

1 Caratteristiche numeriche delle v.a. discrete

Dalla definizione della speranza matematica seguono alcune semplici ma utili proprietà.

Abbiamo già visto che la speranza matematica è *lineare*

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

Un'altra proprietà (che non richiede dimostrazione) è la *positività*.

2) Se $X \geq 0$, allora $E(X) \geq 0$. Inoltre se $E(X) = 0$, allora $X = 0$.

Dalla positività segue immediatamente la *monotonia*, cioè se $X \geq Y$, allora

3) $E(X) \geq E(Y)$. Infatti $X - Y \geq 0$ e dalla linearità segue la tesi.

Dalla doppia disuguaglianza $-|X| \leq X \leq |X|$ segue immediatamente che $-E(|X|) \leq E(X) \leq E(|X|)$ e quindi

4) $|E(X)| \leq E(|X|)$.

5) Se $|X| \leq Y$ e se Y ammette speranza matematica, allora anche X ammette speranza matematica. Questa proprietà dipende dal criterio della serie maggiorante.

Inoltre se la v.a. è costante, cioè se $X = C$, allora $E(X) = C$.

Proprietà della varianza delle v.a. discrete

Ricordiamo che se X ammette speranza matematica, la varianza di X è definita come

$$\text{Var}X = E((X - E(X))^2),$$

sempre che la speranza matematica del quadrato esista.

Sviluppando il quadrato troviamo $E(X^2 - 2XE(X) + E(X)^2)$ e poiché $E(X)$ esiste per ipotesi (e la v.a. costante ammette speranza matematica), la varianza si può definire per quelle v.a. per le quali esiste $E(X^2)$.

Usando la linearità, lo sviluppo formale dà $E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2$. Quindi, se la X ammette varianza, si ha che

$$\text{Var}X = E(X^2) - E(X)^2.$$

Questa si chiama *identità di König*.

Esempio

Sappiamo che la serie $\sum_{n=1}^{\infty} \frac{1}{n^3}$ converge. La sua somma $\alpha \approx 1,202$, detta costante di Apéry, non è nota con esattezza (nello stesso senso in cui π non è noto con esattezza). Nel 1979 il matematico greco-francese Roger Apéry ha dimostrato che α non è razionale.

Non è noto se sia trascendente.

La successione $p_n = \frac{1}{\alpha n^3}$ è la densità di una v.a. numerabile. La speranza matematica di questa v.a. esiste perché la serie

$$\sum_{n=1}^{\infty} \frac{n}{\alpha n^3}$$

converge, ma $E(X^2)$ “sarebbe”

$$\sum_{n=1}^{\infty} \frac{n^2}{\alpha n^3},$$

che diverge.

Dunque l'esistenza della speranza matematica non è sufficiente per assicurare l'esistenza della varianza.

Data un n -upla di v.a. *discrete* X_1, X_2, \dots, X_n esse si dicono indipendenti, se per ogni scelta di n sottoinsiemi della retta reale A_i , $1 \leq i \leq n$ si ha che

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i).$$

In particolare, se scegliamo degli A_i costituiti da un solo punto, questa identità prende la forma di

$$P(X_1 = t_1, X_2 = t_2, \dots, X_n = t_n) = \prod_{i=1}^n P(X_i = t_i),$$

ovvero

$$p(t_1, t_2, \dots, t_n) = \prod_{i=1}^n p_i(t_i)$$

ovvero, in parole, la densità congiunta è il prodotto delle densità marginali.

Le ultime due identità sono soltanto due modi equivalenti di scrivere la stessa cosa, ma non è complicato dimostrare che in effetti tutte e tre le condizioni sono equivalenti e che possiamo, volta per volta, usare la più conveniente.

Assumiamo infatti che l' n -upla X_1, X_2, \dots, X_n verifichi l'ultima identità e che $A_1 = \{t_1^1, t_1^2, \dots, t_1^k\}$. Si ha allora che, per ogni j tra 1 e k

$$p_1(t_1^j) \prod_{i=2}^n p_i(t_i) = p(t_1^j, t_2, \dots, t_n)$$

e, sommando rispetto a j si ottiene

$$\sum_{j=1}^k p_1(t_1^j) \prod_{i=2}^n p_i(t_i) = \sum_{j=1}^k p(t_1^j, t_2, \dots, t_n) =$$

$$P(X_1 \in A_1, X_2 = t_2, \dots, X_n = t_n).$$

Ripetendo un analogo calcolo con $A_2 = \{t_2^1, t_2^2, \dots, t_2^k\}$ e iterando il procedimento fino ad $A_n = \{t_n^1, t_n^2, \dots, t_n^k\}$, si dimostra la prima condizione (che sembra più generale) dall'ultima.

Questa definizione include l'indipendenza tra due o più eventi, concetto che conosciamo già.

Illustriamo questo fatto nel caso semplice di due eventi A e B .

Ai due eventi associamo le v.a. indicatrici $\mathbb{1}_A$ e $\mathbb{1}_B$. La funzione indicatrice di A si definisce come

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A \\ 0, & \text{se } \omega \notin A. \end{cases}$$

Consideriamo due arbitrari sottoinsiemi della retta reale, che indicheremo con E ed F .

Vogliamo valutare $P(\mathbb{1}_A \in E, \mathbb{1}_B \in F)$.

Se $0 \notin E$ e $1 \notin E$ si ha che $(\mathbb{1}_A \in E) = \emptyset$ e quindi anche $(\mathbb{1}_A \in E) \cap (\mathbb{1}_B \in F) = \emptyset$ e la sua probabilità è nulla, come è nullo il prodotto $P(\mathbb{1}_A \in E)P(\mathbb{1}_B \in F)$.

Se invece $0 \in E$ e $1 \in E$ risulta $(\mathbb{1}_A \in E) = \Omega$ e si ha che $(\mathbb{1}_A \in E) \cap (\mathbb{1}_B \in F) = (\mathbb{1}_B \in F)$ e quindi $P(\mathbb{1}_A \in E, \mathbb{1}_B \in F) = P(\mathbb{1}_B \in F) = P(\Omega)P(\mathbb{1}_B \in F) = P(\mathbb{1}_A \in E)P(\mathbb{1}_B \in F)$.

I conti sono del tutto analoghi se al posto di A mettiamo B , al posto di E mettiamo F .

Questi sono i casi banali. Il caso interessante è quando, ad esempio, $1 \in E$ e $0 \notin E$ e allo stesso tempo $1 \in F$ e $0 \notin F$. Allora $(\mathbb{1}_A \in E) = A$ e $(\mathbb{1}_B \in F) = B$ e dunque l'equazione

$$P(\mathbb{1}_A \in E, \mathbb{1}_B \in F) = P(\mathbb{1}_A \in E)P(\mathbb{1}_B \in F)$$

si trasforma in

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

Si noti che tutte le implicazioni sono invertibili.

Le altre possibilità, $0 \in E$ e $1 \notin E$, e così via, si traducono, per gli eventi, in identità che riguardano nei tre modi possibili i complementari di A e B . Ma abbiamo già visto che l'indipendenza di A e B equivale all'indipendenza di A^c e B , di A e B^c ed anche di A^c e B^c .

Ne possiamo concludere che gli eventi A e B sono indipendenti se e solo se lo sono le rispettive funzioni indicatrici.

Una formula moltiplicativa

Vale il seguente

Teorema

Se X ed Y sono due v.a. discrete indipendenti che ammettono speranza matematica, allora anche XY ammette speranza matematica e si ha che

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

Sia $p(x_i, y_j)$ la densità congiunta del vett. a. (X, Y) . Sappiamo ora che $p(x_i, y_j) = p_1(x_i)p_2(y_j)$, dunque il valore $x_i y_j$ viene assunto con probabilità $p_1(x_i)p_2(y_j)$. Pertanto, la speranza matematica di XY è

$$\sum_{i,j} x_i y_j p_1(x_i) p_2(y_j) = \sum_i x_i p_1(x_i) \sum_j y_j p_2(y_j) = E(X)E(Y).$$

Per la verità dovremmo completare la dimostrazione facendo vedere che la serie doppia (se di serie si tratta) è assolutamente convergente se lo sono le due serie che calcolano $E(X)$ ed $E(Y)$.

Momenti di ordine superiore

Nei paragrafi precedenti abbiamo parlato della speranza matematica $E(X)$ e della varianza, che coinvolge $E(X^2)$. In statistica saranno utili anche i *momenti di ordine superiore*. Se $k \in \mathbb{N}$ ed esiste la speranza matematica X^k , si dice che X ammette momento di ordine k .

Valgono i seguenti due risultati:

Proposizione 1

Se X ed Y ammettono momento di ordine $k \in \mathbb{N}$, anche $X + Y$ ammette momento di ordine k .

La dimostrazione segue dalla disuguaglianza

$$|a + b|^k \leq 2^{k-1}(|a|^k + |b|^k)$$

e dalla proprietà 5) della speranza matematica.

Per $k = 2$ otteniamo questo risultato interessante: se X ed Y ammettono varianza, anche $X + Y$ ammette varianza.

Proposizione 2

Se X ammette momento di ordine $k \in \mathbb{N}$ e se h è un intero $h < k$, allora X ammette momento di ordine h .

Anche questo risultato dipende dalla proprietà 5) richiamata prima, oltre che dalla disuguaglianza

$$|a|^h \leq |a|^k + 1.$$

Esercizi e complementi.

Calcolo delle speranze matematiche e varianze di alcune v.a. discrete.

1) Bernoulliana $B(1,p)$

$$\begin{pmatrix} 1 & 0 \\ p & 1 - p \end{pmatrix}$$

La speranza matematica è $1 \cdot p + 0 \cdot (1 - p) = p$.

Poiché $X = X^2$, $\text{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p) = pq$.

Se $A \in \mathcal{A}$ e $P(A) = p$, la funzione indicatrice $\mathbb{1}_A$ di A ha legge $B(1, p)$ e dunque $E(\mathbb{1}_A) = p$.

2) Binomiale $B(n, p)$

$$\begin{pmatrix} 0 & 1 & \dots & k & \dots \\ q^n & npq^{n-1} & \dots & \binom{n}{k} p^k q^{n-k} & \dots \end{pmatrix}$$

La speranza matematica si può calcolare usando la definizione:

$$E(X) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k}.$$

Il primo addendo è nullo e si può far partire la somma da $k = 1$. Il binomiale ha a denominatore $k!$ e questo porta ad una semplificazione, continuando

$$\begin{aligned} &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = \\ &np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1} = np(p+q)^{n-1} = np. \end{aligned}$$

Il procedimento è piuttosto laborioso e richiede qua e là qualche piccola astuzia. C'è però un metodo alternativo che ci sarà particolarmente utile quando si tratterà di calcolare la varianza di una $B(n, p)$.

Possiamo infatti osservare che una $X \sim B(n, p)$ altro non è che la somma di n variabili indipendenti del tipo $B(1, p)$, $X_1 + X_2 + \dots + X_n$. Basta pensare al significato della $B(n, p)$ in relazione ad una sequenza di Bernoulli.

Dunque si ha

$$E(X) = E\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n E(X_k) = np.$$

Calcoliamo ora la varianza di una $B(n, p)$. Possiamo evitare l'aritmetica che abbiamo visto prima, che in questo caso diventa ancora più fastidiosa, utilizzando il fatto che la varianza è additiva se le v.a. sono indipendenti.

$$\text{Var}(X) = \text{Var}(X_1 + X_2 + \dots + X_n) = n\text{Var}(X_1) = npq.$$

3) Varianza della v.a. geometrica.

Usiamo l'identità di König e calcoliamo

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p q^{k-1}$$

Per la speranza matematica abbiamo usato un metodo che potremmo chiamare "aritmetico", ora invece useremo il calcolo differenziale, tanto per consolidare un altro pezzetto della cultura di base.

Consideriamo l'espressione

$$p \sum_{k=1}^{\infty} k^2 q^{k-1}$$

come funzione della variabile q (p si considera una costante moltiplicativa) e riscriviamola in una forma più conveniente.

$$p \sum_{k=1}^{\infty} (k^2 - k + k) q^{k-1} = pq \sum_{k=1}^{\infty} (k-1) k q^{k-2} + p \sum_{k=1}^{\infty} k q^{k-1}$$

Il secondo addendo si riconosce facilmente, si tratta di $E(X)$ che sappiamo essere uguale a $\frac{1}{p}$. Nel primo termine il primo addendo è nullo. Conviene pensare la serie come funzione della variabile reale q e riconoscere nella serie $\sum_{k=2}^{\infty} (k-1) k q^{k-2}$ la derivata seconda di $\sum_{k=0}^{\infty} q^k$ che sappiamo convergere a $\frac{1}{1-q}$.

La derivata seconda della frazione è uguale a $\frac{2}{(1-q)^3}$. Mettendo insieme i vari risultati parziali, otteniamo

$$\text{Var}(X) = \frac{2pq}{(1-q)^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}.$$

4) Varianza della v.a. di Poisson

Se X è una v.a. di Poisson di parametro λ , sappiamo che $E(X) = \lambda$. Calcoliamo $E(X^2)$.

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} (k^2 - k + k) \frac{\lambda^k}{k!} e^{-\lambda} =$$

$$\begin{aligned} \sum_{k=2}^{\infty} (k-1)k \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} &= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda = \\ &= \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} + \lambda = \lambda^2 + \lambda. \end{aligned}$$

Ne segue che

$$\text{Var}(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

5) Da un'urna contenente b palline bianche ed r rosse estraiamo senza rimpiazzo una prima pallina e poi una seconda. Denotiamo con X la v.a. che vale 1 se alla prima estrazione si ottiene una pallina rossa e zero altrimenti. Sia invece Y la v.a. che vale 1 se alla seconda estrazione si ottiene una pallina rossa e zero altrimenti.

Vogliamo calcolare $\rho(X, Y)$.

La v.a. XY assume il valore 1 con probabilità $\frac{r}{b+r} \frac{r-1}{b+r-1}$ ed è nulla altrimenti.

Perciò $E(XY) = \frac{r}{b+r} \frac{r-1}{b+r-1}$. D'altra parte $E(X) = E(Y) = \frac{r}{b+r}$.

Queste considerazioni ci permettono di calcolare la covarianza di X e Y .

$$\begin{aligned} E(XY) - E(X)E(Y) &= \frac{r(r-1)}{(b+r)(b+r-1)} - \frac{r^2}{(b+r)^2} = \\ \frac{(r^2 - r)(b+r) - (b+r-1)r^2}{(b+r)^2(b+r-1)} &= \frac{r^2b + r^3 - rb - r^2 - br^2 - r^3 + r^2}{(b+r)^2(b+r-1)} = \\ &= \frac{-rb}{(b+r)^2(b+r-1)}. \end{aligned}$$

Teniamo ora conto del fatto che $X^2 = X$ e $Y^2 = Y$. Allora

$$\begin{aligned} \text{Var}(X) = \text{Var}(Y) &= E(X^2) - E(X)^2 = E(X) - E(X)^2 = \\ \frac{r}{b+r} - \left(\frac{r}{b+r}\right)^2 &= \frac{r(b+r) - r^2}{(b+r)^2} = \frac{rb}{(b+r)^2}. \end{aligned}$$

Riassumendo,

$$\rho(X, Y) = \frac{\frac{-rb}{(b+r)^2(b+r-1)}}{\frac{rb}{(b+r)^2}} = \frac{-1}{b+r-1}.$$

6) Vediamo come si calcola la densità g della somma di due variabili aleatorie discrete X e Y se conosciamo la densità congiunta $p(i, j)$ del vettore (X, Y) .

Ci proponiamo di dimostrare che

$$P(X + Y = s) = g(s) = \sum_{t \in \mathbb{R}} p(t, s - t).$$

La dimostrazione è pressoché banale: infatti, fissato $s \in \mathbb{R}$, la somma è estesa a tutti i t tali che $t + (s - t) = s$ per i quali $p(t, s - t) \neq 0$.

Se in particolare X e Y sono indipendenti, la formula si semplifica ottenendo

$$g(s) = \sum_{t \in \mathbb{R}} p_1(t)p_2(s - t).$$

7) Applichiamo la formula precedente al caso in cui X e Y sono della variabili di Poisson indipendenti di parametri λ e μ , rispettivamente. Scopriremo che $X + Y$ è una v.a. di Poisson di parametro $\lambda + \mu$.

Per quanto visto al punto precedente, la densità g della somma è uguale a

$$g(i) = \sum_{k=0}^{+\infty} p_1(k)p_2(i - k) = \sum_{k=0}^i p_1(k)p_2(i - k)$$

perché $p_2(i - k) = 0$ se $k > i$. Perciò

$$g(i) = \sum_{k=0}^i \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{i-k}}{(i-k)!} e^{-\mu} = e^{-(\lambda+\mu)} \frac{1}{i!} \sum_{k=0}^i \frac{i!}{k!(i-k)!} \lambda^k \mu^{i-k} = \frac{(\lambda + \mu)^i}{i!} e^{-(\lambda+\mu)}$$

come volevamo dimostrare.

8) Alice e Bruno giocano con due sequenze bernoulliane indipendenti X e Y . Inizia Alice e se X ha successo ($P(X = 1) = p$) ha vinto. Nel caso contrario il gioco passa a Bruno che vince se Y ha successo ($P(Y = 1) = r$), altrimenti il gioco ripassa ad Alice.

I due continuano così fino alla vittoria di uno dei due. Il tratto iniziale dà un vantaggio ad Alice e quindi, affinché il gioco sia equo deve essere $r > p$. Ma di quanto?

Una riflessione un po' più attenta può far sorgere il dubbio che se p è molto grande, r non basta mai. Pensiamo alla situazione limite $p = 1$. È chiaro che Bruno non ha neanche il piacere di partecipare. Viene allora la curiosità di vedere che cosa succede in un intorno sinistro di $p = 1$.

Si osserva che X ed Y sono due variabili aleatorie geometriche di parametri p ed r , rispettivamente, e che Alice vince se e solo se $X \leq Y$.

Per l'indipendenza la densità congiunta del vettore aleatorio (X, Y) , $p(i, j)$ è il prodotto delle densità marginali, dunque

$$p(i, j) = pq^{i-1}rs^{j-1},$$

dove $q = 1 - p$ e $s = 1 - r$. Naturalmente

$$\begin{aligned} P(X \leq Y) &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} pq^{i-1}rs^{j-1} = pr \sum_{i=1}^{\infty} q^{i-1} \sum_{j=i}^{\infty} s^{j-1} = \\ &pr \sum_{i=1}^{\infty} q^{i-1} \frac{s^{i-1}}{1-s} = \frac{pr}{1-s} \sum_{i=1}^{\infty} (qs)^{i-1} = p \frac{1}{1-qs}. \end{aligned}$$

Perché il gioco sia equo deve essere

$$p \frac{1}{1-qs} = \frac{1}{2}$$

ovvero $2p = 1 - qs = 1 - (1 - p)(1 - r) = p + r - pr$.

Semplificando ancora otteniamo l'identità

$$r = \frac{p}{1-p}.$$

Da questa si vede subito che se p è troppo grande, non c'è r che tenga, perché la frazione $\frac{p}{1-p}$ tende all'infinito quando p si avvicina ad 1.

Per lasciare spazio ad r , deve essere $\frac{p}{1-p} < 1$ e quindi $p < \frac{1}{2}$. Se $p = \frac{1}{2}$, deve essere $r = 1$!

Se invece $p = \frac{1}{3}$, $r = \frac{1}{2}$ rende il gioco equo.

9) Nell'esempio precedente il tratto iniziale dà un notevole vantaggio ad Alice.

Per contrasto, vediamo quest'altro esempio.

Da un mazzo di 52 carte francesi Alice e Bruno pescano, a turno, una carta. Vince chi pesca l'asso di cuori e a quel punto ci si ferma, naturalmente. Inizia Alice. Questo le dà un vantaggio?

Ripensando in maniera superficiale all'esempio precedente, si è portati a dire di sì. E invece il gioco è equo!

Il modo più semplice per convincersene è portare la distribuzione delle carte fino alla fine. In questo modo si danno ad Alice e a Bruno 26 carte a testa. Qual è la probabilità che l'asso di cuori sia in una o nell'altra metà del mazzo? Ovviamente $\frac{1}{2}$!

Che cosa succede invece se togliamo dal mazzo una carta (che non sia l'asso di cuori) e ripetiamo il gioco facendo sempre iniziare Alice? Anche in questo caso si capisce meglio il gioco se si distribuiscono tutte le carte. Ma questa volta Alice ne ha 26, mentre Bruno ne ha solo 25. La probabilità che vinca Alice è in questo caso $\frac{26}{51} \sim 0,50980\dots$, poco meno del 51%.

10) Ci sono delle circostanze nelle quali è interessante conoscere la legge di $aX + b$.

Se F è la funzione di ripartizione di X , e se $a < 0$, allora $G(t) = P(aX + b \leq t) = P(X \leq \frac{t-b}{a}) = F(\frac{t-b}{a})$. Naturalmente se $a < 0$, allora $P(aX + b \leq t) = P(X \geq \frac{t-b}{a}) = 1 - P(X < \frac{t-b}{a})$.

Se la v.a. è discreta, da qui procediamo a seconda dei casi.

Se invece la v.a. è assolutamente continua, possiamo calcolare le densità.

Se $a > 0$, si ha

$$g(t) = G'(t) = \frac{1}{a} F' \left(\frac{t-b}{a} \right).$$

se $a < 0$, allora

$$g(t) = -\frac{1}{a} F' \left(\frac{t-b}{a} \right)$$

e le due formule si possono unificare in

$$g(t) = \frac{1}{|a|} F' \left(\frac{t-b}{a} \right) = \frac{1}{|a|} f \left(\frac{t-b}{a} \right) \quad \text{q.o.}$$

Supponiamo che $X \sim N(0, 1)$ e consideriamo la v.a. $Y = \sigma X + \mu$.

La densità di Y è allora

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

e da questo deduciamo che $Y \sim N(\mu, \sigma^2)$.

Questo ci permette anche di calcolare la varianza di una $N(\mu, \sigma^2)$. Infatti

$$\text{Var}(\sigma X + \mu) = \text{Var}(\sigma X) = \sigma^2 \text{Var} X = \sigma^2.$$

Ora sappiamo qual è il significato dei parametri che caratterizzano una $N(\mu, \sigma^2)$.

11) Come nel caso discreto, anche in quello continuo si può trovare un'utile formula per esprimere la legge di $X + Y$ utilizzando la densità congiunta $f(s, t)$ del vettore assolutamente continuo (X, Y) .

La funzione di ripartizione di $Z = X + Y$ è

$$G(t) = P(X + Y \leq t) = \iint_{A_t} f(u, v) du dv,$$

dove A_t è il semipiano $\{(u, v) : u + v \leq t\}$. Per calcolare l'integrale doppio usiamo la formula di riduzione.

$$G(t) = \iint_{A_t} f(u, v) du dv = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{t-u} f(u, v) dv \right) du.$$

La densità $g(t)$ è la derivata di $G(t)$:

$$g(t) = G'(t) = \int_{-\infty}^{+\infty} f(u, t - u) du.$$

Se in particolare le due v.a. X e Y sono indipendenti, l'ultima formula prende la forma seguente:

$$g(t) = G'(t) = \int_{-\infty}^{+\infty} f_1(u) f_2(t - u) du.$$

12) Usiamo subito questa formula per dimostrare il seguente risultato: Se $X \sim \Gamma(\alpha_1, \lambda)$ e $Y \sim \Gamma(\alpha_2, \lambda)$ e se sono indipendenti, allora $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.

Si tratta di calcolare

$$g(t) = \int_{-\infty}^{+\infty} f_1(u) f_2(t - u) du.$$

Nell'esempio discreto abbiamo visto che la serie si riduceva ad una somma finita, perché la densità di Y assumeva solamente valori non negativi. Nel caso che stiamo trattando ora, l'insieme di integrazione si riduce ad un intervallo limitato perché $f_2(t-u) = 0$ per $t-u \geq 0$, cioè per $u \leq t$ ed inoltre $u \geq 0$. Si ha dunque

$$\begin{aligned} g(t) &= \int_0^t f_1(u)f_2(t-u) du = \\ &= \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} \int_0^t u^{\alpha_1-1} e^{-\lambda u} (t-u)^{\alpha_2-1} e^{-\lambda(t-u)} du = \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda t} \int_0^t u^{\alpha_1-1} (t-u)^{\alpha_2-1} du \end{aligned}$$

Sostituiamo ora la variabile di integrazione ponendo $u = st$. Si ottiene così

$$\begin{aligned} g(t) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda t} \int_0^1 (st)^{\alpha_1-1} (t-st)^{\alpha_2-1} t ds = \\ &= \left[\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 s^{\alpha_1-1} (1-s)^{\alpha_2-1} ds \right] t^{\alpha_1+\alpha_2-1} e^{-\lambda t} \end{aligned}$$

La quantità dentro la parentesi quadra è una costante che moltiplica una funzione della t . Nell'espressione della funzione si riconosce la $\Gamma(\alpha_1 + \alpha_2, \lambda)$ e, come abbiamo osservato in altre occasioni, la costante moltiplicativa non può che essere uguale a $\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}$.

Per induzione si dimostra che se $X_i \sim \Gamma(\alpha_i, \lambda)$ per $1 \leq i \leq n$ sono indipendenti, allora $\sum_{i=1}^n X_i \sim \Gamma(\sum_{i=1}^n \alpha_i, \lambda)$.

13) Va trattato ora un altro argomento che riguarda la Γ e cioè la sua speranza matematica e i suoi momenti.

