

Variabili aleatorie

1. Cos'è una variabile aleatoria

Una *variabile aleatoria (reale)* o *casuale* è una variabile che assume un ben determinato valore in funzione dell'esito di un esperimento casuale. Sono esempi di variabili aleatorie “la somma nel lancio di due dadi”, “il numero di automobili che transitano per un dato incrocio nell'arco di una giornata”, “il valore del dollaro”.

Diamo una definizione rigorosa di variabile aleatoria: abbiamo detto che essa assume valore come conseguenza dell'esito, dunque essa sarà una variabile dipendente, che si esprime in funzione degli esiti. Precisamente

Definizione 1.1: Variabile aleatoria, funzioni distribuzione e probabilità

Dato uno spazio di probabilità (Ω, \mathcal{E}, P) una *variabile aleatoria* o *casuale* è una funzione

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

misurabile, ovvero tale che per ogni intervallo $I \subset \mathbb{R}$ l'evento

$$E_I = \{\omega \in \Omega : X(\omega) \in I\}$$

appartiene alla σ -algebra degli eventi misurabili (in simboli $E_I \in \mathcal{E}$).

La funzione che ad ogni intervallo I associa la probabilità $P(E_I)$ si dice *legge* della variabile aleatoria.

Ci interessano principalmente due funzioni ausiliarie di tipo standard (cioè funzioni reali di variabile reale) che descrivono le probabilità inerenti alla variabile aleatoria:

- La funzione reale $F_X : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$F_X(t) := P(E_t) = P(\{X \leq t\}) \quad \text{per ogni } t \in \mathbb{R}$$

si dice *funzione di distribuzione cumulativa* (o anche di *ripartizione*) della variabile X .

- La funzione reale $p_X : \mathbb{R} \rightarrow \mathbb{R}$ definita dalla legge

$$p_X(t) := P(\{X = t\}) \quad \text{per ogni } t \in \mathbb{R}$$

si dice *funzione di probabilità* (o anche di *densità discreta*) della variabile X .

Qual è la differenza tra una variabile aleatoria e le funzioni a cui siamo abituati? In “teoria” nessuna (una variabile aleatoria è una funzione), però

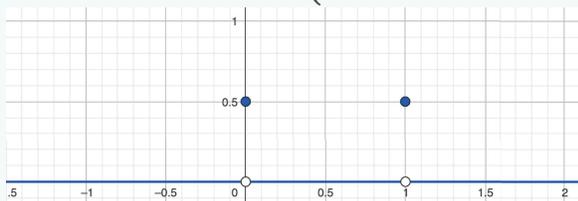
- in generale una variabile aleatoria assume valori reali, ma non è una funzione di variabile reale. Il suo dominio è uno spazio campione, come ad esempio le facce di una moneta o di un dado, o un complicato insieme di eventi macroeconomici se pensiamo al valore del dollaro.
- conoscere la legge di una variabile aleatoria non è soddisfacente. Nelle funzioni standard $f(x)$ la x ha il ruolo di *variabile indipendente* e rappresenta una grandezza che possiamo *scegliere* oppure *osservare*. Pensiamo ad esempio alla funzione che descrive la pressione. Se sono in un laboratorio e sto facendo un esperimento in cui ho tutti i parametri sotto controllo, faccio riferimento all'equazione di stato dei gas perfetti da cui ricavo la legge $P = (nRT)/V$ che descrive pressione P in funzione del numero di moli n , della temperatura T e del volume V (R è una costante che dipende dal gas). Se invece sono interessato alla pressione atmosferica che troverò tra una settimana in una località di villeggiatura, l'insieme di variabili che entrano in ballo non è controllabile né in numero né in rilevanza effettiva, quindi piuttosto che a una legge fisica (deterministica) mi affiderò

alle previsioni metereologiche che guardano la pressione come una variabile aleatoria e conoscono (dalla statistica) la probabilità con cui si distribuisce in un certo range di valori. Le funzioni che ci interessano sono quindi piuttosto la funzione di distribuzione e la funzione di probabilità, e queste sí **sono funzioni reali di variabile reale in senso standard**.

Esempio 1.1: Variabile aleatoria finita

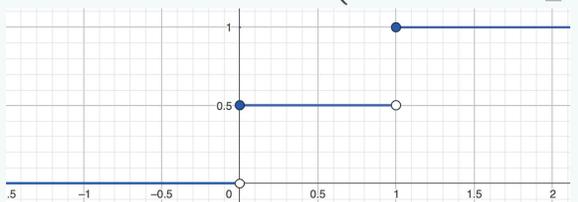
Nello spazio di misura classico associato al lancio di una moneta, definiamo la variabile aleatoria X che vale 1 se esce testa, 0 se esce croce.

La funzione di probabilità è $p_X(t) = \begin{cases} 1/2 & \text{se } t = 0, \\ 1/2 & \text{se } t = 1, \\ 0 & \text{altrimenti} \end{cases}$



e ha grafico

La funzione di distribuzione è $F_X(t) = \begin{cases} 0 & \text{se } t < 0, \\ 1/2 & \text{se } 0 \leq t < 1, \\ 1 & \text{se } t \geq 1 \end{cases}$



e ha grafico

Notiamo che

- la funzione di probabilità ha supporto finito, cioè è diversa da zero solo per un insieme finito di valori di t , precisamente i due valori (0 e 1) che può assumere la variabile aleatoria
- la somma dei valori non nulli di p_X dà 1
- la funzione di distribuzione non è continua: precisamente, ha un andamento a scalini crescenti, con discontinuità di salto in corrispondenza dei punti del supporto di p_X : l'altezza dei salti coincide con i corrispondenti valori di p_X .
- F_X coincide con 0 prima del primo punto del supporto di p_X e con 1 dopo l'ultimo punto.

Esempio 1.2: Variabile aleatoria discreta

Si consideri la variabile aleatoria X che conta le immatricolazioni al Corso di Laurea in Informatica. Per sua stessa definizione, la variabile X assume valori nell'insieme \mathbb{N} dei numeri naturali. Plausibili funzioni di probabilità e di distribuzione sono, rispettivamente,

In questo caso, il supporto di p_X non è finito, ma è comunque discreto perché sono ammessi solo i numeri naturali (e la serie dei valori corrispondenti deve dare 1). Di nuovo, la funzione di distribuzione ha andamento a scalini, con salti nei punti del supporto di p_X di altezza pari al corrispondente valore di p_X . Inoltre F_X vale 0 prima del primo valore ammissibile (che è 0), e ha $\lim_{t \rightarrow +\infty} F_X(t) = 1$.

Esempio 1.3: Variabile aleatoria continua

Si consideri la variabile aleatoria X che esprime il valore del dollaro (in euro). Questa variabile può assumere qualsiasi valore positivo. Una sua plausibile distribuzione è

Ora la funzione di distribuzione è continua, è ancora monotona crescente, vale 0 prima del primo valore ammissibile (che è 0), e ha $\lim_{t \rightarrow +\infty} F_X(t) = 1$.

In questo caso (lo dimostreremo poi rigorosamente) la funzione di probabilità non ci dà alcuna informazione utile, perché è costantemente zero.

I tre esempi appena elencati illustrano i vari tipi di variabili aleatorie che incontreremo: i primi due sono inerenti a **variabili aleatorie discrete** (*finita* nel primo caso, *infinita* nel secondo), mentre il terzo esempio si riferisce ad una **variabile aleatoria continua**. In seguito esamineremo distintamente queste due situazioni. Per ora, mettiamo in evidenza i tratti comuni di tutte le variabili aleatorie.

