiare, se cambiano le informazioni in nostro possesso relativamente a questo evento: se prima di un incontro di tennis tra Agassi e Hewitt si viene a sapere che Agassi ha dolente il braccio con cui serve, i book-makers cambiano la quota di vincita. Il concetto di *probabilità condizionata* dà la probabilità di un evento, valutata sapendo che si è verificato un altro evento:

Definizione. Sia B un evento con $P(B) \neq 0$. Si chiama **probabilità dell'evento A condizionata a B** , il numero

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (\text{da cui } P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \text{ se anche } P(A) \neq 0) \quad (*)$$

ES. Qual è la probabilità che in 10 lanci di moneta, esca tutte le volte croce? $(\frac{1}{2})^{10}$

Qual è la probabilità che in 10 lanci di moneta, esca tutte le volte croce, sapendo che nei primi 9 è uscita croce? $(\frac{1}{2})$

Definizione. Due eventi A e B si dicono **indipendenti** se $P(A) = P(A|B)$ (e quindi $P(B) = P(B|A)$). Come conseguenza se A e B sono indipendenti, da $P(A \cap B) = P(A|B)P(B)$ si ha che $P(A \cap B) = P(A)P(B)$ (*regola del prodotto*). Notiamo che due eventi sono indipendenti quando la conoscenza del verificarsi dell'uno non modifica la probabilità del verificarsi dell'altro: per esempio, nel lancio di due dadi, l'esito relativo al secondo dado è indipendente dal risultato del primo, come i numeri estratti al lotto in un'estrazione sono indipendenti da quelli estratti precedentemente, mentre nell'estrazione di palline da un'urna senza reimbussolamento l'esito della seconda estrazione dipende dall'esito della prima (basti pensare ad un'urna contenente una pallina rossa ed una nera: la probabilità di estrarre alla prima estrazione una rossa è uguale a quella di estrarre una nera, ma alla seconda la probabilità di estrarne una rossa dipende fortemente dal fatto che alla prima si sia estratta una rossa od una nera!). Fare attenzione a non confondere eventi indipendenti con eventi incompatibili: sono nozioni completamente diverse.

$P(-|B)$ ha le proprietà di una probabilità, quindi vale per esempio che $P(\bar{A}|B) = 1 - P(A|B)$.

Se cerchiamo la probabilità di A , può essere utile cercare la probabilità di avere A dato B e la probabilità di avere A dato \bar{B} , in quanto vale che:

$$\boxed{P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

Infatti talvolta è più facile conoscere $P(A|B)$ che $P(A)$, visto che si ha un'informazione in più, e poi trovare $P(A)$ sommando i risultati ottenuti separatamente. È il **Teorema delle probabilità totali**, nel caso particolare in cui si hanno un evento B ed il suo complementare.

Esempio: Supponiamo di avere due urne, la a contenente 3 palline rosse ed 1 nera, la b contenente 2 rosse e 3 nere. Facendo un' estrazione, qual è la probabilità di estrarre una pallina nera? Ovviamente il conteggio sarà diverso se si è scelta l'urna a o l'urna b , quindi dobbiamo tenere conto anche della scelta casuale dell'urna. Siano A l'evento "scelta a ", B l'evento "scelta b ", N l'evento "estratta una nera", R l'evento "estratta una rossa". Possiamo calcolare $P(N)$ tenendo conto delle formula di prima, visto che A e B sono complementari ed avremo che $P(N) = P(N|A)P(A) + P(N|B)P(B)$, con $P(A) = P(B) = \frac{1}{2}$, $P(N|A) = \frac{1}{3}$, $P(N|B) = \frac{3}{5}$. Quindi si ha $P(N) = \frac{1}{3} \cdot \frac{1}{2} + \frac{3}{5} \cdot \frac{1}{2} = \frac{7}{15} = 0.4\bar{6}$.

Dalla formula (*) si ottiene immediatamente una relazione che lega la probabilità di avere A dato B a quella di avere B dato A :

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

e sostituendo in questa relazione $P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$ otteniamo

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

formula nota come **Formula di Bayes**.

Esempio: (TEST CLINICI). Un test clinico serve per vedere, attraverso una data analisi di laboratorio, se un certo individuo ha o no una certa malattia: il test sarà *positivo*, se la malattia è stata riscontrata, *negativo*, nel caso contrario. Ogni test è inevitabilmente soggetto a possibilità di errori: ci saranno *falsi positivi* (individui sani, ma risultati positivi al test) e *falsi negativi* (in realtà malati). Quindi bisognerebbe sapere quanto un test è affidabile per praticarlo in modo sensato. Allora si può sottoporre al test un certo numero di persone (che sappiamo già se sane o malate) e verificare l'accuratezza dei risultati. Indichiamo con M l'evento: "individuo malato", con S l'evento: "individuo sano", con Pos l'evento: "test positivo", con Neg l'evento: "test negativo". In base alle verifiche effettuate possiamo calcolare le probabilità $P(Pos|M)$, detta **sensibilità** del test e $P(Neg|S)$, detta **specificità** del test. Chiaramente un test è buono se ha queste 2 probabilità molto vicine ad 1. Una volta applicato su larga scala, quale sarà la probabilità che un individuo che risulta positivo sia effettivamente malato? Dovremo calcolare quindi $P(M|Pos)$, detto anche *valore predittivo di un esito positivo*, che con la formula di Bayes risulta essere:

$$P(M|Pos) = \frac{P(Pos|M)P(M)}{P(Pos|M)P(M) + P(Pos|S)P(S)}$$

Se conosciamo quindi anche la probabilità $P(M)$ che un individuo contragga la malattia (incidenza o prevalenza della malattia), ottenibile ovviamente con indagini statistiche, possiamo rispondere al quesito, visto che $P(S) = 1 - P(M)$ e $P(Pos|S) = 1 - P(Neg|S)$. In modo del tutto analogo possiamo calcolare il *valore predittivo di un esito negativo* $P(S|Neg)$. Notiamo come i valori predittivi di un test dipendano della incidenza della malattia sulla popolazione in esame.

Supponiamo quindi che un dato test diagnostico per la malattia M abbia sensibilità pari al 98% e specificità pari al 99%. Sapendo che la malattia M ha un'incidenza del 2%, la probabilità che un individuo con esito positivo sia effettivamente malato è:

$$P(M|Pos) = \frac{0,98 \cdot 0,02}{0,98 \cdot 0,02 + (1 - 0,99)0,98} = \frac{\frac{49}{2500}}{\frac{49}{2500} + \frac{1}{100} \frac{98}{100}} = 0.\bar{6} \sim 67\%.$$

Esercizio: Un test con sensibilità e specificità pari al 98% viene applicato ad una popolazione di 10000 soggetti. Calcolare il numero dei positivi, dei negativi, dei falsi positivi e dei falsi negativi previsti, supponendo che la malattia abbia una prevalenza del (a) 1%, (b) del 10%.