Sia β un numero reale positivo e sia X una v.a. di legge $\Gamma(\alpha, \lambda)$. Allora

$$E(X^\beta) = \frac{\lambda^\beta}{\Gamma(\alpha)} \int_0^\infty t^\beta t^{\alpha-1} e^{-\lambda t} dt$$

L'identità si dimostra per approssimazione, come abbiamo fatto con il calcolo di $E(X^2)$. Continuando, troviamo

$$= \frac{\lambda^\beta}{\Gamma(\alpha)} \int_0^\infty t^{\alpha+\beta-1} e^{-\lambda t} dt = \frac{\lambda^\beta}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\lambda^{\alpha+\beta}} \left[\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha + \beta)} \int_0^\infty t^{\alpha+\beta-1} e^{-\lambda t} dt \right] =$$

$$\frac{\Gamma(\alpha + \beta)}{\lambda^\beta \Gamma(\alpha)},$$

perché la quantità in parentesi quadra vale 1, essendo l'integrale su tutto \mathbb{R} della densità di una $\Gamma(\alpha + \beta, \lambda)$.

a) Per $\beta = 1$ si trova

$$E(X) = \frac{\Gamma(\alpha + 1)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

b) Per $\beta = 2$ si ha

$$E(X^2) = \frac{\Gamma(\alpha + 2)}{\lambda^2 \Gamma(\alpha)} = \frac{\alpha(\alpha + 1)}{\lambda^2}.$$

Ne segue che

$$\text{Var } X = E(X^2) - E(X)^2 = \frac{\alpha^2 + \alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

c) Visto che la formula è a portata di mano, calcoliamo anche il momento terzo.

$$E(X^3) = \frac{\Gamma(\alpha + 3)}{\Gamma(\alpha)\lambda^3} = \frac{(\alpha + 2)(\alpha + 1)\alpha}{\lambda^3}.$$

d) Se in particolare $\alpha = 1$, cioè se X ha legge esponenziale di parametro λ , allora si calcola facilmente che

$$E(X) = \frac{1}{\lambda} \text{ e } \text{Var}(X) = \frac{1}{\lambda^2}.$$

2 Variabili aleatorie continue

Nei capitoli precedenti abbiamo studiato le v.a discrete. Esse modellano quegli eventi casuali nei quali gli eventi che si possono verificare sono in numero finito oppure, al più, numerabile. Questi eventi possono essere dei numeri interi, ma anche quando non lo sono (testa/croce, rosso/nero/verde, carte BB/RR/BR etc), conviene alle volte rappresentarli come tali.

Ma ci sono anche situazioni nelle quali il risultato casuale è un numero reale, o un vettore di \mathbb{R}^n , oppure un punto contenuto in una sfera o sulla sua superficie.

È casuale il momento $t \in \mathbb{R}_+$ in cui smette di funzionare un componente elettronico, o si interrompe il collegamento Wi-Fi o viene a mancare in tutta la zona la corrente elettrica, l'istante in cui una massa di materiale radioattivo emette una particella α , il punto di uno schermo che verrà colpito dal prossimo raggio cosmico, l'ipocentro di un terremoto (che può essere profondo fino a oltre 700 km), l'epicentro di un terremoto, il punto colpito da un fulmine, e così via.

Presenteremo ora le variabili aleatorie che chiameremo *continue* per distinguerle da quelle discrete.

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Ricordiamo che una funzione $X : \Omega \rightarrow \mathbb{R}$ si dice una variabile aleatoria, se per ogni $t \in \mathbb{R}$ si ha che $\{\omega : X(\omega) \leq t\} \in \mathcal{A}$.

Nel caso discreto abbiamo sempre avuto a che fare con spazi di probabilità nei quali \mathcal{A} è la famiglia delle parti di Ω . In effetti anche nel caso discreto ci sono esempi patologici, scarsamente interessanti ai fini pratici, nei quali non tutti i sottoinsiemi di Ω appartengono a \mathcal{A} . Vediamo un esempio.

Esempio

Sia $\Omega = \mathbb{N}$ e scegliamo la partizione $\{E_k\}$, per $k \in \mathbb{N}_0$, dove E_0 è l'insieme dei numeri dispari, E_1 è l'insieme dei numeri pari non divisibili per 4 (“la metà dei numeri pari”), E_2 l'insieme dei numeri divisibili per quattro ma non per otto (“la metà dei numeri divisibili per quattro”) e, in generale, E_k è l'insieme dei numeri divisibili per 2^k , ma non per 2^{k+1} .

Sia \mathcal{A} la famiglia di tutte le unioni finite o numerabili di insiemi della partizione. Questa è una σ -algebra e su di essa si può definire una probabilità in maniera molto naturale. I numeri dispari sono, in un certo senso, la metà di tutti gli interi non negativi e quindi è ragionevole porre $P(E_0) = 1/2$. Gli elementi di E_1 sono la “metà” dei numeri pari ed è quindi ragionevole porre $P(E_1) = 1/4$. Gli elementi di E_2 sono a loro volta la “metà” dei numeri rimanenti e quindi è ragionevole la posizione $P(E_2) = 1/8$. E così via: in generale è naturale porre $P(E_k) = 1/2^{k+1}$. Se poi $A = \cup_{k \in I} E_k$ è un'unione finita o numerabile di insiemi della partizione, mediante la posizione $P(A) = \sum_{k \in I} P(E_k)$ si definisce una probabilità su tutto \mathcal{A} .

Abbiamo così ottenuto uno spazio di probabilità numerabile (Ω, \mathcal{A}, P) nel

quale la σ -algebra \mathcal{A} **non** è la famiglia delle parti di Ω .

Per questo motivo non tutte le funzioni di $\Omega = \mathbb{N}$ in \mathbb{R} sono delle variabili aleatorie. Prendiamo ad esempio l'identità $X(n) = n$: l'insieme $(X \leq t)$ è un insieme finito per ogni numero reale t , mentre tutti gli insiemi di \mathcal{A} sono infiniti (a parte l'insieme vuoto).

Fatta chiarezza con le v.a. discrete, affrontiamo la questione ben più profonda delle v.a. continue.

Premettiamo un risultato che ci sarà utile tra poco.

Teorema

Se $\{A_n\}$ è la successione crescente di eventi (cioè $A_n \subset A_{n+1}$), allora

$$\lim_{n \rightarrow \infty} P(A_n) = P(\cup_{n=1}^{\infty} A_n).$$

Introduciamo la successione ausiliaria ponendo $B_k = A_k \setminus A_{k-1}$ (per comodità poniamo $A_0 = \emptyset$).

Si vede facilmente che $\cup_{n=1}^{\infty} B_k = \cup_{n=1}^{\infty} A_k$ ed inoltre che $B_i \cap B_j = \emptyset$ se $i \neq j$.

Abbiamo allora che

$$\begin{aligned} P(\cup_{n=1}^{\infty} A_n) &= P(\cup_{n=1}^{\infty} B_n) = \sum_{k=1}^{\infty} P(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_k \setminus A_{k-1}) = \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n (P(A_k) - P(A_{k-1})) = \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

L'ultima identità segue dalla semplificazione telescopica della somma e dal fatto che $P(A_0) = 0$.

Nell'immediato ci servirà il seguente

Corollario

Se $\{A_n\}$ è una successione decrescente di eventi (cioè $A_{n+1} \subset A_n$), allora

$$\lim_{n \rightarrow \infty} P(A_n) = P(\cap_{n=1}^{\infty} A_n).$$

La dimostrazione si fa per complementazione. Se poniamo $C_n = A_n^c$ segue dal teorema precedente che $\lim_{n \rightarrow \infty} P(C_n) = P(\cup_{n=1}^{\infty} C_n)$ e quindi

$$1 - P(\cup_{n=1}^{\infty} C_n) = P(\cap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} (1 - P(C_n)) = \lim_{n \rightarrow \infty} P(A_n).$$

La condizione $\{\omega : X(\omega) \leq t\} \in \mathcal{A}$ altro non è che la definizione di misurabilità della teoria dell'integrazione alla Lebesgue e in probabilità svolge un ruolo simile, ma con alcuni aspetti che sono specifici.

Le v.a. continue che tratteremo sono definite su $\Omega \subset \mathbb{R}$ e come \mathcal{A} si può prendere la σ -algebra dei boreliani oppure quella, più ampia, degli insiemi misurabili secondo Lebesgue. Sappiamo che esistono insiemi di numeri reali che non sono misurabili secondo Lebesgue. Questo fa sì che non tutte le funzioni di Ω in \mathbb{R} sono delle variabili aleatorie.

Osserviamo intanto che questa condizione permette di definire quella che si chiama *funzione di ripartizione* della v.a. X , cioè

$$F_X(t) = P(X \leq t).$$

Qualche volta, per alleggerire la notazione e se non ci sono dubbi di quale v.a. stiamo parlando, al posto di $F_X(t)$ scriveremo $F(t)$.

Nel caso discreto abbiamo visto che una v.a. discreta è identificata dalla sua densità, cioè dalla funzione $p(t) = P(X = t)$ che è nulla tranne che su un insieme finito o numerabile di valori, che sono poi i valori assunti dalla X .

Se X è una v.a. discreta, allora la funzione di ripartizione si ricava dalla densità nel modo seguente: si ha che $(X \leq t_0) = \bigcup_{t \leq t_0} (X = t)$, dove l'unione di insiemi non vuoti è finita o numerabile (e quindi appartiene alla σ -algebra), dunque $F(t_0) = P(X \leq t_0) = \sum_{t \leq t_0} P(X = t)$ (e la somma è estesa ad un insieme finito o numerabile di indici).

Sempre per le v.a. discrete si può percorrere anche la strada inversa e la funzione di densità si può ricavare dalla funzione di ripartizione. Per far vedere come si fa, abbiamo bisogno di qualche risultato preliminare.

Alcune proprietà della funzione di ripartizione

Una funzione di ripartizione gode delle seguenti proprietà:

- 1) $0 \leq F(t) \leq 1$;
- 2) La F è non decrescente;

- 3) $\lim_{t \rightarrow -\infty} F(t) = 0$ e $\lim_{t \rightarrow +\infty} F(t) = 1$;
 4) la F è continua da destra.

La 1) è ovvia, perché $F(t)$ è la probabilità di un evento.

Anche la 2) è immediata, perché se $t_1 \leq t_2$, allora $(X \leq t_1) \subset (X \leq t_2)$ e, come abbiamo visto, la probabilità è crescente rispetto all'inclusione.

Inseriamo qui una breve parentesi per ricordare tre teoremi sui limiti delle funzioni monotone che ci saranno molto utili.

Teorema 1

Se f è una funzione non decrescente sull'intervallo aperto $]a, b[$ e se $a < c < b$, allora esistono i limiti $\lim_{t \rightarrow c^-} f(t) = f(c^-)$ e $\lim_{t \rightarrow c^+} f(t) = f(c^+)$ e inoltre $f(c^-) \leq f(c) \leq f(c^+)$.

Teorema 2

Se f è una funzione non decrescente sull'intervallo aperto $]a, b[$, se $a < c < b$, e se per una successione strettamente decrescente (crescente) $\{t_n\}$ che converge a c si ha che

$$\lim_{n \rightarrow \infty} f(t_n) = f(c),$$

allora f è continua da destra (sinistra).

Teorema 3

Se f è una funzione non decrescente su \mathbb{R} , e se $\{t_n\}$ è una qualunque successione che converge a $+\infty$ ($-\infty$), allora

$$\lim_{n \rightarrow +\infty} f(t_n) = \lim_{t \rightarrow +\infty} f(t),$$

e un'identità analoga vale se $\{t_n\}$ tende a $-\infty$.

Continuiamo ora con le proprietà delle funzioni di ripartizione. Le dimostrazioni dei due casi della 3) e della 4) sono tra loro simili.

Per calcolare $\lim_{t \rightarrow -\infty} F(t)$ possiamo ad esempio prendere la successione $t_n = -n$ e calcolare il limite per $n \rightarrow +\infty$ di $F(-n) = P(X \leq -n)$. Osserviamo che la successione di eventi $\{(X \leq -n)\}$ è decrescente e che $\bigcap_{n=1}^{\infty} (X \leq -n) = \emptyset$. Dal Corollario dimostrato prima segue la tesi.

In modo analogo si prova che $\lim_{t \rightarrow +\infty} F(t)$ è uguale ad 1.

Anche la continuità da destra si dimostra con questo ragionamento. Basta osservare che la successione di eventi $\{(X \leq t_0 + \frac{1}{n})\}$ è decrescente e che l'intersezione è $(X \leq t_0)$.

Non è detto che la funzione di ripartizione sia continua. Essendo la continuità da destra garantita, una funzione di ripartizione (spesso si usa l'abbreviazione f.r) è continua se e solo se è continua da sinistra. I teorema successivo caratterizza la continuità di una f.r..

Teorema

Una funzione di ripartizione X è continua in t_0 se e solo $P(X = t_0) = 0$.

Si ha che $(X = t_0) = \bigcap_{n=1}^{\infty} (t_0 - \frac{1}{n} < X \leq t_0)$ e quindi

$$\begin{aligned} P(X = t_0) &= P\left(\bigcap_{n=1}^{\infty} (t_0 - \frac{1}{n} < X \leq t_0)\right) = \lim_{n \rightarrow \infty} P\left(t_0 - \frac{1}{n} < X \leq t_0\right) = \\ &= \lim_{n \rightarrow \infty} \left(P(X \leq t_0) - P(X \leq t_0 - \frac{1}{n})\right) = P(X \leq t_0) - \\ &\quad \lim_{n \rightarrow \infty} \left(P(X \leq t_0 - \frac{1}{n})\right) = F(t_0) - \lim_{n \rightarrow \infty} F\left(t_0 - \frac{1}{n}\right) \end{aligned}$$

e la differenza è nulla se e solo se F è continua da sinistra (e quindi continua).

Merita notare che la F è continua su tutto \mathbb{R} se e solo se $P(X = t) = 0$ per ogni $t \in \mathbb{R}$.

Le v.a. discrete che hanno qualche interesse nel calcolo delle probabilità hanno funzioni di ripartizione “a gradini”.

Per capire meglio, possiamo dare un'occhiata ad un caso particolare, alla v.a. $B(2, \frac{1}{2})$ (che è quella che abbiamo associato al lancio di due dadi, per esempio). Il grafico della sua f.r. è composto da tre segmenti e due semirette: la semiretta $\{(x, y) : -\infty \leq x < 0, y = 0\}$, dai due segmenti chiusi a sinistra ed aperti a destra $\{(x, y) : k \leq x < k + 1, y = y_k\}$, (dove $y_0 = \frac{1}{4}$, $y_1 = \frac{3}{4}$) e dalla semiretta $\{(x, y) : 2 \leq x < +\infty, y = 1\}$.

La f.r. della v.a geometrica di parametro $\frac{1}{2}$ ha come grafico una scalinata i cui gradini partono da $x = 1$, hanno “pedata” uguale a 1 e “alzata” sempre più bassa. Il k -esimo gradino ha quota $\sum_{i=1}^k \frac{1}{2^i}$ e la quota che questi gradino

ci fa guadagnare è $\frac{1}{2^k}$. Non si raggiunge mai la quota 1, in altre parole $F(t) < 1$ per ogni $t \in \mathbb{R}$.

Questi esempi ci fanno capire perché è più comodo definire la legge di una v.a. discreta mediante la densità.

Però, come dicevamo, anche dalla funzione di ripartizione è possibile risalire alla densità di una v.a. discreta.

Poiché

$$(X = t_k) = \bigcap_{n=1}^{\infty} \left(t_k - \frac{1}{n} < X \leq t_k + \frac{1}{n} \right)$$

e la successione degli insiemi a destra è decrescente,

$$\begin{aligned} P(x = t_k) &= \lim_{n \rightarrow \infty} P \left(t_k - \frac{1}{n} < X \leq t_k + \frac{1}{n} \right) = \\ &= \lim_{n \rightarrow \infty} \left(F \left(t_k + \frac{1}{n} \right) - F \left(t_k - \frac{1}{n} \right) \right) = F(t_k) - \lim_{n \rightarrow \infty} F \left(t_k - \frac{1}{n} \right). \end{aligned}$$

Dunque $P(x = t_k) = p_k$ rappresenta il “salto” che la funzione di ripartizione fa in corrispondenza al valore t_k , assunto dalla X . La cardinalità di questi salti è finita o numerabile. Questo lo sappiamo dallo studio delle v.a. discrete, ma anche dal fatto che una funzione monotona non può avere più di un’infinità numerabile di punti di discontinuità.

Fatte queste premesse che servono per ricollegarci a quanto abbiamo visto nella prima parte del corso, parliamo finalmente delle v.a. continue.

Definizione

Diremo che una variabile aleatoria X è continua, se è continua la sua funzione di ripartizione F_X .

Vediamo qualche primo esempio.

Variabile aleatoria uniformemente distribuita su $[0, 1]$

La v.a. X si dice uniformemente distribuita su $[0, 1]$ se la sua funzione di ripartizione F è definita nel modo seguente:

$$F_X(t) = \begin{cases} 0, & \text{se } t \leq 0 \\ t, & \text{se } 0 \leq t \leq 1 \\ 1, & \text{se } t \geq 1. \end{cases}$$

Questa v.a. si chiama così perché la probabilità è “spalmata uniformemente” su $[0, 1]$, nel senso che qualunque sia la posizione dell’intervallo $[a, b] \subset [0, 1]$, si ha che $P(a < X \leq b) = b - a$, cioè la probabilità dipende solo dalla lunghezza dell’intervallo e non dalla sua posizione.

Poiché la F_X (e quindi la v.a.) è continua, $P(X = a) = 0$ e quindi $P(a < X \leq b) = P(a \leq X \leq b)$.

C’è una perfetta analogia con il caso classico. Se Y rappresenta l’estrazione di un numero dall’insieme $\{1, 2, 3, \dots, N\}$ con $P(k) = \frac{1}{N}$, allora $P(a < Y \leq b) = \frac{b-a}{N}$, qualunque siano a e b interi compresi tra 1 e N e quindi la probabilità dipende solo dal numero dei numeri estratti, non dalla loro posizione.

Una v.a. simile è quella *uniformemente distribuita su* $[a, b]$ che ha come funzione di ripartizione

$$F_X(t) = \begin{cases} 0, & \text{se } t \leq a \\ \frac{t-a}{b-a}, & \text{se } a \leq t \leq b \\ 1, & \text{se } t \geq b. \end{cases}$$

In questo caso la probabilità è uniformemente spalmata sull’intervallo $[a, b]$.

Un altro esempio di v.a. continua è quella *esponenziale* di parametro λ . La sua funzione di ripartizione è

$$F_X(t) = \begin{cases} 0, & \text{se } t \leq 0 \\ 1 - e^{-\lambda t}, & \text{se } t \geq 0. \end{cases}$$

Questa v.a. fa parte di un’importante e più ampia famiglia di v.a., quella delle leggi Gamma.

In un certo senso, essa è la controparte continua della v.a. geometrica. La rassomiglianza sta nel seguente

Teorema

La variabile aleatoria esponenziale non ha memoria.

In termini di probabilità condizionale la tesi si scrive nel seguente modo:

$$P(X > s + t | X > s) = P(X > t).$$

Il primo termine è uguale a

$$\frac{P(X > s+t, X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)} = \frac{1 - P(X \leq s+t)}{1 - P(X \leq s)} =$$

$$\frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - (1 - e^{-\lambda t}) = P(X > t).$$

Variabili aleatorie assolutamente continue

Vedremo in seguito altri esempi di variabili aleatorie continue. Ora è venuto il momento di parlare della *densità* delle (di certe) v.a. continue. Il fatto che, se X è una v.a. continua, si ha che $P(X = t) = 0$ per ogni $t \in \mathbb{R}$ rappresenta, apparentemente, un ostacolo insormontabile. Ma non è così.

Definizione di densità

Sia X una v.a. continua. Una funzione $p_X : \mathbb{R} \rightarrow \mathbb{R}$ si dice *densità* della X , se

- 1) $p_X(t) \geq 0$ per ogni $t \in \mathbb{R}$;
- 2) $p_X(t)$ è integrabile su \mathbb{R} ;
- 3) $F_X(t) = \int_{-\infty}^t p_X(s) d(s)$.

Dall'ultima condizione segue che $\int_{-\infty}^{+\infty} p_X(s) d\lambda(s) = 1$.

Quando sarà chiaro di quale v.a. stiamo parlando useremo la scrittura meno pesante $p(t)$.

Dalla 3) segue anche che $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b p(s) ds$.

L'integrale può essere inteso alla Lebesgue, che al momento è uno strumento esageratamente sofisticato. Le densità che incontreremo sono integrabili secondo Riemann, anzi sono continue tranne che in un insieme finito di punti. Avremo bisogno di qualcosina di più quando parleremo della *speranza matematica* delle v.a. continue.

Si conoscono variabili aleatorie continue che non ammettono densità.

Sappiamo che se la X ammette densità, allora $F(t) = \int_{-\infty}^t p(s) d(s)$ è una funzione continua.

Se la p è continua, allora sappiamo, per il teorema fondamentale del calcolo integrale, che la F è derivabile e che $F'(t) = p(t)$. Se la p è semplicemente integrabile, le cose sono un po' più complicate e ne parleremo più avanti.

3 Rappresentazione dei numeri reali positivi in base b

In questa prima sezione tratteremo brevemente la rappresentazione dei numeri reali positivi in base b , con $b \in \mathbb{N}$, $n \geq 2$.

Si tratta di un argomento un po' trascurato che però ha svariate applicazioni ed è comunque un aspetto fondamentale della matematica.

Una delle applicazioni è la descrizione dell'insieme e della funzione di Cantor.

In internet si trova poco o niente, perché Google privilegia i siti e documenti più consultati e non c'è dubbio che la rappresentazione in base b interessi più gli informatici che i matematici. Ma gli informatici sono interessati quasi esclusivamente alla rappresentazione dei *numeri interi*, e con una preferenza netta per la base $b = 2$ (detta rappresentazione binaria) e a quella in base $b = 16$ (base esadecimale), più raramente a quella "ottale" ($b = 8$).

Noi siamo talmente abituati alla rappresentazione decimale da confondere, molto spesso, il *numero* razionale (o reale) con la sua *rappresentazione*.

Che cos'è π ? Beh, è $3,14159\dots$. Che cosa è il numero di Nepero e ? La risposta è analoga: $e = 2,718\dots$.

Ma riflettendo sulla questione ci rendiamo conto che il numero e la sua rappresentazione sono due cose ben distinte.

Noi conosciamo la rappresentazione dei numeri interi usata dai Romani. È chiaro che "XXI" e "21" rappresentano lo stesso numero (di giorni, di orci di olio, di soldati, per dire). Questo numero che duemila anni fa i Romani rappresentavano in un certo modo, viene oggi rappresentato in un modo diverso. Altri popoli contemporanei ai Romani (i Greci, ad esempio) usavano sistemi diversi.

Gli stessi Arabi, dai quali abbiamo preso il sistema oggi usato da tutto il mondo, usano il sistema decimale, ma si servono di altri simboli per denotare le cifre da 0 a 9.

Vedasi <http://web.tiscali.it/istitutodeamicis/sitoscueladue/arabo/cifre.htm>.

Ma prima di discutere della rappresentazione dei numeri reali in base dieci ed in altre basi, facciamo qualche riflessione sul periodo "preistorico" che precede l'invenzione della scrittura.

L'umanità, nelle varie lingue, ha come prima cosa espresso i numeri (e le altre parole) *verbalmente* e solo il progresso ci ha portati alla rappresentazione simbolica di parole e numeri.

In quella fase preistorica, che precede la creazione dei sistemi di scrittura, le varie culture hanno preferito la base dieci, probabilmente perché abbiamo dieci dita e con esse ci possiamo aiutare sia per dare una rappresentazione (che usiamo ancora oggi, e che potremmo definire corporale) dei numeri, sia per fare alcune semplici operazioni.

I Babilonesi usavano però un sistema sessagesimale. Poiché le loro scoperte in astronomia erano molto avanzate, tracce importanti di questo sistema sono arrivate fino a noi, passando attraverso altre civiltà. Infatti ogni ora è divisa in sessanta minuti, i quali sono a loro volta divisi in sessanta secondi. Lo stesso sistema viene usato anche per misurare gli angoli.

Altri popoli, che per contare utilizzavano anche le dita dei piedi -tutti erano scalzi a quei tempi- usavano la base venti. In francese le parole per denotare settanta, ottanta e novanta sono espresse con parole di origine latina, ma corrispondono ad un'altra logica. Si usano infatti le parole *soixante-dix*, *quatre-vingt* e *quatre-vingt-dix*. Si tratta di un reperto archeologico della lingua dei Galli che ritroviamo nella lingua neolatina.

Altri popoli usavano la base dodici, che è il numero delle falangi delle quattro dita parallele della mano (pollice contrapposto escluso). Gli antichi Greci usavano per i pesi e le lunghezze un sistema a base dodici.

Di questo sono rimaste altre tracce: abbiamo dodici ore, *ante meridiem* e *post meridiem*, dodici mesi, dodici segni zodiacali. La dozzina si usa ancora per contare certi beni (le uova) e si era anche conservata nel cosiddetto sistema imperiale britannico (uno scellino era composto da 12 pence, inoltre dodici pollici sono un piede e dodici once sono una libbra, misure queste ancora usate negli USA).

La lingua inglese conserva tracce del sistema in base dodici. Infatti *eleven* e *twelve* sono formati in maniera diversa dai numeri da tredici a diciannove. L'etimologia più accreditata è che le due parole significhino "uno (e due) oltre". Poi arrivano "thirteen" e così via fino a "nineteen". La stessa cosa succede nella lingua tedesca: undici e dodici si dicono "elf" e "zwölf" e poi la struttura cambia e si continua con "dreizehn", eccetera, fino a "neunzehn".

Alla fine il sistema decimale ha però preso il sopravvento su tutti gli altri.

Ma naturalmente la *rappresentazione decimale* è un'altra cosa rispetto al *sistema decimale*. Esso arrivò in Europa attraverso gli Arabi, i quali la

appresero dagli Indiani. L'idea geniale è quella della *virgola* e dello zero.

Ripensiamo alla rappresentazione decimale dei numeri reali non negativi. L'abituale scrittura $\pi = 3,14159\dots$ (che ci è familiare dalle elementari) è equivalente alla notazione che implicitamente coinvolge l'idea delle serie

$$\pi = 3 \cdot 10^0 + \frac{1}{10} + \frac{4}{10^2} + \frac{1}{10^3} + \frac{5}{10^4} + \frac{9}{10^5} \dots$$

Ogni numero reale positivo x si può esprimere come

$$x = \sum_{n=-\infty}^{\infty} \frac{a_n}{10^n},$$

dove $a_n = 0$ per tutti gli n minori di un certo intero n_0 e le cifre a_n sono degli interi compresi tra 0 e 9.

Ad esempio,

$$321,1234567\dots = 3 \cdot 10^2 + 2 \cdot 10^1 + 1 \cdot 10^0 + \frac{1}{10} + \frac{2}{10^2} + \frac{3}{10^3} + \frac{4}{10^4} + \frac{5}{10^5} \dots$$

Come è ben noto, alcuni numeri reali ammettono una doppia rappresentazione. Questo dipende dal fatto che la serie

$$\sum_{n=1}^{\infty} \frac{9}{10^n}$$

converge ad 1, che si può scrivere quindi anche come $0,\bar{9}$.

Per evitare la doppia rappresentazione ed avere una corrispondenza bi-univoca tra i numeri reali e le loro rappresentazioni, possiamo decidere (a seconda di quello che ci è più utile) di scegliere o la rappresentazione con un numero finito di cifre decimali oppure quella di periodo 9. Nessuno dirà che la probabilità di ottenere con un dado regolare un numero pari è $0,4\bar{9}$. Però un modo frequentemente usato per descrivere l'insieme di Cantor è proprio quello di utilizzare le rappresentazioni periodiche.

