

Variabili aleatorie

Iniziamo con una definizione (capiremo fra poco la sua utilità):

DEFINIZIONE DI VARIABILE ALEATORIA

Una variabile aleatoria (in breve v.a.) X è una funzione che ha come dominio Ω e come codominio \mathbb{R} . In formule:

$$X : \Omega \rightarrow \mathbb{R}.$$

A voler essere precisi non tutte le funzioni possibili sono v.a. ma questo è un problema tecnico di cui non ci curiamo (è veramente molto difficile trovare funzioni che non siano v.a.).

A cosa servono?

Una funzione fa corrispondere ad ogni valore nel dominio un valore nel codominio. Sembrerebbe non esserci nulla di aleatorio in questo! (aleatorio=casuale da *alea* in latino)

Dobbiamo pensare che il caso peschi dal nostro mondo dei casi possibili Ω un elemento $\omega \in \Omega$ (dunque vediamo uno dei casi possibili, se ne *realizza* uno solo).

Può accadere che noi non conosciamo esattamente chi sia ω ma ne conosciamo $X(\omega)$ dove X è una particolare funzione. Ad esempio se una persona va dal medico, quest'ultimo ha davanti a sé un "caso possibile" fra tutti i casi "essere umano" possibili.

Il medico non può sapere **tutto** del paziente (ad esempio lo stato di tutte le molecole del suo corpo!), ma può osservare il valore di alcune v.a. come X = pressione massima, Y = pressione minima, Z = glicemia a digiuno, etc etc.

Variabili aleatorie

La roulette

Valori possibili

La probabilità

La legge di una
variabile aleatoria

V.a. discrete e continue

La densità di una v.a. discreta

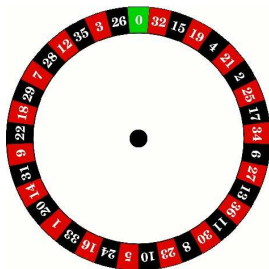
Proprietà di f_X

Esempi

Approfondiamo

Esempio: la roulette

Supponiamo di giocare alla roulette, puntando 10 € sul rosso, 15 € sul pari e 5 € sul 19; quello che vinciamo dopo che la ruota ha finito di girare è una v.a. X .



Infatti qui $\Omega = \{0, 1, \dots, 36\}$ (ω possibili = interi fra 0 e 36) e (salvo trucchi) ogni caso ha uguale probabilità. X dipende dall' ω che esce quindi è una **funzione** definita su Ω a valori in \mathbb{R} .

Le vincite possibili

Quel che ci interessa a questo punto non è tanto che ω (cioè il numero della roulette) esca, quanto $X(\omega)$ = vincita risultante (in €).

La vincita è diversa se esce un rosso pari (RP), un rosso dispari (RD), un nero pari (NP), un nero dispari (ND), zero. Infatti

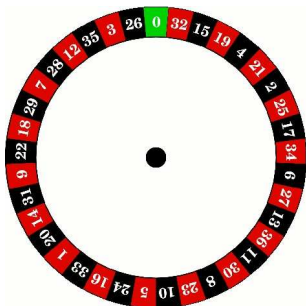
$$X(\omega) = 20 \text{ se } \omega \in RD \setminus \{19\}$$

$$X(\omega) = 30 \text{ se } \omega \in NP$$

$$X(\omega) = 50 \text{ se } \omega \in RP$$

$$X(\omega) = 200 \text{ se } \omega = 19$$

$$X(\omega) = 0 \text{ se } \omega \in ND \cup \{0\}$$



La probabilità delle vincite

La v.a. X può quindi assumere i valori: $\{0, 20, 30, 50, 200\}$.

Ricordando che la roulette “pesca” con uguale probabilità uno dei 37 numeri possibili, abbiamo

$$\mathbb{P}(X = 0) = \frac{9}{37} \text{ (gli } \omega \text{ corrispondenti sono } 0, 15, 17, 13, 11, 33, 31, 29, 35)$$

$$\mathbb{P}(X = 20) = \frac{9}{37} \text{ (gli } \omega \text{ corrispondenti sono } 21, 25, 27, 23, 5, 1, 9, 7, 3)$$

$$\mathbb{P}(X = 30) = \frac{10}{37} \text{ (gli } \omega \text{ corrispondenti sono } 4, 2, 6, 8, 10, 24, 20, 22, 28, 26)$$

$$\mathbb{P}(X = 50) = \frac{8}{37} \text{ (gli } \omega \text{ corrispondenti sono } 32, 34, 36, 30, 16, 14, 18, 12)$$

$$\mathbb{P}(X = 200) = \frac{1}{37} \text{ (l' } \omega \text{ corrispondente è 19)}$$