(a) $P(Pos) = P(Pos|S)P(S) + P(Pos|M)P(M) = (1 - P(Neg|S))(1 - P(M)) + P(Pos|M)P(M) = 0,02 \cdot 0,99 + 0,98 \cdot 0,01 = (198 + 98)10^{-4} = 296 \cdot 10^{-4}$, da cui il numero dei positivi previsti su 10000 è di 296 ed il numero dei negativi previsti sarà di 9704.

Il numero dei falsi positivi previsti su 10000 test è di: $P(Pos \cap S) \cdot 10000 = P(Pos|S) \cdot P(S) \cdot 10000 = (1 - P(Neg|S)) \cdot (1 - P(M)) \cdot 10000 = 0,02 \cdot 0,99 \cdot 10000 = 2 \cdot 99 = 198$.

Il numero dei falsi negativi previsti è di $P(M \cap Neg) \cdot 10000 = P(Neg|M) \cdot P(M) \cdot 10000 = (1 - P(Pos|M)) \cdot P(M) \cdot 10000 = 0,02 \cdot 0,01 \cdot 10000 = 2 \cdot 1 = 2$.

Variabili aleatorie

Definizione: Se Ω è uno spazio campione, si chiama **variabile aleatoria** (v.a.) una qualunque funzione $X : \Omega \rightarrow \mathbb{R}$. Se Ω è discreto, la variabile si dirà discreta, continua in caso contrario.

Una variabile aleatoria svolge il compito di trasferire i possibili esiti di un esperimento aleatorio (che possono essere di varia natura, da un colore ad una sequenza di lettere, come nel caso del DNA), al campo dei numeri reali, assegnando ad ogni esito ϖ un numero reale $X(\varpi) = x \in \mathbb{R}$. Quello che interessa di una v.a. non sono solo i valori reali che assume, ma anche la probabilità che essa assuma dati valori. In generale indicheremo con $(X = a)$, $(X \leq b)$, $(X > a)$, $(X \in I)$, rispettivamente gli eventi: “ X assume il valore a ”, “ X assume valori minori od uguali a b ”, “ X assume valori maggiori di b ”, “ X assume un valore appartenente ad I ”, (in altri termini, X “cade” dentro I) e così via. Se abbiamo a disposizione uno spazio di probabilità, cioè conosciamo la probabilità di ogni evento $A \in \mathbb{P}(\Omega)$, possiamo conoscere la **legge (o distribuzione)** della v.a. X , legge che associa ad ogni intervallo $I \subseteq \mathbb{R}$ la probabilità che X assuma un valore reale appartenente ad I . Avere la legge con cui si distribuisce la probabilità significa allora conoscere

$$P(X \in I) = P(\varpi \in \Omega : X(\varpi) = x \in I).$$

Nel caso di v.a. finite sappiamo che i valori assunti sono finiti e, se indicati in ordine crescente, saranno quindi del tipo x_1, x_2, \dots, x_n . Allora

$$P(X = x) = \begin{cases} 0, & \text{se } x \neq x_k, \text{ per ogni } k = 1, \dots, n \\ p_k, & \text{se } x = x_k, \text{ per qualche } k = 1, \dots, n \end{cases}$$

Chiamiamo **funzione di probabilità o densità (discreta)** della v.a. X la funzione p_x che associa ad ogni valore assunto la probabilità che X lo assuma:

$$p_x(x_k) = p_k$$

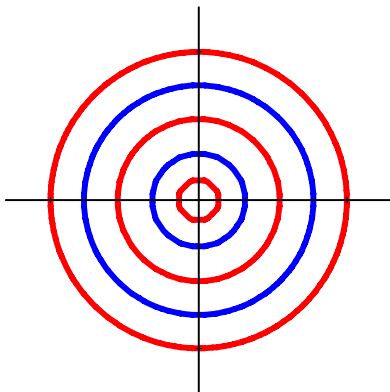
Da notare che $\sum_{k=1}^n p_k = 1$. Attraverso tale funzione possiamo quindi avere la distribuzione di X , in quanto $P(X \in I) = \sum_{x_k \in I} p_k$. In conclusione una v.a. aleatoria X finita è data quando sono dati i valori assunti, ognuno con la probabilità di essere assunto da X , quindi attraverso una tabella del tipo:

X	x_1	x_2	\dots	x_n
p_x	p_1	p_2	\dots	p_n

ad esempio:

X	-3	-1	0	1/3	1	5/4	2
p_x	0.02	0.14	0.33	0.21	0.05	0.15	0.1

Invece nel caso continuo, abbiamo che $P(X = a) = 0$, comunque si scelga il valore reale a . Si pensi, per esempio all'esperimento seguente: si disegni su un foglio un cerchio e si tratteggino delle corone circolari, poi si lanci una freccetta:



La probabilità di centrare una corona circolare sarà proporzionale ovviamente all'area della stessa, quindi la probabilità di prendere il centro è nulla! (mentre non è nulla, anche se bassa, la probabilità di centrare il cerchio più piccolo). Questo vale ovviamente anche per gli altri punti del cerchio: si può avere una probabilità diversa da 0 solo se cerchiamo la probabilità che la freccetta cada in una certa regione di area non nulla e la probabilità sarà tanto più grande quanto più sarà grande la regione che prendiamo. Se indichiamo con X la v.a. che dà la distanza dal centro del punto di arrivo della freccetta, la probabilità che la freccetta cada in una data corona circolare definita dai raggi r_1, r_2 sarà data da $P(r_1 < X < r_2)$. Da notare che, poiché $(r_1, r_2] = (r_1, r_2) \cup \{r_2\}$, per la proprietà della probabilità dell'unione, si ha che $P(r_1 < X \leq r_2) = P(r_1 < X < r_2) + P(X = r_2) = P(r_1 < X < r_2)$ ed analogamente per $P(r_1 \leq X \leq r_2)$ e $P(r_1 \leq X < r_2)$. Allora nel caso di una v.a. continua *serve conoscere la probabilità che X cada in un certo intervallo I , senza preoccuparsi se gli estremi appartengano o meno ad I . Notiamo che per conoscere ciò, ci basta conoscere la probabilità che X cada in semirette sinistre, $P(X \leq b)$, in quanto $P(X > b) = 1 - P(X \leq b)$ e $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$.*

La legge di una v.a. X continua è determinata dall'assegnazione di una **funzione di densità (di probabilità)** $f : \mathbb{R} \rightarrow \mathbb{R}$, con le proprietà

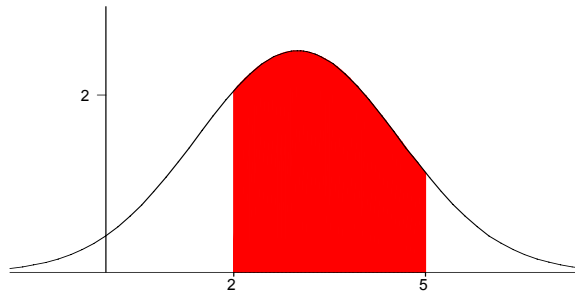
$$f(x) \geq 0 \text{ e } \int_{-\infty}^{+\infty} f(x) dx = 1$$

attraverso il calcolo di un integrale, in quanto:

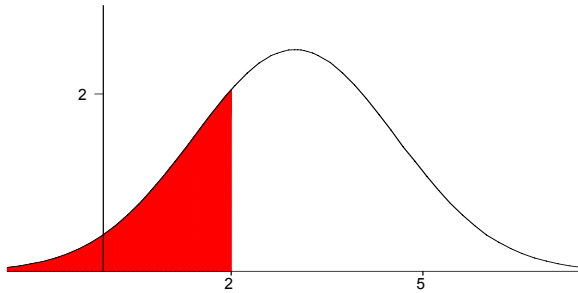
$$P(a < X \leq b) = \int_a^b f(x) dx,$$

$$P(X \leq b) = \int_{-\infty}^b f(x) dx$$

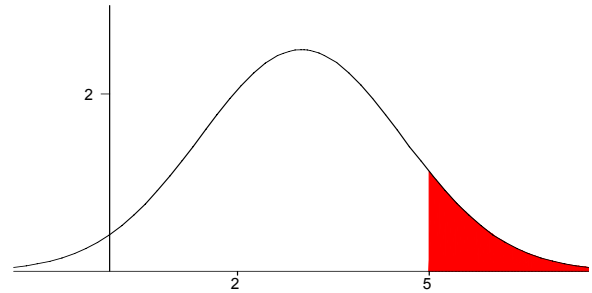
$$\text{e } P(a < X) = \int_a^{+\infty} f(x) dx$$



$$P(2 < X \leq 5) = \int_2^5 f(x)dx$$



$$P(X \leq 2) = \int_{-\infty}^2 f(x)dx$$



$$P(5 < X) = \int_5^{+\infty} f(x)dx$$

Ad esempio, $f(x) = \frac{1}{\pi(1+x^2)}$ (densità di Cauchy), $f(x) = \begin{cases} 0, & \text{per } x < 0 \\ \frac{3}{(x+3)^2}, & \text{per } x \geq 0 \end{cases}$. Verifichiamo

che la prima sia una funzione di densità, calcolando $\int_{-\infty}^{+\infty} f(x)dx$:

$$\int_{-\infty}^{+\infty} \frac{1}{\pi(1+x^2)} dx = 2 \int_0^{+\infty} \frac{1}{\pi(1+x^2)} dx = 2 \lim_{b \rightarrow +\infty} \int_0^b \frac{1}{\pi(1+x^2)} dx = 2 \lim_{b \rightarrow +\infty} \left[\frac{1}{\pi} \arctan b \right]_0^b = 2 \frac{1}{\pi} \frac{\pi}{2} - 0 = 1.$$

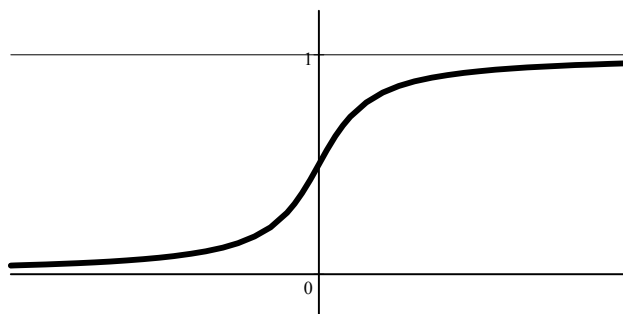
Notiamo che f definisce attraverso un integrale improprio una funzione che si chiama la **funzione di distribuzione (o di ripartizione)** $F: \mathbb{R} \rightarrow \mathbb{R}$ di X , data da

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

funzione crescente (debolmente) che ha le proprietà:

- $F(x) \geq 0$
- $\lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} P(X \leq x) = P(X < +\infty) = 1$
- $\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P(X \leq x) = P(X < -\infty) = 0$

Negli esempi di prima, $F(x) = \int_{-\infty}^x \frac{1}{\pi(1+t^2)} dt = \frac{\arctan x}{\pi} + \frac{1}{2}$, distribuzione di Cauchy, con grafico dato da:



Per calcolare delle probabilità, la funzione di distribuzione è molto conveniente, in quanto:

$$\boxed{P(X \leq b) = F(b)}, \quad \boxed{P(a \leq X) = 1 - F(a)}, \quad \boxed{P(a < X \leq b) = F(b) - F(a)}.$$

Nel caso di f continua, il teorema fondamentale del calcolo integrale ci dice che $F'(x) = f(x)$. Notiamo che in questo caso, avere F o f dal punto di vista teorico è assolutamente equivalente. Il problema pratico è dato dal fatto che ci sono funzioni continue che ammettono primitive, ma queste non sono esprimibili in modo elementare, cioè come composizione, somma, prodotto, quoziente di funzioni elementari. Allora calcolare un integrale (finito o improprio) di tali funzioni non è possibile in modo elementare: si ricorre quindi a valori approssimati dell'integrale, che vanno bene per il calcolo di probabilità (che sono numeri), ma non per la determinazione della funzione di distribuzione.

Esempi di variabili aleatorie

CASO DISCRETO: La distribuzione binomiale (o di Bernoulli) e la distribuzione ipergeometrica.

Si dice *esperimento o prova di Bernoulli* un esperimento aleatorio con due soli esiti possibili che chiamiamo convenzionalmente "successo" ed "insuccesso", il primo con probabilità p ed il secondo quindi di probabilità $1 - p$. Tale p si dice *parametro* della prova.

Esempi di prove bernoulliane sono: il lancio di una moneta ($p = 0.5$), l'estrazione di una pallina bianca da un'urna contenente 3 palline bianche, 2 rosse, 4 nere ($p = \frac{1}{3}$), ect.

EX: fare 3 esempi di prove di Bernoulli con parametro $p \neq 0.5$.

Un **processo binomiale (o di Bernoulli)** è una sequenza di prove di Bernoulli tra di loro indipendenti, tutte di ugual parametro p , (per esempio, 10 lanci di monete, 15 giocate a pari e dispari, ect).

Possiamo descrivere dati concernenti una singola prova di Bernoulli attraverso una variabile aleatoria X_1 che assume valore 1 in caso di successo, e valore 0 in caso contrario, e che si chiama *bernoulliana (o binomiale) di parametro p* :

X_1	1	0
p_{X_1}	p	$1 - p$

Se abbiamo una sequenza di n prove di parametro p , l'esito di ogni prova sarà descritto da una v.a. di questo tipo: indichiamo con X_k la v.a. che descrive l'esito della prova k . La v.a. X che conta il numero dei successi ottenuti in n prove si chiama **distribuzione binomiale di parametri n e p** ed è data dalla somma delle v.a. descrittive le singole n prove, cioè $X = X_1 + X_2 + \dots + X_n$, e può assumere tutti e soli i valori $0, 1, 2, \dots, n$.