Il numero 10 ha di particolare solamente la nostra struttura anatomica. Topolino e Paperino hanno otto dita, quindi per loro sarebbe stato naturale sviluppare un sistema di scrittura dei numeri basata sul numero 8 e i dollari contati da Paperon dei Paperoni dovrebbero avere il valore espresso in base 8.

In effetti, per rappresentare i numeri reali un qualunque numero intero b maggiore o uguale a 2 può andar bene, perché ogni numero reale positivo si può esprimere come

$$x = \sum_{n=-\infty}^{\infty} \frac{a_n}{b^n},$$

dove $a_n = 0$ per tutti gli n minori di un certo intero n_0 e le cifre a_n sono degli interi compresi tra 0 e $b - 1$. Se $b > 10$ avremo bisogno, per la rappresentazione, di simboli con cui denotare le cifre da 10 a $b - 1$.

Il sistema dozzinale usa di solito il simbolo X per il dieci ed il simbolo E per l'undici (si veda su Wikipedia "Sistema numerico dozzinale"). Così il numero (scritto in base 12) $[E6, X]_{12}$ si scrive in base 10 come $11 \cdot 12^1 + 6 \cdot 12^0 + 10 \cdot 12^{-1} = 132 + 6 + \frac{10}{12} = 138,8\bar{3}$.

Poiché la scrittura in base b degli **interi** (positivi) è stata vista in altri corsi, ci soffermeremo su alcuni dettagli relativi alla scrittura in base b dei numeri dell'intervallo $[0, 1]$. Lo faremo in relazione all'insieme di Cantor, per il quale sono interessanti i casi $b = 2$ e $b = 3$.

4 L'insieme e la funzione di Cantor

L'insieme di Cantor è un sottoinsieme di \mathbb{R} definito in modo ricorsivo togliendo via via degli intervalli aperti all'intervallo $C_0 = [0, 1]$. Diceva Michelangelo: la scultura si trova già dentro il blocco di marmo, per farla emergere basta togliere il superfluo.

Da C_0 rimuoviamo l'intervallo aperto $] \frac{1}{3}, \frac{2}{3} [$. Si ottiene così il compatto C_1 , che è unione di due intervalli chiusi di lunghezza $\frac{1}{3}$.

In base 3 gli estremi dell'intervallo tolto hanno coordinate $\frac{1}{3} = [0, 1]_3 = [0, 0\bar{2}]_3$ e $\frac{2}{3} = [0, 2]_3$. Per rappresentare il primo estremo preferiremo la notazione periodica che non usa la cifra 1.

Si noti che i punti tolti sono quelli che hanno in base 3, dopo la virgola, la cifra 1. La notazione $[0, 0\bar{2}]_3$ "salva" l'estremo destro della prima componente di C_1 .

Al secondo passo rimuoviamo da ciascuno degli intervalli che compongono C_1 il segmento aperto centrale avente lunghezza $\frac{1}{9}$. Si ottiene così un compatto C_2 che è composto da quattro intervalli chiusi, ciascuno di lunghezza $\frac{1}{9}$. Gli estremi dei due intervalli tolti al secondo passo hanno, in base 3, coordinate $[0, 00\bar{2}]_3$, $[0, 02]_3$, $[0, 20\bar{2}]_3$, e $[0, 022]_3$.

Abbiamo nuovamente utilizzato le notazioni periodiche per denotare gli estremi sinistri degli intervalli tolti.

Osserviamo ancora una volta che la notazione che abbiamo scelto per gli estremi sinistri degli intervalli tolti ci permette di dire che **non** fanno parte di C_2 esattamente quei numeri reali che, in base 3, non presentano né al primo né al secondo posto, dopo la virgola, la cifra 1.

Nella stessa maniera si procede per passare da C_n a C_{n+1} : C_n è costituito da 2^n intervalli chiusi, ciascuno di lunghezza $\frac{1}{3^n}$. Da ciascuno di essi togliamo il segmento aperto centrale di lunghezza $\frac{1}{3^{n+1}}$. Adottando la convenzione di rappresentare tutti gli estremi sinistri degli intervalli tolti con la notazione periodica, C_{n+1} contiene tutti i numeri reali che in base 3 non presentano la cifra 1 nei primi $n + 1$ posti dopo la virgola.

L'insieme di Cantor C è l'intersezione di tutti in C_n :

$$C = \bigcap_{n=0}^{\infty} C_n.$$

Adottando la più volte richiamata convenzione che evita l'uso della cifra 1, gli elementi $x \in C$ sono tutti e soli i punti che si lasciano scrivere, in base 3, come $x = [0, a_1 a_2 a_3 \dots a_n \dots]_3$ con $a_n \in \{0, 2\}$.

Questo fatto ci permette di stabilire la cardinalità di C . Infatti l'applicazione $x \mapsto [0, b_1 b_2 b_3 \dots b_n \dots]_2$, dove $b_n = \frac{a_n}{2}$, è un'applicazione suriettiva φ di C su $[0, 1]$ che conviene pensare ora rappresentato in base 2. La φ è suriettiva, perché ogni $y \in [0, 1]$ rappresentato in base 2 da $[0, b_1 b_2 b_3 \dots b_n \dots]_2$, è immagine del punto $[0, 2b_1 2b_2 b_3 \dots 2b_n \dots]_3$. Invece non è iniettiva perché ad esempio il punto $y = [0, 0\bar{1}]_2 = [0, 1]_2$ è immagine due punti distinti di C , $[0, 0\bar{2}]_3$ e $[0, 2]_3$.

Dunque la cardinalità di C è quella di $[0, 1]$, cioè quella del continuo.

Per calcolare la misura di C , basta sommare "il superfluo" che è stato tolto:

$$\frac{1}{3} + \frac{2}{9} + \dots + \frac{2^{n-1}}{3^n} + \dots = \frac{1}{3} \left(1 + \frac{2}{3} + \left(\frac{2}{3}\right)^n \dots \right) = 1.$$

Quindi la misura di Lebesgue di C è uguale a zero.

L'insieme di Cantor è "piccolo" dal punto di vista della misura, ma, come abbiamo visto, è "grande" dal punto di vista della cardinalità.

Costruiamo ora la funzione di Cantor, chiamata anche *scala del diavolo*.

Come molte funzioni “strane” si costruisce come limite di una opportuna successione di funzioni continue $\{f_n\}$.

La prima funzione della successione, f_1 , suggerisce il processo che si segue per costruire le funzioni successive.

Il suo grafico è una spezzata a zig-zag composta da tre segmenti. Il primo è il tratto ripido che collega i punti $(0,0)$ e $(\frac{1}{3}, \frac{1}{2})$, seguito da un segmento orizzontale che collega $(\frac{1}{3}, \frac{1}{2})$ a $(\frac{2}{3}, \frac{1}{2})$ ed il terzo è ancora un tratto ripido che da $(\frac{2}{3}, \frac{1}{2})$ raggiunge $(1,0)$.

La f_2 si costruisce in maniera *analoga*. Evitiamo la parola “simile” perché in matematica essa ha un significato che in questo caso ci porterebbe fuori pista. Il grafico della f_2 coincide con quello della f_1 dove questa è costante, mentre i due tratti ripidi vengono sostituiti da due spezzate a zig-zag. Il primo zig-zag è costituito da tre segmenti. Il primo collega l’origine al punto $(\frac{1}{9}, \frac{1}{4})$, il secondo segmento è orizzontale e collega $(\frac{1}{9}, \frac{1}{4})$ a $(\frac{2}{9}, \frac{1}{4})$ e il terzo segmento collega quest’ultimo punto a $(\frac{1}{3}, \frac{1}{2})$.

Il secondo zig-zag è una traslazione del primo e collega $(\frac{2}{3}, \frac{1}{2})$ al punto $(1,1)$.

Si procede così per analogia. La f_{n+1} coincide con la f_n dove questa è localmente costante (cioè sul complementare di C_n), mentre negli altri intervalli il tratto ripido viene sostituito dagli zig-zag simili a quelli visti nel definire la f_2 .

La successione $\{f_n\}$ converge uniformemente ad una funzione limite f (ovviamente anch’essa continua) tale che $f(0) = 0$ e $f(1) = 1$.

La f è continua e crescente in $[0,1]$. Essa è derivabile, con derivata nulla in tutti i punti che non appartengono all’insieme di Cantor e quindi tranne che su un insieme di misura nulla.

Per questo motivo è sempre più frequentemente usata per questa funzione il nome di “scala del diavolo”, che è diabolica perché consente di salire da quota 0 a quota 1 su una scalinata che è quasi ovunque orizzontale.

Per ogni $x \in [0, 1]$ si ha

$$\int_0^x f'(t)d\lambda(t) = 0 \neq f(x)$$

e quindi per questa funzione non vale il teorema fondamentale del calcolo integrale.

Una condizione sufficiente per la validità di questo teorema è la continuità della derivata f' .

Cantor aveva pubblicato l'esempio nel 1884 ed immediatamente il matematico Scheeffer gli aveva segnalato che esso costituiva un controesempio ad una generalizzazione (errata) del teorema fondamentale del calcolo integrale pubblicato da poco dal noto matematico Harnack.

In effetti il teorema fondamentale del calcolo integrale può essere esteso ad una classe più ampia di funzioni, quelle *assolutamente continue*. Questa elegante generalizzazione è dovuta a Lebesgue.

Una funzione $f : [a, b] \rightarrow \mathbb{R}$ si dice assolutamente continua su $[a, b]$, se per ogni $\varepsilon > 0$ esiste un $\delta > 0$ tale che per ogni famiglia finita di intervalli disgiunti $I_k = [a_k, b_k]$, $1 \leq k \leq n$, contenuti in $[a, b]$ e tali che $\sum_{k=1}^n (b_k - a_k) < \delta$ si ha che $\sum_{k=1}^n |f(b_k) - f(a_k)| < \varepsilon$.

Vale il seguente teorema:

Teorema (Lebesgue)

Le due seguenti condizioni sono equivalenti:

- (1) La f è assolutamente continua su $[a, b]$;
- (2) La f è derivabile quasi ovunque, la sua derivata f' è integrabile secondo Lebesgue su $[a, b]$ e inoltre per ogni $x \in [a, b]$ si ha

$$f(x) - f(a) = \int_a^x f'(t) d\lambda(t).$$

Esempi

Le due v.a. aleatorie che abbiamo visto precedentemente, e cioè la v.a. uniformemente distribuita su $[a, b]$ e quella esponenziale, sono in effetti *assolutamente continue*.

La dimostrazione è semplice, ma può essere istruttiva. Vediamo la v.a. uniformemente distribuita su $[0, 1]$. La derivata F' della f.r. si calcola subito

$$F'_X(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1, & \text{se } 0 < t < 1 \\ 0, & \text{se } t > 1. \end{cases}$$

Si noti che F' non esiste in 0 e in 1, ma poco male, l'integrale di Riemann (e quello di Lebesgue) non se ne accorgono. Se $0 \leq a \leq t \leq 1$, allora

$$\int_a^t F'(s) ds = t - a = F(t) - F(a).$$

Se invece $a < 0 < t \leq 1$,

$$\int_a^t F'(s)ds = \int_0^t F'(s)ds = t = F(t) - F(a).$$

Un calcolo del tutto analogo si può fare se $t > 1$.

Fare la verifica per la v.a. esponenziale può essere un esercizio istruttivo.

Variabile aleatoria normale

La v.a. più importante nel calcolo delle probabilità è la cosiddetta *normale*.

Come abbiamo visto prima, una variabile aleatoria assolutamente continua può essere assegnata tramite la sua funzione di ripartizione oppure tramite la sua densità. La normale viene assegnata mediante la sua densità e vedremo tra poco perché.

La densità della v.a. normale di parametri $\sigma > 0$ e $\mu \in \mathbb{R}$ è la funzione

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

Dobbiamo chiederci, prima di andare avanti, se questa è la densità di una v.a. La positività, la continuità e il suo comportamento all'infinito promettono bene, resta da vedere se

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} = 1.$$

Con qualche sostituzione di variabili, il calcolo si riduce alla dimostrazione che

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}.$$

La dimostrazione è un bell'esercizio che combina integrali doppi, integrali impropri e il cambiamento di variabile con il calcolo in coordinate polari. Se non lo avete visto in qualche corso di analisi, date un'occhiata alla voce *Gaussian integral* su Wikipedia.

La v.a. normale di parametri σ e μ si denota con $N(\mu, \sigma^2)$.

Sappiamo che la funzione e^{-x^2} non ammette una primitiva esprimibile con funzioni elementari e che resiste a tutti i cambiamenti di variabili, integrazioni

per parti e altre diavolerie. Per questo motivo la funzione di ripartizione F_X non è esprimibile come una funzione elementare.

Ma noi abbiamo bisogno di conoscere, almeno approssimativamente, i valori della funzione di ripartizione delle $N(0, 1)$, cioè di

$$F_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2}} ds.$$

Questi valori si trovano ad esempio sul sito

www.mathsisfun.com/data/standard-normal-distribution-table.html

Se abbiamo bisogno dei valori della f.r di una $N(\mu, \sigma^2)$, facciamo una sostituzione di variabili.

Variabili aleatorie gamma

Iniziamo a studiare la *funzione* gamma. Si tratta di una funzione della variabile α , definita per $\alpha > 0$ e a valori reali positivi nel modo seguente:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx.$$

Tranne che per valori particolari di α l'integrale non è calcolabile in termini elementari.

È facile fare il calcolo se α è un numero naturale. Intanto

$$\Gamma(1) = \int_0^{+\infty} e^{-x} dx = 1.$$

Per $\alpha > 0$ con un'integrazione per parti si ottiene

$$\Gamma(\alpha + 1) = \int_0^{+\infty} x^\alpha e^{-x} dx = -x^\alpha e^{-x} \Big|_0^{+\infty} + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha).$$

Combinando i due risultati si trova che $\Gamma(2) = 1$ e, per induzione, $\Gamma(n) = (n-1)!$. Abbiamo così trovato una funzione analitica che assume per $\alpha = n$ il valore $(n-1)!$.

In seguito calcoleremo altri valori della Γ , ma ora introduciamo le v.a. gamma. Dati due parametri positivi $\alpha > 0$ e $\lambda > 0$, la densità è

$$f(t) = \begin{cases} 0, & \text{se } t \leq 0 \\ \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda t}, & \text{se } t > 0. \end{cases}$$

Si noti che per $\alpha = 1$ ritroviamo la v.a. esponenziale.

Osservazione

Per trovare altri valore interessanti per i quali la Γ è calcolabile in termini esatti, facciamo qualche premessa.

Se X è una v.a., allora anche X^2 lo è. Infatti se $t < 0$, $(X^2 \leq t) = \emptyset$, $(X^2 = 0) = (X = 0)$, se invece $t > 0$, allora $(X^2 \leq t) = (-\sqrt{t} \leq X \leq \sqrt{t})$ e in conclusione $(X \leq t) \in \mathcal{A}$ per ogni $t \in \mathbb{R}$.

Per le v.a. discrete abbiamo tralasciato questa discussione perché superflua.

Supponiamo ora che X sia assolutamente continua e che la sua funzione di ripartizione sia F .

Facciamo vedere che anche X^2 è una v.a. assolutamente continua e troviamo la sua funzione di ripartizione.

La sua funzione di ripartizione è nulla per $t < 0$, mentre per $t \geq 0$ si ha $G(t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = F(\sqrt{t}) - F(-\sqrt{t})$.

La funzione $\psi(t) = \sqrt{t}$ è assolutamente continua su \mathbb{R}^+ . Si tratta di un classico esempio di funzione assolutamente continua che non è lipschitziana.

Sia infatti $\{[a_k, b_k]\}$ una famiglia di n intervalli a due a due disgiunti contenuti in \mathbb{R}^+ . Possiamo sempre supporre che $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq a_3 < \dots \leq a_n < b_n$.

Osserviamo che se $a < b$ e se $0 \leq h \leq a < b$, allora, per la concavità,

$$\sqrt{b} - \sqrt{a} \leq \sqrt{b-h} - \sqrt{a-h}.$$

La prima differenza è infatti uguale all'ampiezza dell'intervallo moltiplicata per la derivata della funzione in un punto ξ compreso tra a e b . La seconda differenza è invece uguale all'ampiezza dell'intervallo (che è la stessa) moltiplicata per la derivata della funzione in un punto η compreso tra $a-h$ e $b-h$. La disuguaglianza segue dal fatto che $\eta < \xi$ e che la derivata è decrescente.

“Spingiamo” ora gli intervalli a sinistra conservandone l'ampiezza ma accostandoli l'uno all'altro. Otteniamo così gli intervalli $\{[a'_k, b'_k]\}$ tali che $a'_k - b'_k = a_k - b_k$, $1 \leq k \leq n$, inoltre

$$0 = a'_1 < b'_1 = a'_2 < b'_2 = a'_3 < \dots = a'_n < b'_n.$$

Per la disuguaglianza ottenuta prima, si ha, per la proprietà telescopica,

$$\sum_{k=1}^n (\sqrt{b_k} - \sqrt{a_k}) \leq \sum_{k=1}^n (\sqrt{b'_k} - \sqrt{a'_k}) = \sqrt{b'_n} =$$

$$= \sqrt{(b_1 - a_1) + (b_2 - a_2) + \cdots + (b_n - a_n)}.$$

Fissato dunque ε , basta prendere $\delta = \varepsilon^2$ e si vede che $\psi(t) = \sqrt{t}$ è assolutamente continua.

Un'altra proposizione che ci serve è la seguente:

Se φ e ψ sono assolutamente continue, allora anche la loro composta $\varphi \circ \psi$ lo è.

La dimostrazione è semplice. Fissiamo $\varepsilon > 0$ e sia δ il corrispondente numero positivo relativo all'assoluta continuità della φ . In corrispondenza a δ sia ora η il numero positivo che corrisponde all'assoluta continuità della ψ . Se allora $\{[a_k, b_k]\}$ è una famiglia di n intervalli disgiunti del dominio di ψ , da $\sum_{k=1}^n (b_k - a_k) < \eta$ segue che $\sum_{k=1}^n |\psi(b_k) - \psi(a_k)| < \delta$, da cui si ha $\sum_{k=1}^n |\varphi(\psi(b_k)) - \varphi(\psi(a_k))| < \varepsilon$, cioè la tesi.

Esempio

Se $X \sim N(0, \sigma^2)$, allora X^2 è una $\Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$.

Abbiamo imparato che X^2 è una v.a. assolutamente continua e che la sua funzione di ripartizione G è uguale a $F(\sqrt{t}) - F(-\sqrt{t})$ (dove F è la funzione di ripartizione di X).

Esplicitamente, per $t > 0$,

$$G(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\sqrt{t}} e^{-\frac{s^2}{2\sigma^2}} ds - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-\sqrt{t}} e^{-\frac{s^2}{2\sigma^2}} ds.$$

Ricordando che la derivata di $F(\sqrt{t})$ è uguale a $\frac{1}{2\sqrt{t}}F'(\sqrt{t})$, si vede che la densità $g(t)$ è

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{2\sqrt{t}} \left(2 e^{-\frac{t}{2\sigma^2}} \right) = \frac{1}{\sqrt{2\pi}\sigma} t^{\frac{1}{2}-1} e^{-\frac{t}{2\sigma^2}}.$$

A parte la costante moltiplicativa, si vede che si tratta di una v.a. $\Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$.

Ma la costante moltiplicativa deve necessariamente coincidere con $\frac{\lambda^\alpha}{\Gamma(\alpha)}$ e da

$$\alpha = \frac{1}{2} \quad \lambda = \frac{1}{2\sigma^2} \quad \text{e} \quad \frac{\frac{1}{\sqrt{2\sigma}}}{\Gamma(\frac{1}{2})} = \frac{1}{\sqrt{2\pi}\sigma}$$

si deduce che $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Dall'equazione $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ possiamo calcolarci ricorsivamente i valori di $\Gamma(\frac{2k+1}{2})$ per ogni numero naturale k .

5 Caratteristiche numeriche delle v.a. assolutamente continue

Per una variabile aleatoria assolutamente continua che ha densità $f(t)$ la speranza matematica si definisce come

$$E(X) = \int_{-\infty}^{+\infty} t f(t) dt,$$

se l'integrale converge assolutamente (o, equivalentemente, se esiste l'integrale nel senso di Lebesgue).

Per l'integrale di Riemann si definisce il concetto di *integrale improprio* che potrebbe confondere le idee.

L'integrale improprio della g su \mathbb{R} esiste se è finito il limite

$$\lim_{a \rightarrow -\infty, b \rightarrow +\infty} \int_a^b f(t) dt,$$

(con ovvie modifiche nel caso degli integrali su semirette). La funzione

$$g(t) = \frac{\sin t}{t}$$

è integrabile *in senso improprio* sulla semiretta $[1, +\infty[$. Ma non è integrabile alla Lebesgue, perché non è integrabile $|g|$.

Conviene rileggersi la discussione che è stata fatta sull'esistenza della speranza matematica delle v.a. discrete e la serie armonica a segni alterni, della quale la funzione considerata sopra è una stretta parente.

Faremo ora vedere quanto sia naturale la definizione di speranza matematica delle v.a. assolutamente continue approssimandole con delle v.a. discrete.

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità e sia X una variabile aleatoria assolutamente continua avente densità $f(t)$.

Per motivi tecnici (facilmente superabili, ma non molto istruttivi e per questo li tralasciamo) sarà comodo supporre che la X sia limitata, esista cioè una costante C tale che $|X(\omega)| < C$ per ogni $\omega \in \Omega$.

Fissiamo un $n \in \mathbb{N}$ e suddividiamo l'intervallo $] - C, C]$ in 2^{n+1} parti di lunghezza uguale $I_k =] - C + k \frac{C}{2^n}, -C + (k+1) \frac{C}{2^n}]$, con $0 \leq k \leq 2^{n+1} - 1$.

Per comodità scriveremo $I_k =]t_k, t_{k+1}]$

Siccome X è una v.a., $(X \in I_k) \in \mathcal{A}$.

Consideriamo ora, al variare di n , le v.a. discrete $X'_n =$

$$\begin{pmatrix} t_0 & t_1 & \dots & t_k & \dots & t_{2^n-2} \\ \dots & \dots & \dots & P(X \in I_k) & \dots & \dots \end{pmatrix}$$

e $X''_n =$

$$\begin{pmatrix} t_1 & t_2 & \dots & t_{k+1} & \dots & t_{2^n-1} \\ \dots & \dots & \dots & P(X \in I_k) & \dots & \dots \end{pmatrix}$$

Naturalmente si ha che $X'_n(\omega) \leq X(\omega) \leq X''_n(\omega)$ per ogni $\omega \in \Omega$ e per ogni $n \geq 1$.

Calcoliamo le speranze matematiche di queste v.a. discrete.

$$\begin{aligned} E(X'_n) &= \sum_{k=0}^{2^{n+1}-1} t_k P(X \in I_k) = \sum_{k=0}^{2^{n+1}-1} t_k \int_{t_k}^{t_{k+1}} f(t) dt = \sum_{k=0}^{2^{n+1}-1} \int_{t_k}^{t_{k+1}} t_k f(t) dt \leq \\ &\leq \sum_{k=0}^{2^{n+1}-1} \int_{t_k}^{t_{k+1}} t f(t) dt = \int_{-C}^C t f(t) dt = \int_{-\infty}^{+\infty} t f(t) dt. \end{aligned}$$

Analogamente si calcola

$$\begin{aligned} E(X''_n) &= \sum_{k=0}^{2^{n+1}-1} t_{k+1} P(X \in I_k) = \sum_{k=0}^{2^{n+1}-1} t_{k+1} \int_{t_k}^{t_{k+1}} f(t) dt = \\ &= \sum_{k=0}^{2^{n+1}-1} \int_{t_k}^{t_{k+1}} t_{k+1} f(t) dt \geq \sum_{k=0}^{2^n-1} \int_{t_k}^{t_{k+1}} t f(t) dt = \int_{-C}^C t f(t) dt = \int_{-\infty}^{+\infty} t f(t) dt. \end{aligned}$$

D'altro canto

$$\begin{aligned} E(X''_n) - E(X'_n) &= \sum_{k=0}^{2^n-1} (t_{k+1} - t_k) P(X \in I_k) = \sum_{k=0}^{2^n-1} \frac{C}{2^n} P(X \in I_k) = \\ &= \frac{C}{2^{n+1}} \sum_{k=0}^{2^n-1} P(X \in I_k) = \frac{C}{2^n}. \end{aligned}$$

In conclusione, essendo, per ogni $n \in \mathbb{N}$,

$$E(X_n'') \geq E(X) \geq E(X_n'),$$

la definizione di $E(X) = \int_{-\infty}^{+\infty} tf(t) dt$ è l'unica scelta possibile.

Anche nel caso assolutamente continuo si pone il problema dell'esistenza della speranza matematica. Il controesempio è una v.a. che ha qualche interesse nel calcolo delle probabilità. Si tratta della v.a. *di Cauchy* la cui densità è data da

$$f(t) = \frac{1}{\pi} \frac{1}{1+t^2}.$$

Si vede facilmente che l'integrale della f su \mathbb{R} è 1, ma si vede altrettanto facilmente che l'integrale

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{s}{1+s^2} ds$$

non esiste (neanche come integrale improprio).

Si può dimostrare che se X ed Y sono indipendenti ed hanno legge $N(0, 1)$, allora $\frac{X}{Y}$ ha legge di Cauchy.

Linearità della speranza matematica

Il ragionamento che abbiamo visto prima sull'approssimazione della speranza matematica di una v.a. assolutamente continua mediante speranze matematiche di successioni di v.a. discrete è alla base di diverse dimostrazioni che riguardano le v.a. assolutamente continue.

Vediamo un accenno alla dimostrazione dell'additività della speranza matematica. Se X e Y sono due v.a. assolutamente continue (e per semplicità limitate), allora esistono successioni crescenti di v.a. discrete $\{X_n'\}$ e $\{Y_n'\}$ (analoghe a quelle viste nella costruzione precedente) tali che $E(X_n') \rightarrow E(X)$ e $E(Y_n') \rightarrow E(Y)$. Ne segue che $E(X_n') + E(Y_n') = E(X_n' + Y_n') \rightarrow E(X + Y)$. Ma $E(X_n') + E(Y_n')$ tende a $E(X) + E(Y)$, da cui la tesi.

Lo stesso ragionamento ci consente di dimostrare l'omogeneità: se c è una costante reale e se $E(X_n') \rightarrow E(X)$, $E(cX_n') \rightarrow E(cX)$. Ma $E(cX_n') = cE(X_n') \rightarrow cE(X)$.