Dalla definizione di funzione di distribuzione discende che:

Proposizione 1.2: Proprietà assiomatiche della distribuzione

Sia X una variabile aleatoria e $F_X(t) = P(\{X \leq t\})$ la sua funzione di distribuzione. Allora

- i) La distribuzione assume valori compresi fra 0 e 1, cioè $0 \leq F_X(t) \leq 1$ per ogni $t \in \mathbb{R}$.
- ii) La distribuzione è monotona crescente, cioè $F_X(t) \leq F_X(s)$ se $t \leq s$.
- iii) La distribuzione ha limiti $\lim_{t \rightarrow -\infty} F_X(t) = 0$ e $\lim_{t \rightarrow +\infty} F_X(t) = 1$.
- iv) La distribuzione è continua da destra, cioè $\lim_{s \rightarrow t} F_X(s) = F_X(t)$ per ogni $t \in \mathbb{R}$.

Nota la funzione di distribuzione, è facile calcolare alcune probabilità inerenti alla variabile aleatoria:

Proposizione 1.3: Regole di calcolo con la distribuzione

Sia X una variabile aleatoria e $F_X(t) = P(\{X \leq t\})$ la sua funzione di distribuzione. Allora

- (1.1) $P(\{X > t\}) = 1 - F_X(t)$ per ogni $t \in \mathbb{R}$,
- (1.2) $P(\{s < X \leq t\}) = F_X(t) - F_X(s)$ per ogni $s < t$,
- (1.3) $p_X(t) = P(\{X = t\}) = F_X(t) - \lim_{s \rightarrow t^-} F_X(s)$ per ogni $t \in \mathbb{R}$.

Come conseguenza della proprietà (1.3) segue che

Corollario 1.4

Sia X una variabile aleatoria e $F_X(t) := P(\{X \leq t\})$ la sua funzione di distribuzione. Se F_X è una funzione continua, allora $p_X(t) = 0$ per ogni $t \in \mathbb{R}$.

Pertanto la funzione di probabilità si rivela utile solo nell'analisi delle variabili aleatorie discrete. Vediamo quali caratteristiche discendono dalla sua stessa definizione.

Proposizione 1.5: Proprietà assiomatiche della funzione di probabilità

Sia X una variabile aleatoria che assume valori in un insieme **discreto**, cioè

$$X \in \{x_1, x_2, \dots, x_N\} \quad (\text{caso finito}),$$

oppure

$$X \in \{x_1, x_2, \dots, x_n, \dots\} \quad \text{con} \quad \lim_{n \rightarrow \infty} x_n = +\infty \quad (\text{caso infinito}).$$

Sia poi $p_X = P(\{X = t\})$ la sua funzione di probabilità. Allora

- i) La funzione di probabilità è non negativa, cioè $p_X(t) \geq 0$ per ogni $t \in \mathbb{R}$.
- ii) La funzione di probabilità $p_X(t)$ è positiva solo se $t = x_n$, con $n = 1, \dots, N$ nel caso finito o $n \in \mathbb{N}$ nel caso infinito.
- iii) La funzione di probabilità ha discontinuità eliminabili nei punti $t = x_n$, per $n = 1, 2, \dots, N$ nel caso finito o $n \in \mathbb{N}$ nel caso infinito.
- iv) La somma dei valori assunti dalla funzione di probabilità è 1, cioè

$$\sum_{n=1}^N p_X(x_n) = 1 \quad (\text{caso finito}),$$

$$\sum_{n=1}^{\infty} p_X(x_n) = 1 \quad (\text{caso infinito}).$$

Nota la funzione di probabilità, è facile calcolare alcune probabilità inerenti alla variabile aleatoria:

Proposizione 1.6: Regole di calcolo con la funzione di distribuzione

Sia X una variabile aleatoria discreta e $p_X(t) = P(\{X = t\})$ la sua funzione di distribuzione. Allora

$$(1.4) \quad F(t) = P(\{X \leq t\}) = \sum_{x_n \leq t} p_X(x_n) \quad \text{per ogni } t \in \mathbb{R},$$

$$(1.5) \quad P(\{X > t\}) = \sum_{x_n > t} p_X(x_n) \quad \text{per ogni } t \in \mathbb{R},$$

$$(1.6) \quad P(\{s < X \leq t\}) = \sum_{s < x_n \leq t} p_X(x_n) \quad \text{per ogni } s < t,$$

$$(1.7) \quad P(\{s \leq X \leq t\}) = \sum_{s \leq x_n \leq t} p_X(x_n) \quad \text{per ogni } s < t.$$

Osservazione 1.7

La somme che compaiono in (1.5)–(1.7) hanno sempre un numero finito di addendi nel caso di variabili discrete finite. Se, invece, X è una variabile discreta infinita, le somme in (1.5) e (1.6) possono indicare delle serie con infiniti termini. Esse saranno sempre convergenti in virtù della proprietà iv) della Proposizione 1.5.

Nel caso di variabili aleatorie continue, è possibile definire un analogo della funzione di probabilità.

Definizione 1.8: Densità

Sia X una variabile aleatoria continua, e F_X la sua funzione di distribuzione. Se F_X è derivabile, definiamo **funzione di densità** di X la funzione reale $f_X : \mathbb{R} \rightarrow \mathbb{R}$ definita dalla legge

$$f_X(t) = \frac{d}{dt}F_X(t).$$

Dalle proprietà assiomatiche della funzione di distribuzione segue

Proposizione 1.9: Proprietà assiomatiche della densità

Sia X una variabile aleatoria continua che ammette densità $f_X(t)$. Allora

- i) La densità è non negativa, cioè $f_X(t) \geq 0$ per ogni $t \in \mathbb{R}$.
- ii) La densità è integrabile in senso improprio su \mathbb{R} nel senso che

$$\int_{-\infty}^{+\infty} f_X(t) dt \left(:= \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b f_X(t) dt \right) = 1.$$

La funzione di densità ci permette di calcolare alcune probabilità inerenti alla variabile aleatoria:

Proposizione 1.10: Regole di calcolo con la densità

Sia X una variabile aleatoria di densità $f_X(t)$. Allora

$$(1.8) \quad F_X(t) = P(\{X \leq t\}) = \int_{-\infty}^t f_X(r) dr \quad \text{per ogni } t \in \mathbb{R},$$

$$(1.9) \quad P(\{X > t\}) = \int_t^{+\infty} f_X(r) dr \quad \text{per ogni } t \in \mathbb{R},$$

$$(1.10) \quad P(\{s < X \leq t\}) = P(\{s \leq X \leq t\}) = \int_s^t f_X(r) dr \quad \text{per ogni } s < t.$$

2. Esempi notevoli di variabili aleatorie discrete

Passiamo ora in rivista alcune delle variabili aleatorie che più sono utilizzate nelle applicazioni. La più semplice di esse, la variabile Bernoulliana, altro non è che una generalizzazione della variabile introdotta nell'Esempio 1.1.

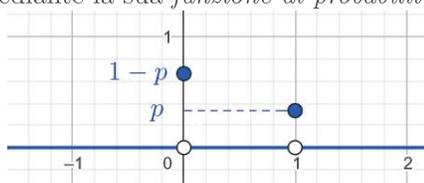
2.1. Variabile bernoulliana $\text{Bern}(p)$. Un esperimento aleatorio si dice *bernoulliano* se ha solo due possibili esiti: successo o insuccesso. Ad esempio, il lancio di una moneta costituisce un esperimento bernoulliano (in cui il successo può essere, per chiarire, l'uscita di testa). Anche il lancio di un dado a sei facce può essere visto come un esperimento bernoulliano se, ad esempio, interpretiamo come successo l'uscita del 6, e come insuccesso qualunque altro esito. Chiamiamo allora *parametro di un esperimento Bernoulliano* il numero p (compreso fra 0 e 1) che esprime la probabilità di successo. Il lancio di una moneta è dunque un esperimento bernoulliano di parametro $1/2$, mentre il lancio di un dado a sei facce dell'esempio precedente costituisce un esperimento bernoulliano di parametro $1/6$.