Variabili
aleatorie

La roulette

Valori possibili

La probabilità

La legge di una
variabile aleatoria

V.a. discrete e
continue

La densità di
una

v.a. discreta

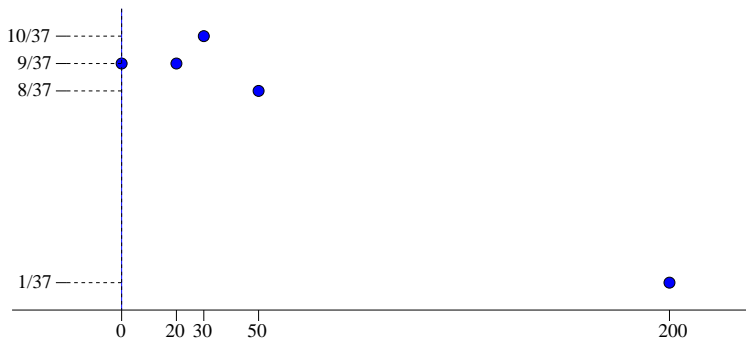
Proprietà di f_X

Esempi

Approfondiamo

La probabilità in un grafico

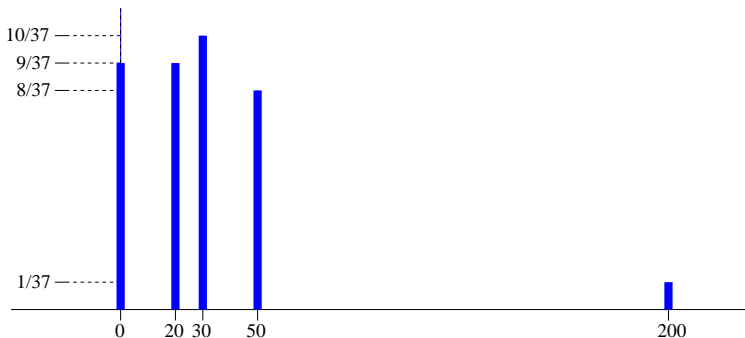
Rappresentiamo i valori che X può assumere sull'asse delle x e la probabilità con cui li assume sull'asse delle y .



L'altezza (y) di un punto (x, y) è la probabilità con cui X assume il corrispondente valore x .

La probabilità in un grafico

Per evidenziare meglio i punti sostituiamoli con delle linee verticali:



Notate che la somma delle altezze di queste righe fa 1 (perché?)

Un po' di nomenclatura

Nota bene

Le caratteristiche interessanti di una v.a. X sono due:

- l'insieme dei suoi (possibili) valori;
- la probabilità con cui assume uno o più valori.

I valori che le v.a. (che noi tratteremo) possono assumere sono numeri reali: le v.a. in questo caso si dicono anche **v.a. numeriche**.

Le v.a. numeriche che tratteremo si suddividono in due tipi: le **v.a. discrete** e le **v.a. continue**. (Questi due casi NON esauriscono tutte le v.a. numeriche)

La legge di una variabile aleatoria

DEFINIZIONE DI LEGGE

Data una variabile aleatoria X la misura di probabilità \mathbb{P}_X così definita

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)), \quad A \subseteq \mathbb{R}$$

si dice **legge** di X .

Non tutti i sottoinsiemi di \mathbb{R} sono ammissibili nella definizione precedente, ma questo è un problema tecnico di cui non ci occuperemo.

V.a. discrete e continue

DEFINIZIONE DI V.A. DISCRETA O CONTINUA

Una v.a. $X : \Omega \rightarrow \mathbb{R}$ avente come immagine in \mathbb{R} l'insieme V (= insieme dei valori che X può assumere) si dice

- discreta se V è un insieme finito oppure infinito numerabile;
- continua se V è un insieme infinito continuo (= più “numeroso” del numerabile).

La X dell'esempio della roulette ha $V = \{0, 20, 30, 50, 200\}$ quindi è discreta.

Ricordo che esempi di insiemi continui sono \mathbb{R} o suoi intervalli (non confondete finito con limitato!); ad esempio la variabile che mi dice dopo quanto si brucerà una lampadina che accendessi adesso è una v.a. continua.

La densità di una v.a. discreta

La seconda caratteristica importante delle v.a. è la probabilità di assumere i valori (uno o più di essi).

Questa probabilità è la funzione di cui abbiamo visto il grafico poco fa (prima con i punti, poi con le righe).

DEFINIZIONE DI DENSITÀ DISCRETA DI V.A.