Sono utili anche le seguenti proprietà.

- 1) Se $P(X \geq 0) = 1$, allora $E(X) \geq 0$.

2) Se $P(X \geq Y) = 1$, allora $E(X) \geq E(Y)$ e l'uguaglianza vale se e solo se $P(X = Y) = 1$.

3) $|E(X)| \leq E(|X|)$.

Esempi

a) Se X è la v.a. uniformemente distribuita su $[a, b]$, allora

$$E(X) = \int_a^b \frac{t-a}{b-a} dt = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

b) Se $X \sim N(0, 1)$, la sua speranza matematica è nulla, poiché la funzione integranda è dispari e l'insieme di integrazione è simmetrico rispetto all'origine:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt = 0$$

La speranza matematica di una $N(\mu, \sigma^2)$ invece è uguale a μ . Basta fare la sostituzione di variabili $t = s - \mu$.

c) Della gamma ci occuperemo più tardi (incluso anche il caso dell'esponenziale).

Varianza e momenti di ordine superiore

Se X è una v.a. assolutamente continua e $f(t)$ è la sua densità, si definisce *varianza* di X il valore dell'integrale (se è assolutamente convergente)

$$\text{Var}(X) = \int_{-\infty}^{+\infty} f(t) (t - E(X))^2 dt.$$

Non è difficile vedere che ci sono v.a. assolutamente continue che ammettono speranza matematica, ma non varianza.

Possiamo considerare ad esempio la funzione $r(t) = \frac{1}{(|t|+1)(1+t^2)}$. Si tratta di una funzione pari. Usando la formula di Hermite si calcola l'integrale di $\frac{1}{(t+1)(1+t^2)}$ tra 0 e $+\infty$ e si considera

$$\gamma = 2 \int_0^{+\infty} \frac{dt}{(t+1)(1+t^2)}.$$

La funzione

$$\frac{1}{\gamma} \frac{1}{(|t|+1)(1+t^2)}$$

è la densità di una v.a. assolutamente continua X che ammette speranza matematica, ma non ammette varianza.

Sviluppando il quadrato e sfruttando la linearità dell'integrale, si riottiene l'identità di König

$$\text{Var}(X) = \int_{-\infty}^{+\infty} t^2 f(t) dt - E(X)^2 = E(X^2) - E(X)^2.$$

La varianza gode di diverse proprietà interessanti.

- 1) $\text{Var}(aX) = a^2 \text{Var}(X)$.
- 2) $\text{Var}(X + c) = \text{Var}(X)$.
- 3) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E((X - E(X))(Y - E(Y)))$.

L'ultimo addendo si chiama *covarianza* di X e Y e si denota con $\text{Cov}(X, Y)$. Ne parleremo in maniera più estesa in seguito. Il suo segno indica se i due fenomeni descritti da X e da Y hanno la stessa "tendenza" o la tendenza opposta. Ad esempio, se osserviamo su una popolazione le caratteristiche dell'essere sovrappeso e l'aver problemi cardiaci, la covarianza sarà positiva e diremo che i due fenomeni sono positivamente correlati. Se invece si confronta l'esito dell'esame di CALPS e l'anno in corso del candidato, si osserverà covarianza negativa e i due fenomeni sono negativamente correlati. Magari ci potranno essere delle eccezioni, ma in media chi fuori corso da tre anni, per dire, avrà una resa inferiore allo studente in corso.

4) Se X ammette varianza e se $|Y| \leq X$, allora anche Y ammette varianza. Segue dai vari teoremi del confronto funzione-funzione, funzione serie e serie-funzione.

Come nel caso discreto, anche per le v.a. assolutamente continue ha senso definire i *momenti di ordine superiore*. Per $k \geq 2$, si dice momento di ordine k il valore di $E(X^k)$.

Le stesse disuguaglianze che abbiamo usato nel caso discreto implicano anche nel caso assolutamente continuo che se X e Y ammettono momento di ordine k , anche $X + Y$ lo ammette e inoltre, se $2 \leq h < k$ e se X ammette momento k -esimo, allora ammette anche momento h -esimo.

Inoltre è facile costruire esempi di v.a. che ammettono momento di ordine k , ma non quello di ordine $k + 1$. Basta prendere una densità proporzionale a $\frac{1}{1+|t|^{k+2}}$.

Ad esempio, si ha che $\int_{-\infty}^{+\infty} \frac{1}{1+t^4} dt = \frac{\pi\sqrt{2}}{2}$
 Ovviamente

$$\frac{2}{\pi\sqrt{2}(1+t^4)}$$

è la densità di una v.a. assolutamente continua che ammette varianza, ma non il momento di ordine 3.

Vettori aleatori

Se (Ω, \mathcal{A}, P) è uno spazio di probabilità un'applicazione $X : \Omega \rightarrow \mathbb{R}^m$ si dice *vettore aleatorio assolutamente continuo*, se le componenti di X , X_i , con $1 \leq i \leq m$, sono assolutamente continue.

La funzione $F(t_1, t_2, \dots, t_m) = P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_m \leq t_m)$ si chiama *funzione di ripartizione* (congiunta) del vettore aleatorio X .

Se esiste una funzione $f(t_1, t_2, \dots, t_m)$ integrabile e non negativa su \mathbb{R}^m tale che

$$F(t_1, t_2, \dots, t_m) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_m} f(s_1, s_2, \dots, s_m) ds_1 ds_2 \dots ds_m,$$

la f si dice densità (congiunta) di X . Dall'identità precedente segue che l'integrale di f su \mathbb{R}^m è uguale a 1.

Le funzioni di ripartizione marginali sono, come al solito, $F_i(t_i) = P(X_i \leq t_i)$, mentre densità f_i delle coordinate X_i si chiamano densità *marginali*.

Le f.r. e densità marginali si calcolano nel modo seguente.

$$F_i(t_i) = P(X_i \leq t_i) = P(X_1 \in \mathbb{R}, \dots, X_i \leq t_i, \dots, X_m \in \mathbb{R}) = \int_{-\infty}^{t_i} \left(\int_{\mathbb{R}^{m-1}} f(u_1, \dots, v, \dots, u_m) du_1 \dots du_m \right) dv$$

e quindi la densità marginale i -esima è, quasi ovunque,

$$\int_{\mathbb{R}^{m-1}} f(u_1, \dots, t_i, \dots, u_m) du_1 \dots du_m$$

In dimensione 2 o superiori l'analogo del teorema fondamentale del calcolo integrale non ammette una formulazione così elegante e conclusiva, perciò in questo caso ci dovremo accontentare di ragionamenti un po' alla buona,

sorvolando sul rigore matematico (che sarebbe in effetti in questo caso sproporzionato) e fidarci della plausibilità dei risultati. Gli esempi che considereremo utilizzano densità sufficientemente “docili” che giustificano questo approccio pragmatico.

Poiché non c'è nessuna differenza sostanziale tra il caso bidimensionale ed il caso in cui $m > 2$, ci limiteremo per il momento al caso $m = 2$, salvo enunciare poi i risultati nel caso generale.

Supponiamo allora che il vettore aleatorio $X = (X_1, X_2)$ ammetta densità e che quindi si ha

$$F(s, t) = \int_{-\infty}^s \int_{-\infty}^t f(u, v) \, dudv$$

se deriviamo questa identità rispetto ad s , otteniamo quasi ovunque

$$\frac{\partial F(s, t)}{\partial s} = \int_{-\infty}^t f(s, v) \, dv$$

e, con un'ulteriore derivazione (sempre quasi ovunque)

$$\frac{\partial^2 F(s, t)}{\partial t \partial s} = f(s, t).$$

Quest'ultima identità vale, nei casi che si presenteranno, con l'eccezione di un insieme di misura nulla.

Più in generale, in dimensione m , si ha che

$$f = \frac{\partial^m F}{\partial t_1, \dots, \partial t_m}.$$

Indipendenza

Parlando di v.a. *discrete*, abbiamo detto che data un m -upla di v.a. X_1, X_2, \dots, X_m esse si dicono indipendenti, se per ogni scelta di m sottoinsiemi della retta reale A_i , $1 \leq i \leq m$ si ha che

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_m \in A_m) = \prod_{i=1}^m P(X_i \in A_i).$$

Con le v.a. (assolutamente) continue, dobbiamo essere più restrittivi. e non possiamo scegliere dei sottoinsiemi arbitrari della retta reale.

Ricordiamo che quando si è parlato, nella teoria dell'integrazione alla Lebesgue, delle funzioni misurabili, si è dimostrato il seguente risultato

Teorema

Se (Ω, \mathcal{A}) è uno spazio misurabile, per una funzione g a valori reali definita su Ω le seguenti due proprietà sono equivalenti:

- A) $\{\omega : g(\omega) \leq t\} = g^{-1}(]-\infty, t]) \in \mathcal{A}$ per ogni $t \in \mathbb{R}$;
- B) $g^{-1}(B) \in \mathcal{A}$ per ogni sottoinsieme boreliano B della retta reale.

Ricordiamo che la σ -algebra dei boreliani \mathcal{B} è la più piccola σ -algebra contenente gli insiemi aperti.

Poiché sappiamo che ci sono sottoinsiemi della retta reale che non appartengono a \mathcal{B} , la condizione imposta alle v.a. continue [e cioè che l'insieme $X^{-1}(]-\infty, t])$ appartenga ad \mathcal{A}] ci costringe a limitarci ai boreliani nella scelta degli insiemi che compaiono della definizione di indipendenza.

L'indipendenza di m variabili continue si esprime dunque con la condizione

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_m \in B_m) = \prod_{i=1}^m P(X_i \in B_i),$$

che deve valere per ogni scelta dei boreliani B_i , $1 \leq i \leq m$.

In particolare (visto che le semirette sono dei boreliani) si ha che

$$P(X_1 \leq t_1, X_2 \leq t_2, \dots, X_m \leq t_m) = \prod_{i=1}^m P(X_i \leq t_i),$$

ovvero

$$F(t_1, t_2, \dots, t_m) = \prod_{i=1}^m F_i(t_i),$$

Usando il metodo visto prima per calcolare la densità congiunta, considerando la derivata parziale m -esima mista, si ottiene che

$$f(t_1, t_2, \dots, t_m) = \prod_{i=1}^m f_i(t_i) \text{ q.o.}$$

Il percorso inverso è percorribile, ma un po' laborioso. L'ultima identità in effetti caratterizza l'indipendenza delle v.a. X_1, X_2, \dots, X_m .

Un teorema di rappresentazione

Iniziamo con delle considerazioni preliminari che sono di per sé importanti.

Sia $Z = (X, Y)$ un vettore aleatorio di densità $f(s, t)$. Consideriamo un rettangolo $R = I \times J$ e calcoliamo $P(Z \in R)$. Indichiamo con a e b e, rispettivamente, c e d gli estremi di I e J . Allora

$$\begin{aligned} P(X \in I, Y \in J) &= P(X \leq b, Y \in J) - P(X \leq a, Y \in J) = \\ &P(X \leq b, Y \leq d) - P(X \leq b, Y \leq c) - P(X \leq a, Y \leq d) + P(X \leq a, Y \leq c) = \\ &F(b, d) - F(b, c) - F(a, d) + F(a, c) = \iint_{A_{b,d}} f(s, t) ds dt - \iint_{A_{b,c}} f(s, t) ds dt - \\ &\iint_{A_{a,d}} f(s, t) ds dt + \iint_{A_{a,c}} f(s, t) ds dt, \end{aligned}$$

dove abbiamo indicato con $A_{b,d}$ il quadrante di vertice (b, d) , $\{(s, t) : s \leq b, t \leq d\}$, con analogo significato per gli altri tre insiemi di integrazione.

Si vede facilmente che l'ultima espressione è uguale all'integrale della f su R , riassumendo:

$$P(Z \in R) = \iint_R f(s, t) ds dt.$$

Può essere utile ricordare che la teoria dell'integrale di Lebesgue permette di concludere che dall'essere vera questa identità per ogni rettangolo R , segue che

$$P(Z \in B) = \iint_B f(s, t) ds dt$$

perché ogni boreliano B è approssimabile con unioni finite di rettangoli.

Ci proponiamo di dimostrare il seguente teorema:

Teorema

Se X e Y sono due v.a. assolutamente continue che ammettono varianza ed hanno densità congiunta $f(s, t)$, allora esiste $E(XY)$ e inoltre

$$E(XY) = \iint_{\mathbb{R}^2} st f(s, t) ds dt.$$

L'esistenza della speranza matematica di XY segue dall'identità

$$XY = \frac{1}{2}((X + Y)^2 - X^2 - Y^2)$$

e dall'esistenza delle varianze.

Supponiamo, per comodità, che $|X|$ e $|Y|$ siano limitate da C . Suddividiamo $[-C, C]$ in 2^{n+1} intervallini I_k come fatto in precedenza e consideriamo

$$\begin{aligned} \sum_{j,k} t_j^n t_k^n P(\in I_j \times I_k) &= E(X'_n Y'_n) = \sum_{j,k} t_j^n t_k^n \iint_{R_{j,k}} f(s, t) ds dt = \\ &= \sum_{j,k} \iint_{R_{j,k}} t_j^n t_k^n f(s, t) ds dt, \end{aligned}$$

dove $R_{j,k} = I_j \times I_k$.

Valutiamo ora la differenza

$$\begin{aligned} \left| \sum_{j,k} \iint_{R_{j,k}} t_j^n t_k^n f(s, t) ds dt - \sum_{j,k} \iint_{R_{j,k}} st f(s, t) ds dt \right| &\leq \\ &= \sum_{j,k} \iint_{R_{j,k}} |t_j^n t_k^n - st| f(s, t) ds dt. \end{aligned}$$

Poiché $|t_j^n t_k^n - st| \leq |t_j^n t_k^n - t_j^n t + t_j^n t - st| \leq |t_j^n t_k^n - t_j^n t| + |t_j^n t - st| \leq C|t_k^n - t| + C|t_j^n - s| \leq 2C \frac{1}{2^{n+1}} = \frac{C}{2^n}$ e quest'ultima maggiorazione tende a zero per n che tende all'infinito, l'ultima espressione si lascia maggiorare da

$$\frac{C}{2^n} \sum_{j,k} \iint_{R_{j,k}} f(s, t) ds dt = \frac{C}{2^n} \iint_{\mathbb{R}^2} f(s, t) ds dt = \frac{C}{2^n},$$

da cui la tesi.

Una formula moltiplicativa

La formula appena dimostrata ci permette di dimostrare per le v.a. assolutamente continue una proprietà simile a quella che abbiamo visto per le v.a. discrete, e cioè

Se X e Y sono due v.a. assolutamente continue indipendenti che ammettono speranza matematica, allora

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

Si ha infatti

$$E(XY) = \iint_{\mathbb{R}^2} stf(s,t) dsdt = \iint_{\mathbb{R}^2} stf_1(s)f_2(t) dsdt = \\ \int_{\mathbb{R}} sf_1(s) ds \int_{\mathbb{R}} tf_2(t) dt = E(X)E(Y).$$

Il penultimo passaggio segue dal teorema di Fubini.

Merita osservare che, per il teorema di Fubini, in questo caso la formula vale se X e Y ammettano speranza matematica e l'ipotesi sulle varianze è superflua.

L'indipendenza implica la formula moltiplicativa, ma il viceversa non è vero.

Esempio

Siano $\Omega =]0, 1]$, $\mathcal{A} = \mathcal{B}$ e $P = \lambda$. Dividiamo $]0, 1]$ in tre parti uguali $J_k =]\frac{k-1}{3}, \frac{k}{3}]$, per $1 \leq k \leq 3$, e definiamo $X = -\mathbb{1}_{J_1} + \mathbb{1}_{J_3}$. Si ha ovviamente $X^2 = \mathbb{1}_{J_1 \cup J_3} = Y$ e $X^3 = XY = X$.

Si ha dunque $E(XY) = E(X^3) = 0$ ed allo stesso tempo $E(X)E(Y) = 0$ perché il primo fattore è nullo. Vale dunque per questa coppia di v.a. la formula moltiplicativa, ma X e Y non sono indipendenti, essendo addirittura funzionalmente dipendenti.

Volendo essere più formali e volendo verificare la definizione, si ha che

$$P(X = 1, Y = 1) = P(X = 1) = \frac{1}{3} \text{ mentre } P(Y = 1) = \frac{2}{3}$$

$$\text{e dunque } P(X = 1)P(Y = 1) = \frac{2}{9}.$$

Osservazione

Sappiamo che $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. Ricordiamo che

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y).$$

Ne segue che se $E(XY) = E(X)E(Y)$, allora la varianza è additiva per quella coppia di v.a..

Definizione

Due variabili si dicono *non correlate* se $E(XY) = E(X)E(Y)$.

Detto in altre parole, l'indipendenza implica la non correlazione e la varianza è additiva se le v.a. sono non correlate e in particolare se sono indipendenti.

Vettori aleatori uniformemente distribuiti

Parleremo ora di una classe molto significativa di vettori aleatori. Si tratta dei vettori aleatori *uniformemente distribuiti su un insieme* A .

Se A è un sottoinsieme sufficientemente regolare di \mathbb{R}^n , diremo che un vettore aleatorio X è uniformemente distribuito su A , se la sua densità f è costante su A e nulla fuori di A , cioè $f = c \mathbb{1}_A$. Negli esempi che tratteremo, A sarà misurabile secondo Peano-Jordan, anzi sarà assegnato descrivendone la frontiera con tratti di curve continue.

Poiché

$$\int_{\mathbb{R}^n} c \mathbb{1}_A d\lambda_n = c \int_{\mathbb{R}^n} \mathbb{1}_A d\lambda_n = c \lambda_n(A) = 1,$$

si ha necessariamente che

$$c = \frac{1}{\lambda_n(A)}.$$

Vediamo ora qualche esempio.

1) Sia $Z = (X, Y)$ il vett. a. uniformemente distribuito sul disco unitario $\{(s, t) : s^2 + t^2 \leq 1\}$.

Le domande che si pongono sono:

- a) determinare la densità congiunta di X ;
- b) determinare le densità marginali;
- c) vedere se X e Y sono non correlate;
- d) vedere se X e Y sono indipendenti.

Affrontiamo le singole domande.

a) Poiché $\lambda_2(A) = \pi$, la densità congiunta è

$$f(s, t) = \begin{cases} \frac{1}{\pi} & \text{se } (s, t) \in A \\ 0 & \text{altrimenti.} \end{cases}$$

b) La densità (marginale) di X si calcola con la formula vista in precedenza. Se $|t| > 1$, la retta verticale $x = t$ non interseca il disco, quindi la f ,

ristretta ad essa, è identicamente nulla e quindi $f_1(t) = 0$. Se invece $|t| \leq 1$, la retta interseca il disco nel segmento $[-\sqrt{1-t^2}, \sqrt{1-t^2}]$ e dunque

$$f_1(t) = \begin{cases} 0 & \text{se } |t| > 1 \\ \int_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} \frac{1}{\pi} dt = \frac{2}{\pi} \sqrt{1-t^2} & \text{se } |t| \leq 1. \end{cases}$$

Per simmetria si ha che

$$f_2(s) = \begin{cases} 0 & \text{se } |s| > 1 \\ \frac{2}{\pi} \sqrt{1-s^2} & \text{se } |s| \leq 1. \end{cases}$$

c) Si ha che $E(X) = \frac{2}{\pi} \int_{-1}^1 t \sqrt{1-t^2} dt = 0$, poiché la funzione integrabile è dispari e l'intervallo è simmetrico rispetto all'origine. Per lo stesso motivo $E(Y) = 0$.

Resta da calcolare

$$E(XY) = \iint_{\mathbb{R}^2} stf(s,t) ds dt = \frac{2}{\pi} \int_{-1}^1 t \left(\int_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} s ds \right) dt = 0.$$

L'integrale doppio è nullo, poiché l'integrale interno è nullo, essendo ancora una volta la funzione integrabile dispari e l'intervallo di integrazione è simmetrico rispetto all'origine.

Dunque $E(XY) = E(X)E(Y)$ e le due coordinate sono non correlate.

Fossero state correlate, avremmo finito di soffrire, perché avremmo potuto concludere, in base al teorema dimostrato prima, che sono dipendenti. Così invece dobbiamo ancora rispondere alla domanda d).

La risposta non è complicata (anzi, di solito richiede meno calcoli che la domanda riguardante la non correlazione), ma richiede una maggiore attenzione alla teoria. Quindi meno calcoli, ma più ragionamenti.

Conviene richiamare qualche nozione appresa quando si è studiato l'integrale di Lebesgue.

Due funzioni integrabili f e g si dicono *equivalenti* se $f(x) = g(x)$ quasi ovunque, ovvero se l'insieme $\{x : f(x) \neq g(x)\}$ ha misura nulla. Le funzioni integrabili secondo Lebesgue hanno questa proprietà di essere inafferrabili, per così dire, nel senso che data f e scelto un qualunque punto x del suo dominio, il valore $f(x)$ non è definito, perché esistono infinite funzioni equivalenti ad f e ciascuna di queste può assumere in x un valore arbitrario o

addirittura non essere definita. Pensiamo ad esempio ad F' , la densità della v.a. uniformemente distribuita su $[0, 1]$, che non è definita in 0 e in 1.

Per dire dunque che due funzioni f e g sono *diverse*, la misura dell'insieme $\{x : f(x) \neq g(x)\}$ deve essere strettamente positiva.

Vediamo ora di rispondere alla domanda d). Quanto diremo ha valenza generale e conviene prestare molta attenzione ai ragionamenti che seguono che non riguardano il particolare esercizio, ma si applicano in tutte le circostanze analoghe.

La funzione di ripartizione congiunta è uguale a $\frac{1}{\pi}$ nel disco e zero altrove, mentre $f_1(s)f_2(t) = \frac{4}{\pi^2} \sqrt{(1-s^2)(1-t^2)}$. Si vede ictu oculi che le due funzioni sono diverse, ma per dimostrarlo in maniera soddisfacente ci sono almeno due modi.

Il primo è quello di osservare che $f_1(s)f_2(t) > 0$ in tutti i punti del quadrato $] - 1, 1[\times] - 1, 1[$, mentre $f(s, t) > 0$ solamente nel disco. Ci sono dunque i quattro “spicchi” accanto ai vertici del quadrato in cui $f_1(s)f_2(t) > f(s, t)$, e questo è un insieme di misura positiva.

Un altro metodo è quello di notare che $f(0, 0) = \frac{1}{\pi} < f_1(0)f_2(0) = \frac{4}{\pi^2}$ e che entrambe le funzioni sono continue nell'origine. Esiste allora un intorno (ad esempio circolare) U dove $f(s, t) < f_1(s)f_2(t)$. La conclusione segue ora dal fatto che $\lambda_2(U) > 0$.

Il coefficiente di correlazione

Il coefficiente di correlazione tra due v.a. che ammettono varianza si definisce nel modo seguente:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var X Var Y}}.$$

Rispetto alla covarianza ha il vantaggio di essere insensibile al cambiamento di scala. Se abbiamo due fenomeni che vengono misurati, ad esempio, in gradi centigradi (temperatura nel centro di Cosenza) e concentrazione di ozono in milligrammi su metro cubo (sempre nel centro di Cosenza, nello stesso periodo), il valore del coefficiente di correlazione non cambia se la temperatura la misuriamo in Fahrenheit e la concentrazione in libbre su piede cubo. Analiticamente,

$$|\rho(aX, bY)| = |\rho(X, Y)|.$$

Abbiamo messo il valore assoluto perché se $ab < 0$, il coefficiente di correlazione cambia di segno.

Proposizione (disuguaglianza di Schwarz)

Se X ed Y sono due v.a. che ammettono varianza, allora

$$E(XY)^2 \leq E(X^2)E(Y^2).$$

Consideriamo la v.a. $(|X| + t|Y|)^2$. Essendo non negativa, anche la sua speranza matematica lo è. Dunque

$$0 \leq E((|X| + t|Y|)^2) = E(X^2 + 2t|XY| + t^2Y^2) = E(X^2) + 2tE(|XY|) + t^2E(Y^2).$$

Vista come funzione della t , l'ultima espressione rappresenta una parabola con la concavità rivolta rispetto verso l'alto ed essendo sempre positiva o nulla, il discriminante deve essere negativo (o nullo), cioè

$$4E(|XY|)^2 - 4E(X^2)E(Y^2) \leq 0,$$

ovvero

$$E(|XY|)^2 \leq E(X^2)E(Y^2),$$

e dalla disuguaglianza $|E(XY)| \leq E(|XY|)$ segue immediatamente la tesi.

Corollario

Si ha $|\rho(X, Y)| \leq 1$.

Infatti

$$\rho(X, Y)^2 \leq \frac{E[(X - E(X))(Y - E(Y))]^2}{E((X - E(X))^2)E((Y - E(Y))^2)}$$

Questa disuguaglianza si ottiene applicando la disuguaglianza di Schwarz alle due v.a. $X - E(X)$ e $Y - E(Y)$, rispettivamente.

Dunque, per il corollario precedente, il valore del coefficiente di correlazione è sempre, in valore assoluto, minore o uguale ad uno. Esso vale zero se X ed Y sono non correlate.

La proposizione seguente chiarirà invece quale sia il rapporto tra X ed Y quando il coefficiente di correlazione vale 1 o -1 .

Proposizione. Se $|\rho(X, Y)| = 1$, allora $Y = aX + b$ quasi certamente.

Siano \bar{X} e \bar{Y} le normalizzazioni di X ed Y rispettivamente.

[Cioè $\bar{X} = \frac{X-E(X)}{\sigma(X)}$, e stessa cosa per \bar{Y} . La normalizzazione associa ad una v.a che ammette varianza una v.a. dello stesso tipo che ha speranza matematica nulla e varianza uguale a 1.

La normalizzazione di una v.a. $N(\mu, \sigma^2)$ è una $N(0, 1)$.]

Osserviamo intanto che

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = E\left(\frac{X - E(X)}{\sigma(X)} \frac{Y - E(Y)}{\sigma(Y)}\right) = E(\bar{X}\bar{Y}).$$

Supponiamo ora che $\rho(X, Y) = 1$. Allora

$$E[(\bar{X} - \bar{Y})^2] = E(\bar{X}^2) - 2E(\bar{X}\bar{Y}) + E(\bar{Y}^2) = 0.$$

Ma essendo $(\bar{X} - \bar{Y})^2 \geq 0$, ciò implica che $(\bar{X} - \bar{Y})^2 = 0$ quasi certamente e quindi che $\bar{X} - \bar{Y} = 0$ quasi certamente. Quest'ultima identità si può d'altronde riscrivere come

$$Y = \frac{\sigma(Y)}{\sigma(X)}X - \frac{\sigma(Y)}{\sigma(X)}E(X) + E(Y),$$

cioè la tesi.