Dato un esperimento bernoulliano di parametro p , la **variabile bernoulliana** associata (che indicheremo nel seguito con $\text{Bern}(p)$) è la variabile che vale 1 se l'esito dell'esperimento è successo, 0 se è insuccesso.

Si tratta chiaramente di una variabile discreta che può prendere un numero finito di valori (0 o 1)

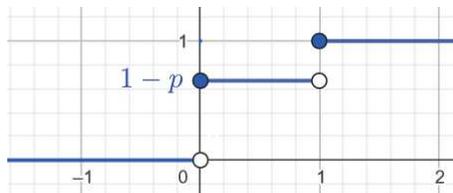
e che possiamo descrivere completamente mediante la sua *funzione di probabilità*:

$$p(t) = \begin{cases} p & \text{se } t = 1, \\ 1 - p & \text{se } t = 0, \\ 0 & \text{altrimenti} \end{cases}$$



La *distribuzione di probabilità* è

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ 1 - p & \text{se } 0 \leq t < 1, \\ 1 & \text{se } t \geq 1 \end{cases}$$



Esercizio 2.1

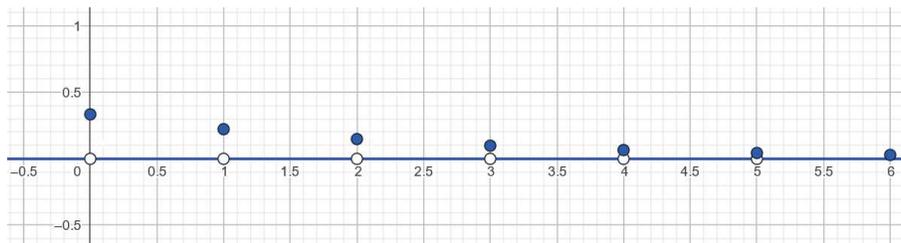
Ripetiamo 3 volte un esperimento bernoulliano di parametro $p = 2/5$, in modo che ogni nuovo esperimento sia indipendente dai risultati precedenti. Qual è la probabilità che si verifichino

- esattamente due insuccessi,
- due insuccessi nei primi due tentativi e poi un successo,
- esattamente 2 successi,
- almeno 2 successi,
- non più di 2 successi,

2.2. Variabile geometrica $\text{Geom}(p)$. Ripetiamo (in modo indipendente) un esperimento bernoulliano di parametro p fino a che non otteniamo un successo. La *variabile geometrica* conta il numero di insuccessi prima del primo successo; essa può dunque assumere il valore 0 (successo al primo tentativo), 1 (insuccesso al primo tentativo e successo al secondo), 2, e così via: si tratta cioè di una variabile discreta che può assumere infiniti valori.

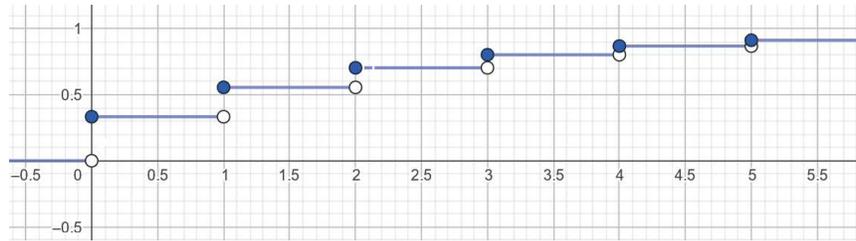
Generalizzando quanto visto nell'Esercizio 2.1, si calcola la sua *funzione di probabilità*

$$p(t) = \begin{cases} (1-p)^n p & \text{se } t = n \text{ è un intero} \\ 0 & \text{altrimenti,} \end{cases}$$



e di conseguenza la *distribuzione di probabilità*

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ p \sum_{k=0}^n (1-p)^k = 1 - (1-p)^{n+1} & \text{se } n \leq t < n+1 \end{cases}$$



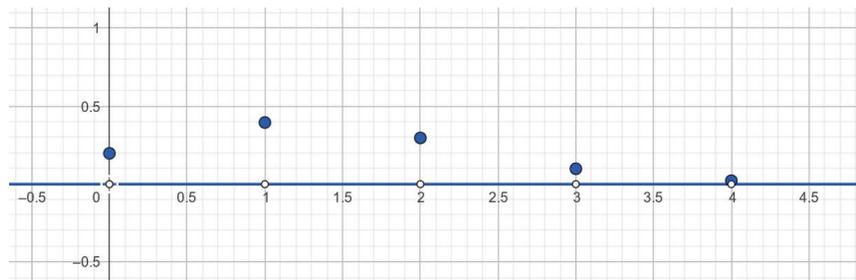
Esercizio 2.2

Verificare che la funzione $p(t)$ appena introdotta è effettivamente una funzione di probabilità, cioè che soddisfa le proprietà assiomatiche elencate nella Proposizione 1.5.

Preso poi un esperimento bernoulliano di parametro $p = 1/6$, calcolare la probabilità di avere successo entro il terzo tentativo (ovvero $P(\{\text{Geom}(1/6) \leq 3\})$) e la probabilità di avere successo fra il terzo e il quinto tentativo (ovvero $P(\{3 \leq \text{Geom}(1/6) \leq 5\})$).

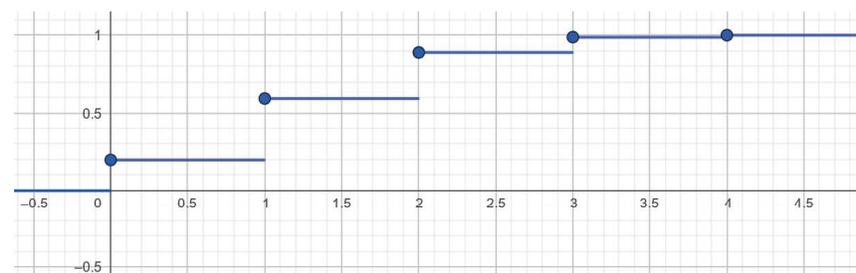
2.3. Variabile binomiale $\text{Bin}(n, p)$. Ripetiamo n volte un esperimento bernoulliano in modo indipendente e contiamo quanti successi abbiamo ottenuto. La **variabile binomiale $\text{Bin}(n, p)$** fa proprio questo, a titolo di esempio vale 2 se abbiamo avuto esattamente 2 successi e $n - 2$ insuccessi. Si tratta dunque di una variabile discreta che può assumere come valore ogni numero intero compreso fra 0 e n . Abbiamo già calcolato la sua funzione di probabilità: essa infatti è data dalla formula di Bernoulli (Teorema 2.8)

$$p(t) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{se } x = k \text{ con } k = 0, 1, \dots, n, \\ 0 & \text{altrimenti} \end{cases}$$



Da questa ricaviamo la *distribuzione di probabilità*

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} & \text{se } k \leq t < k+1, \text{ con } k = 0, 1, \dots, n-1 \\ 1 & \text{se } t \geq n. \end{cases}$$



Esercizio 2.3

Verificare che la funzione $p(t)$ appena introdotta è effettivamente una funzione di probabilità.

Esercizio 2.4: Prove ripetute

Ripetiamo un esperimento bernoulliano di parametro p in modo che ogni nuovo esperimento sia indipendente dai precedenti. Calcoliamo la probabilità degli eventi

A: almeno un successo su n prove,

B: esattamente k successi su n prove,

C: primo successo all' n -ma prova.