Sia $X : \Omega \rightarrow \mathbb{R}$ una v.a. discreta. La funzione

$$f_X : \mathbb{R} \rightarrow [0, 1],$$

definita da $f_X(x) = \mathbb{P}(X = x)$ è la funzione di densità (discreta) della v.a. X .

Dunque f_X associa ad ogni valore x la probabilità che X assuma quel valore.

Se x non è uno dei valori possibili, allora $f_X(x) = 0$, mentre se x è uno dei valori possibili, allora $f_X(x)$ è la probabilità che la v.a. X assuma il valore x .

Come può essere V nel caso discreto

Ricordiamo ancora

Ciò che ci interessa di una v.a. discreta è:

- l'insieme V dei valori possibili;
- la densità f_X .

L'insieme V può essere

- finito, e allora lo indichiamo genericamente con $V = \{x_1, x_2, \dots, x_n\}$ (qui V ha n elementi);
- oppure numerabile, e allora lo indichiamo come una *successione* di elementi $V = \{x_i\}_{i=1}^{\infty}$.

Due proprietà di f_X

Una funzione di densità discreta

- 1 assume solo valori ≥ 0
- 2 la somma dei suoi valori è 1, ovvero

$$\sum_{x_i \in V} f_X(x_i) = 1.$$

La cosa non deve stupirci in quanto i valori di f_X altro non sono che probabilità di eventi e

$$\begin{aligned} \sum_{x_i \in V} f_X(x_i) &= \sum_{x_i \in V} \mathbb{P}(X = x_i) \\ &= \mathbb{P}\left(\bigcup_{x_i \in V} (X = x_i)\right) = \mathbb{P}(\Omega) = 1. \end{aligned}$$

Nell'esempio di X = vincita alla roulette,

$$V = \{0, 20, 30, 50, 200\}$$

$$f_X(0) = \frac{9}{37}, f_X(20) = \frac{9}{37}, f_X(30) = \frac{10}{37}, f_X(50) = \frac{8}{37},$$

$$f_X(200) = \frac{1}{37} \text{ (} f_X(x) = 0 \text{ per gli altri } x \in \mathbb{R} \text{)}.$$

Nota bene

Abbiamo ricavato la densità f_X **assumendo implicitamente** che la roulette segua un modello: quello in cui tutti i numeri della ruota hanno la stessa probabilità.

In molti casi capire quanto vale la probabilità dei singoli valori possibili non è banale. Si deve o supporre che la variabile segua un certo modello noto (vedremo esempi di modelli), oppure procedere come nel prossimo esempio.

Le covate del *Passer Italix*

Vogliamo analizzare la biologia riproduttiva del *Passer Italix*
(un ibrido del *Passer Domesticus* e del *Passer Hispaniolensis*).



<http://www.pbase.com/bracciluca>

(Un articolo del 1993 di Brichetti, Caffi, Gandini, ha raccolto e
analizzato molti dati a riguardo.)

Le covate del *Passer Italiae*

Variabili aleatorie

La roulette

Valori possibili

La probabilità

La legge di una
variabile aleatoria

V.a. discrete e continue

La densità di
una

v.a. discreta

Proprietà di f_X

Esempi

Approfondiamo

Consideriamo la variabile X = numero di uova per covata: a priori non sappiamo dire né chi sia V né f_X .

Idea: raccogliamo (molti) dati e poniamo V = insieme dei valori osservati; probabilità di un valore = frequenza relativa con cui è stato osservato.

Questo metodo ha delle limitazioni:

- alcuni valori possibili potrebbero non venire osservati;
- la frequenza osservata potrebbe anche essere assai diversa dalla “vera” probabilità.

Il **buon senso** ci dice che questi inconvenienti sono meno probabili se raccogliamo **molti** dati (ad esempio: 5 dati non va bene, 200 dati meglio!).

Sono stati raccolti 230 dati:

Numero di uova	Freq. assoluta	Freq. relativa
2	12	0.0522
3	15	0.0652
4	21	0.0913
5	82	0.3565
6	96	0.4174
7	3	0.0130
10	1	0.0044

Ricordiamo che probabilità \approx frequenza relativa è solo una **approssimazione** (ma in questo caso è il meglio che possiamo fare).