Se $|\rho(X, Y)| = -1$ si rifanno considerazioni analoghe valutando $E[(\bar{X} + \bar{Y})^2]$. Si può osservare che in questo caso il coefficiente di X risulta negativo.

6 Simulazione

Siamo circondati, ogni giorno di più, da congegni elettronici che simulano il comportamento di eventi casuali.

Un qualunque videogioco, anche il più semplice, lancia dadi, distribuisce carte, propone tabelle di sudoku, fa giocare a video poker, roulette, bridge, etc. Tutte queste operazioni simulano giochi veri, in cui i dadi vengono lanciati da avversari in carne e ossa, le carte vengono spartite dagli avversari, la ruota della roulette viene fatta girare da croupier.

Naturalmente stiamo parlando ora solamente dell'aspetto "casuale", la grafica è un'altra cosa ed è di solito la cosa più difficile da realizzare.

I videogiochi dispongono dunque di un software che simula il gioco vero. La qualità del gioco (e il suo prezzo) dipende da quanto questa simulazione

riesce convincente. La simulazione sarebbe perfetta se il giocatore non si accorgesse (neanche con l'ausilio di tecniche statistiche) che chi mescola le carte o lancia due dadi etc non è una persona, ma un computer.

In videogiochi più complessi è importante che quando il giocatore “cammina” per un corridoio pieno di insidie di tutti i tipi, il mostro, o il personaggio armato di lanciafiamme, o la voragine che si apre improvvisamente, appaiano a caso e senza una regola riconoscibile e ripetitiva, altrimenti non c'è gusto.

Probabilmente ne sapete più di me, ma giusto per avere un semplice riferimento in comune, potete vedere qualche esempio sul sito gratuito 247games. Se provate a giocare a bridge, vedrete che il software distribuisce a voi e agli altri tre giocatori tredici carte a testa. Almeno all'apparenza lo fa senza ripetizione e senza distribuzioni assurde. Stessa cosa con il Black Jack o con il poker.

C'è da dire che la simulazione che più si conosce è quella dei giochi, ma che è stata inventata, ed è ancora usata, per risolvere problemi scientifici.

In questo capitolo cercheremo di capire con alcuni esempi quali sono i metodi che vengono utilizzati per la simulazione e concluderemo con un cenno al metodo di Monte Carlo.

Sulla generazione di numeri casuali

Per simulare delle variabili aleatorie c'è innanzitutto bisogno di disporre di un generatore di numeri casuali. Con questo si intende che bisogna disporre di un procedimento (ce ne sono tantissimi) che ci dia delle successioni di numeri che si comportano come se fossero dei numeri casuali uniformemente distribuiti nell'intervallo $[0, 1]$.

Chiameremo “successione pseudo-casuale” una successione di questo genere.

Il primo ad inventare un procedimento (chiamato middle-square method) per la generazione di numeri pseudo-casuali fu il grande matematico von Neumann che lo usò per il metodo di “Monte Carlo” (al quale dedicheremo qualche cenno più avanti), creato dal suo collega e collaboratore Stanislaw Ulam.

L'algoritmo proposto era il seguente: si prenda un numero di dieci cifre, ad esempio

5772156649

(che identifichiamo con 0,577215664) e ne facciamo il quadrato, ottenendo un numero di venti cifre:

33317|7923805949|09201.

Prendiamo ora le dieci cifre di mezzo ed otteniamo così un altro numero di dieci cifre e precisamente

7923805949,

da identificarsi con 0,7923805949.

Il quadrato di 7923805949 è

72214|3436800353|90611,

da identificarsi con 0,3436800353.

Deve essere chiaro che c'è una bella differenza tra una successione di v.a. uniformemente distribuite nell'intervallo $[0, 1]$ ed una successione di numeri pseudo-casuali. Questi ultimi vengono infatti generati con un procedimento deterministico, e perciò dato il primo termine, tutti gli altri sono univocamente determinati. Questo era ben presente a von Neumann, il quale scrisse nel 1951: **“Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin”**.

È chiaro il pericolo insito nel metodo di von Neumann (e in tutti i metodi aritmetici), e cioè che la successione generata potrebbe accumularsi più da una parte che dall'altra, o convergere o perfino stabilizzarsi.

È noto il seguente esempio che riguarda numeri di quattro cifre. Partiamo da 3792. Il suo quadrato è 14379264 e dunque i quattro numeri centrali danno nuovamente 3792. Il procedimento si ferma. Non sappiamo che cosa succede con dieci cifre. In mancanza di meglio, von Neumann usava questo metodo e sembra che i risultati fossero soddisfacenti, ma non è mai stato dimostrato che fosse corretto. Nel frattempo è stato abbandonato a favore di altri metodi, ma anche questi vengono continuamente testati, perché c'è sempre il pericolo che qualcosa vada storto.

Il famoso matematico americano Donald Knuth usa dire: **“I numeri casuali non possono essere generati da un metodo scelto a caso”**. Il metodo di von Neumann è descritto nel suo famoso libro “The art of computer programming”.

I vari metodi per generare numeri casuali possono essere schematizzati nel seguente modo:

Si sceglie un insieme (finito) Ω di numeri [von Neumann ha preso tutti i numeri da 0 a 10^{10}], si considera un'applicazione $\varphi : \Omega \rightarrow \Omega$ [von Neumann ha considerato l'applicazione che abbiamo descritto(*)] e poi si itera la funzione. Si parte da un numero $\omega_1 \in \Omega$ (chiamato "seme" del procedimento) e poi, di seguito, si considera la successione $\omega_2 = \varphi(\omega_1)$, $\omega_3 = \varphi(\omega_2)$, e, in generale $\omega_{n+1} = \varphi(\omega_n)$.

[(*) La φ di von Neumann non è difficile da programmare. Dopo aver quadrato il seme ω e ottenuto un numero di 20 cifre (eventualmente si aggiungono degli zeri all'inizio, se il seme è un numero minore di 10^{10}), dividiamo il risultato per 10^{15} e prendiamo la mantissa del risultato. In questo modo ci siamo liberati delle prime cinque cifre. Poi moltiplichiamo il risultato per 10^5 e ne prendiamo la parte intera. Fatto! Abbiamo così ottenuto $\varphi(\omega)$.]

A parte preoccupazioni di carattere statistico, certamente importanti, un problema che può sorgere è che risulti $\omega_n = \varphi(\omega_n)$ per un certo n .

Ma questo, se la φ è scelta "a caso", è abbastanza probabile. Quando abbiamo studiato il gioco del "treize" abbiamo visto che un'applicazione "scelta a caso" di $\{1, 2, 3 \dots, N\}$ in sé ha almeno un punto fisso (e quindi già $\omega = \varphi(\omega)$, con un seme scelto a caso) con probabilità prossima a $1 - \frac{1}{e} = 0,63212 \dots$, e quindi abbastanza elevata.

Attualmente i metodi per la generazione di successioni pseudo-casuali utilizzano metodi di teoria dei numeri. In un manuale di un piccolo personal computer di qualche anno fa era riportato il seguente metodo. A partire da un "seme" x_0 , la seguente iterazione definita ricorsivamente forniva la successione pseudo-uniforme $\{y_n\}$:

$$x_{n+1} = (7^5 x_n) \pmod{(2^{31} - 1)}$$

$$y_{n+1} = x_{n+1} \pmod{1} .$$

Uno dei difetti di questo tipo di sequenze è che sono cicliche: la lunghezza massima del ciclo dell'esempio precedente è $2^{31} - 1$, perché una volta che tutti i numeri sono esauriti (se non prima) l'iterazione ricomincia a fornire la stessa sequenza di numeri.

La questione della creazione di successioni di numeri pseudo-casuali è tutt'altro che risolta. In un articolo del 1992 apparso sulla rivista Physical Review Letters un gruppo di fisici e di ricercatori dell'IBM Corporation's Supercomputing Center dava notizia che cinque metodi aritmetici per la

generazione di numeri pseudo-casuali, che avevano superato numerosi test statistici e che erano commercializzati da anni, avevano prodotto dei risultati errati in una simulazione che riguardava il comportamento di un atomo in un cristallo magnetico.

Quando le terne consecutive di numeri, interpretate come coordinate di un punto nello spazio, prodotte da uno di questi sistemi sono state riportate nello spazio tridimensionale, si è notato che i punti, invece di creare una “nuvola”, tendevano a distribuirsi intorno ad un numero finito di piani.

Esistono referenze anche molto recenti di problemi con i generatori di numeri pseudo-casuali.

È stato proposto di utilizzare per la generazione di numeri casuali il decadimento radioattivo degli atomi. L'istante in cui un atomo radioattivo si scinde è completamente imprevedibile, perché obbedisce alle leggi statistiche della meccanica quantistica. Ma il famoso probabilista americano, Persi Diaconis, vede anche in questo metodo dei possibili pericoli. Per registrare infatti il decadimento di un atomo occorrerebbe utilizzare qualche congegno, il quale potrebbe produrre degli errori sistematici. Un altro esperto ha sottolineato inoltre che quando viene effettuata una simulazione, questa deve essere controllata da altri, i quali devono sapere quale successione di numeri pseudo-casuali è stata utilizzata. Le successioni prodotte dal decadimento radioattivo non sono invece ripetibili.

Per concludere questa discussione sui numeri pseudo-casuali, sottolineiamo ancora una volta la necessità di utilizzare dei test statistici per valutare “quanto casuale” sia la successione. Il nostro istinto non è affidabile, in quanto tendiamo a confondere troppo casualità e disordine.

Test psicologici hanno dimostrato che se si chiede ad una persona di scrivere una sequenza “casuale” di cento cifre tra 0 e 9, questa sequenza passerà il primo test, cioè ci saranno più o meno lo stesso numero di cifre 0, 1, ..., 9, ma non passerà un secondo test, quello dell'indipendenza, nel senso che in genere le persone evitano di ripetere lo stesso numero (ed anche altre sequenze che sembrano troppo regolari, come ad esempio 012). Eppure in cento estrazioni a sorte di numeri da 0 a 9 la probabilità di trovare coppie di cifre consecutive uguali è piuttosto elevata, perché la probabilità che una cifra si ripeta è $\frac{1}{10}$ e su novantanove esperimenti in media dovremmo avere circa un decimo, quindi circa dieci, coppie di cifre uguali consecutive. Inoltre dovremmo aspettarci, in media, che in una sequenza di quella lunghezza ci sia una terna di cifre uguali consecutive.

Siccome nell'estrazione del lotto tutte le cifre hanno la stessa probabilità, si potrebbero dare a chi intende giocare a Super Enalotto due consigli. Il primo è : non fatelo. Il secondo è: se proprio lo volete fare, giocate combinazioni regolari in modo da minimizzare la probabilità che altri facciano la stessa giocata (e che siate quindi costretti a dividere il monte premi con altri, nel fortunato e improbabile caso di vincita). La combinazione 1, 2, 3, 4, 5, 6 ha la stessa probabilità della combinazione 7, 15, 24, 55, 63, 78, eppure a molti la seconda appare immensamente più probabile della prima.

L'ingrediente principale (per così dire la materia prima) con cui si può simulare una v.a. è una successione $\{u_n\}$ che simula la v.a. uniformemente distribuita sull'intervallo $[0, 1]$.

Due osservazioni importanti

a) Per la simulazione di vettori aleatori ma anche di certe v.a. discrete conviene notare questo semplice fatto: se le v.a. uniformemente distribuite U_i , $1 \leq i \leq n$ sono indipendenti, allora il vettore aleatorio $V = (U_1, U_2, \dots, U_n)$ è uniformemente distribuito sull'ipercubo $[0, 1]^n$.

Per l'indipendenza, la densità congiunta è il prodotto delle densità marginali e quindi è proprio uguale a 1 su $[0, 1]^n$ e nulla fuori dall'ipercubo.

b) Nella simulazione di una v.a. si distinguono due passi che vogliamo mettere in evidenza. Il primo passo consiste nello esprimere la legge di X in funzione della v.a. uniformemente distribuita U , il secondo è quello di simulare U e quindi anche X .

Variabili aleatorie discrete

1) Simulazione della v.a. X di Bernoulli $B(1, p)$.

Consideriamo la v.a. $\left\lfloor \frac{U}{1-p} \right\rfloor$. Se $U < 1 - p$, allora $\frac{U}{1-p} < 1$ (evento che ha probabilità $1 - p$) e la sua parte intera è zero. Se invece $\frac{U}{1-p} > 1$, la parte intera è uguale a 1 e questo accade con probabilità p . Poiché $P\left(\frac{U}{1-p} = 1\right) = 0$, abbiamo ottenuto una v.a. che ha legge $B(1, p)$.

Per *simulare* una $B(1, p)$ prendiamo ora una successione pseudo-casuale $\{u_n\}$ che simula la U e consideriamo la successione

$$\left\{ \left\lfloor \frac{u_n}{1-p} \right\rfloor \right\}.$$

Questa simula la X .

2) Simulazione della v.a. binomiale $B(n, p)$

Ricordiamo che per ottenere una $B(n, p)$, basta sommare n copie indipendenti di una $B(1, p)$. Per quanto abbiamo appena visto, supponiamo che $\{U_k\}$ sia una successione di v.a. indipendenti uniformemente distribuite su $[0, 1]$. Allora

$$Y_k = \sum_{i=(k-1)n+1}^{kn} \left\lfloor \frac{U_i}{1-p} \right\rfloor$$

ha legge $B(n, p)$. La simulazione si ottiene come prima.

3) Vediamo ora come si simula Z , una v.a. uniforme sull'insieme $\{1, 2, \dots, N\}$. Suddividiamo l'intervallo $[0, 1]$ in N parti uguali $]\frac{k-1}{N}, \frac{k}{N}]$, con $1 \leq k \leq N$.

Poiché $(\frac{k-1}{N} \leq U \leq \frac{k}{N}) = (k-1 \leq NU \leq k)$ si ha che

$$P(\lfloor NU \rfloor + 1 = k) = \frac{1}{N}$$

e così abbiamo trovato la legge di Z in funzione di U .

Per la simulazione si procede come al solito.

4) Simulazione della v.a. geometrica.

La funzione di ripartizione della v.a. geometrica X di parametro p è uguale a

$$F(i) = \sum_{k=1}^i pq^{k-1} = 1 - P(X > i) = 1 - \sum_{k=i+1}^{\infty} pq^{k-1} = 1 - p \frac{q^{i+1}}{1-q} = 1 - q^{i+1}.$$

Ne segue che $X = i$ se e solo se

$$1 - q^{i-1} \leq U \leq 1 - q^i$$

ovvero se

$$q^i \leq 1 - U \leq q^{i-1}.$$

Ma se U è uniformemente distribuita su $[0, 1]$, anche $1 - U$ lo è, quindi l'ultima disuguaglianza si può anche scrivere come

$$q^i \leq U \leq q^{i-1}.$$

ovvero ancora come

$$e^{i \log q} \leq U \leq e^{(i-1) \log q}.$$

Il logaritmo preserva la doppia disuguaglianza e se dividiamo per $\log q$, che è negativo, otteniamo

$$i - 1 \leq \frac{\log U}{\log q} \leq i.$$

Ne segue che

$$X = \left\lfloor \frac{\log U}{\log q} \right\rfloor + 1.$$

A questo punto simuliamo U ed avremo ottenuto una simulazione di X .

Il metodo della funzione inversa

Se X è una variabile aleatoria la cui densità è strettamente positiva nell'intervallo $]a, b[$ e nulla fuori, allora la sua funzione di ripartizione F è strettamente crescente in $]a, b[$ (e a valori in $]0, 1[$). Se l'inversa $G = F^{-1}$ è esprimibile analiticamente, allora $G(U) = F^{-1}(U)$ ha la legge di X e la successione $\{G(u_n)\}$ simula la v.a. X .

Risulta infatti $P(G(U) \leq t) = P(U \leq F(t)) = F(t)$.

Esempi

1) V.a. uniformemente distribuita su $[a, b]$.

Abbiamo già visto la funzione di ripartizione di questa v.a. che è uguale a $\frac{t-a}{b-a}$ nell'intervallo aperto $]a, b[$. Si vede facilmente che $F^{-1}(u) = a + (b-a)u$ per $u \in]0, 1[$. Dunque la successione $\{a + (b-a)u_n\}$ simula la v.a. in questione.

2) V.a. esponenziale.

La densità è positiva per $t > 0$. La funzione di ripartizione è uguale a $1 - e^{-\lambda t}$. L'inversa è $F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u)$. Per il teorema precedente la successione $\{-\frac{1}{\lambda} \log(1 - u_n)\}$ simula la v.a. esponenziale.

3) V.a. di Cauchy

La densità è positiva su tutto \mathbb{R} e la funzione di ripartizione F è uguale a

$$\frac{1}{\pi} \int_{-\infty}^t \frac{ds}{1 + s^2} = \frac{1}{\pi} \arctan t + \frac{1}{2}.$$

La sua inversa è $F^{-1}(u) = \tan \pi(u - \frac{1}{2})$ e l'ovvia conclusione è che la successione $\{\tan \pi(u_n - \frac{1}{2})\}$ simula la v.a. di Cauchy, se $\{u_n\}$ simula la v.a. uniformemente distribuita in $[0, 1]$.

Simulazione dei vettori aleatori

1) Nelle lezioni precedenti abbiamo presentato i vettori aleatori uniformemente distribuiti su un boreliano *limitato* B .

Abbiamo visto prima come si ottiene il vettore aleatorio V uniformemente distribuito sull'ipercubo $[0, 1]^m$. Facciamo ora in piccolo (ma importante) passo in avanti: supponiamo per ora che il boreliano sia $R = \prod_{i=1}^m [a_i, b_i]$.

Possiamo allora avvalerci di un risultato unidimensionale: la $U'_i = a_i + (b_i - a_i)U_i$ ha la legge della v.a. uniformemente distribuita su $[a_i, b_i]$. Ripetendo allora il ragionamento fatto prima, $V' = (V'_1, V'_2, \dots, V'_m)$ ha la legge del vett. a. uniformemente distribuito sull'iperrettangolo $R = \prod_{i=1}^m [a_i, b_i]$.

2) Supponiamo ora che B sia un boreliano limitato di \mathbb{R}^m .

Il primo passo consiste nel prendere un iperrettangolo R che contiene B . Per motivi che appariranno chiari tra un'attimo, conviene prendere R il più piccolo possibile.

Fatto questo applichiamo quello che si chiama *il metodo del rigetto*. Se la successione $\{z_n\}$ simula il vettore aleatorio V' uniformemente distribuito su R , iniziamo a generare gli z_n *rigettando* quelli che non appartengono a B . Questo equivale a considerare il vettore aleatorio $\frac{1}{\lambda(B)}V\mathbb{1}_B$, il quale è uniformemente distribuito su B . Sappiamo infatti che se V è uniformemente distribuito in R , allora

$$P(V \in A) = \frac{\lambda_m(A)}{\lambda_m(R)}$$

per ogni boreliano $A \subset [0, 1]^m$.

Pertanto, se C è un boreliano contenuto in B ,

$$P(V\mathbb{1}_B \in C) = \frac{\lambda_m(B \cap C)}{\lambda_m(B)}.$$

Ripensiamo al vettore aleatorio uniformemente distribuito sul disco unitario $B = \{(s, t) : s^2 + t^2 \leq 1\}$. Prendiamo il quadrato $R = [-1, 1] \times [-1, 1]$ e simuliamo con la successione $\{z_n\}$ il vettore uniformemente distribuito nel quadrato. Se $z_1 \notin B$, lo scartiamo e passiamo a z_2 . Se $z_2 \in B$, ce lo teniamo. E così via. Si ottiene in questa maniera una sottosuccessione $\{z_{n_k}\}$ di punti

del disco. Quelli che stanno nel quadrato, ma non nel disco, sono eliminati da questa procedura.

La successione $\{z_{n_k}\}$ simula il vett.a. uniformemente distribuito nel disco.

Questo succede in generale. Se infatti V è uniformemente distribuito in R , sappiamo che

$$P(V \in A) = \lambda_m(A)$$

per ogni boreliano $A \subset [0, 1]^m$.

È semplice immaginare come questa procedura si generalizzi a un boreliano qualunque in una dimensione arbitraria: Si prende $R \supset B$, si prende $\{z_n\}$ che simula il vett.a. uniformemente distribuito e si scartano, quando compaiono, gli z_n che non appartengono a B .

L'opportunità di "economizzare" nella scelta di R è ovvia. Più piccolo si riesce a scegliere R , meno z_n si sprecano. La proporzione dei punti sprecati è uguale alla misura di $R \setminus B$ rispetto alla misura di R .

Nel caso del disco, $\frac{4-\pi}{4} \sim 0,215$ e quindi vanno sprecati circa il 21,5% dei punti.

3) Vediamo ora un metodo scoperto relativamente di recente (Box e Muller, 1958) per simulare la $N(0, 1)$, anzi coppie di v.a. indipendenti aventi questa legge.

Se (X, Y) è una coppia di v.a. indipendenti di legge $N(0, 1)$, la densità congiunta è ovviamente data da

$$f(s, t) = \frac{1}{2\pi} e^{-(s^2+t^2)/2}.$$

Per la simmetria radiale della densità conviene pensare alle coordinate polari (ρ, ϑ) con $0 \leq \rho \leq 1$ e $0 \leq \vartheta \leq 2\pi$.

Ovviamente la v.a. Θ è uniformemente distribuita su $0 \leq \vartheta \leq 2\pi$ e per la sua simulazione possiamo usare $2\pi U_1$.

La funzione di ripartizione R la v.a. relativa al raggio (usiamo la lettera R perché la ρ maiuscola si scrive P e vogliamo evitare confusione nella formula che segue) è

$$P(R \leq r) = F_R(r) = \iint_{D_r} f(s, t) ds dt =$$

(dove D_r è il disco di raggio r e centrato nell'origine)

$$= \int_0^r \int_0^{2\pi} \frac{1}{2\pi} e^{-s^2/2} s ds d\vartheta = \int_0^r e^{-s^2/2} s ds$$

e con un cambio di variabili $u = s^2/2$, $s ds = du$ si ottiene

$$F_R(r) = 1 - e^{-r^2/2}.$$

A questa funzione di ripartizione si applica il teorema della funzione inversa visto prima. Poiché se U_2 è uniformemente distribuita su $[0, 1]$, anche $1 - U_2$ lo è e dunque conviene risolvere l'equazione

$$1 - e^{-r^2/2} = 1 - U_2,$$

ottenendo in conclusione

$$\Theta = 2\pi U_1 \text{ e } R = \sqrt{-2 \log U_2}$$

e quindi

$$X = R \cos \Theta \text{ e } Y = R \sin \Theta.$$

La cosa che può sembrare sorprendente è che X e Y sono indipendenti, visto che usano la stessa R e la stessa Θ . Ma lo dimostrano i conti fatti, eravamo partiti da due v.a. X e Y indipendenti!

7 Funzioni caratteristiche

In questa sezione considereremo delle particolari variabili aleatorie a valori in \mathbb{C} .

Un'applicazione $Z = X + iY$ di Ω in \mathbb{C} si dice una variabile aleatoria se X e Y lo sono. Essa ammette speranza matematica, se la ammettono X e Y e $E(Z) = E(X) + iE(Y)$.

Si noti che Z ha speranza matematica se e solo se ce l'ha $|Z|$. Questo succede perché $|Z| = \sqrt{X^2 + Y^2} \leq |X| + |Y|$. Viceversa, $|X| \leq |Z|$ e $|Y| \leq |Z|$.

È facile dimostrare che $E(\bar{Z}) = \overline{E(Z)}$.

Ricordiamo ancora che la funzione esponenziale nel campo complesso è definita come somma della serie $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ ed è l'unica funzione analitica che si riduce per i valori reali di z alla funzione esponenziale e^x .

È importante la seguente identità di Eulero: $e^z = e^{x+iy} = e^x(\cos y + i \sin y)$.

Sia ora X un vettore aleatorio, $X : \Omega \rightarrow \mathbb{R}^m$. La sua *funzione caratteristica* è la funzione $\Phi : \mathbb{R}^m \rightarrow \mathbb{C}$ definita da

$$\Phi(\vartheta) = E(e^{i\langle \vartheta, X \rangle}),$$

dove \langle , \rangle denota il prodotto scalare in \mathbb{R}^m .

La funzione caratteristica è definita su tutto \mathbb{R}^m , perché il modulo di e^{it} è uguale a 1 per ogni $t \in \mathbb{R}$.

Dimostriamo ora due lemmi.

Lemma

Se $X : \Omega \rightarrow \mathbb{R}^m$ è un vettore aleatorio e se $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}^k$ è una funzione boreliana, allora $\Psi(X)$ è un vettore aleatorio.

Il teorema è vero se, in particolare, la Ψ è continua.

Lemma

Se X e Y sono due v.a. indipendenti e se $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione boreliana, allora $\Psi(X)$ e $\Psi(Y)$ sono indipendenti.

Infatti,

$$P(\Psi(X) \in A, \Psi(Y) \in B) = P(X \in \Psi^{-1}(A), Y \in \Psi^{-1}(B)) =$$

$$P(X \in \Psi^{-1}(A))P(Y \in \Psi^{-1}(B)) = P(\Psi(X) \in A)P(\Psi(Y) \in B).$$

Enunciamo ora due teoremi, uno per le v.a. discrete e il secondo per le v.a. assolutamente continue.

Teorema

Se X è una vett.a. discreto a valori in \mathbb{R}^m avente densità congiunta $p(t)$ e se $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}$ è una funzione boreliana, allora

$$E(\Psi(X)) = \sum_{t \in \mathbb{R}^m} \Psi(t)p(t),$$

sempre se la serie converge assolutamente.

Teorema

Se X è una vett.a. assolutamente continuo a valori in \mathbb{R}^m avente densità congiunta $f(t)$ e se $\Psi : \mathbb{R}^m \rightarrow \mathbb{R}$ è una funzione boreliana, allora

$$E(\Psi(X)) = \int_{\mathbb{R}^m} \Psi(t)f(t)dt,$$

sempre che il valore assoluto sia integrabile.

Il primo teorema è abbastanza facile, basta osservare che vengono sommati esattamente i valori di $\Psi(X)$ moltiplicati per la loro densità $p(t)$.

Il secondo teorema si dimostra per approssimazione. In precedenza abbiamo usato (dimostrandolo in dettaglio) un caso particolare di questo teorema con $\Psi(t) = t^2$.

Dimostriamo ora un altro importante risultato.