Svolgimento. È possibile svolgere questo esercizio anche senza conoscere le variabili aleatorie, usando solo le proprietà degli eventi indipendenti e il calcolo combinatorio. Tuttavia le variabili aleatorie ci aiutano perché rispondono in modo “automatico” a quesiti di questo tipo senza dover ricavare ogni volta una nuova formula, ma scegliendo la variabile aleatoria notevole che fa al caso nostro.

A: Possiamo calcolare la probabilità dell’evento A in almeno due modi diversi.

1° modo: Utilizziamo la variabile binomiale $\text{Bin}(n, p)$, che conta quanti successi si ottengono in n prove indipendenti. L’evento A si verifica quando $\text{Bin}(n, p)$ è strettamente maggiore di 0, dunque

$$P(A) = P\{\text{Bin}(n, p) > 0\} \stackrel{(1.1)}{=} 1 - F_{\text{Bin}(n, p)}(0) = 1 - \binom{n}{0} p^0 (1-p)^{n-0} = 1 - (1-p)^n.$$

2° modo: Può convenire calcolare la probabilità della negazione di A, cioè di $\neg A$: “nessun successo in n prove indipendenti” = “ n insuccessi in n prove indipendenti”. Poiché la probabilità di insuccesso è $1-p$, il numero di insuccessi è ancora regolato da una variabile binomiale, che avrà però parametro $1-p$ anziché p . Ora $\neg A$ si verifica quando $\text{Bin}(n, 1-p)$ è uguale a n , quindi dalla Definizione 1.1 sappiamo che

$$P(\neg A) = P\{\text{Bin}(n, 1-p) = n\} = p_{\text{Bin}(n, 1-p)}(n) = \binom{n}{n} (1-p)^n p^{n-n} = (1-p)^n.$$

Ricaviamo infine $P(A) = 1 - P(\neg A) = 1 - (1-p)^n$.

B: L’evento B si verifica quando $\text{Bin}(n, p)$ è uguale a k , dunque dalla Definizione 1.1 sappiamo che

$$P(B) = P\{\text{Bin}(n, p) = k\} = p_{\text{Bin}(n, p)}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

C: In questo caso ci serve la variabile geometrica, che conta gli insuccessi consecutivi prima del primo successo. L’evento C si verifica se $\text{Geom}(p) = n-1$, quindi

$$P(C) = P\{\text{Geom}(p) = n-1\} = p_{\text{Geom}(p)}(n-1) = (1-p)^{n-1} p.$$

□

Esercizio 2.5

Sapendo che il 15% della popolazione è mancina, calcolare la probabilità che

A: Bisogna fermare almeno 20 persone per trovare un mancino

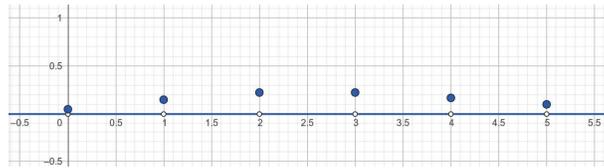
B: Bisogna fermare esattamente 20 persone per trovare un mancino

C: Bisogna fermare esattamente 20 persone per trovare 3 mancini

2.4. Variabile di Poisson $\mathcal{P}(\lambda)$. La [variabile di Poisson](#) è una variabile discreta infinita che assume valori nell’insieme \mathbb{N} dei numeri naturali; essa viene utilizzata per “contare” gli eventi rari. Se $\lambda > 0$ rappresenta il numero di eventi che si verificano mediamente in un intervallo unitario di tempo, la variabile di Poisson si definisce a partire dalla sua funzione di probabilità

$$p(t) = \begin{cases} e^{-\lambda} \frac{\lambda^n}{n!} & \text{se } t = n \in \mathbb{N}, \\ 0 & \text{altrimenti.} \end{cases}$$

il cui grafico è

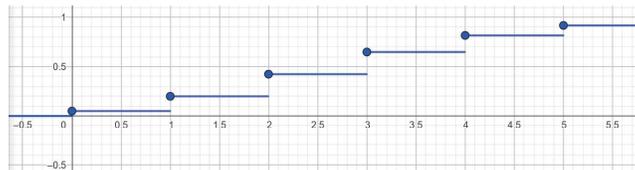


Si noti che p è effettivamente una funzione di probabilità, precisamente $e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = 1$, poichè la serie di potenze $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$ è la serie di Mc Laurin della funzione $f(\lambda) = e^\lambda$.

Si ricava poi la funzione di distribuzione

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!} & \text{se } n \leq t < n+1 \text{ con } n \in \mathbb{N} \end{cases}$$

il cui grafico è



3. Esempi notevoli di variabili aleatorie continue

Si vedano, per ora, le dispense “Variabili aleatorie elementari”

4. Distribuzioni congiunte e variabili indipendenti

Talvolta è utile considerare congiuntamente due variabili aleatorie. Per farlo, ci servono alcune funzioni ausiliarie.

Definizione 4.1: distribuzione, probabilità e densità congiunte

Se X e Y sono due variabili aleatorie, chiamiamo funzione di **distribuzione congiunta** di X e Y la funzione $F_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ definita mediante

$$F_{XY}(t, s) = P(\{X \leq t\} \wedge \{Y \leq s\}) \quad \text{per ogni } (t, s) \in \mathbb{R}^2.$$

Se X e Y sono variabili discrete, risulta utile introdurre la funzione di **probabilità congiunta**, $p_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ definita dalla legge

$$p_{XY}(t, s) = P(\{X = t\} \wedge \{Y = s\}) \quad \text{per ogni } (t, s) \in \mathbb{R}^2.$$

Se X e Y sono variabili continue, definiamo invece la funzione di **densità congiunta**, $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ che soddisfa la proprietà

$$P(\{a \leq X \leq b\} \wedge \{c \leq Y \leq d\}) = \int_a^b \int_c^d f_{XY}(t, s) dt ds \quad \text{per tutti i numeri reali } a < b \text{ e } c < d.$$

Se X e Y sono variabili discrete, diciamo per chiarire $X \in \{x_1, x_2, \dots, x_N\}$ e $Y \in \{y_1, y_2, \dots, y_M\}$, la probabilità congiunta p_{XY} è sempre nonnegativa, e positiva solo nei punti di (x_i, y_j) con $i = 1, \dots, N$ e $j = 1, \dots, M$, inoltre

$$\sum_{i=1}^N \sum_{j=1}^M p_{XY}(x_i, y_j) = 1.$$

La funzione di probabilità della sola variabile X può essere ricostruita mediante la formula

probabilità **marginale**:
$$p_X(x_i) = \sum_{j=1}^M p_{XY}(x_i, y_j),$$

poiché $p_X(x_i) = P(\{X = x_i\}) = P(\{X = x_i\} \wedge \{-\infty < Y < +\infty\}) = \sum_{j=1}^M p_{XY}(x_i, y_j)$. Formula analoga vale per la variabile Y .

Se X e Y sono variabili continue, la densità congiunta f_{XY} è sempre nonnegativa e

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(t, s) dt ds = 1.$$

La funzione di densità della sola variabile X può essere ricostruita mediante la formula

densità **marginale**:
$$f_X(t) = \int_{-\infty}^{+\infty} f_{XY}(t, s) ds.$$

Formula analoga vale per la variabile Y .