Essendo le prove indipendenti, la probabilità di avere k successi in una data sequenza di n prove coincide con il prodotto $p^k(1-p)^{n-k}$:

1	2	3	4	5	...	$(n-1)$	n
S	I	I	S	S	...	I	S
p	$(1-p)$	$(1-p)$	p	p		$(1-p)$	p

ed ogni sequenza con k successi coincide con una combinazione, quindi il numero di sequenze con k successi coincide con $C_{n,k} = \binom{n}{k}$. La **funzione di probabilità di una distribuzione binomiale** n, k è data quindi da:

$$P_n(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Es. Calcoliamo la probabilità che il Rosso esca per 5 volte di fila alla roulette. Dobbiamo prima stabilire qual è il parametro p , cioè la probabilità che il Rosso esca una volta: $\frac{18}{37}$ (lo 0 non ha colore). Essendo nel caso di una distribuzione binomiale con $n = 5$ e $p = \frac{18}{37}$, basta calcolare $P_5(X = 5) = \binom{5}{5} \cdot \left(\frac{18}{37}\right)^5 \left(\frac{19}{37}\right)^0 = \left(\frac{18}{37}\right)^5 \sim 0.03$.

Vogliamo calcolare la probabilità che su 3 estrazioni del lotto, esca almeno una volta il numero 8.

Lo schema generale è quello della distribuzione binomiale con $n = 3$, con parametro $p = \frac{\binom{89}{4}}{\binom{90}{5}} = \frac{\frac{89!}{4!85!}}{\frac{90!}{5!85!}} = \frac{89! 85! 5}{85! 90!} = \frac{5}{90} = \frac{1}{18}$, quindi $P_3(X \geq 1) = 1 - P_3(X = 0) = 1 - \binom{3}{0} \cdot \left(\frac{1}{18}\right)^0 \left(\frac{17}{18}\right)^3 = 1 - \left(\frac{17}{18}\right)^3 \sim 0.16$.

Un tipico esempio di processo bernoulliano è dato dall'estrazione *con reimbussolamento* di n oggetti da un'urna che ne contiene N . Lo schema dell'estrazione *senza reimbussolamento* dà una v.a. X che conta il numero di oggetti "privilegiati" che si trovano negli n estratti. Tale X segue una legge diversa, detta **ipergeometrica**, che dipende da 3 parametri: il numero N degli oggetti contenuti nell'urna, il numero K degli oggetti "privilegiati" ed il numero n di estrazioni. La **funzione di probabilità della distribuzione ipergeometrica** è data da:

$$P_{N,K,n}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Es. Contiamo qual è la probabilità di fare terno giocando 4 numeri al lotto. Poichè nel gioco del lotto si estraggono 5 numeri senza reimbussolamento da un'urna contenente 90 numeri (dall'1 al 90), siamo nello schema di una distribuzione ipergeometrica con $N = 90, K = 4$ (i 4 numeri giocati), $n = 5$. Per fare terno, devono esserci tra i 5 estratti 3 di quelli giocati, quindi $k = 3$. Allora avremo

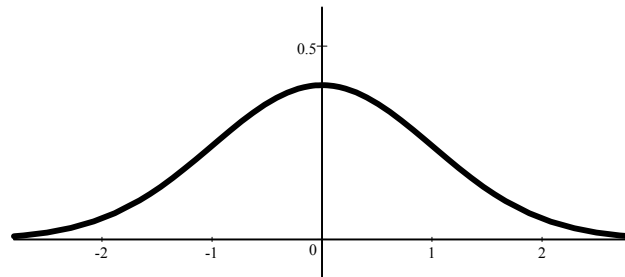
$$P_{90,4,5}(X = 3) = \frac{\binom{4}{3} \binom{86}{2}}{\binom{90}{5}} = \frac{\frac{4!}{1!3!} \frac{86!}{2!84!}}{\frac{90!}{5!85!}} = 2 \frac{86! 5! 85!}{84! 90!} = 2 \frac{5! 85}{90 \cdot 89 \cdot 88 \cdot 87} \sim 3.3 \times 10^{-4}$$

Distribuzioni normali o gaussiane.

È assolutamente la classe di distribuzioni più importante, per motivi a cui accenneremo più avanti. Una funzione di densità **normale o gaussiana di parametri** μ, σ^2 è una funzione del tipo:

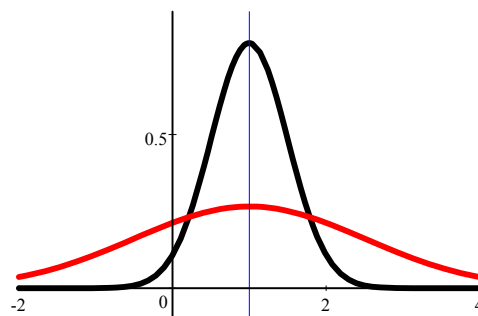
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e la v.a. continua relativa è detta **normale** μ, σ^2 ed indicata con $N(\mu, \sigma^2)$. Se $\mu = 0, \sigma^2 = 1$ si ha la **normale standard** $N(0, 1)$, la cui densità $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ ha grafico dato da:



Le densità normali hanno *grafici simmetrici rispetto a $x = \mu$, indice di centro*.

Il parametro σ è indice di dispersione dei dati rispetto al centro, in quanto per σ piccolo, il valore di massimo è più grande, la regione sotto il grafico è più concentrata verso il centro, quindi c'è più probabilità di trovare valori vicini a μ , cioè meno dispersione dei dati:

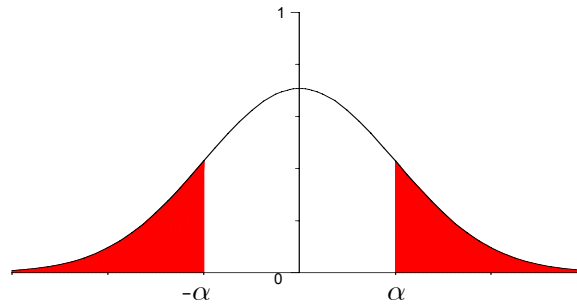


nero: $N(1, \frac{1}{2})$, rosso: $N(1, \frac{3}{2})$

Il problema delle distribuzioni normali è che le loro densità non sono integrabili in modo elementare, quindi *i valori degli integrali sono dati approssimati*. Fortunatamente basta conoscere i dati relativi alla normale standard, in quanto si ha il teorema:

Se X è una v.a. che si distribuisce $N(\mu, \sigma^2)$, allora la v.a. $Y = \frac{X - \mu}{\sigma}$ si distribuisce in modo standard, cioè $N(0, 1)$.

Per calcolare i valori relativi a $N(0, 1)$ si utilizza una tabella, da usare con attenzione, in quanto la tabella riporta solo i valori (approssimati) della funzione di distribuzione normale standard $\Phi(x) = P(X \leq x)$ per i valori di $x \geq 0$. Nel caso $N(0, 1)$, la funzione di densità ha un grafico simmetrico rispetto all'asse y e sfruttando questa simmetria otteniamo che $\Phi(-\alpha) = 1 - \Phi(\alpha)$, in quanto l'area della regione di piano sottesa alla curva prima di $-\alpha$ coincide con l'area della regione di piano sottesa alla curva dopo il simmetrico α :



area che coincide con la probabilità $P(X \geq \alpha) = 1 - P(X \leq \alpha)$. Questo ci dà anche la possibilità di leggere sulla tabella le probabilità che la distribuzione normale standard ha di cadere in una semiretta destra.