Teorema

Se X e Y sono indipendenti allora

$$\Phi_{X+Y}(\vartheta) = \Phi_X(\vartheta)\Phi_Y(\vartheta).$$

Infatti

$$\begin{aligned}\Phi_{X+Y}(\vartheta) &= E(e^{i\langle\vartheta, X+Y\rangle}) = E(e^{i(\langle\vartheta, X\rangle + \langle\vartheta, Y\rangle)}) = \\ &= E(e^{i\langle\vartheta, X\rangle} e^{i\langle\vartheta, Y\rangle}) = E(e^{i\langle\vartheta, X\rangle})E(e^{i\langle\vartheta, Y\rangle}) = \Phi_X(\vartheta)\Phi_Y(\vartheta).\end{aligned}$$

Vediamo altre due proprietà che ci serviranno in seguito.

Proposizione

$$\Phi_{-X}(\vartheta) = \overline{\Phi_X(\vartheta)}.$$

Infatti, $\Phi_{-X}(\vartheta) = E(e^{i\langle\vartheta, -X\rangle}) = E(e^{-i\langle\vartheta, X\rangle}) = E(\overline{e^{i\langle\vartheta, X\rangle}}) = \overline{E(e^{i\langle\vartheta, X\rangle})}$.

Proposizione

$$\begin{aligned}\Phi_{aX+b}(\vartheta) &= e^{i\langle\vartheta, b\rangle}\Phi_X(a\vartheta). \\ \Phi_{aX+b}(\vartheta) &= E(e^{i\langle\vartheta, aX+b\rangle}) = E(e^{i\langle\vartheta, aX\rangle} e^{i\langle\vartheta, b\rangle}) = \\ &= e^{i\langle\vartheta, b\rangle} E(e^{i\langle a\vartheta, X\rangle}) = e^{i\langle\vartheta, b\rangle}\Phi_X(a\vartheta).\end{aligned}$$

Esempi

1) $X \sim B(n, p)$. Allora

$$\Phi(\vartheta) = \sum_{k=1}^n \binom{n}{k} p^k (1-p)^{n-k} e^{i\vartheta k} = \sum_{k=1}^n \binom{n}{k} (pe^{i\vartheta})^k (1-p)^{n-k} = (1-p + pe^{i\vartheta})^n.$$

2) Sia X la v.a. geometrica di parametro p . Allora

$$\Phi(\vartheta) = \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{i\vartheta k} = e^{i\vartheta} p \sum_{k=1}^{\infty} ((1-p)e^{i\vartheta})^{k-1} = e^{i\vartheta} p \frac{1}{1 - (1-p)e^{i\vartheta}}.$$

3) Sia X la v.a. di Poisson di parametro λ . Allora

$$\Phi(\vartheta) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{i\vartheta k} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{i\vartheta})^k}{k!} = e^{-\lambda} e^{\lambda e^{i\vartheta}} = e^{\lambda(e^{i\vartheta} - 1)}.$$

4) Sia ora X la v.a. esponenziale di parametro λ . Allora

$$\Phi(\vartheta) = \lambda \int_0^{+\infty} e^{-\lambda t} e^{i\vartheta t} dt = \lambda \int_0^{+\infty} e^{t(i\vartheta - \lambda)} dt.$$

$$\frac{\lambda}{i\vartheta - \lambda} e^{t(i\vartheta - \lambda)} \Big|_0^{+\infty} = \frac{\lambda}{i\vartheta - \lambda}$$

5) Sia $X \sim N(0, 1)$, la normale standard. Allora

$$\Phi(\vartheta) = e^{-\frac{\vartheta^2}{2}}$$

Si ha

$$\Phi(\vartheta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\vartheta t} e^{-\frac{t^2}{2}} dt$$

e derivando otteniamo

$$\Phi'(\vartheta) = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\vartheta t} t e^{-\frac{t^2}{2}} dt =$$

$$\frac{i}{\sqrt{2\pi}} \left(\left| -e^{i\vartheta t} e^{-\frac{t^2}{2}} \right|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} i\vartheta e^{i\vartheta t} e^{-\frac{t^2}{2}} dt \right) =$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \vartheta e^{i\vartheta t} e^{-\frac{t^2}{2}} dt = -\vartheta \Phi(\vartheta).$$

Abbiamo così ottenuto per la $\Phi(\vartheta)$ un'equazione differenziale lineare del primo ordine. La soluzione generale è $\Phi(\vartheta) = ce^{-\frac{\vartheta^2}{2}}$. Tenendo conto della condizione iniziale $\Phi(0) = 1$ si ottiene che $\Phi(\vartheta) = e^{-\frac{\vartheta^2}{2}}$.

In seguito utilizzeremo altre due proprietà importanti delle funzioni caratteristiche, enunciate nei due risultati seguenti.

Teorema

Se X e Y sono due v.a. che hanno la stessa funzione caratteristica, allora hanno la stessa legge.

Teorema

La Φ_X è continua. Se inoltre X ha momento k -esimo (finito), allora Φ_X è k volte derivabile e si ha

$$\frac{d^k}{d\vartheta^k} \Phi_X = E \left(\frac{d^k}{d\vartheta^k} e^{i\vartheta X} \right) = E \left((iX)^k e^{i\vartheta X} \right) = i^k E \left(X^k e^{i\vartheta X} \right).$$

Un'importante conseguenza è che $\Phi_X^{(k)}(0) = i^k E(X^k)$.

8 Convergenza

Disuguaglianza di Čebišëv

Questo capitolo che sarà dedicato a risultati che riguardano la ripetizione di esperimenti e sequenze di variabili aleatorie (come la legge debole dei grandi numeri e il Teorema Centrale del Limite).

Iniziamo con un'interessante e utile disuguaglianza che dà una stima, in termini di varianza, di quanto si può discostare una v.a. dalla sua media.

Teorema

Se X è una v.a. che ammette varianza, allora per ogni $\eta > 0$ si ha che

$$P(|X - E(X)| \geq \eta) \leq \frac{\text{Var}(X)}{\eta^2}.$$

Indichiamo con A l'evento

$$\{\omega : |X(\omega) - E(X)| \geq \eta\} = \{\omega : (X(\omega) - E(X))^2 \geq \eta^2\}$$

e consideriamo la v.a. $\eta^2 \mathbb{1}_A(\omega)$.

Ci proponiamo di far vedere che $(X(\omega) - E(X))^2 \geq \eta^2 \mathbb{1}_A(\omega)$.

Dividiamo Ω in due parti: A e A^c e facciamo vedere che la disuguaglianza vale in entrambe le parti.

Se $\omega \in A$, per definizione di A risulta $(X(\omega) - E(X))^2 \geq \eta^2 = \eta^2 \mathbb{1}_A(\omega)$. Se invece $\omega \in A^c$, si ha comunque $(X(\omega) - E(X))^2 \geq 0$, mentre $\eta^2 \mathbb{1}_A(\omega) = 0$.

Utilizziamo ora la monotonia della speranza matematica. La disuguaglianza tra le due v.a. implica che

$$[\text{Var}(X) =] E((X - E(X))^2) \geq E(\eta^2 \mathbb{1}_A) = \eta^2 P(A) = \eta^2 P(|X - E(X)| \geq \eta).$$

Dividendo per η^2 si ottiene la tesi.

La disuguaglianza mette in confronto la probabilità che la v.a. X si discosti dalla sua media attesa di più di η e un opportuno multiplo la varianza della X . Alle volte la disuguaglianza può essere anche un po' larga, ma in molti casi si rivela, come vedremo, molto efficace.

Tipi di convergenza di successioni di v.a.

Capita in diversi contesti che tra funzioni dello stesso tipo si considerino diversi tipi di convergenza. Un esempio classico sono la convergenza puntuale e quella uniforme di una successione di funzioni continue. Se il dominio non è un insieme compatto, una successione può essere puntualmente convergente ma non convergente in maniera uniforme.

Un esempio più avanzato è quello delle funzioni integrabili secondo Lebesgue per le quali, oltre alla convergenza puntuale, si studiano la convergenza quasi certa, quella in misura, quella in media, quella in media quadratica ed altre ancora.

Parleremo ora di alcuni tipi più significativi di convergenza delle successioni di variabili aleatorie.

Incominciamo con la *convergenza quasi certa*.

Diremo che una successione $\{X_n\}$ di variabili aleatorie definite su Ω e a valori reali converge quasi certamente (q.c.) a X , se

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1.$$

Nell'integrazione alla Lebesgue questo tipo di convergenza corrisponde alla convergenza quasi ovunque.

Un secondo tipo di convergenza che si rivela molto utile è la *convergenza in probabilità*.

Diremo che una successione $\{X_n\}$ di variabili aleatorie definite su Ω e a valori reali converge in probabilità a X , se per ogni $\eta > 0$ si ha che

$$\lim_{n \rightarrow \infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \eta\}) = 0.$$

A parole, comunque si fissi un $\eta > 0$, la probabilità che X_n e X differiscano per più di η tende a zero.

Nell'integrazione alla Lebesgue questo tipo di convergenza corrisponde alla convergenza in misura, un concetto tutto sommato poco usato, a differenza del suo corrispondente probabilistico.

Teorema

Se $\{X_n\}$ converge ad X quasi certamente, allora converge anche in probabilità.

L'ipotesi si può anche riformulare dicendo che l'evento

$$C = \{\omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq X(\omega)\}$$

ha probabilità nulla.

Fissiamo ora un $\eta > 0$ e poniamo $A_n = \bigcup_{m \geq n} (|X_m - X| \geq \eta)$.

La successione è ovviamente decrescente. Denotiamo con A l'intersezione di tutti gli A_n . Per un teorema precedente sappiamo che la successione $\{P(A_n)\}$ decresce e tende a $P(A)$.

Prendiamo un qualunque $\omega \notin C$. Per come è stato definito C ,

$$\lim_{m \rightarrow \infty} X_m(\omega) = X(\omega)$$

e quindi esiste un intero $N_\eta(\omega)$ tale che, per $m \geq N_\eta(\omega)$, $|X_m(\omega) - X(\omega)| < \eta$ e perciò $\omega \notin A_n$ per tutti gli n sufficientemente grandi da cui deduciamo che $\omega \notin A$.

In altre parole $A \subset C$ e pertanto $P(A) = 0$.

Per il risultato richiamato prima, $P(|X_n(\omega) - X(\omega)| \geq \eta)$ tende a zero, per n che tende all'infinito, ovvero c'è convergenza in probabilità.

Presenteremo ora un controesempio, facendo vedere che ci sono successioni di v.a. che convergono in probabilità, ma non quasi certamente.

La successione della tastiera

Come spazio di misura prendiamo $\Omega =]0, 1]$ dotato della σ -algebra dei boreliani \mathcal{B} , su cui è definita la misura di Lebesgue λ .

Costruiamo ora la seguente successione di v.a.: le prime due siano $X_1 \equiv 1$ e $X_2 \equiv 0$. Le due successive incominciano ad essere interessanti: dividiamo $]0, 1]$ in due parti uguali, $I_1^1 =]0, \frac{1}{2}]$ e $I_2^1 =]\frac{1}{2}, 1]$ e consideriamo le v.a. $X_3 = 1 - \mathbb{1}_{I_1^1}$ e $X_4 = 1 - \mathbb{1}_{I_2^1}$.

Procediamo in questa maniera, dividendo al passo successivo $]0, 1]$ in quattro parti uguali, $I_1^2 =]0, \frac{1}{4}]$, $I_2^2 =]\frac{1}{4}, \frac{1}{2}]$, $I_3^2 =]\frac{1}{2}, \frac{3}{4}]$ e $I_4^2 =]\frac{3}{4}, 1]$. Le quattro v.a. successive della successione, i **quattro tasti**, X_5, X_6, X_7 e X_8 sono le v.a. $1 - \mathbb{1}_{I_j^2}$ per $1 \leq j \leq 4$. (Si capisce ora perché abbiamo preso tempo con le prime due v.a., spendendo per loro gli indici 1 e 2: è per scrivere in maniera più semplice gli indici delle v.a. successive.)

Le otto v.a. successive (**gli otto tasti**) X_k , per $9 \leq k \leq 16$, si ottengono come $1 -$ le v.a. indicatrici degli otto intervalli di lunghezza uguale in cui viene suddiviso l'intervallo al passo successivo.

Continuando in questo modo si divide all' n -esimo passo l'intervallo $]0, 1]$ in 2^n parti uguali e si ottengono le v.a. X_k , per $2^{n-1} + 1 \leq k \leq 2^n$ (**i 2^n tasti**), come $1 -$ le indicatrici dei 2^n intervalli.

Si vede ora che tale successione non converge in *nessun* punto di $]0, 1]$, perché qualunque sia $\omega \in \Omega$, delle 2^n v.a. costruite all' n -esimo passo esattamente una assume in ω il valore 0, mentre tutte le altre vi assumono valore 1, quindi la successione $\{X_n(\omega)\}$ contiene infiniti "zeri" e infiniti "uni", pertanto non è convergente.

Però $\{X_n\}$ è convergente in probabilità alla v.a. X identicamente uguale a 1. Fissato infatti $\eta > 0$ (gli η significativi sono quelli minori di 1), si ha che $|X - X_k| = X - X_k \geq \eta$ su un insieme di misura $\frac{1}{2^n}$ tutte le volte che $2^{n-1} < k \leq 2^n$, cioè $n - 1 < \log_2 k \leq n$. Ne segue che quando k tende all'infinito, anche n tende all'infinito e quindi $P(|X - X_n| \geq \eta)$ tende a 0.

La successione della tastiera tende in probabilità al silenzio assoluto.

Teorema

Se $\{X_n\}$ converge in probabilità a X e a Y , allora $X = Y$ quasi certamente.

Si ha che

$$(|X(\omega) - Y(\omega)| \geq 2\eta) \subset (|X_n(\omega) - X(\omega)| \geq \eta) \cup (|X_n(\omega) - Y(\omega)| \geq \eta)$$

[Infatti, se $|X_n(\omega) - X(\omega)| < \eta$, allora necessariamente $|X_n(\omega) - Y(\omega)| \geq \eta$.]
 Ne segue che

$$P(|X - Y| \geq 2\eta) \leq P(|X_n - X| \geq \eta) + P(|X_n - Y| \geq \eta)$$

e siccome entrambi gli addendi a destra tendono a zero, necessariamente per ogni $\eta > 0$

$$P(|X - Y| \geq 2\eta) = 0.$$

Ma

$$(X(\omega) \neq Y(\omega)) = \bigcup_{k=1}^{\infty} \left(|X(\omega) - Y(\omega)| \geq \frac{1}{k} \right),$$

e tutti gli eventi dell'unione numerabile a destra hanno probabilità nulla, pertanto

$$P(X(\omega) \neq Y(\omega)) = 0.$$

Un terzo tipo di convergenza che ci sarà utile nel seguito è quello della *convergenza in legge*.

Diremo che una successione $\{X_n\}$ di variabili aleatorie a valori reali converge in legge a X , se (denotate con F_n ed F le funzioni di ripartizione delle X_n e della X), risulta

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

fatta eventualmente eccezione per i punti di discontinuità della F .

Merita fare alcune osservazioni.

Ad una prima lettura può essere sfuggito il fatto che le X_n non sono necessariamente definite sullo stesso insieme campione Ω . Vedremo in seguito qualche esempio.

Una seconda osservazione riguarda il fatto già citato a proposito delle funzioni monotone le quali hanno al più un'infinità numerabile di punti di discontinuità. Quindi il limite di cui sopra potrebbe non esistere per al più un'infinità numerabile di $t \in \mathbb{R}$.

Una terza osservazione riguarda la posizione di questo tipo di convergenza rispetto agli altri due già introdotti. Ebbene questo tipo di convergenza è tra i tre il più debole. Si può infatti dimostrare che la convergenza in probabilità implica la convergenza in legge, ma che il viceversa non è vero.

Faremo ora vedere che la convergenza in probabilità implica la convergenza in legge.

Dimostriamo prima un lemma.

Lemma

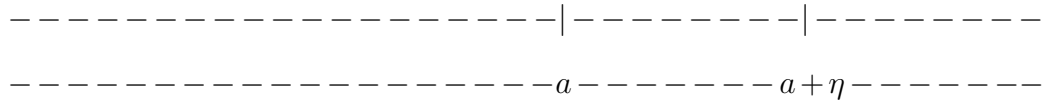
Per ogni $a \in \mathbb{R}$ e per ogni $\eta > 0$ si ha

$$P(Y \leq a) \leq P(X \leq a + \eta) + P(|X - Y| \geq \eta).$$

Si ha

$$(Y \leq a) \subset (X \leq a + \eta) \cup (|X - Y| > \eta).$$

Infatti se $Y(\omega) \leq a$, e se ω non appartiene al primo dei due eventi a destra, allora appartiene al secondo.



In termini più analitici, se $Y(\omega) \leq a$ e $X(\omega) \geq a + \eta$, allora $X(\omega)$ e $Y(\omega)$ distano tra loro di più di η .

La tesi segue poi dalla subadditività della probabilità.

Teorema

Se $\{X_n\}$ converge a X in probabilità, allora converge ad X anche il legge.

Sia t_0 un punto di continuità della F . Fissiamo $\delta > 0$. Dal lemma precedente otteniamo

$$P(X_n \leq t_0) \leq P(X \leq t_0 + \delta) + P(|X_n - X| > \delta)$$

ed anche, scambiando X con X_n e sostituendo t_0 con $t_0 - \delta$,

$$P(X \leq t_0 - \delta) \leq P(X_n \leq t_0) + P(|X_n - X| > \delta).$$

Dalla seconda disuguaglianza otteniamo che

$$P(X \leq t_0 - \delta) - P(|X_n - X| > \delta) \leq P(X_n \leq t_0) = F_n(t_0),$$

termine che ora maggioriamo usando la prima disuguaglianza

$$\leq P(X \leq t_0 + \delta) + P(|X_n - X| > \delta).$$

Quando n tende all'infinito, $P(|X_n - X| > \delta)$ tende a zero, per ipotesi, e quindi si ottiene che

$$F(t_0 - \delta) \leq \liminf_{n \rightarrow \infty} F_n(t_0) \leq \limsup_{n \rightarrow \infty} F_n(t_0) \leq F(t_0 + \delta).$$

D'altra parte, per la monotonia della F ,

$$F(t_0 - \delta) \leq F(t_0) \leq F(t_0 + \delta).$$

Sapendo che la F è continua in t_0 , per ogni $\varepsilon > 0$ esiste un $\delta > 0$ tale che $F(t_0 + \delta) - F(t_0 - \delta) < \varepsilon$. Dunque $F(t_0)$, $\liminf_{n \rightarrow \infty} F_n(t_0)$ e $\limsup_{n \rightarrow \infty} F_n(t_0)$ appartengono allo stesso intervallo di ampiezza ε , che è peraltro arbitrario e dunque da questo si deduce che i tre numeri coincidono, ovvero

$$\liminf F_n(t_0) = \limsup F_n(t_0) [= \lim_{n \rightarrow \infty} F_n(t_0)] = F(t_0),$$

da cui la tesi.

Ci si può chiedere se la convergenza in legge implichi quella in probabilità. La risposta è negativa e quasi immediata. Abbiamo dimostrato prima che per la convergenza in probabilità c'è l'unicità del limite q.c.. Con un semplice esempio si può vedere che non è così con la convergenza in legge.

Esempio

Sia $\Omega = [0, 1]$, \mathcal{B} la σ -algebra dei boreliani e P la misura di Lebesgue. Sappiano che se U è uniformemente distribuita su $[0, 1]$, anche $V = 1 - U$ ha la stessa legge. Se prendiamo allora la successione costante $X_n = U$, questa converge in legge sia a U che a V .

Rivediamo ora, alla luce delle definizioni introdotte, la variabile aleatoria di Poisson e il modo con il quale è stata da noi introdotta.

Ci proponiamo di dimostrare che la successione $X_n \sim B(n, \frac{\lambda}{n})$ converge *in legge* alla v.a. di Poisson di parametro λ .

La funzione di ripartizione della v.a di Poisson X di parametro λ è discontinua negli interi non negativi. Il "salto" che fa nel punto k , ($0 \leq k < +\infty$) è pari a $\frac{\lambda^k}{k!} e^{-\lambda}$.

Prendiamo un numero reale positivo t che non sia un intero. La funzione di ripartizione di X in t vale

$$F(t) = \sum_{k=0}^{\lfloor t \rfloor} \frac{\lambda^k}{k!} e^{-\lambda},$$

mentre, per $n > \lfloor t \rfloor$ la f.r. di X_n è uguale a

$$F_n(t) = \sum_{k=0}^{\lfloor t \rfloor} \binom{n}{k} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Abbiamo dimostrato, quando abbiamo introdotto la v.a di Poisson, che $\binom{n}{k} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$ tende, per n che tende all'infinito, a $\frac{\lambda^k}{k!} e^{-\lambda}$. I termini della seconda somma tendono quindi, per ogni k , ai termini della prima. Ma la somma è finita e quindi possiamo utilizzare la linearità del limite e concludere con la tesi che volevamo dimostrare.

Nota: volta per volta gli spazi Ω_n su cui sono definite le X_n cambiano. Ma se questo dà fastidio, in questo esempio questo aspetto è eliminabile, perché si potrebbe pensare tutte le X_n definite su $\Omega = \mathbb{N}_0$.

Vediamo ora un altro esempio.

Sia $\Omega_n = \{\omega = \frac{k}{n}, 1 \leq k \leq n\}$, con la P_n uniformemente distribuita sugli eventi elementari. La σ -algebra è ovviamente la famiglia delle parti di Ω_n .

La f.r. di X_n è nulla per $t < \frac{1}{n}$ e vale 1 se $t \geq 1$. Per gli altri valori di t si ha invece

$$F_n(t) = P(X_n \leq t) = \sum_{k=1}^{\lfloor nt \rfloor} \frac{1}{n} = \frac{\lfloor nt \rfloor}{n}.$$

Dalla disuguaglianza $t - \frac{1}{n} \leq \frac{\lfloor nt \rfloor}{n} = F_n(t) \leq t$, segue che, per $0 \leq t \leq 1$, $\lim_{n \rightarrow \infty} F_n(t) = t$.

Quindi la successione X_n converge in legge alla v.a. uniformemente distribuita su $[0, 1]$, come era prevedibile.

In questo caso gli spazi (finiti) Ω_n sono inevitabilmente diversi tra di loro e lo spazio $\Omega = [0, 1]$, su cui è definita la v.a. limite, ha cardinalità del continuo.

Vedremo l'importanza della convergenza in legge quando affronteremo il tema del Teorema Centrale del Limite (TCL).

Ora vedremo l'enunciato di un notevole teorema dovuto a Paul Lévy e che lega la convergenza in legge alla convergenza delle funzioni caratteristiche.

Teorema di Paul Lévy

Se $\{X_n\}$ è una successione di v.a., allora essa converge in legge a X se e solo se la successione delle corrispondenti funzioni caratteristiche $\{\Phi_n\}$ converge puntualmente su tutto \mathbb{R} alla funzione caratteristica Φ della X .

9 I teoremi dei grandi numeri

Se $\{X_n\}$ è una successione di v.a. indipendenti, la legge dei grandi numeri descrive il comportamento della media empirica $M_n = \frac{1}{n} \sum_{k=1}^n X_k$.

$\{X_n\}$ potrebbe ad esempio rappresentare una sequenza di lanci di un dado e la (una) domanda è come si comporta M_n al tendere di n all'infinito?

Questa è *la simulazione* una sequenza di venti lanci di un dado (ottenuto con un iphone).

3, 1, 6, 1, 6, 6, 1, 5, 5, 2, 1, 3, 4, 3, 4, 1, 1, 2, 2, 4, ...

La media è $M_{20} = 2,7$.

Che cosa prevediamo succeda se continuiamo a far crescere il numero dei lanci?

Il matematico italiano Gerolamo Cardano (1501-1576) fu il primo ad accorgersi, sperimentalmente, di questo fatto curioso: mentre un singolo evento casuale non si può assolutamente prevedere, si possono invece fare delle previsioni "collettive" su come si comporterà un insieme di eventi.

Il risultato del lancio di un dado è imprevedibile, ma se lo lanciamo molte volte, il punteggio medio "tenderà" a 3.5. Naturalmente quando Cardano fece queste osservazioni, non era ancora stato formalizzato nemmeno il concetto di limite di una successione!

Il primo a dimostrare un teorema dei grandi numeri fu Jacob Bernoulli (1654-1705). Lo pubblicò in un suo articolo del 1689 ed esso fu ripresentato nel suo famoso libro *Ars Conjectandi*, pubblicato postumo nel 1713.

Il teorema che vedremo adesso è molto vicino alla sua formulazione.

Legge debole dei grandi numeri

Se $\{X_n\}$ è una successione di v.a. indipendenti, $X_n : \Omega \rightarrow \mathbb{R}$, che hanno la stessa legge e ammettono varianza σ^2 (e speranza matematica μ), allora

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mu \text{ in probabilità.}$$

Useremo la disuguaglianza di Čebišëv. Notiamo intanto che $E(\frac{1}{n} \sum_{k=1}^n X_k) = \mu$. Questo segue immediatamente dalla linearità della speranza matematica. Ma allora, per ogni $\eta > 0$,

$$P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| \geq \eta \right) \leq \frac{\text{Var}(\frac{1}{n} \sum_{k=1}^n X_k)}{\eta^2} = \frac{1}{n^2} \frac{\sum_{k=1}^n \text{Var} X_k}{\eta^2} = \frac{\sigma^2}{n \eta^2}$$

e questo è un infinitesimo, da cui la tesi.

Bernoulli diede una stima esplicita di quanto debba essere grande n perché la probabilità sia inferiore a $\varepsilon > 0$

Bernoulli ha dimostrato anche altre cose interessanti, ad esempio che il segno di $M_n - \mu$ cambia infinite volte, al crescere di n .

La legge dei grandi numeri è stata studiata successivamente da molti matematici che cercarono di indebolire le ipotesi e di rafforzare la tesi. Uno dei risultati più facili da capire (ma non facili da dimostrare) è il seguente.

Legge forte dei grandi numeri

Se $\{X_n\}$ è una successione di v.a. indipendenti, $X_n : \Omega \rightarrow \mathbb{R}$, che hanno la stessa legge e ammettono varianza σ^2 (e speranza matematica μ), allora

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mu \text{ quasi certamente.}$$

Le ipotesi sono le stesse, la tesi è più forte.

Storia e storie

Wikipedia (ed altre fonti) attribuiscono a Gerolamo Cardano (1501–1576) l'osservazione che “l'accuratezza delle statistiche empiriche tendono a migliorare con il crescere del numero dei tentativi”. Riporto la citazione da Wikipedia, perché finora non sono riuscito a trovare citazioni, neanche di seconda mano, dell'originale.