Esempio 4.1: lancio di un dado e di una moneta

Immaginiamo di lanciare un dado e una moneta. La variabile X legge la faccia del dado, la variabile Y prende il valore -1 se sulla moneta esce testa e $+1$ se esce croce. Possiamo considerare congiuntamente le variabili nello spazio campione $\Omega = \{(j, k) : j = 1, \dots, 6 \text{ e } k = \pm 1\}$, composto da 12 esiti equiprobabili. Dunque $p_{XY}(j, k) = 1/12$ per qualunque $j = 1, \dots, 6$ e $k = \pm 1$. Possiamo ricostruire la probabilità marginale di X e Y calcolando

$$p_X(j) = p_{XY}(j, +1) + p_{XY}(j, -1) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$$

$$p_Y(\pm 1) = \sum_{j=1}^6 p_{XY}(j, \pm 1) = \frac{6}{12} = \frac{1}{2}.$$

Abbiamo in precedenza parlato di *eventi* indipendenti. Vogliamo ora estendere alle variabili aleatorie la nozione di indipendenza. Intuitivamente, due variabili si diranno indipendenti se gli eventi che riguardano solo una variabile sono indipendenti dagli eventi che riguardano solo l'altra. Non è necessario prendere in esame tutti gli eventi, bensì basta considerare solo gli eventi del tipo $\{X \leq t\}$. Si ha dunque

Definizione 4.2: Variabili indipendenti

Due variabili aleatorie X e Y si dicono **indipendenti** se, e solo se, per ogni $t, s \in \mathbb{R}$ gli eventi $\{X \leq t\}$ e $\{Y \leq s\}$ sono eventi indipendenti. Ricordando (2.2), ciò è equivalente ad affermare

$$P(\{X \leq t\} \wedge \{Y \leq s\}) = P(\{X \leq t\}) \cdot P(\{Y \leq s\}).$$

È possibile caratterizzare l'indipendenza mediante le funzioni ausiliarie.

Proposizione 4.3: Condizioni necessarie e sufficienti d'indipendenza

Siano X e Y due variabili aleatorie e indichiamo con F_X e F_Y le rispettive distribuzioni. Le variabili X e Y sono indipendenti se, e solo se,

$$F_{XY}(t, s) = F_X(t) \cdot F_Y(s) \quad \text{per ogni } (t, s) \in \mathbb{R}^2.$$

Se poi X e Y sono due variabili discrete con funzione di probabilità rispettive p_X e p_Y , esse sono indipendenti se, e solo se,

$$p_{XY}(t, s) = p_X(t) \cdot p_Y(s) \quad \text{per ogni } (t, s) \in \mathbb{R}^2.$$

Se invece X e Y sono due variabili continue con densità rispettive f_X e f_Y , esse sono indipendenti se, e solo se,

$$f_{XY}(t, s) = f_X(t) \cdot f_Y(s) \quad \text{per ogni } (t, s) \in \mathbb{R}^2.$$

Esempio 4.2: lancio di un dado e di una moneta - continua

È evidente che le variabili aleatorie dell'esempio Esempio 4.1 sono indipendenti. Dalla definizione vediamo che, comunque scelti $j = 1, \dots, 6$ e $k \pm 1$, si ha

$$p_{XY}(j, k) = \frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2} = p_X(j)p_Y(k).$$

Esempio 4.3: incidenza di una malattia a seconda del genere

Su un campione di 20000 persone, 200 sono state riscontrate affette da una certa malattia. Dunque l'incidenza della malattia è $200/20000 = 0.1\%$. Per studiare l'effetto del genere sulla malattia, il campione è stato composto con 10000 uomini e 10000 donne. Dei 200 malati, 136 sono uomini e 64 donne. Notiamo dunque che c'è una correlazione tra la probabilità di sviluppare la malattia e il genere. Possiamo inquadrare la questione in termini di variabili aleatorie. Definiamo

- X vale 1 se si è malati e 0 se si è sani. Interpretando l'incidenza della malattia nel nostro campione come una probabilità, abbiamo $p_X(1) = 0.1\%$, $p_X(0) = 99.9\%$.
- Y vale 1 se si è uomini e 0 se si è donne. Interpretando di nuovo la frequenza nel nostro campione come probabilità abbiamo $p_Y(1) = p_Y(0) = 50\%$.

Guardando congiuntamente le variabili, dai nostri dati ricaviamo che

$$p_{XY}(1, 1) = \frac{\# \text{ uomini malati}}{\# \text{ totale}} = \frac{136}{20000} = 0.68\%, \quad p_{XY}(1, 0) = \frac{\# \text{ donne malate}}{\# \text{ totale}} = \frac{64}{20000} = 0.32\%,$$

$$p_{XY}(0, 1) = \frac{\# \text{ uomini sani}}{\# \text{ totale}} = \frac{9864}{20000} = 99.32\%, \quad p_{XY}(0, 0) = \frac{\# \text{ donne sane}}{\# \text{ totale}} = \frac{9936}{20000} = 99.68\%.$$

Queste variabili **non** sono indipendenti, infatti $p_X(1) \cdot p_Y(0) = 1\% \cdot 50\% = \frac{50}{10000} = 0.5\%$ è diverso da $p_{XY}(1, 0) = 0.32\%$.

5. Valore atteso, varianza e scarto quadratico medio

Spendiamo due parole sul concetto di **media** in *statistica*. Se vogliamo conoscere il numero medio di figli in una famiglia italiana, il buon senso ci suggerisce di fare il rapporto tra il numero di figli (n) e il numero di famiglie (N). In *statistica*, si prende un campione di N famiglie, e le si suddivide in base al numero di figli. Se chiamiamo N_0, N_1, N_2, \dots il numero di famiglie con 0, 1, 2, ... figli, allora il numero totale di famiglie è $N = N_0 + N_1 + N_2 + \dots$, mentre il numero totale di figli è $n = 0 \cdot N_0 + 1 \cdot N_1 + 2 \cdot N_2 + \dots$. Il numero medio di figli è dato allora da

$$\bar{n} = \frac{n}{N} = \frac{0 \cdot n_0 + 1 \cdot n_1 + 2 \cdot n_2 + \dots}{N} = 0 \cdot \frac{n_0}{N} + 1 \cdot \frac{n_1}{N} + 2 \cdot \frac{n_2}{N} + \dots.$$

Nell'ultima espressione, le quantità $\frac{N_0}{N}, \frac{N_1}{N}, \dots$ rappresentano la *frequenza* di famiglie con $0, 1, \dots$ figli. Se poi immaginiamo di avere un campione infinito (cioè estrapoliamo il limite per $N \rightarrow \infty$), tali frequenze approssimano la probabilità dei rispettivi eventi, secondo la concezione frequentista.

Possiamo allora riformulare il discorso in termini *probabilistici*. Consideriamo la variabile aleatoria X che esprime il numero di figli in una famiglia italiana scelta a caso; si tratta di una variabile discreta che prende valori in \mathbb{N} . La sua funzione di probabilità p_X è plausibilmente espressa dal limite delle frequenze

$$p_X(i) = \lim_{N \rightarrow +\infty} \frac{n_i}{N} \quad \text{per } i = 0, 1, \dots$$

Il corrispondente limite della media statistica viene interpretato come **valor medio** della variabile X ; scriviamo in formule

$$E(X) = \lim_{N \rightarrow +\infty} \bar{n} = \lim_{N \rightarrow +\infty} \left(0 \cdot \frac{n_0}{N} + 1 \cdot \frac{n_1}{N} + 2 \cdot \frac{n_2}{N} + \dots \right) = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + \dots$$

Introduciamo ora una definizione rigorosa, alla luce di questo esempio.