EX: 1) Sia X una v.a. distribuita $N(0, 1)$. Calcolare le probabilità: $P(X \leq 1.23)$, $P(X \leq -0.37)$, $P(X \geq 1.23)$, $P(0.67 \leq X \leq 0.78)$, $P(X \geq -0.15)$, $P(0.5 \leq X \leq 1)$.

2) Supponiamo che il peso (in Kg) di sacchi di grano venduti da un'azienda agricola sia rappresentato da una v.a. X . che si distribuisce normalmente con $\sigma = 2$ e $\mu = 30$. Qual è la probabilità che un sacco pesi più di 31 Kg?

Esempi di distribuzioni normali.

La misura di una qualsiasi grandezza fisica è necessariamente affetta da un errore di misurazione, quindi si può pensare come la somma $m + X$ tra il valore "vero" m della grandezza (m numero reale costante) e dell'errore di misurazione X , che è una v.a., in quanto misure diverse forniranno valori diversi. Questa v.a. "errore di misurazione" si distribuisce normalmente con parametri μ e σ , dove, se $\mu \neq 0$, si chiama *errore sistematico* e σ che rappresenta l'*inaccuratezza* delle misura.

Anche molte grandezze che rappresentano caratteristiche di una popolazione (omogenea), come l'altezza, il peso, ect, sono rappresentate da distribuzioni normali, come pure grandezze con significato fisiologico (pressione arteriosa) o biologico.

Altro caso in cui si presentano distribuzioni gaussiane è quello in cui c'è una produzione in serie di oggetti che si vorrebbero fare tutti identici, ma che risultano in effetti con piccole differenze: per esempio, la v.a. che rappresenta il volume effettivo di liquido contenuto in una bottiglia da 33 ml di bibita, o il peso effettivo di fustini da 5 kg di detersivo e così via. Una motivazione teorica del perchè la distribuzione normale rappresenti tutti queste v.a. è data dal teorema del limite centrale, per la cui formulazione abbiamo bisogno di alcune definizioni di grande importanza probabilistica:

Definizione: a) Sia X una v.a. finita. con possibili valori assunti x_1, x_2, \dots, x_n e con funzione di probabilità $p_x(x_k) = p_k$, per $k = 1, 2, \dots, n$. Si definisce allora **media o valor medio o valor atteso o speranza matematica** di X il numero

$$\mu = E(X) = \sum_{k=1}^n x_k p_k.$$

b) Sia X una v.a. continua con funzione di densità $f(x)$. Allora la **media o valor medio o valor atteso o speranza matematica** di X sarà dato da

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx, \text{ se tale integrale improprio è convergente.}$$

Proprietà del valor medio

- $E(X + c) = E(X) + c$ (in particolare, per $Z = c$, avremo $E(Z) = c$)
- $E(cX) = cE(X)$
- $E(X + Y) = E(X) + E(Y)$
- Se X, Y sono v.a. indipendenti, allora $E(XY) = E(X)E(Y)$.

Come conseguenza, se chiamiamo **varianza** di X il numero $\sigma^2 = Var(X) = E((X - \mu)^2)$ per X v.a. finita o continua di valor medio $E(X) = \mu$, avremo che

$$\begin{aligned}\sigma^2 &= E((X - \mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + E(\mu^2) = \\ &= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2\end{aligned}$$

Indicheremo poi con $\sigma = \sqrt{Var(X)}$ la **deviazione standard** della v.a. X .

Proprietà della varianza

- $Var(cX) = c^2 Var(X)$
- Se X, Y sono v.a. indipendenti, allora $Var(X + Y) = Var(X) + Var(Y)$.

È facile vedere che il valor medio e la deviazione standard di una distribuzione normale coincidono con i parametri μ e σ della sua funzione di densità: nel caso $N(0, 1)$, si ha per esempio che

$$E(X) = \int_{-\infty}^{+\infty} x \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}x^2} dx = 0,$$

integrale convergente e nullo, in quanto si integra una funzione dispari (grafico simmetrico rispetto all'origine) su un intervallo simmetrico rispetto all'origine.

Il valor medio e la varianza di una distribuzione bernoulliana di parametri n, p saranno:

$$E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) = np,$$

$$\sigma^2 = \sigma^2(X_1 + X_2 + \dots + X_n) = np(1 - p),$$

in quanto in ogni singola prova si ha valor medio $E(X_i) = 0(1 - p) + 1p = p$ e varianza $\sigma^2 = E(X_i^2) - \mu^2 = p - p^2 = p(1 - p)$.

Esempi: Calcolare il valor medio e la varianza della v.a. X data da:

X	-3	-1	0	1/3	1	5/4	2
p_x	0.02	0.14	0.33	0.21	0.05	0.15	0.1

$$E(X) = -3 \frac{2}{100} - 1 \frac{14}{100} + 0 \frac{33}{100} + \frac{1}{3} \frac{21}{100} + 1 \frac{5}{100} + \frac{5}{4} \frac{15}{100} + 2 \frac{10}{100} = -\frac{169}{400} = -.4225$$

$$\sigma^2 = E(X^2) - \mu^2 = 9\frac{2}{100} + 1\frac{14}{100} + 0\frac{33}{100} + \frac{1}{9}\frac{21}{100} + 1\frac{5}{100} + \frac{25}{16}\frac{15}{100} + 4\frac{10}{100} - \left(-\frac{169}{400}\right)^2 \sim 15.7,$$

$$\sigma = \sqrt{15.7} \sim 3.96.$$

Calcolare il valor medio di X con funzione di densità $f(x) = \begin{cases} 0 & \text{per } x < 1 \\ \frac{2}{x^3} & \text{per } x \geq 1 \end{cases}$:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = 0 + \int_1^{+\infty} x\frac{2}{x^3}dx = [(-2x^{-1})]_1^{+\infty} \\ &= \lim_{x \rightarrow +\infty} (-2x^{-1}) - (-2 \cdot 1^{-1}) = 2. \end{aligned}$$

Il seguente risultato dà una spiegazione dell'uso del valor medio e della deviazione standard:

Diseguaglianza di Chebychev: Se X è una v.a. con valor medio μ e deviazione standard σ , allora per ogni $t > 1$ si ha che

$$P(|X - \mu| < t\sigma) \geq 1 - \frac{1}{t^2}$$

In particolare

$$\begin{aligned} P(X \in [\mu - \frac{3}{2}\sigma, \mu + \frac{3}{2}\sigma]) &\geq 0.55 \\ P(X \in [\mu - 2\sigma, \mu + 2\sigma]) &\geq 0.75 \\ P(X \in [\mu - 3\sigma, \mu + 3\sigma]) &\geq 0.98 \end{aligned}$$

Questo risultato ci dà una limitazione alla probabilità che X si discosti dal valor medio per più di $t\sigma$ o, detto in altri termini, *questa diseguaglianza ci permette di avere una stima della probabilità che i valori della v.a. cadano in un intervallo di centro il valor medio e ampiezza $t\sigma$, cioè in $[\mu - t\sigma, \mu + t\sigma]$.*

Se il valor medio di X è 5 e la deviazione standard è 0.5, la probabilità che X si discosti dal valor medio per più di $2\sigma = 1$ è $\leq \frac{1}{4} = 0.25$, quindi almeno il 75% dei valori cade in $[5 - 1, 5 + 1] = [4, 6]$.