I risultati che ottenne in algebra ed anche in meccanica (giunto cardanico) sono ritenuti nei siti internet di importanza prevalente rispetto a questa geniale intuizione probabilistica.

Per spiegare con un esempio, se nel lancio di dadi (o di una o due o più monete) la probabilità di un evento è p , allora aumentando a sufficienza il numero n dei lanci e denotato con k il numero di volte in cui l'evento si ripresenta, k/n “non è molto lontano” da p . Cardano ebbe questa ed altre intuizioni geniali, ma era figlio del suo tempo.

È stato una personalità molto controversa. Profondamente religioso, ma ad un certo punto finì davanti all'inquisizione. Accanito giocatore d'azzardo, negli anni tardi sembrò pentirsi, scrivendo “merito di esser biasimato per essermi dato al gioco degli scacchi e dei dadi senza alcun freno”. Era un

grande scienziato, ma al contempo si vantava di aver spiegato “che cosa sia il fato e come si esplichì”.

La LGN ci dice possiamo essere ragionevolmente sicuri che la media sperimentale, che calcoliamo a partire da un numero sufficiente di campioni, sia sufficientemente vicina alla media “vera”.

Cosa vuol dire numero sufficientemente grande e sufficientemente vicino è difficile da decidere.

Però possiamo fare come suggeriva De Finetti: facciamo degli esperimenti.

In un vasetto mettiamo 100 monetine da un centesimo. Lo agitiamo e rovesciamo sul tavolo il contenuto.

Possiamo farci alcune domande a priori e poi verificare l’esito dell’esperimento.

- 1) È plausibile che tutte le monetine mostrino la stessa faccia?
- 2) Che teste e croci siano equamente divise 50:50?
- 3) Che si ottengano più di 70 teste?
- 4) Che la sequenza di lanci sia del tipo TCTCTCTCTC.....

Riportiamo di seguito l’esperimento fatto con il lancio di 1000 monete (T=numero, C= Castel del Monte).

Anzi, ad essere sinceri, abbiamo agitato e vuotato dieci volte il vasetto con cento monetine da un centesimo.

Per questo mancano (non sono mai esistiti) i risultati intermedi.

Testa	Croce	Frequenza cumulativa T.	Freq. cum. relativa
45	55	45	0.45
48	52	93	0.465
47	53	140	0.46̄
50	50	190	0.47
54	46	244	0.488
51	49	295	0.4916̄
55	45	350	0.5
43	57	393	0,49125
42	58	435	0.483̄
43	57	478	0.478
totale 478	totale 522		

Che cosa possiamo dedurre dall'esperimento?

1) Che ad ogni lancio di 100 monete, le Teste e le Croci si distribuiscono abbastanza equamente e in ogni lancio la frequenza relativa è abbastanza vicina a 0.5. Che sia proprio uguale a 0.5 è più un'eccezione che la regola.

2) Al crescere del numero dei lanci la frequenza cumulativa relativa sembra essere attratta da 0.5. È come un tappo di sughero catturato da un vortice. Un po' si avvicina, un po' si allontana, ma, a lungo andare, si agita sempre più vicino al centro del vortice.

3) Si osservi che dopo 700 lanci la frequenza relativa è proprio 0.5. Poi si è allontanata, ma non di molto. Questi sbalzi capitano spesso.

4) La media non si può mai stabilizzare su 0.5. Se ad un certo momento si raggiunge 0.5 (cosa, come abbiamo visto, non impossibile), il lancio della moneta successiva "sporca" l'equilibrio.

5) Nell'esempio in ogni singolo lancio di 100 monete le teste, o le croci, non hanno mai superato le 60 unità. È un caso. Sul web ho visto riportato un esperimento analogo, nel quale i risultati erano simili ai nostri, ma con qualche 61 e 63 a favore di testa o croce.

Come abbiamo visto, un pareggio 50:50 è possibile, ma più raro di quanto si possa pensare. È un po' meno dell'8%.

Uno sbilanciamento del tipo 70/30 (o peggio) è molto raro, la probabilità è circa del 0,5%.

Ottenere 100 croci è possibile, ma estremamente poco probabile.

La probabilità è pari a $2^{-100} \sim 10^{-30}$. Per dare un senso a questo numero, pensiamo a una sfera che abbia raggio 6000 km, riempita di granelli di sabbia delle dimensioni di un millimetro cubo. Tutti bianchi, tranne un granello nero. La probabilità di estrarre a sorte il granello nero è circa 10^{-30} .

La legge dei grandi numeri mette in evidenza che il risultato di un lancio di una moneta è assolutamente imprevedibile, mentre invece riusciamo a dire qualcosa di molto significativo sul loro comportamento collettivo.

In uno scenario completamente diverso si nota lo stesso fenomeno: a meno di diagnosi infauste già certificate, per un settantenne medio non sappiamo

dire se l'anno prossimo sarà ancora vivo. Ma (dati ISTAT raccolti qualche anno fa) la probabilità che non ce la faccia è dell'1,3% circa.

Gli esperimenti di Kerrich

Il matematico inglese John Edmund Kerrich (1903-1985) stava visitando nel 1940 dei suoi parenti in Danimarca e durante la visita avvenne l'invasione nazista.

Essendo cittadino inglese, fu internato dai nazisti in un campo di concentramento danese, dove rimase fino alla fine della guerra. Come prigioniero civile ebbe parecchio tempo libero a disposizione.

Durante quel lungo periodo fece numerosi esperimenti con monete e palline di ping-pong e altri oggetti, per dimostrare la validità empirica di varie leggi probabilistiche.

Dopo la guerra pubblicò un libricino intitolato "An Experimental Introduction to the Theory of Probability", nel quale è riportato un esperimento del lancio di una moneta ripetuto 10,000 volte, nel quale aveva ottenuto 5067 teste.

Nel "nostro" esperimento siamo distanti di 0.022 da 0.5 mentre nell'esperimento di Kerrich la distanza è di 0.0067.

Anche questo esperimento risultò sia la LGN sia la lentezza del miglioramento dell'approssimazione.

Il grande probabilista Persi Diaconis (1945-) ha voluto effettuare un esperimento su una sua congettura riguardante il lancio di una moneta. In un articolo del 2004 ha dimostrato che il gioco non è, malgrado quello che in genere si pensa, un gioco equo.

La sua intuizione era che nel lancio di una moneta il suo movimento rotatorio crei un effetto giroscopico la cui dinamica influisce sull'equilibrio dei due risultati e che ci possa essere un (leggero) vantaggio per la faccia che, al momento del lancio, sta sopra.

Per avere anche una conferma sperimentale di questa sua idea (e per poter distinguere la deviazione che si aspettava dalle oscillazioni statistiche), chiese a due suoi studenti Susan Holmes e Richard Montgomery di effettuare, durante un semestre, 20.000 lanci di una moneta ciascuno.

Per uniformare i lanci e "sterilizzare" l'effetto di un errore sistematico dovuto al modo in cui una specifica persona effettua i lanci, era stato creato un apposito congegno che li eseguiva

La legge dei numeri enormi.

Il matematico sovietico Boris Vladimirovich Gnedenko (1912-1995) ha pubblicato nel suo libro il seguente esempio:

Prendiamo due vasi comunicanti A e B aventi lo stesso volume e li riempiamo di una mole di ossigeno ($\sim 6 \cdot 10^{23}$ molecole). Dopo 24 ore ciascuna molecola si trova nel vaso A con probabilità $\frac{1}{2}$. Quindi è come se noi facessimo $6 \cdot 10^{23}$ lanci di una moneta (sorvoliamo sull'indipendenza).

La domanda è: qual è la probabilità che in uno dei due vasi ci sia la centomiliardesima parte di molecole in più rispetto all'altro?

Si può dimostrare che questa probabilità è inferiore a 10^{-100} .

Le legge dei numeri enormi (il nome è un'invenzione dello scrivente e non si trova nella letteratura) entra in gioco quando studiamo fenomeni microscopici che coinvolgono un numero enorme di particelle (molecole, elettroni, neutroni etc), i quali osservati a livello macroscopico non presentano fluttuazioni strumentalmente rilevabili.

Le informazioni ottenute con un grande numero di esperimenti sono alla base della statistica che utilizza la legge dei grandi numeri alla rovescia.

Supponiamo di essere interessati ad un evento E , ma di non conoscerne la probabilità.

Allora ripetiamo numerose volte l'esperimento e se notiamo che la sua frequenza relativa "ronza" intorno ad un certo numero, è ragionevole assumere quel numero come un'approssimazione dell'ignota probabilità di E .

Questo utilizzo del TGN era noto almeno nel diciassettesimo secolo, quindi prima della sua dimostrazione.

Vediamo qualche esempio.

Sarà maschio o femmina?

L'inglese John Graunt (1620-1674) è considerato il padre della demografia, cioè della scienza che ha per oggetto lo studio delle popolazioni umane, che tratta del loro ammontare, della loro composizione, della mortalità, del loro sviluppo e dei loro caratteri generali, considerati principalmente da un punto di vista quantitativo. (da Wikipedia)

Graunt fu il primo ad osservare che nascono più maschi che femmine, contrariamente alla credenza comune.

John Arbuthnot (1667-1735) fu uno statistico, medico e scrittore scozzese che dimostrò l'osservazione di Graunt utilizzando dati sui battesimi che ottenne dalle parrocchie londinesi (82 anni dal 1629 al 1710) e scoprì che tutti gli anni i maschi erano [troppo] più numerosi delle femmine.

Ma, ragionò lui, se la nascita di un maschio e di una femmina fossero ugualmente probabili, la prevalenza degli uni o degli altri avrebbe probabilità $\frac{1}{2}$. Il risultato da lui raccolto avrebbe allora la stessa probabilità di ottenere 82 volte testa con una moneta equilibrata, cioè $\frac{1}{2^{82}} \sim 2 \times 10^{-24}$. Ne concluse che “la moneta” era squilibrata, cioè che la nascita dei maschi è più frequente.

In effetti i dati raccolti da Arbuthnot dimostrano, al di là di ogni dubbio, che molte femminucce non furono mai battezzate e che finirono, tragicamente, nel Tamigi.

A cavallo dell' '800 in alcune città europee venne messo in piedi il servizio anagrafico, che registrava nascite, morti e stato civile della popolazione.

Laplace descrisse nel suo libro “Essai philosophique sur les probabilités”, pubblicato nel 1814, un episodio curioso.

I dati statistici provenienti da Londra, Berlino e San Pietroburgo e che si riferivano a diversi decenni, davano come rapporto di nascite maschi/femmine un numero abbastanza vicino a 0.517.

Invece i dati raccolti a Parigi erano vicini piuttosto a 0.510. La differenza era piccola (sette millesimi), ma significativa, vista la grande quantità dei dati raccolti.

Laplace, incuriosito da questa differenza, cercò di capirne il motivo.

Uno studio attento degli archivi parigini mostrò che venivano contati tra i nati a Parigi anche i bambini abbandonati e portati in città dalle campagne circostanti.

Siccome le bambine erano, più dei maschietti, considerati un peso economico per la famiglie povere, c'era, tra i bambini abbandonati, una tale prevalenza di femminucce da influenzare i dati statistici.

Contando solamente i bambini nati a Parigi, il dato statistico risultò compatibile con i dati provenienti dalle altre tre capitali.

Un'interpretazione errata della LGN

Discuteremo ora di uno degli errori più diffusi e anche più pericolosi (delle persone si sono rovinare e suicidate per questo motivo) riguardante la LGN.

Se cercate su Google la voce “numeri ritardatari” troverete ben 726.000 risultati.

Come mai tutto questo interesse per i numeri ritardatari?

Ci sono delle persone (che chiamerò in seguito “ritardatari”) che dalla LGN traggono la conclusione che il ritardo di un certo risultato “favorisca” l’evento ritardatario negli esperimenti successivi.

Il lotto è molto popolare e la quantità di soldi che fa girare è enorme e quindi credo che c’è anche un interesse peloso a rinfocolare questa credenza.

Ad oggi (3 maggio 2020) il numero più “ritardatario” è il 45 sulla ruota di Bari, che non è stato estratto da 109 sorteggi consecutivi.

Possiamo essere sicuri che un numero consistente di giocatori punti su questo numero.

È difficile giustificare un’idea nella quale non si cede, ma ci proverò.

Dal fatto che la frequenza teorica del 45 è $\frac{1}{18} = 0,05\bar{5}$, se la frequenza osservata si allontana, c’è chi che crede che qualche forza soprannaturale spinga il 45 ad uscire per rimettere le cose a posto, perché altrimenti la frequenza relativa vagherebbe senza meta!

Ma nella LGN non c’è nulla di magico. Per semplicità osserviamo che cosa succede con il lancio di una moneta. A contare il numero delle frequenze possibili è il triangolo di Tartaglia.

Limitiamoci a sei lanci di una moneta equilibrata. Sono pochi, ma già indicano una tendenza.

Per la formula di Bernoulli, la probabilità che escano k teste è, per $0 \leq k \leq 6$

$$P(k) = \binom{6}{k} \cdot 2^{-6}.$$

Dunque la probabilità che il numero dei successi, k , differisca dalla media (che è 3) di al più un’unità è

$$P(2 \leq k \leq 4) = P\left(\left|\frac{k}{6} - \frac{1}{2}\right| \leq \frac{1}{6}\right) = \frac{50}{2^6} = 0,78125,$$

quasi l’ottanta per cento.

Se facciamo 12 lanci, si trova che

$$P\left(\left|\frac{k}{12} - \frac{1}{2}\right| \leq \frac{1}{6}\right) = 0,854.$$

Il gioco del lotto in parlamento

Nel 1998 era successo che a Milano era stata appena sgominata una banda che truccava le estrazioni su quella ruota. Con la complicità di funzionari del Ministero delle Finanze e dando istruzioni ad un bambino bendato male, la banda era riuscita a fare alcuni bei colpi.

Nello stesso periodo (ma si è trattato di una coincidenza) ci sono state delle forti vincite al lotto. Queste erano dovute al fatto che erano usciti dei numeri che ritardavano da parecchio tempo.

Per questi motivi, il 20 gennaio 1999 è stato udito dalla sesta commissione del Parlamento italiano il direttore generale dell'amministrazione autonoma dei Monopoli di Stato, che gestiva allora il gioco del lotto.

Il Direttore ha fatto una relazione e ha poi risposto a varie domande dei parlamentari.

Ecco un brano interessante della sua relazione.

“Notiamo [...] vincite di grosse entità rispetto ai numeri ritardatari, con giocate di un certo impegno. L'utilizzo del mezzo informatico e il proliferare di agenzie e pseudo esperti che danno i numeri sulla base di calcoli informatici fanno sì che sempre di più si restringa l'area dell'alea. [...] sapendo che la legge dei grandi numeri prevede un massimo di 202 settimane di mancata esclusione di un numero, se si cominciano a giocare i numeri che ritardano oltre una certa settimana ci si avvicina con una certa approssimazione alla vincita.”

L'unica cosa giusta che dice il Direttore dei Monopoli è che le grandi vincite capitano quando finalmente esce qualche numero ritardatario. È chiaro che, con il passare delle estrazioni, sono sempre più numerose e sempre più consistenti le puntate che si concentrano su questi numeri e quando finalmente questi numeri escono, lo stato sborsa più soldi del solito.

Soldi abbondantemente compensati dalle puntate delle estrazioni precedenti che hanno costretto molti giocatori, per mancanza di risorse, a ritirarsi dalla corsa al rialzo.

Poi però il direttore dei Monopoli spara una cavolata dopo l'altra.

Sostiene che l'utilizzo scientifico del gioco del lotto e dei calcoli informatici ridurrebbe l'area dell'alea. Ma quali sarebbero questi calcoli informatici?

Se esistesse un modo scientifico per vincere al lotto, le spiagge dei Caraibi sarebbero affollate di probabilisti e informatici, trasferitisi lì per godersi i soldi

fregati allo stato. (E lo stato, invece di aumentare il numero delle estrazioni, abolirebbe il gioco del lotto.)

Il Direttore si contraddice: quelli che trovano i numeri “buoni” usando il calcolo scientifico sarebbero pseudo esperti. Ma se il metodo funzionasse davvero dovrebbero essere chiamati super esperti!

Ma vediamo ora la perla di tutto il discorso.

Dice il Direttore dei Monopoli: “sapendo che la legge dei grandi numeri prevede un massimo di 202 settimane di mancata esclusione di un numero, se si cominciano a giocare i numeri che ritardano oltre una certa settimana ci si avvicina con una certa approssimazione alla vincita.” Bernoulli si rivolta nella tomba.

Naturalmente la LGN non fissa il massimo ritardo di un numero del lotto! Nel '99 il massimo ritardo registrato nella storia del lotto è stato di 201 settimane. La tabella aggiornata è questa:

Numero	Ruota	Ritardo	Data
53	Nazionale	257	24.1.2017
76	Cagliari	210	27.1.2018
34	Cagliari	203	1.4.2006
8	Roma	201	23.8.1941
21	Bari	201	25.7.2017
82	Firenze	198	4.12.2018
55	Bari	196	12.3.1960
82	Bari	193	27.11.1943
71	Cagliari	191	26.6.1971
67	Venezia	191	18.10.1924

Il Direttore è stato smentito ben tre volte, dopo il '99.

In futuro succederà ancora che i ritardi supereranno le 201 ed anche le 257 estrazioni.

Già Bernoulli sapeva che comunque fissiamo un limite (201, 257, 500, 1000, ... prima o poi il ritardo di qualche numero del lotto supererà il limite fissato.

Lo stesso teorema però dice anche che più grande è questo limite, più lunga sarà, probabilmente, l'attesa.

Per superare il ritardo di 201 estrazioni, che risale al 1941, si è dovuto aspettare fino al 2006 per avere un ritardo di 203 estrazioni.

Chi ha creduto agli atti della Commissione Parlamentare, arrivato alla duecentoduesima estrazione avrebbe giocato tutti i suoi averi e si sarebbe indebitato con banche e strozzini, perché il Direttore aveva affermato che, sicuro come la morte, il 34 sulla ruota di Cagliari sarebbe uscito.

Il “ritardo” in letteratura e nel cinema

Edgar Alan Poe (1809-1849)

Uno dei “ritardatari” più famosi è il grande scrittore americano Edgar Alan Poe, inventore del racconto poliziesco, della letteratura dell’orrore e del giallo psicologico (Wikipedia).

Nel racconto “Il mistero di Marie Rogêt” scrive:

Niente [...] è più difficile che convincere il lettore medio che se un giocatore di dadi tira due volte di seguito il sei in una partita, questo basta per scommettere con buon margine di sicurezza che il sei non uscirà una terza volta. [...] L’errore [...] non pretendo di chiarirlo nei limiti che mi sono posti, e con i logici non è necessario chiarirlo.

Come dire che le persone intelligenti lo capiscono da sole e con gli ignoranti non merita sprecare parole.

Poe era un grande scrittore, ma la probabilità la doveva lasciare a quelli del mestiere.

Tom Stoppard

Lo scrittore inglese Tom Stoppard ha scritto nel 1964 la tragicomedia intitolata “Rosencrantz e Guildenstern sono morti”. Da essa è stato tratto nel 1990 un film.

I due personaggi del titolo sono, nel dramma di Shakespeare, amici di infanzia del principe Amleto. L’assurdo della storia sta nella trasformazione della prospettiva, vista questa volta dalla parte di due personaggi marginali della vicenda.

Quando la scena si apre, i due stanno giocando a testa e croce (forse si tratta di un richiamo all’”Essere o non essere”) e a creare l’atmosfera surreale contribuisce anche il fatto che ottengono Testa per cento volte consecutive (probabilità = $2^{-100} \sim 0,8 \cdot 100^{-30}$).

Il metodo del raddoppio

La tecnica del raddoppio è un’altra illusione di chi spera di vincere contro lo stato o contro i casinò.

L'idea sarebbe questa: al casinò punto 1 euro sul rosso. Se esce, bene, intasco l'euro e continuo a giocare.

Se non esce, al turno successivo punto 2 euro sul rosso. Se esce, vinco 2 euro. Così mi rifaccio della perdita del primo giro e resto con una vincita di 1 euro.

Ma magari non esce neanche al secondo giro. E allora raddoppio nuovamente la puntata: punto 4 euro sul rosso. Se esce, vinco 4 euro, mi rifaccio delle perdite (1+2) dei giri precedenti e mi resta sempre la vincita di 1 euro.

Facciamolo ancora una volta: se il rosso non esce neanche la terza volta, le mie perdite ammontano a $1+2+4=7$ euro. Punto allora 8 euro e se vinco, mi rifaccio delle perdite ed ho una vincita netta di 1 euro.

La tecnica prevede che si continui così, con un raddoppio della scommessa se si perde, e quando esce finalmente il rosso il giocatore si è rifatto di tutte le perdite e vince 1 euro netto.

Splendido! Però . . . il raddoppio inizialmente non si fa sentire molto, ma se il rosso tarda troppo, può diventare pesante.

È notissima la leggenda del principe indiano al quale un mercante chiese, per avergli fatto conoscere il gioco degli scacchi, come compenso un chicco di grano per la prima casella della scacchiera, due per la seconda, quattro per la terza, e così via, raddoppiando ogni volta. La richiesta sembrava ragionevole, ma quando il principe fece i conti, scoprì che il suo regno non avrebbe mai potuto produrre tutto il grano che il mercante gli chiedeva.

Da una ricerca effettuata nei siti di lingua inglese, sembra che il record di uscite dello stesso colore alla roulette sia avvenuta nel 1943 in un casinò americano, quando uscì per ben 32 volte di fila il colore rosso.

Chi, eventualmente, aveva incominciato a giocare con la tecnica del raddoppio, quando finalmente è uscito il rosso, era rovinato da un bel po' perché per rifarsi e guadagnarsi il sospiratissimo dollaro iniziale, avrebbe dovuto puntare oltre 4 miliardi di dollari (di allora, 60 miliardi di oggi). Tra l'altro, nessun casinò accetta puntate così alte.

E comunque, la sua vincita sarebbe stata di 1 dollaro!

Il casinò con maggior tradizione è senza dubbio quello di Monte Carlo e la documentazione relativa alla sua gestione è molto precisa.

Il maggior ritardo del rosso in quel casinò risale alla sera del 18 agosto 1913, quando il nero è uscito 26 volte consecutive.

Chi aveva (eventualmente) iniziato a puntare 1 franco giocando al raddoppio, al ventisettesimo giro doveva puntare, per riprendersi il suo franco, una somma corrispondente, oggi, a 4 milioni di euro.

Riassumendo, il metodo del raddoppio presenta due problemi:

1) Bisogna avere a disposizione un capitale notevole per far fronte alle possibili lunghe serie negative. Al confronto anche del giocatore ricchissimo il banco (casinò, stato) dispone di capitali praticamente illimitati.

2) È vero che la probabilità di perdere è piccola, ma questa relativa certezza è bilanciata, in negativo, dal modestissimo guadagno finale e dalla somma enorme che uno rischia di perdere.

È la situazione opposta a quella che si ha, ad esempio, con la Lotteria Italia: il prezzo di un biglietto era nel 2019 di 5 euro, mentre il primo premio era di 5.000.000 di euro. Piccola spesa (e piccola probabilità) e grande premio.

Anche nel gioco, se uno proprio non sa farne a meno, conviene attenersi alla modica quantità.

La LGN dei “ritardatari” ha la memoria corta

I “ritardatari” fanno affidamento ad una LGG dalla memoria corta.

Osservando circa un anno fa i dati relativi alla ruota “Nazionale”, creata nel maggio del 2005, avevo notato che il numero che in quel momento ritardava di più su quella ruota era allo stesso tempo anche il numero che era stato estratto più volte nel corso dei 14 anni.

Per uno che intende la LGN come una sorta di legge di compensazione, sostenuta da una forza soprannaturale che cerca di ripianare il numero delle estrazioni dei vari numeri, si trova, come l’asino di Buridano, davanti a due spinte ugualmente (per lui) allettanti: dovrebbe giocare sul numero ritardatario, ma per le stesse ragioni dovrebbe anche evitare quel numero perché essendo uscito tante volte, ora deve cedere il passo ad altri.

Ma nessuno dei “ritardatari” tiene conto di tutta la storia di un numero, sarebbe troppo faticoso perché richiederebbe un’analisi di tutte le estrazioni dal 1871 in poi e tutti suggeriscono si appigliano al passato più recente e consigliano di puntare sul numero ritardatario.

Si tratta anche di una scelta furbesca, più attrattiva dal punto di vista delle vendite. I numeri meno estratti sulle varie ruote rimangono probabilmente gli stessi per mesi, anni e forse decenni. Il gioco diventerebbe allora

noioso. Così invece, uscito un numero ritardatario, c'è subito un altro che prende il suo posto e la caccia al numero “buono” si riapre.

I ritardatari preferiscono credere che questa forza soprannaturale, in cui E. A. Poe credeva così fermamente, ha memoria, ma non troppa.

Attualmente, sempre sulla ruota Nazionale, il numero 88 è stato finora estratto solamente 98 volte, (quasi il 25% in meno rispetto alla media di circa 130 estrazioni per gli altri numeri). Ma il suo ritardo è di sole 29 estrazioni! Che fare?

I ritardatari non hanno dubbi: evitare l'88 come la peste!

Ma finiamola qui con questi ragionamenti. Oscar Wilde diceva: Mai discutere con un idiota, ti trascina al suo livello e ti batte con l'esperienza.

Metodo di Monte Carlo

Il metodo di Monte Carlo era il nome in codice del metodo inventato da Ulam e usato da von Neumann ed altri per simulazioni riguardanti la costruzione della bomba all'idrogeno.

L'idea venne a Ulam nel 1946, quando stava riprendendosi da una grave malattia e, per passare il tempo, giocava a “Canfield solitaire”, un solitario che si gioca con un mazzo di 52 carte.

Egli cercò in un primo momento di calcolare la probabilità di completare il solitario con successo. Dopo aver provato a lungo e inutilmente, gli venne l'idea di fare un centinaio di tentativi e di assumere, come risposta approssimata, il risultato di questi tentativi.

Parlandone con von Neumann, venne l'idea di utilizzare metodi casuali per il calcolo di certi integrali multidimensionali che erano essenziali per il completamento del progetto.

I due scoprirono che la legge dei grandi numeri fornisce un modo probabilistico per il calcolo degli integrali.

Proposizione

Se f è una funzione definita su $[0, 1]$ e a valori reali integrabile secondo Lebesgue e se $\{U_n\}$ è una successione di variabili aleatorie indipendenti uniformemente distribuite in $[0, 1]$, allora

$$\frac{1}{n} \sum_{k=1}^n f(u_k) \rightarrow \int_0^1 f(t) d\lambda(t) \text{ q.c. ,}$$

dove u_k sono i valori assunti dalle U_k .