Definizione 5.1: Valore atteso o medio

Se X è una variabile aleatoria discreta finita, $X \in \{x_1, x_2, \dots, x_N\}$, e p_X la sua funzione di probabilità, si definisce **valor medio** o **valore atteso** o **speranza matematica** di X il numero reale

$$E(X) := x_1 p_X(x_1) + x_2 p_X(x_2) + \dots + x_N p_X(x_N).$$

Se X una variabile aleatoria discreta infinita, $X \in \{x_1, x_2, \dots, x_n, \dots\}$, e p_X la sua funzione di probabilità, la somma precedente diventa una serie. Precisamente, quando la serie di termine generico $x_n p_X(x_n)$ è assolutamente convergente, il valore di tale serie è detto **valor medio** o **valore atteso** o **speranza matematica** di X , in formule

$$E(X) := \sum_{n=1}^{\infty} x_n p_X(x_n).$$

Infine, se X è una variabile aleatoria continua e f_X la sua funzione di densità, la somma diventa un integrale. Precisamente, quando la funzione $t f_X(t)$ è assolutamente integrabile in senso improprio su $(-\infty, +\infty)$, il valore di tale integrale è detto **valor medio** o **valore atteso** o **speranza matematica** di X , in formule

$$E(X) := \int_{-\infty}^{+\infty} t f_X(t) dt.$$

Esempio 5.1

Sia X la variabile aleatoria che esprime l'esito del lancio di un dado perfetto a 6 facce, che prende i valori $\{1, 2, \dots, 6\}$ e ha funzione di probabilità

$$p_X(t) = \begin{cases} 1/6 & \text{se } t = 1, 2, \dots, 6 \\ 0 & \text{altrimenti.} \end{cases}$$

Il suo valor medio è

$$E(X) = 1 p_X(1) + 2 p_X(2) + \dots + 6 p_X(6) = \frac{1 + 2 + \dots + 6}{6} = \frac{21}{6} = \frac{7}{2}.$$

Si noti che il valor medio non rientra tra i valori effettivamente assunti dalla variabile aleatoria.

Esempio 5.2

Calcoliamo il valore atteso della variabile bernoulliana introdotta nel paragrafo 2.1. Si ha

$$E(\text{Bern}(p)) = (1-p)0 + p1 = p.$$

Esempio 5.3

Calcoliamo il valore atteso della variabile geometrica $\text{Geom}(p)$ definita nel paragrafo 2.2. Per definizione

$$E(\text{Geom}(p)) = \sum_{n=0}^{\infty} n(1-p)^n p = p \sum_{n=0}^{\infty} n(1-p)^n = p \sum_{n=1}^{\infty} n(1-p)^n.$$

Ma per ogni n fissato si ha $n = (n-1) + 1$ e dunque

$$\sum_{n=1}^{\infty} n(1-p)^n = \sum_{n=1}^{\infty} ((n-1) + 1)(1-p)^n = \sum_{n=1}^{\infty} (n-1)(1-p)^n + \sum_{n=1}^{\infty} (1-p)^n.$$

Calcoliamo separatamente queste due serie. Ponendo $m = n-1$ si ottiene

$$\sum_{n=1}^{\infty} (n-1)(1-p)^n = \sum_{m=0}^{\infty} m(1-p)^{m+1} = (1-p) \sum_{m=0}^{\infty} m(1-p)^m = \frac{1-p}{p} E(\text{Geom}(p)).$$

D'altra parte

$$\sum_{n=1}^{\infty} (1-p)^n = \sum_{n=0}^{\infty} (1-p)^n - 1 = \frac{1}{1-(1-p)} - 1 = \frac{1}{p} - 1 = \frac{1-p}{p},$$

ricordando la serie geometrica di ragione $1-p$. Ricapitolando

$$E(\text{Geom}(p)) = p \sum_{n=1}^{\infty} n(1-p)^n = p \left(\frac{1-p}{p} E(\text{Geom}(p)) + \frac{1-p}{p} \right) = (1-p) E(\text{Geom}(p)) + 1-p.$$

Esaminiamo l'uguaglianza fra primo e ultimo membro e ricaviamo

$$(1-1+p) E(\text{Geom}(p)) = 1-p,$$

da cui $E(\text{Geom}(p)) = (1-p)/p$.

Il seguente risultato afferma che il valore atteso è una grandezza lineare.

Proposizione 5.2: Linearità del valore atteso

Siano X e Y due variabili aleatorie e $a, b, c \in \mathbb{R}$. Allora

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

Osserviamo esplicitamente che, come è facile verificare dalla definizione, il valore atteso di una costante è pari alla costante.

Esempio 5.4

Calcoliamo il valor medio della variabile binomiale introdotta nel paragrafo 2.3. Tale variabile altro non è che la somma di n variabili bernoulliane di parametro p (una per ogni ripetizione della prova), dunque

$$E(\text{Bin}(n, p)) = n E(\text{Bern}(p)) = np.$$

Esempio 5.5

I valori attesi delle principali variabili aleatorie continue sono calcolati nelle dispense “Variabili aleatorie elementari”.

Il valore atteso è lineare rispetto alla *somma* di variabili aleatorie, ma non rispetto al *prodotto*: non è vero che, in generale, $E(XY) = E(X)E(Y)$. L'uguaglianza vale solo per le variabili indipendenti, come afferma il teorema seguente.

Teorema 5.3: Valore atteso del prodotto di variabili indipendenti

Se X e Y sono due variabili indipendenti, allora

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

Dimostrazione. Faremo la dimostrazione nel caso in cui X e Y sono variabili discrete, ovvero $X \in \{x_1, \dots, x_n\}$ e $Y \in \{y_1, \dots, y_m\}$, e indicheremo con p_X e p_Y le rispettive funzioni di probabilità. La variabile XY è ancora una variabile discreta che può assumere i valori $\{x_i y_j : i = 1, \dots, n \text{ e } j = 1, \dots, m\}$ e per definizione

$$(5.1) \quad E(XY) = \sum_{t \in \mathbb{R}} t q(t),$$

dove q designa la sua funzione di probabilità $q(t) = P\{XY = t\}$ ¹. Se $t = 0$, $q(0)$ non dà alcun contributo al calcolo del valore atteso secondo la formula (5.1). Se, invece, $t \neq 0$, possiamo esprimere $q(t)$ in termini delle due funzioni p_X e p_Y . Cominciamo con l'osservare che l'evento $\{XY = t\}$ si può decomporre nella somma logica di eventi mutuamente esclusivi

$$\{XY = t\} = \bigcup_{s \in \mathbb{R} \setminus \{0\}} \{X = s \text{ e } Y = t/s\},$$

pertanto

$$(5.2) \quad q(t) = P\{XY = t\} = \sum_{s \in \mathbb{R} \setminus \{0\}} P\{X = s \text{ e } Y = t/s\}.$$

D'altronde $\{X = s \text{ e } Y = t/s\} = \{X = s\} \wedge \{Y = t/s\}$ e, poiché le variabili X e Y sono indipendenti, si ha

$$P(\{X = s\} \wedge \{Y = t/s\}) = P\{X = s\} P\{Y = t/s\} = p_X(s) p_Y(t/s).$$

Sostituendo questa uguaglianza nella formula (5.2) si ottiene

$$q(t) = \sum_{s \in \mathbb{R} \setminus \{0\}} p_X(s) p_Y\left(\frac{t}{s}\right),$$

e quindi, tornando alla formula (5.1),

$$E(XY) = \sum_{t \in \mathbb{R} \setminus \{0\}} \sum_{s \in \mathbb{R} \setminus \{0\}} t p_X(s) p_Y\left(\frac{t}{s}\right).$$

¹si osservi che la somma scritta è finita, poiché $q(t) \neq 0$ solo per un numero finito di valori di t

Ora, moltiplicando e dividendo per s si ottiene

$$\begin{aligned} E(XY) &= \sum_{t \in \mathbb{R} \setminus \{0\}} \sum_{s \in \mathbb{R} \setminus \{0\}} s p_X(s) \frac{t}{s} p_Y\left(\frac{t}{s}\right) \\ &= \sum_{s \in \mathbb{R} \setminus \{0\}} s p_X(s) \sum_{t \in \mathbb{R} \setminus \{0\}} \frac{t}{s} p_Y\left(\frac{t}{s}\right) \\ &= \sum_{s \in \mathbb{R} \setminus \{0\}} s p_X(s) \sum_{r \in \mathbb{R} \setminus \{0\}} r p_Y(r) = E(X) E(Y). \end{aligned}$$

