

Capitolo 2

Induzione statistica

2.1 Introduzione

Il processo di apprendimento e di creazione di nuova conoscenza è possibile grazie a procedimenti di carattere deduttivo e induttivo. L'acquisizione per via deduttiva permette, partendo da delle premesse generali la cui validità è garantita, di giungere a delle conclusioni particolari. Un esempio di ragionamento deduttivo è il seguente:

- Premessa maggiore: Tutti gli uomini sono mortali
- Premessa minore: Socrate è un uomo
- Conclusione: Socrate è mortale

Per contro, l'apprendimento per via induttiva consiste nel trovare delle leggi universalmente valide partendo da osservazioni particolari. Il processo di estensione dal particolare al generale è chiamato inferenza induttiva. Questo tipo di estensione della conoscenza è tuttavia soggetto ad incertezza: l'induzione non dimostra niente, essa è corretta solo nella totalità dei casi in cui è confermata la sua validità.

In statistica ci troviamo tipicamente nella situazione di dover inferire qualcosa circa la

1. *distribuzione* di una o più caratteristiche di una certa popolazione obiettivo, dove per *popolazione obiettivo* si intende la totalità degli elementi (individui, aziende, famiglie, cantoni, regioni, ...) per i quali desideriamo ottenere tali informazioni. In questo caso la popolazione obiettivo è finita ed indicheremo con la lettera N il numero dei suoi elementi.
2. legge o distribuzione di probabilità di un esperimento aleatorio.

Per quanto riguarda il caso di una popolazione finita, avete visto nel corso di statistica I come è possibile ottenere la distribuzione di un determinato carattere della popolazione in termini assoluti o relativi (percentuali) a partire dall'osservazione di ogni singolo elemento o unità della popolazione stessa. La conoscenza della distribuzione statistica è fondamentale in quanto essa descrive in maniera univoca e completa la caratteristica sotto esame. Purtroppo però solo in rare occasioni l'informazione disponibile è rappresentata dalla distribuzione statistica di *tutta* la popolazione obiettivo. Anzi è interessante osservare come ci siano situazioni dove la numerosità stessa della popolazione obiettivo non è conosciuta. Si pensi ad esempio ad un'analisi sul numero di sigarette che un fumatore consuma giornalmente. Quanto vale N , il numero di fumatori? Non solo la distribuzione del numero giornaliero di sigarette consumate da un fumatore è sconosciuta ma anche il numero stesso di fumatori che in questo esempio costituiscono la popolazione obiettivo.

Nel caso di un esperimento aleatorio formalizzato tramite uno spazio di probabilità (Ω, \mathcal{E}, P) l'interesse è posto sulla legge o distribuzione di probabilità P . Essa racchiude l'informazione disponibile sull'esperimento e le sue caratteristiche aleatorie. Vista la natura casuale dell'esperimento il risultato o esito $\omega \in \Omega$ dell'esperimento non è conosciuto a priori: quello che possiamo fare è assegnare delle probabilità agli eventi $E \in \mathcal{E}$. La conoscenza della legge di probabilità P permette di quantificare l'incertezza sul risultato dell'esperimento.

Per studiare la legge di probabilità di un dato esperimento aleatorio è necessario ripetere l'esperimento ed osservarne un certo numero di risultati. La teoria della stima si occupa di come utilizzare l'informazione disponibile per risalire alla legge di probabilità P o equivalentemente¹ alla funzione di ripartizione F . Se invece l'attenzione è rivolta allo studio di una certa caratteristica di una popolazione obiettivo, allora sarà necessario osservare un certo numero di individui/unità della popolazione ed in seguito stimare la distribuzione dell'intera popolazione obiettivo. È questo dunque il carattere induttivo della statistica inferenziale: da un numero *limitato* di osservazioni ottenere delle conclusioni di carattere *generale*.

2.2 Modelli parametrici

La stima di F risulta particolarmente difficile in quanto occorre stimare un'intera funzione: è necessario conoscere non solo il valore che F assume in un

¹Nell'arco di questo corso tratteremo unicamente spazi di probabilità ed esperimenti aleatori per i quali l'esistenza di una funzione di ripartizione è assicurata.

determinato punto, ma in tutti i punti dello spazio campione! È questa un'impresa estremamente difficile. Sappiamo che F ha una caratteristica particolare: è una funzione monotona crescente. Tuttavia esiste un'infinità di funzioni del genere. Per tale motivo lo statistico spesso restringe la sua ricerca ad un sottoinsieme ben preciso di funzioni di ripartizione. Ad esempio, si potrebbe pensare di limitarsi a quelle funzioni di ripartizione F che sono continue, o a quelle che posseggono una funzione di densità f (chiamate distribuzioni assolutamente continue, vedi Definizione 7)

$$F(x) = \int_{-\infty}^x f(t)dt \quad \text{con } f \text{ non negativa.} \quad (2.1)$$