Data una qualunque v.a. X , almeno il 55% dei valori cade in $[\mu - \frac{3}{2}\sigma, \mu + \frac{3}{2}\sigma]$, almeno il 75% dei valori cade in $[\mu - 2\sigma, \mu + 2\sigma]$, almeno il 98% dei valori cade in $[\mu - 3\sigma, \mu + 3\sigma]$.

Passiamo ora ad affrontare il problema dell'approssimazione di variabili aleatorie, introducendo dapprima la nozione di media campionaria.

Definizione: date n v.a. X_1, X_2, \dots, X_n mutuamente indipendenti, di valor medio μ e varianza σ^2 , la v.a. data dalla media aritmetica delle n v.a.

$$M_n = \frac{1}{n} \sum_{k=1}^n X_k = \frac{1}{n} S_n$$

è una nuova variabile aleatoria, con valor medio:

$$E(M_n) = E\left(\frac{1}{n}S_n\right) = \frac{1}{n}E(S_n) = \frac{1}{n}E(X_1 + X_2 + \dots + X_n) = \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) = \frac{1}{n}n\mu = \mu$$

e varianza:

$$\begin{aligned} \text{Var}(M_n) &= \text{Var}\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}\text{Var}(S_n) = \frac{1}{n^2}\text{Var}(X_1 + X_2 + \dots + X_n) = \\ &= \frac{1}{n^2}(\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

Quindi la media campionaria ha valor medio uguale alle v.a. di partenza e varianza che diminuisce al crescere di n . Nel caso in cui le v.a. di partenza siano ugualmente distribuite, la media campionaria avrà generalmente una nuova distribuzione, con uguale valor medio μ e con probabilità di avere valori vicini a μ sempre più alte al crescere di n .

Nel caso particolare in cui le X_i siano tutte distribuite normalmente $N(\mu, \sigma^2)$ anche la media campionaria M_n sarà distribuita normalmente $N(\mu, \frac{\sigma^2}{n})$.

In generale la media campionaria sarà distribuita con una legge (che non conosciamo) di tipo diverso da quella comune alle X_i , ma il **Teorema del limite centrale** ci dice sostanzialmente che se n v.a. hanno tutte la stessa legge, di qualunque tipo non normale, senza essere normale M_n sarà comunque approssimata al crescere di n da una legge normale $N(\mu, \frac{\sigma^2}{n})$, dove μ e σ^2 sono il valor medio e la varianza comuni a tutte le v.a. X_i .

Se X_1, X_2, \dots, X_n sono mutuamente indipendenti e identicamente distribuite, con valor medio μ e varianza σ^2 , la media campionaria $M_n = \frac{1}{n} \sum_{k=1}^n X_k$ al crescere di n tende a distribuirsi con legge normale $N(\mu, \frac{\sigma^2}{n})$ (M_n è approssimata, per valori di n grandi, da $N(\mu, \frac{\sigma^2}{n})$).

Di conseguenza $S_n = X_1 + X_2 + \dots + X_n = nM_n$ sarà approssimata, per valori di n grandi, da $N(n\mu, n\sigma^2)$.

Anche una distribuzione binomiale può quindi avere un'approssimazione normale, che è buona se $np > 5$ e $n(1-p) > 5$. In tal caso, poiché una distribuzione binomiale X di parametri n e p si ottiene come somma $X_1 + X_2 + \dots + X_n$ di distribuzioni zero-uno ognuna con $\mu = p$ e $\sigma^2 = p(1-p)$, si avrà che $X = S_n \approx N(np, np(1-p))$, mentre per la media campionaria (frequenza di successi su n prove) si avrà: $M_n \approx N(p, \frac{p(1-p)}{n})$.

Esempi.

Da un'indagine svolta su un campione di neonati, il peso alla nascita è risultato rappresentabile con una v.a. normale con media $\mu = 3.2$ Kg e deviazione standard $\sigma = 0.6$ Kg.

i) Su un campione casuale di 1000 bambini, quanti sono attesi avere un peso compreso tra 3 e 3.2 Kg?

ii) Considerando i pesi medi calcolati su 25 bambini in 1000 ospedali, in quanti casi ci si attende una media compresa tra 3 e 3.2 Kg?

Alla prima domanda rispondiamo usando la v.a. X che si distribuisce $N(3.2, 0.6^2)$ e quindi calcolando $P(3 < X < 3.2) = P(\frac{3-3.2}{0.6} = -\frac{2}{3} \simeq -0.34 < \frac{X-3.2}{0.6} < \frac{3.2-3.2}{0.6} = 0) = P(Y < 0) - P(Y < -0.34) = 0.5 - (1 - P(Y < 0.34)) = 0.5 - 1 + P(Y < 0.34) \simeq -0.5 + 0.63307 \simeq 0.133$

quindi su 1000 bambini ci si aspetta che circa 133 abbiano un peso compreso tra 3 e 3.2 Kg. Per la seconda prendiamo la media campionaria \bar{X} che in questo caso si distribuisce proprio normalmente con media $\mu = 3.2$ Kg e deviazione standard $= \frac{0.6}{\sqrt{25}} = \frac{0.6}{5} = 0.12$ e quindi otteniamo $P(3 < \bar{X} < 3.2) = P(\frac{3-3.2}{0.12} < \frac{\bar{X}-3.2}{0.12} < \frac{3.2-3.2}{0.12}) = P(-\frac{0.2}{0.12} \simeq -1.67 < Y < 0) = P(Y < 0) - P(Y < -1.67) = 0.5 - (1 - P(Y < 1.67)) = 0.5 - 1 + P(Y < 1.67) \simeq -0.5 + 0.95254 \simeq 0.453$, quindi su 1000 ospedali ci si aspetta che in circa 453 abbiano rilevato un peso medio compreso tra 3 e 3.2 Kg.

L'altezza X della popolazione femminile adulta di un certo comune si distribuisce normalmente con $\mu = 162$ cm e $\sigma = 5$ cm. Qual è la probabilità che una ragazza presa a caso abbia un'altezza inferiore a 160cm? Qual è la probabilità che 16 ragazze prese a caso abbiano in media un'altezza inferiore a 160cm?

Alla prima domanda rispondiamo usando la v.a. X che si distribuisce $N(162, 5^2)$ e quindi calcolando

$$P(X < 160) = P\left(\frac{X - 162}{5} < \frac{160 - 162}{5}\right) = P\left(Y < -\frac{2}{5}\right) = 1 - P\left(Y < \frac{2}{5} = 0.4\right) \sim 1 - 0.65542 = .34458.$$

Per la seconda prendiamo la media campionaria M_{16} che in questo caso si distribuisce proprio normalmente $N(162, \frac{5^2}{16})$ e quindi otteniamo

$$P(M_{16} < 160) = P\left(\frac{M_{16} - 162}{\frac{5}{4}} < \frac{160 - 162}{\frac{5}{4}}\right) = P\left(Y < -\frac{8}{5}\right) = 1 - P\left(Y < \frac{8}{5} = 1.6\right) \sim 1 - 0.94520 = .05480$$

Esempi.