□

Come abbiamo già sottolineato, il valore atteso ci dà un'idea di qual è il valore *medio* intorno a cui gli effettivi valori assunti dalla variabile aleatoria si distribuiscono. Non è in grado di dirci, però, quanto la variabile aleatoria può discostarsi da questo valore. Consideriamo due variabili aleatorie X e Y definite come segue:

X vale sempre zero. Quindi $p_X(0) = 1$, $p_X(t) = 0$ se $t \neq 0$ e $E(X) = 0 \cdot 1 = 0$

Y può prendere come valore qualunque numero intero $k \in \mathbb{Z}$, $k \neq 0$, con $p_Y(k) = c/k^4$, dove $c = \left(2 \sum_{k=1}^{\infty} \frac{1}{k^4}\right)^{-1}$. Quindi

$$E(Y) = \sum_{k \in \mathbb{Z} \setminus \{0\}} k \cdot \frac{c}{k^4} = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{c}{k^3} = \sum_{k=1}^{\infty} \frac{c}{k^3} + \sum_{k=1}^{\infty} \frac{c}{(-k)^3} = \sum_{k=1}^{\infty} \frac{c}{k^3} - \sum_{k=1}^{\infty} \frac{c}{k^3} = 0.$$

Anche se Y ha valore atteso pari a zero, è impossibile che prenda il valore zero, e c'è invece probabilità non nulla che assuma valori arbitrariamente grandi.

Per meglio descrivere una variabile aleatoria, quindi, serve anche sapere “quanto” può discostarsi dal suo valore atteso. La grandezza che misura questo scostamento è la *varianza* viene definito come segue.

Definizione 5.4: Varianza

Sia X una variabile aleatoria con valor medio finito e poniamo $\mu = E(X) \in \mathbb{R}$. Si definisce *varianza* di X il numero reale nonnegativo

$$\sigma^2(X) = E((X - \mu)^2).$$

La radice quadrata della varianza è spesso indicata come *deviazione standard* o *scarto quadratico medio*.

Con un po' di lavoro si può ricavare un'utile formula di calcolo

Proposizione 5.5: Formula per il calcolo della varianza

Sia X una variabile aleatoria di valore atteso finito; indichiamo con $\mu = E(X)$ il suo valore atteso. Allora

$$\sigma^2(X) = \begin{cases} \sum_{n=1}^N (x_n - \mu)^2 p_X(x_n) & \text{nel caso discreto finito,} \\ \sum_{n=1}^{\infty} (x_n - \mu)^2 p_X(x_n) & \text{nel caso discreto infinito,} \\ \int_{-\infty}^{+\infty} (t - \mu)^2 f_X(t) dt & \text{nel caso continuo.} \end{cases}$$

Se ora torniamo alle variabili X e Y di prima, vediamo che

$$\sigma^2(X) = (0 - 0) \cdot 1 = 0,$$

$$\sigma^2(Y) = \sum_{k \in \mathbb{Z} \setminus \{0\}} (k - 0)^2 \cdot \frac{c}{k^4} = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{c}{k^2} = 2c \sum_{k=1}^{\infty} \frac{1}{k^2} > 0$$

Esempio 5.6

Calcoliamo la varianza della variabile bernoulliana $\text{Bern}(p)$ definita nel paragrafo 2.1. Sappiamo che $E(\text{Bern}(p)) = p$, poi

$$\sigma^2(\text{Bern}(p)) = (0 - p)^2(1 - p) + (1 - p)^2p = p(1 - p).$$

Per chiarirci il significato della varianza vediamo una nota disuguaglianza.

Proposizione 5.6: Disuguaglianza di Chebychev

Se X è una variabile aleatoria di valore atteso μ e varianza σ^2 , allora

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

per ogni $\epsilon > 0$.

Dimostrazione. Consideriamo solo il caso in cui X è una variabile continua che ammette densità $f(t)$. Poiché $\{|X - \mu| \geq \epsilon\} = \{X \leq \mu - \epsilon\} \vee \{X \geq \mu + \epsilon\}$ si ha

$$(5.3) \quad P(|X - \mu| \geq \epsilon) = \int_{-\infty}^{\mu - \epsilon} f(t) dt + \int_{\mu + \epsilon}^{+\infty} f(t) dt.$$

D'altra parte

$$\begin{aligned} \frac{\sigma^2}{\epsilon^2} &= \int_{-\infty}^{+\infty} \frac{(t - \mu)^2}{\epsilon^2} f(t) dt \\ &= \int_{-\infty}^{\mu - \epsilon} \frac{(t - \mu)^2}{\epsilon^2} f(t) dt + \int_{\mu - \epsilon}^{\mu + \epsilon} \frac{(t - \mu)^2}{\epsilon^2} f(t) dt + \int_{\mu + \epsilon}^{+\infty} \frac{(t - \mu)^2}{\epsilon^2} f(t) dt. \end{aligned}$$

Ora, il secondo integrale è maggiore o uguale di zero perché tale è l'integrando. Inoltre, nel primo e nel secondo intervallo di integrazione si ha $|t - \mu| \geq \epsilon$ e dunque $(t - \mu)^2/\epsilon^2 \geq 1$. Ne segue che

$$\frac{\sigma^2}{\epsilon^2} \geq \int_{-\infty}^{\mu - \epsilon} f(t) dt + \int_{\mu + \epsilon}^{+\infty} f(t) dt = P(|X - \mu| \geq \epsilon)$$

per la (5.3). □

Elenchiamo alcune utili proprietà della varianza

Proposizione 5.7

Se X è una variabile aleatoria di valor medio finito e $a, b \in \mathbb{R}$, allora

$$(5.4) \quad \sigma^2(X) = E(X)^2 - (E(X))^2,$$

$$(5.5) \quad \sigma^2(X) = 0 \quad \text{se, e solo se, } X \text{ è costante (con probabilità 1),}$$

$$(5.6) \quad \sigma^2(aX + b) = a^2 \sigma^2(X).$$

Dimostrazione. Per verificare la proprietà (5.4), poniamo $\mu = E(X)$ e calcoliamo

$$\sigma^2(X) = E((X - \mu)^2) = E(X^2 - 2\mu X + \mu^2)$$

Applicando la Proposizione 5.2 si ha

$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

□

Definizione 5.8: Variabile standardizzata

Se X è una variabile casuale di valor medio μ e deviazione σ , chiamiamo **variabile standardizzata** associata a X la variabile casuale

$$X^* = \frac{X - \mu}{\sigma}.$$

Utilizzando le proprietà elencate nelle Proposizione 5.2 e Proposizione 5.7 si verifica facilmente che

$$E(X^*) = 0, \quad \sigma^2(X^*) = 1.$$

A titolo di esempio, vediamo come è fatta la bernoulliana standardizzata. Poiché sappiamo che $E(\text{Bern}(p)) = p$ e $\sigma^2(\text{Bern}(p)) = p(1-p)$, la standardizzata sarà $\frac{1}{\sqrt{p(1-p)}}(\text{Bern}(p) - p)$, cioè la variabile che può prendere il valore $\sqrt{\frac{1-p}{p}}$ con probabilità p o il valore $-\sqrt{\frac{p}{1-p}}$ con probabilità $1-p$.

In generale, la varianza non è lineare. Tuttavia, la varianza della somma di due variabili indipendenti è la somma delle varianze.