Variabili
aleatorie

La roulette

Valori possibili

La probabilità

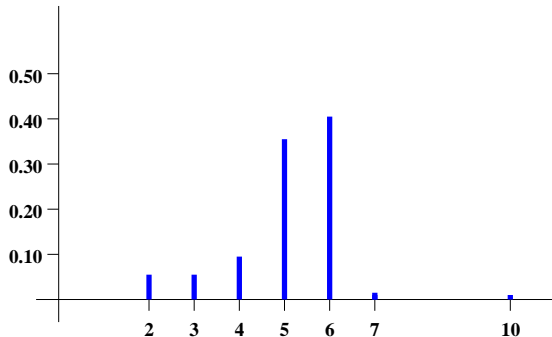
La legge di una
variabile aleatoriaV.a. discrete e
continueLa densità di
una

v.a. discreta

Proprietà di f_X

Esempi

Approfondiamo



Trattandosi di un'approssimazione sperimentale, il buon senso deve guidarci nel suo utilizzo. Ad esempio Brichetti, Caffi, Gandini osservano che il valore 10 è da considerarsi frutto del concorso di due coppie nello stesso nido, quindi andrebbe scartato come errore sperimentale.

Statistica descrittiva e inferenziale

Con l'analisi delle covate abbiamo iniziato a fare della statistica:

- descrittiva: analizza con grafici i dati osservati;
- inferenziale: deduce dai dati proprietà del modello.

Con la descrittiva abbiamo fatto un grafico delle frequenze osservate; con l'inferenziale abbiamo **stimato** la probabilità dei singoli valori con la frequenza osservata.

Quest'ultimo passaggio per ora è giustificato solo dal buon senso, vedremo che ha una giustificazione profonda.

A questo punto potrebbe sorgere un dubbio: abbiamo definito

$$f_X(x_i) = \mathbb{P}(X = x_i),$$

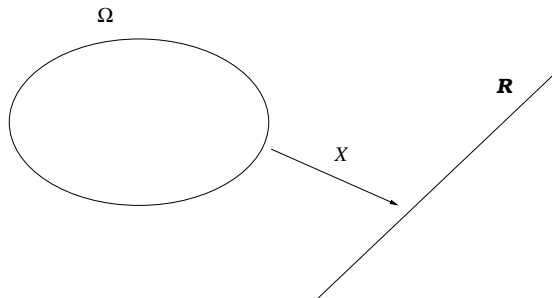
dove $x_i \in V$. Ma la probabilità si calcola di **eventi**: siamo sicuri che $(X = x_i)$ sia un evento?

Quando X assume valore x_i ? Quando il caso che si realizza ω è tale che $X(\omega) = x_i$.

Detto altrimenti, quando $\omega \in X^{-1}(x_i)$ e $X^{-1}(x_i)$ è effettivamente un evento (= sottoinsieme di Ω).

Rappresentiamo la funzione X

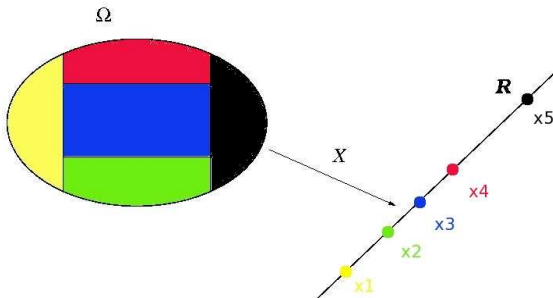
Vediamo una figura “insiemistica” della funzione X :



X associa ad ogni elemento di ω un elemento di \mathbb{R} .

Rappresentiamo la funzione X

Supponiamo che X possa assumere solo i valori x_1, x_2, x_3, x_4, x_5 .



Vediamo nei rispettivi colori: $(X = x_1)$ è la parte gialla di Ω , $(X = x_2)$ quella verde, $(X = x_3)$ quella blu, $(X = x_4)$ quella rossa e $(X = x_5)$ quella nera.

La parte gialla di Ω è fatta da quegli ω per cui $X(\omega) = x_1$, etc.

Ecco un altro dubbio: per il problema della vincita alla roulette potevamo anche scegliere:

$$\Omega = \{0, 20, 30, 50, 200\}$$

con $\mathbb{P}(0) = \frac{9}{37}$, $\mathbb{P}(20) = \frac{9}{37}$, $\mathbb{P}(30) = \frac{10}{37}$, $\mathbb{P}(50) = \frac{8}{37}$,
 $\mathbb{P}(200) = \frac{1}{37}$.

Sembrerebbe che le v.a. non servano veramente e che ci bastino invece i nostri spazi campionari con la loro probabilità.

Perché le v.a.

Sono utili per due motivi:

- 1 in genere non serve costruire Ω , basta chiedersi quali sono i valori possibili e le loro probabilità;
- 2 spesso sullo stesso Ω “vivono” più di una v.a. e il “trucco” che abbiamo usato per l'esempio della roulette (costruire un Ω = valori possibili) potrebbe risultare difficile.