Viene ipotizzato che la durata di vita delle piantine di tulipano in Nuova Zelanda sia rappresentabile con una v.a. di valore atteso 2 anni e deviazione standard 2 anni. Piantando 400 piantine, calcolare valore atteso e deviazione standard per la durata media di vita delle 400 piantine. Adottando una approssimazione normale, calcolare la probabilità p che la durata media di vita delle 400 piantine sia inferiore a 2.1 anni.

Soluzione : $E(M_{400}) = 2$ anni e $\sigma(M_{400}) = \frac{2}{\sqrt{400}} = \frac{1}{10}$ di anno. $P(M_{400} < 2.1) = P\left(\frac{M_{400}-2}{\frac{1}{10}} < \frac{2.1-2}{\frac{1}{10}} = 1\right) = P(Y < 1) \sim 0.84134$

Supponiamo che la durata di funzionamento di una torcia sia rappresentabile con una v.a. di valore atteso 4 ore e deviazione standard 30 minuti. Usando 100 torce di fila (appena si spegne una ne accendo un'altra), qual è approssimativamente la probabilità p di non rimanere al buio prima di 17 giorni?

Sia Y la somma dei punteggi di 20 dadi equilibrati. Calcolare l'approssimazione normale per $P(60 < Y < 75)$.

Supponiamo che X abbia distribuzione binomiale con parametri $n = 50$ e $p = 0.3$. Calcola l'approssimazione normale per $P(12 < X < 16)$

Supponiamo che X_1, X_2, \dots, X_{30} sia un campione casuale di dimensione 30 estratto da una distribuzione uniforme su $(0, 1)$. Sia $S = X_1 + X_2 + \dots + X_{30}$. Trova le approssimazioni normali a $P(13 < S < 18)$.

Sia M la media campionaria di un campione casuale di dimensione 50 tratto da una distribuzione con funzione di densità $f(x) = 3x^{-4}, x > 1$. Trovare le approssimazioni di $P(M > 1.6)$.

Stime e stimatori

Nella raccolta di dati statistici, con il termine *popolazione* si intende di solito l'insieme delle osservazioni possibili, mentre il *campione statistico* si riferisce all'insieme delle osservazioni effettivamente utilizzate. Per poter arrivare a conclusioni attendibili, ci si deve basare su una scelta "casuale" del campione e si parla quindi di un campione casuale. Noi comunque non tratteremo i vari metodi che possono essere utilizzati per estrarre un campione casuale.

Mentre nella probabilità si studiano problemi di cui si conosce (o si può conoscere) la distribuzione di probabilità, nella statistica si incontrano spesso problemi di cui non si conosce il tipo di distribuzione probabilistica oppure si conosce il tipo, ma non tutti i parametri che la definiscono, per esempio di una binomiale non si sa la probabilità p di successo o di una normale non si conosce il valor medio μ o la varianza σ^2 . Come possiamo utilizzare un campione casuale per trovare una stima di un parametro incognito?

Sia $\{x_1, x_2, \dots, x_n\}$ un campione casuale di n variabili aleatorie X_1, X_2, \dots, X_n indipendenti e identicamente distribuite. Supponiamo che la tale distribuzione comune contenga un parametro α incognito. Per usare il campione casuale per avere una stima del parametro, cerchiamo una funzione $t(x_1, x_2, \dots, x_n)$ del campione (per esempio la somma od il prodotto) tale che il valore $t = t(x_1, x_2, \dots, x_n)$ calcolato sul campione sia una *stima* del parametro α . Notiamo che il numero t può essere considerato come un'osservazione della nuova variabile aleatoria

$$T = T(X_1, X_2, \dots, X_n).$$

Definizione: tale v.a. si dice *stimatore* di α .

Essendo T una v.a., è definito il suo valor medio $E(T)$.

Definizione: uno stimatore di un parametro α si dice *corretto* quando il suo valor medio è proprio α , cioè quando $E(T) = \alpha$.

Inoltre se ci sono più stimatori corretti, migliore sarà quello con varianza più piccola, cioè quello in cui la probabilità di trovare valori lontani dal valor medio (che è α) è la più bassa.

Si dirà quindi che lo stimatore corretto T è il *più efficiente* se $\sigma^2(T) \leq \sigma^2(T')$, per ogni altro stimatore corretto T' .

STIMA DEL VALOR MEDIO: siano X_1, X_2, \dots, X_n n v.a. indipendenti di uguale distribuzione con valor medio μ incognito. Come stima di μ di solito si usa la *media del campione* $\{x_1, x_2, \dots, x_n\}$, per cui il relativo stimatore è la media campionaria M_n . Abbiamo visto che $E(M_n) = \mu$, quindi la *media campionaria* è uno stimatore corretto del valor medio.

STIMA DELLA VARIANZA con valor medio noto: siano X_1, X_2, \dots, X_n n v. a. indipendenti di uguale distribuzione con valor medio μ noto e varianza ignota. Poiché $E[(X_i - \mu)^2] = \sigma^2$,

$$E\left[\frac{1}{n}\sum_{i=1}^n(X_i - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^n E[(X_i - \mu)^2] = \frac{1}{n}n\sigma^2 = \sigma^2, \text{ da cui otteniamo che } \frac{1}{n}\sum_{i=1}^n(X_i - \mu)^2 \text{ è uno}$$

stimatore corretto della varianza e valutato sul campione dà una stima $\frac{1}{n}\sum_{i=1}^n(x_i - \mu)^2$ detta *varianza campionaria*, che coincide con la varianza dei dati raccolti.

STIMA DELLA VARIANZA con valor medio incognito: poiché ora anche il valor medio è incognito, la scelta più ovvia per stimare la varianza sarebbe quella di sostituire nello stimatore precedente

μ con il suo stimatore noto, cioè M_n per ottenere $\frac{1}{n}\sum_{i=1}^n(X_i - M_n)^2$. Ma in questo modo troviamo

uno stimatore non corretto (provare per credere!). Uno stimatore corretto è dato invece da $S^2 =$

$$\frac{1}{n-1}\sum_{i=1}^n(X_i - M_n)^2.$$

STIMA DELLA PROPORZIONE p : sappiamo che se ripetiamo n prove indipendenti, ognuna con probabilità di successo p , la v.a. X data dal numero dei successi ha distribuzione binomiale con $E(X) = np$. Di conseguenza la variabile aleatoria $\frac{X}{n}$ che conta la frequenza di successi avrà $E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$, quindi risulta essere uno stimatore corretto di p (ed è anche lo stimatore più efficiente). La frequenza di successi sul campione darà quindi una stima di p .

Determinante di una matrice quadrata

Per calcolare il determinante ($|A|$ o $\det A$) di una matrice quadrata $A = (a_{ij})_{i=1\dots n}^{j=1\dots n}$ di ordine n , si procede nel modo seguente:

- se $n = 1$, ossia se $A = \begin{pmatrix} a_{11} \end{pmatrix}$, allora $|A| = a_{11}$;
- se $n = 2$, ossia se $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, allora $|A| = a_{11}a_{22} - a_{12}a_{21}$.