Teorema 5.9: Varianza della somma di variabili indipendenti

Se X e Y sono due variabili indipendenti, allora

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y).$$

Dimostrazione. Per la formula si ha

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

dopo aver utilizzato la linearità del valore atteso Proposizione 5.2. Raccogliendo abbiamo

$$\begin{aligned} \sigma^2(X + Y) &= \left(E(X^2) - (E(X))^2\right) + \left(E(Y^2) - (E(Y))^2\right) + 2(E(XY) - E(X)E(Y)) \\ &= \sigma^2(X) + \sigma^2(Y). \end{aligned}$$

Nell'ultima uguaglianza abbiamo usato di nuovo la formula (5) per il calcolo di $\sigma^2(X)$ e $\sigma^2(Y)$; inoltre abbiamo utilizzato che $E(XY) = E(X)E(Y)$ per il Teorema 5.3. □

Esempio 5.7

Calcoliamo la varianza della variabile binomiale $\text{Bin}(n, p)$ introdotta nell'Esempio 2.3. Poiché essa è la somma di n variabili bernoulliane di parametro p mutuamente indipendenti, segue dal Teorema 5.9 che

$$\sigma^2(\text{Bin}(n, p)) = n\sigma^2(\text{Bern}(p)) = np(1-p).$$

6. Convergenza di variabili aleatorie: la legge dei grandi numeri e il teorema del limite centrale

Secondo la nozione frequentista di probabilità, ripetendo un gran numero di volte un esperimento, la frequenza con cui un dato evento A si verifica è “all’incirca” uguale alla sua probabilità. Questa affermazione è nota come “legge empirica dei grandi numeri”. Andiamo ora ad enunciare un teorema che formalizza questa legge empirica.

Dato un qualunque evento A , possiamo definire una variabile che vale 1 se l’evento si verifica, 0 altrimenti. È questa una variabile di tipo bernoulliano di parametro $p = P(A)$. Immaginiamo ora di ripetere il nostro esperimento n volte in modo indipendente e di contare quante volte l’evento si verifica. Otteniamo così una variabile binomiale che indicheremo in seguito con

$$X_n = \text{Bin}(n, p).$$

La frequenza (statistica) con cui l’evento A si verifica è ora data da

$$\text{fq}_n(A) = \frac{X_n}{n}.$$

La legge empirica dei grandi numeri afferma dunque che

$$\lim_{n \rightarrow \infty} \frac{X_n}{n} = P(A).$$

Ma cosa significa questa scrittura? Come va inteso il limite di una successione di variabili aleatorie? Intuitivamente, possiamo dire che una successione di variabili X_n tende a X se, per grandi valori di n , siamo pressoché certi che X_n e X abbiano valori vicini (o, viceversa, è pressoché impossibile che abbiano valori lontani). Formalmente, si ha la seguente definizione

Definizione 6.1: Convergenza in probabilità

Una successione di variabili casuali $(X_n)_{n \in \mathbb{N}}$ **converge in probabilità** alla variabile X se per ogni $\epsilon > 0$ si ha

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1,$$

o equivalentemente

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0.$$

In simboli scriveremo $X_n \xrightarrow{P} X$.

Siamo ora pronti ad enunciare

Teorema 6.2: Legge dei grandi numeri

Sia A un evento. Se X_n è il numero di volte che l’evento A si verifica in n prove indipendenti, allora

$$\frac{X_n}{n} \xrightarrow{P} P(A).$$

Dimostrazione. Detta $p = P(A)$, dobbiamo dimostrare che per ogni $\epsilon > 0$ si ha

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_n}{n} - p\right| \geq \epsilon\right) = 0.$$

Poiché, come abbiamo osservato in precedenza, X_n è una variabile di tipo binomiale di parametri n e p , si ha che

$$\begin{aligned}\mu &= E\left(\frac{X_n}{n}\right) = \frac{E(X_n)}{n} = \frac{np}{n} = p, \\ \sigma^2 &= \sigma^2\left(\frac{X_n}{n}\right) = \frac{\sigma^2(X_n)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.\end{aligned}$$

Applichiamo ora la disuguaglianza di Chebichev alla variabile X_n/n , si ottiene

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

D'altra parte per costruzione $P(|X_n/n - p| \geq \epsilon) \geq 0$. Segue allora dal teorema dei due carabinieri che

$$0 \leq \lim_{n \rightarrow \infty} P\left(\left|\frac{X_n}{n} - p\right| \geq \epsilon\right) \leq \lim_{n \rightarrow \infty} \frac{p(1-p)}{n\epsilon^2} = 0,$$

cioè appunto la nostra tesi. \square

Passiamo ora a trattare un altro teorema fondamentale, noto come “teorema del limite centrale”. Si tratta anche in questo caso di un teorema di approssimazione, ma si fa ricorso ad un altro concetto di limite.

Definizione 6.3: Convergenza in legge

Una successione di variabili casuali $(X_n)_{n \in \mathbb{N}}$ **converge in legge** alla variabile X se convergono puntualmente le rispettive funzioni di distribuzione, precisamente se

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t)$$

per ogni t in cui F_X è continua. In simboli scriveremo $X_n \xrightarrow{L} X$.

Teorema 6.4: Teorema del limite centrale

Sia X_n una successione di variabili casuali mutuamente indipendenti, tutte con la stessa funzione di distribuzione F . In particolare, esse hanno uguale valor medio $E(X_n) = \mu$ e ugual varianza $\sigma^2(X_n) = \sigma^2$ per ogni $n \in \mathbb{N}$. Definiamo poi la variabile

$$Z_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}}.$$

Allora Z_n converge in legge alla variabile normale standard $N(0, 1)$, ovvero

$$Z_n \xrightarrow{L} N(0, 1).$$

Consideriamo la somma parziale delle X_n , cioè $\sum_{k=1}^n X_k$; si osservi che, poichè le X_n sono indipendenti, si ha

$$E\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n E(X_k) = n\mu, \quad \sigma^2\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \sigma^2(X_k) = n\sigma^2.$$

Pertanto Z_n è la variabile standardizzata di tale somma parziale. Se le X_n sono tutte variabili normali distribuite secondo $N(\mu, \sigma^2)$, allora Z_n è la variabile normale standard per ogni $n \in \mathbb{N}$. Il teorema del limite centrale afferma che, in generale, anche se le X_n non sono gaussiane, la loro somma standardizzata approssima la variabile normale standard.

Dal punto di vista applicativo, se una variabile casuale può essere vista come somma di molte variabili casuali equidistribuite (cioè con la stessa funzione di distribuzione), allora la sua distribuzione è “quasi” normale. Ecco perché l’uso della distribuzione normale è molto frequente nei problemi di tipo statistico.

Esempio 6.1: Approssimazione normale della distribuzione binomiale

Abbiamo osservato in precedenza che la variabile binomiale $\text{Bin}(n, p)$ è la somma di n variabili bernoulliane $\text{Bern}(p)$ indipendenti fra loro: indichiamole con X_1, X_2, \dots, X_n (dove l’indice rappresenta la prova a cui si riferiscono). Sappiamo che $\mu = E(X_k) = p$, $\sigma = \sigma(X_k) = \sqrt{p(1-p)}$, $E(\text{Bin}(n, p)) = np$, $\sigma(\text{Bin}(n, p)) = \sqrt{np(1-p)}$. Dunque la variabile standardizzata della binomiale è

$$Z_n = \frac{\text{Bin}(n, p) - np}{\sqrt{np(1-p)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}}$$

e il Teorema del Limite Centrale afferma che, per n grande, essa può essere approssimata dalla variabile gaussiana standard.

In modo equivalente, per n grande possiamo approssimare la funzione della binomiale $\text{Bin}(n, p)$ con la distribuzione normale $N(np, \sqrt{np(1-p)})$.