La dimostrazione segue dalla legge forte dei grandi numeri che ci dice che M_n converge quasi certamente a $E(f(U))$ che è uguale a $\int_0^1 f(t) d\lambda(t)$.

A questo punto prendiamo una successione $\{s_k\}$ che simula la v.a. U e consideriamo la successione $\{\frac{1}{n} \sum_{k=1}^n f(s_k)\}$. Per n che tende all'infinito (se la simulazione è stata ben scelta) otteniamo un'approssimazione dell'integrale che desideriamo calcolare.

Il metodo si estende immediatamente al caso pluridimensionale. Se f è, diciamo, continua sull'ipercubo $[0, 1]^m$, allora

$$E(f(V)) = \int_{[0,1]^m} f(t_1, t_2, \dots, t_m), d\lambda_m,$$

dove V è il vettore uniformemente distribuito su $[0, 1]^m$.

Si prende allora una successione di punti $\{z_n\}$ che simula il vettore V e si procede come nel caso unidimensionale.

Naturalmente a nessuno verrebbe in mente di usare questo risultato per il calcolo di un integrale unidimensionale. I metodi numerici (trapezi, Simpson, Gauss) sono più molto più precisi e veloci.

Le cose cambiano se passiamo ad integrali multidimensionali. I metodi numerici si appesantiscono molto, perché richiedono che la funzione integranda venga calcolata in molti punti e questa necessità cresce con la dimensione e il metodo di Monte Carlo diventa competitivo e addirittura più efficace.

Una variante del metodi di Monte Carlo si utilizza per il calcolo della misura m -dimensionale di insiemi A limitati e misurabili secondo Lebesgue.

Prendiamo un iperrettangolo $R \supset A$ e sia V il vett. a. uniformemente distribuito su A .

Sia $\{z_n\}$ una successione che simula il vettore V e contiamo, dato n , quanti sono i punti che stanno in A (diciamo k) e quanti sono gli altri. Ebbene, $\frac{k}{n}$ è un'approssimazione per la misura di A .

Consideriamo ad esempio l'insieme A contenuto nel primo quadrante costituito dai punti (s, t) tali che $t \geq s^3$ e $t \leq \sin s$. Si tratta di una specie di losanga con un vertice nell'origine e l'altro che non si riesca a calcolare esplicitamente. Numericamente troviamo, con tre cifre decimali, il punto $(0.929, 0.801)$. Ci convinciamo facilmente che $A \subset [0, 1] \times [0, 1]$.

L'area di A è circa uguale a

$$\int_0^{0.929} (\sin t - t^3) dt,$$

che possiamo calcolare esplicitamente perché la primitiva della funzione integranda è una funzione elementare, salvo osservare che l'estremo destro di integrazione non è preciso.

In questo esempio il modo di procedere migliore è proprio questo e l'errore dovuto all'imprecisione dell'intervallo di integrazione è trascurabile (e si può rendere meno sensibile calcolando con maggior precisione l'intersezione delle due curve che delimitano A).

Ma non è difficile immaginare esempi più complicati già nel caso bidimensionale, per non parlare di dimensioni superiori.

Si può allora pensare di prendere una successione che simuli la distribuzione uniforme in $[0, 1] \times [0, 1]$ e utilizzare il fatto che è facile decidere se un punto $z_n = (x_n, y_n)$ sta in A oppure no: nel nostro caso si tratta di verificare le due disuguaglianze $y_n \geq x_n^3$ e $y_n \leq \sin x_n$.

L'ago di Buffon

Quello dell'ago di Buffon (1707 –1788) è un risultato isolato che precede il metodo di Monte Carlo di più di due secoli!

Supponiamo di avere su un tavolo una tovaglia a righe, distanziate di un'unità di lunghezza.

Prendiamo un ago di lunghezza pari alla distanza tra due rette contigue e lo lanciamo sul tavolo facendolo ruotare nell'aria. La nostra distanza dal tavolo sia tale che ci impedisca di "mirare" e la posizione in cui l'ago atterra e si ferma si possa assumere casuale.

Ci poniamo la domanda: qual è la probabilità che l'ago intersechi una delle righe?

Possiamo "localizzare" il problema: il centro dell'ago cadrà tra due righe adiacenti e anzi, possiamo limitarci a considerare solamente la retta a cui il centro è più vicino. L'ago certamente non sfiora la riga adiacente.

Indichiamo con x la distanza del centro dell'ago dalla linea più vicina. $\vartheta \in [0, \pi]$ è l'angolo tra l'ago e la direzione delle rette parallele.

Quindi l'ago interseca la retta (più vicina) se e solo se $x \leq \frac{1}{2} \sin \vartheta$.

Noi possiamo assumere che x stia tra 0 e $\frac{1}{2}$ e che ϑ stia tra 0 e π .

La probabilità cercata è il rapporto tra l'area dell'insieme A dove $x \leq \frac{1}{2} \sin \vartheta$ e l'area del rettangolo $[0, \frac{1}{2}] \times [0, \pi]$ che contiene A cioè $\frac{\pi}{2}$.

Invece l'area dell'insieme A è

$$\frac{1}{2} \int_0^\pi \sin \vartheta \, d\vartheta = -\frac{1}{2} \cos \vartheta \Big|_0^\pi = \frac{1}{2}(1 - (-1)) = 1.$$

In definitiva la probabilità cercata è $\frac{2}{\pi}$.

Ma l'idea che ci riporta al metodo di Monte Carlo risale all' '800 e consiste nel ricavare delle approssimazioni per π mediante ripetuti lanci dell'ago. Se su n lanci l'ago interseca una delle rette k volte, il rapporto $\frac{k}{n}$ approssima $\frac{2}{\pi}$. In altre parole, $\pi \approx \frac{2n}{k}$.

Naturalmente si tratta di un metodo poco pratico per approssimare π e la convergenza è molto lenta.

Nella letteratura e anche in internet trovate alcuni esperimenti effettivamente fatti in passato. (Tesi Rizzo)

Esempio di Steinhaus (1887 - 1972)

Supponiamo di avere un allevamento di trote. In una piscina ne abbiamo qualche migliaio.

Vogliamo farci un'idea almeno approssimativa di quante ce ne sono.

Ne peschiamo 100 a caso, le marchiamo e le ributtiamo nella piscina. Se N è il numero delle trote, $\frac{100}{N}$ è la proporzione delle marchate rispetto al totale.

Le lasciamo lì per un po' di tempo in modo che si rimescolino tra di loro e sia ragionevole pensare che le marchate siano uniformemente distribuite nella vasca e poi ne peschiamo, ad esempio, 100.

Tra queste troveremo alcune trote marchate (per esempio 5) ed altre che invece non lo sono. Il rapporto $\frac{5}{100}$ è, per la legge dei grandi numeri, non lontano da $\frac{100}{N}$.

Dunque $\frac{100}{N} \sim \frac{5}{100}$, ovvero $N \sim 2000$.

Ora siamo in grado di dare risposta ad eventuali acquirenti.

Una variante

Una variante di questo problema potrebbe essere quella di mettersi nelle condizioni di utilizzare il *teorema dei numeri enormi*.

Supponiamo di avere in uno stagno la riserva d'acqua potabile per il nostro bestiame. Vorremmo sapere quanti metri cubi di acqua abbiamo a disposizione.

Prendiamo una sostanza solubile nell'acqua (ovviamente non tossica e che non renda comunque l'acqua inservibile) e ne sciogliamo un quantitativo noto. Lo zucchero, per esempio, potrebbe andar bene. La diffusione fa il suo lavoro: dopo un po' lo zucchero sarà uniformemente sciolto nell'acqua. Il numero di molecole dello zucchero e il numero delle molecole dell'acqua sono enormi. A livello macroscopico il fenomeno cessa di essere probabilistico e in base alla concentrazione dello zucchero in un campione raccolto, diciamo, il giorno dopo, possiamo farci un'idea molto precisa del volume dell'acqua nello stagno.

Vediamo un esempio numerico.

Buttiamo nello stagno 200 kg di zucchero disciolto in 100 litri di acqua.

Il giorno dopo facciamo un prelievo e misuriamo 1 grammo di zucchero ogni 100 litri di acqua. Ne abbiamo immesso 200.000 grammi, dunque lo stagno contiene 100×200.000 litri di acqua, ovvero $20.000 m^3$.

Per concludere questo capitolo che riguarda la convergenza delle successioni di v.a., richiamiamo, in una forma leggermente generalizzata, un lemma riguardante un limite notevole.

Un lemma sull'esponenziale

Proposizione. Se $\varphi(x) : \mathbb{R}_+ \rightarrow \mathbb{R}$ è tale che $\lim_{x \rightarrow \infty} \varphi(x) = 0$, allora per ogni $a \in \mathbb{R}$

$$\lim_{x \rightarrow \infty} \left(1 + \frac{a + \varphi(x)}{x} \right)^x = e^a.$$

Per $|y| < 1$ risulta

$$\begin{aligned} \log(1 + y) &= y - \frac{y^2}{2} + \frac{y^3}{3} - \frac{y^4}{4} + \dots + (-1)^{n+1} \frac{y^n}{n} + \dots = \\ &= y - \frac{y^2}{2} + y^3 \left(\frac{1}{3} - \frac{y}{4} + \dots + (-1)^{n+1} \frac{y^{n-3}}{n} + \dots \right) = \\ &= y - \frac{y^2}{2} + y^3 g(y), \end{aligned}$$

dove la g è continua per $|y| < 1$ e $g(0) = 1/3$.

Si ha allora che

$$\log \left[\left(1 + \frac{a + \varphi(x)}{x} \right)^x \right] = x \log \left(1 + \frac{a + \varphi(x)}{x} \right) =$$

$$= x \left(\frac{a + \varphi(x)}{x} - \frac{(a + \varphi(x))^2}{x^2} + \frac{(a + \varphi(x))^3}{x^3} g \left(\frac{a + \varphi(x)}{x} \right) \right) =$$

$$a + \varphi(x) - \frac{(a + \varphi(x))^2}{x} + \frac{(a + \varphi(x))^3}{x^2} g \left(\frac{a + \varphi(x)}{x} \right).$$

Il limite di quest'ultima espressione, per $x \rightarrow \infty$, è a . Ne segue quindi la tesi.

Affrontiamo ora un risultato molto importante, il Teorema Centrale del Limite (TCL).

Teorema Centrale del Limite

Enunciamo il teorema e facciamo qualche commento preliminare.

Sia $\{X_n\}$ una successione di v.a. indipendenti aventi la stessa legge, con media μ e varianza σ^2 .

Allora, posto

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}},$$

si ha che la successione $\{S_n^\}$ converge in legge ad una v.a. $Z \sim N(0, 1)$.*

Osservazioni

Le variabili aleatorie $\{S_n^*\}$ sono le normalizzazioni delle medie $M_n = \frac{1}{n} \sum_{k=1}^n X_k$.

Infatti, $E(M_n) = E(X_1) = \mu$ e $\text{Var}(M_n) = \frac{1}{n^2} n \text{Var}(X_1) = \frac{1}{n} \sigma^2$, quindi lo scarto quadratico medio è $\frac{\sigma}{\sqrt{n}}$.

Questo calcolo è semplice e non serve nella dimostrazione, ma aiuta a digerire e memorizzare l'enunciato perché spiega la presenza di quella radice di n a denominatore.

Vediamo ora una “mappa” della dimostrazione che richiede diversi risultati sparsi nella materia trattata finora.

Il primo passo consiste nel trovare la funzione caratteristica Ψ_n di S_n^* . Useremo per questo qualcuna delle proprietà delle funzioni caratteristiche e il lemma sulla funzione esponenziale appena richiamato. Questa è la parte che richiede più calcoli.

Il passo successivo (e finale) consiste nel far vedere che la successione delle funzioni caratteristiche converge a $e^{-\frac{t^2}{2}}$, che è, come abbiamo visto, la funzione caratteristica di una $N(0, 1)$.

Procediamo ora con la dimostrazione.

Calcoliamo la funzione caratteristica di $S_n^* = \sum_{k=1}^n \frac{1}{\sqrt{n}} Y_k$, dove $Y_k = \frac{X_k - \mu}{\sigma}$.

Se indichiamo con Ψ la funzione caratteristica di Y_k , allora, ricordando il risultato sui valori di Ψ e delle sue derivate per $\vartheta = 0$, si ottiene che $\Psi(0) = 1$, $\Psi'(0) = 0$ (poiché $E(Y_k) = 0$) e $\Psi''(0) = -\text{Var}(Y_k) = -1$, pertanto (usando la formula di Taylor) abbiamo che

$$\Psi_{S_n^*}(\vartheta) = \Psi\left(\frac{\vartheta}{\sqrt{n}}\right)^n = \left(1 - \frac{\vartheta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

Qui abbiamo usato il fatto che la funzione caratteristica della somma è il prodotto delle funzioni caratteristiche e la proposizione che permette di calcolare $\Psi(a\vartheta)$.

Per il lemma precedente quest'ultima espressione tende a $e^{-\frac{\vartheta^2}{2}}$ e così la tesi è dimostrata.

Approssimazione normale

Da ora in poi (questa notazione verrà conservata anche nei capitoli che seguono) indichiamo con Φ la funzione di ripartizione della $N(0, 1)$. Questa tradizionale notazione è usata dalla maggior parte dei testi.

Abbiamo utilizzato apposta, nel teorema precedente, la lettera Ψ per indicare le funzioni caratteristiche in gioco, per evitare che le notazioni ci traggano in inganno.

Per il TCL, la successione delle funzioni di ripartizione $\{F_{S_n^*}\}$ converge in legge a Φ . Ne deduciamo che, per n sufficientemente grande,

$$P(X_1 + X_2 + \dots + X_n \leq t) = P\left(S_n^* \leq \frac{t - n\mu}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{t - n\mu}{\sigma\sqrt{n}}\right).$$

Non ci sono risultati teorici che ci dicano quando l'approssimazione del termine a sinistra con il valore della f.r. della normale è giustificata, ma l'esperienza consiglia un valore $n \geq 30$.

Esempi

1) Lanciamo 200 volte una moneta regolare e ci chiediamo qual è la probabilità che il numero delle teste non superi 110.

Abbiamo a che fare con una somma di 200 v.a. di legge $B(1, \frac{1}{2})$. Sappiamo che $\sigma^2 = pq = \frac{1}{4}$, $\sigma = \frac{1}{2}$.

Si ha allora per n sufficientemente grande,

$$P(X_1 + X_2 + \dots + X_{200} \leq 110) = P\left(S_n^* \leq \frac{110 - 200\frac{1}{2}}{\frac{1}{2}\sqrt{200}}\right) =$$

$$P\left(S_n^* \leq \frac{20}{14.14}\right) = P(S_n^* \leq 1.414) \approx \Phi(1.41).$$

Si tratta ora di trovare il valore di $\Phi(1.41)$. Per questo ci sono le tabelle. Tutti i libri di probabilità le riportano alle ultime pagine (insieme a tabelle relative ad altre v.a) e si trovano anche in rete. Il sito indicato sotto ha anche un'animazione che illustra graficamente la Φ .

www.mathsisfun.com/data/standard-normal-distribution-table.html

Dalla tabella ricaviamo che

$$P(X_1 + X_2 + \dots + X_{200} \leq 110) \approx 0.9207.$$

2) Qual è la probabilità di ottenere almeno 29 teste con 50 lanci di una moneta equilibrata?

Le X_k sono nuovamente delle $B(1, \frac{1}{2})$. Si tratta di calcolare questa volta

$$P(X_1 + X_2 + \dots + X_{50} > 28).$$

Questo esempio è interessante perché la disuguaglianza è rovesciata, ma approfitteremo dell'occasione anche per illustrare quella che si chiama la "correzione di continuità".

È evidente infatti che mentre l'approssimazione fornita dalla Φ è una funzione continua dell'argomento, la funzione $P(X_1 + X_2 + \dots + X_n \leq t)$ è una funzione che è discontinua nei valori interi di t , che sono proprio quelli nei quali interessa calcolarla. Per questo motivo l'approssimazione migliora, se si approssima $P(\sum_{k=1}^n X_k \leq t)$ con

$$\Phi\left(\frac{t + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right).$$

Nel nostro caso questo ci porta a

$$P(X_1 + X_2 + \dots + X_{50} > 28) = 1 - P(X_1 + X_2 + \dots + X_{50} \leq 28) \approx$$

$$1 - \Phi\left(\frac{28 + \frac{1}{2} - 50\frac{1}{2}}{\frac{1}{2}\sqrt{50}}\right) \approx 1 - \Phi(0.99) = 1 - 0.83891 \approx 0.161.$$

3) Un nuovo medicinale dà beneficio al 70% delle persone affette da una certa malattia. Supponiamo di curare con questo medicinale 200 pazienti. Qual è la probabilità che più di 130 pazienti migliorino?

In questo caso le v.a. hanno legge $B(1, 0.7)$, pertanto $\sigma^2 = 0.7 \cdot 0.3 = 0.21$ e $\sigma = \sqrt{0.21} = 0.428$. Quindi

$$P(X_1 + X_2 + \dots + X_{200} \geq 130, 5) = 1 - P(X_1 + X_2 + \dots + X_{200} < 130, 5) \approx$$

$$1 - \Phi\left(\frac{130.5 - 200 \cdot 0.7}{\sqrt{200} \cdot 0.458}\right) = 1 - \Phi\left(\frac{-9.5}{14.14 \cdot 0.458}\right) = 1 - \Phi(-1.47).$$

In questo esempio si presenta un problema nuovo: l'argomento di Φ è negativo.

Ma ricordiamoci che $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2}} ds$ e che la funzione integranda è pari. Ne segue che $\Phi(-\alpha) = 1 - \Phi(\alpha)$.

Possiamo ora concludere il nostro esempio.

$$P(X_1 + X_2 + \dots + X_{200} \geq 130, 5) \approx 1 - (1 - \Phi(1, 47)) =$$

$$\Phi(1, 47) = 0.9292.$$

Funzione di ripartizione empirica

Sia $X_1, X_2, \dots, X_n, \dots$, una successione di v.a. indipendenti aventi la stessa legge di X .

Fissiamo un intero n e un $t \in \mathbb{R}$ e consideriamo la variabile aleatoria definita come

$$F_n^*(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{[X_k, +\infty[}(t).$$

Si noti che $\{F_n^*\}$ è una famiglia di successione di variabili aleatorie che dipendono dalla variabile reale t .

Se eseguiamo le “osservazioni” X_1, X_2, \dots, X_n , e x_1, x_2, \dots, x_n sono i valori osservati, denotiamo con

$$\bar{F}_n(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{[x_k, +\infty[}(t)$$

quella che si chiama *funzione di ripartizione empirica*. Si tratta in questo caso di una funzione della variabile reale t .

Si noti la differenza tra $\{F_n^*\}$ e $\{\bar{F}_n\}$: nel primo caso abbiamo una successione di variabili aleatorie (dipendenti da t), nel secondo caso una funzione reale dipendente da t , ottenuta sostituendo alle v.a. X_k i valori osservati x_k .

Vedremo un risultato che collega la probabilità alla statistica. Esso è più debole di altri analoghi (e che portano lo stesso nome) che si trovano nella letteratura per adeguarci alle tecniche che sono alla nostra portata.

Teorema (Glivenko-Cantelli, 1933)

La successione $\{F_n^(t)\}$ converge in probabilità alla funzione di ripartizione $F(t)$ della v.a. X .*

Denotiamo con Y_k la v.a. funzione indicatrice dell'intervallo $[X_k, +\infty[$. Essa assume il valore 1 se $X_k(\omega) \leq t$, il valore 0 altrimenti. Trattandosi di una v.a. bernoulliana, la sua speranza matematica è uguale a $P(X_k \leq t) = F(t)$.

Si ha inoltre $\text{Var } Y_k = (P(X_k \leq t))(1 - P(X_k \leq t))$ che, fissato t , è una costante il cui valore non supera $\frac{1}{4}$.

Sono verificate dunque le ipotesi della legge debole dei grandi numeri e possiamo concludere che

$$F_n^*(t) = \frac{1}{n} \sum_{k=1}^n Y_k \rightarrow F(t)$$

in probabilità.

Il risultato ci dice che facendo un numero sufficiente di prove, riusciamo ad approssimare la funzione di ripartizione della variabile aleatoria X che è oggetto delle nostre osservazioni.

Ma che cosa è questa probabilità?

Molte delle nostre considerazioni sono state fatte nell'ambito della probabilità classica. Essa è un'idealizzazione di un mondo che in effetti non esiste.

Il parallelo con la geometria euclidea calza perfettamente.

Uno dei teoremi di Euclide dice che la somma degli angoli interni di un triangolo è l'angolo pieno. Ma di che tipo di triangolo stiamo parlando? Se ne disegniamo uno sul pavimento del nostro soggiorno, neanche gli strumenti più sensibili riusciranno a smentire Euclide. Ma se i vertici li prendiamo in tre punti situati, rispettivamente, a Roma, Parigi e San Pietroburgo, le cose cambiano e la somma sarà maggiore dell'angolo pieno.

Il triangolo con un vertice al polo nord e gli altri due sull'equatore, distanti tra loro di 10.000 km (un quarto dell'equatore) ha ai vertici tre angoli retti.

Per tornare alla probabilità: noi siamo circondati da situazioni in cui parliamo di probabilità e che non solo non rientrano nella probabilità classica, ma che abbiamo qualche volta difficoltà a classificare. Ecco qualche esempio:

Qual è la probabilità che la Juventus vinca il prossimo campionato?

Qual è la probabilità che domani piova?

Qual è la probabilità che l'automobilista A si renda responsabile di un incidente stradale e che la sua assicurazione dovrà rifondere i danni? E qual è la probabilità che l'esborso per l'automobilista sarà, nell'arco di un anno, superiore a 5000 euro?

Qual è la probabilità che un cliente sessantenne che stipula oggi con un'assicurazione una polizza sulla vita sia ancora vivo e vegeto a settant'anni?

Qual è la probabilità che un dato medicinale abbia effetto benefico sul paziente Tizio? O che lo stesso medicinale causi uno degli effetti collaterali descritti dal "bugiardello"?

Il paziente Caio risulta positivo ad un test genetico. Qual è la probabilità che si tratti di un falso positivo?

Una valutazione assoluta è impossibile. Il grande probabilista De Finetti (1906-1985) diceva che "la probabilità non esiste".

Con questo voleva sottolineare che mentre le quantità fisiche (tempo, lunghezza, peso, temperatura, quantità di sostanza, intensità di corrente elettrica, intensità luminosa e le quantità fisiche derivate), possono essere calcolate con l'approssimazione della strumentazione disponibile, ci sono delle circostanze nelle quali la probabilità non può *in generale* essere calcolata e nelle quali non è credibile che ci sia qualcosa da calcolare.

Se viene scoperto un nuovo corpo celeste che potrebbe essere un pianeta simile alla terra, magari non siamo in grado di calcolarne (ora) la massa,

ma sappiamo che la sua massa esiste e si tratta di una quantità ben precisa, anche se in questo momento ignota. Essa potrà però essere stimata in futuro in maniera sempre più precisa con il progredire della tecnologia e con esplorazioni spaziali che porteranno navette cariche di strumenti sempre più vicine al corpo celeste che ci interessa.

L'approccio frequentista

Valutare la probabilità di un evento in base alla frequenza con cui esso si verifica è istintivo e lo facciamo quotidianamente senza, forse, fare troppi ragionamenti.

Se telefoniamo a tre impiegati e troviamo che uno risponde sempre, l'altro quasi mai e il terzo ha il telefono quasi sempre fuori posto, ci facciamo un'idea abbastanza precisa sul diverso ritmo di lavoro dei tre.

Se troviamo che, "grazie" ad una nuova rotonda, la solita strada per tornare a casa ci fa perdere un quarto d'ora, proviamo un paio di volte un'alternativa, per vedere se, in media, risulta più scorrevole.

Se lanciamo una puntina da disegno 100 volte e questa atterra 62 volte con la punta in su e 38 volte con la punta in giù, ci facciamo un'idea (ammesso che la cosa interessi a qualcuno) sulla probabilità dell'evento "punta in su".

Un matematico austriaco naturalizzato statunitense, Richard von Mises (1883-1953), cercò di definire la probabilità di un evento come limite, per n che tende all'infinito, delle sue frequenze relative. Ma la sua teoria si è rivelata logicamente incoerente.

Malgrado ciò il già citato matematico russo Andrey Kolmogorov scrisse "La base dell'applicabilità dei risultati della teoria matematica della probabilità a fenomeni aleatori reali deve dipendere in un modo o nell'altro dal concetto di probabilità come frequenza."

In ogni caso la teoria di von Mises, oltre alle sue contraddizioni interne, è impotente davanti al fatto che ci sono tanti esperimenti che non si possono ripetere, almeno non nelle stesse condizioni: se il cinquantenne Sempronio, che si è assicurato sulla vita, viene a mancare sei mesi dopo la stipula della polizza, è finita lì, il povero Sempronio vorrebbe anche fare una seconda prova, ma la natura non lo consente.

Approccio soggettivista

Spesso la valutazione della probabilità è frutto di una valutazione personale che dipende a sua volta dall'esperienza e dalle informazioni a disposizione di chi osserva il fenomeno.

Queste valutazioni possono differire da persona a persona, ma alle volte possono trovare il consenso di più persone o anche l'unanimità. Nella probabilità classica si registra solitamente un'identità di valutazioni.

Il grande probabilista italiano, Bruno De Finetti, propose di basare la probabilità sul concetto di gioco equo.

Supponiamo che due giocatori convengano sulla probabilità di un evento A (di quelli rognosi elencati sopra) e del suo opposto $B = A^c$.

Il primo giocatore è disposto a puntare la somma a sull'evento A , se il secondo giocatore è disposto a puntare la somma b sull'evento B . Questo vuol dire che entrambi reputano equo il gioco

$$\begin{pmatrix} b & a \\ p & 1-p \end{pmatrix}$$

cioè sono d'accordo che $bp - a(1-p) = 0$, ovvero

$$\frac{a}{p} = \frac{b}{1-p},$$

ovvero ancora

$$p = \frac{a}{a+b}.$$

In queste considerazioni stiamo idealizzando i due giocatori e presumiamo anche che sono disposti a scambiarsi le parti, come chi gioca a testa e croce e, non avendo particolari informazioni sulla moneta, accetta di puntare indifferentemente su testa o croce.

Le puntate dei giocatori sono proporzionali alle probabilità di A e B , rispettivamente.

Se in particolare $a + b = 1$, allora $p = a$, cioè la probabilità p che le due parti assegnano ad A è la somma che il primo giocatore (ma anche il secondo) sono disposti a puntare, *contro un monte premi unitario*, sull'evento A .