Esempio 1 Date le matrici $A = \begin{pmatrix} -1 & 5 \\ 0 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 3 & -2 \\ 1 & -1 \end{pmatrix}$
 si ha: $|A| = (-1) \cdot 2 - 0 \cdot 5 = -2$; $|B| = 3 \cdot (-1) - (-2) \cdot 1 = -1$.

Per poter calcolare il determinante di una matrice quadrata di ordine $n \geq 3$, dobbiamo introdurre le seguenti definizioni.

- Data la matrice quadrata A di ordine $n \geq 2$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & & \dots & a_{nn} \end{pmatrix}$$

si definisce

- *Minore complementare dell'elemento a_{ij}* : il determinante della matrice quadrata di ordine $n - 1$, A_{ij} , ottenuta da A cancellando la riga i e la colonna j .
- *Complemento algebrico di a_{ij}* : il numero reale $(-1)^{i+j} |A_{ij}|$

Esempio 2 Sia $A = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ allora

$$A_{11} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}, \text{ e il complemento algebrico dell'elemento } a_{11} \text{ è: } (-1)^{1+1} |A_{11}| = 5$$

$$A_{21} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}, \text{ e il complemento algebrico dell'elemento } a_{21} \text{ è: } (-1)^{2+1} |A_{21}| = -2$$

$$A_{31} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}, \text{ e il complemento algebrico dell'elemento } a_{31} \text{ è: } (-1)^{3+1} |A_{31}| = 1$$

- Siamo ora in grado di calcolare il determinante di una qualunque matrice quadrata di ordine $n \geq 3$:

Il determinante di A si ottiene sommando i prodotti degli elementi di una qualunque linea (riga o colonna) di A per i loro complementi algebrici.

NOTA Per calcolare il determinante conviene, ovviamente, fissare una linea della matrice che contiene più zeri.

Esempio 3 Sia A la matrice dell'esempio precedente, per calcolarne il determinante fissiamo (ad esempio) la prima colonna, allora

$$|A| = 2 \cdot 5 + (-1) \cdot (-2) + 0 \cdot 1 = 12.$$

Fissando invece la terza riga si ottiene, come vediamo, lo stesso risultato:

$$|A| = 0 \cdot \begin{vmatrix} 1 & 0 \\ 3 & 1 \end{vmatrix} - 1 \cdot \begin{vmatrix} 2 & 0 \\ -1 & 1 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 1 \\ -1 & 3 \end{vmatrix} = -1 \cdot 2 + 2 \cdot (6 + 1) = 12.$$

Esempio 4 Per calcolare il determinante della matrice

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 2 & 1 \\ 2 & 0 & 1 & 0 \\ 0 & -2 & 1 & 1 \end{pmatrix},$$

fissiamo, ad esempio, la prima colonna e otteniamo:

$$|A| = 1 \cdot \begin{vmatrix} 3 & 2 & 1 \\ 0 & 1 & 0 \\ -2 & 1 & 1 \end{vmatrix} - 0 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 0 & 1 & 0 \\ -2 & 1 & 1 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 3 & 2 & 1 \\ -2 & 1 & 1 \end{vmatrix} - 0 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 3 & 2 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

Fissando ora la seconda riga nella prima matrice e la prima colonna nella terza otteniamo:

$$\begin{aligned} |A| &= 1 \cdot \begin{vmatrix} 3 & 1 \\ -2 & 1 \end{vmatrix} + 2 \cdot \left(2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} - 3 \cdot \begin{vmatrix} 3 & 4 \\ 1 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix} \right) \\ &= 5 + 2 \cdot [2 + 3 + 10] = 35. \end{aligned}$$

► *Alcune proprietà del determinante.*

- ◇ Se A è triangolare, allora $|A| = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn}$.
- ◇ Se A ha una riga o una colonna di zeri, allora $|A| = 0$.
- ◇ Se A ha due righe o due colonne uguali, allora $|A| = 0$.
- ◇ Scambiando due righe o due colonne il determinante cambia segno.
- ◇ (*Teorema di Binet*) Se A e B sono matrici quadrate dello stesso ordine, allora

$$|AB| = |A| \cdot |B|.$$

Matrici inverse

Sia A matrice quadrata di ordine n .

► A è *invertibile* se esiste una matrice A^{-1} (detta *inversa di A*), quadrata di ordine n , tale che

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

dove I è la matrice *identica* di ordine n , cioè

$$I = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Esempio 5 La matrice inversa di $A = \begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix}$ è la matrice $A^{-1} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}$.

Infatti si verifica che:

$$AA^{-1} = \begin{pmatrix} 3 \cdot 1 - 1 \cdot 2 & 3 \cdot 1 - 1 \cdot 3 \\ -2 \cdot 1 + 1 \cdot 2 & -2 \cdot 1 + 1 \cdot 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{e anche } A^{-1}A = I.$$

Esempio 6 La matrice inversa **non** sempre esiste. Ad esempio, la matrice $A = \begin{pmatrix} 3 & -1 \\ 0 & 0 \end{pmatrix}$ non ha inversa.

Infatti se si cerca una matrice $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ tale che $AB = I$, dovrebbe essere

$$AB = \begin{pmatrix} 3a - c & 3b - d \\ 0 & \boxed{0} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \boxed{1} \end{pmatrix} = I,$$

da cui $0 = 1$, ma questo è impossibile.

Vale il seguente

Teorema. Sia $A = (a_{ij})$ una matrice quadrata di ordine n . La matrice A è invertibile se e solo se $|A| \neq 0$.

In tal caso gli elementi b_{ij} della matrice inversa A^{-1} sono dati da

$$b_{ij} = \frac{1}{|A|} (-1)^{i+j} |A_{ji}|.$$

In particolare, se $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ con $ad - bc \neq 0$, allora $A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Esempio 7 Calcolare, se possibile, la matrice inversa delle seguenti matrici

$$i) A = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}; \quad ii) A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}; \quad iii) A = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 1 & 2 \\ -1 & 1 & 1 \end{pmatrix}.$$

$$i) |A| = 5 \text{ e } A^{-1} = \frac{1}{5} \begin{pmatrix} 3 & 1 \\ -2 & 1 \end{pmatrix};$$

ii) $|A| = 0$ quindi non esiste l'inversa;

$$iii) |A| = -5 \text{ e } A^{-1} = \frac{1}{5} \begin{pmatrix} 1 & 2 & -3 \\ 2 & -1 & 4 \\ -1 & 3 & -2 \end{pmatrix}.$$

Bibliografia e Testi consigliati

Marco Bramanti - “Calcolo delle probabilità e statistica” , Progetto Leonardo, Bologna.

G. Naldi, L. Pareschi, G. Aletti - “ Matematica I” , McGraw-Hill.

A. Camussi, F. Möller, E. Ottaviano, M. Sari Gorla - “Metodi statistici per la sperimentazione biologica” , Zanichelli.

S. Invernizzi, M. Rinaldi, A. Sgarro - “Moduli di Matematica e Statistica” , Zanichelli.