Questo in parte può aiutare a semplificare la scelta (stima) di F ma non risolve tutti i problemi. La funzione di ripartizione di esperimento aleatorio il cui spazio campione è discreto, ad esempio, non è continua e tanto meno può essere rappresentata tramite un integrale del tipo (2.1). Inoltre, pur restringendo l'insieme dal quale stimare F , la scelta risulta ancora assai vasta. Una strada diversa e a volte complementare a quella appena esposta consiste nel limitare la ricerca di F ad un particolare insieme di distribuzioni parametriche. Come trattato nella Sezione 1.4, una famiglia parametrica di distribuzioni è un insieme di distribuzioni. Da ora in poi noteremo con la lettera \mathcal{P} una qualsiasi famiglia parametrica di distribuzioni. I parametri sconosciuti associati a tale famiglia saranno indicati² con $\theta_1, \theta_2 \dots \theta_K$. Utilizzeremo semplicemente la notazione θ quando desideriamo indicare l'insieme dei K parametri posti in una lista ordinata che chiameremo semplicemente vettore $\theta = (\theta_1, \theta_2, \dots, \theta_K)$. La funzione di ripartizione verrà quindi indicata con $F(\cdot; \theta)$ e se esiste la funzione di densità (caso continuo) o di probabilità (caso discreto) scriveremo analogamente $f(\cdot; \theta)$. In generale si assumerà che a due diversi parametri θ e $\tilde{\theta}$ corrispondano due diverse leggi di probabilità. Infine, l'insieme dei valori ammissibili per θ è indicato col simbolo Θ . Come osservato in precedenza, nel corso introduttivo di Statistica I avete incontrato diverse famiglie di distribuzioni parametriche sia discrete che continue (si vedano le dispense riguardanti alcune delle distribuzioni di probabilità parametriche più comuni). Utilizzeremo la seguente notazione per definire una famiglia parametrica di distribuzioni

$$\mathcal{P} = \{F(\cdot; \theta) : \theta \in \Theta\}$$

oppure, nel caso in cui la famiglia parametrica venga definita indirettamente tramite la funzione di probabilità (caso discreto) o la funzione di densità

²In generale e a dipendenza della distribuzione si utilizzano diversi altri simboli, quali ad esempio μ , σ o λ .

(caso continuo)

$$\mathcal{P} = \{f(\cdot; \theta) : \theta \in \Theta\}.$$

Esempio 11. La distribuzione continua uniforme è caratterizzata da due parametri che notiamo con a e b e che corrispondono rispettivamente all'estremo inferiore e superiore della funzione di densità. Formalmente questa famiglia può essere scritta come

$$\mathcal{P} = \left\{ f(x; \theta) = \frac{1}{b-a} I_{[a,b]}(x); \theta = (a, b) \in \mathbb{R}^2, \text{ con } a < b \right\}.$$

In questo esempio a qualsiasi coppia di numeri reali (a, b) tali per cui $a < b$ associamo una funzione di densità definita da

$$f(x; \theta) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{altrimenti} \end{cases}. \quad (2.2)$$

Nel caso della distribuzione uniforme $A = [a, b]$. È facile dimostrare applicando la definizione di funzione indicatrice che la funzione di densità (2.2) è riscrivibile più semplicemente come

$$f(x; \theta) = \frac{1}{b-a} I_{[a,b]}(x).$$

L'insieme dei possibili valori di θ , notato in precedenza con Θ , è dunque uguale al sottoinsieme dei punti del piano la cui prima coordinata è minore della seconda, formalmente

$$\Theta = \{(a, b) \in \mathbb{R}^2 \mid a < b\}.$$

Esercizio 1. Consideriamo la distribuzione parametrica di Poisson. Rispondete alle seguenti domande.

1. La distribuzione di Poisson è una distribuzione discreta o continua?
2. Qual è o quali sono i parametri che la caratterizzano. Che valori possono assumere i/il parametro/i di una distribuzione di Poisson? Definite il contenuto di θ nonché l'insieme Θ . Definite \mathcal{P} .
3. Assegnate dei valori al/ai parametro/i. In Excel o Openoffice eseguite i grafici della funzione di densità e della funzione di ripartizione. Calcolate la probabilità degli eventi $E_1 = [0, 1]$ ed $E_2 = (1, 2)$.

4. Scegliete nuovi valori per il/i parametri e col medesimo programma confrontate i grafici delle due funzioni di densità e delle due funzioni di ripartizione. Sia S il sottoinsieme di \mathbb{R} definito da quei punti per cui la funzione di densità è maggiore di 0, cioè

$$S = \{x \in \mathbb{R} \mid f(x) > 0\}.$$

Rispetto al valore assegnato al punto precedente cambia l'insieme S con i nuovi valori? Siete in grado di generalizzare la vostra conclusione?

5. Cosa implica un cambiamento dei valori assegnati al/ai parametro/i rispetto alle probabilità degli eventi E_1 ed E_2 definiti in precedenza? Notate delle differenze? Spiegate?

Esercizio 2. Consideriamo la distribuzione parametrica di Pareto. Rispondete alle seguenti domande.

1. La distribuzione di Pareto è una distribuzione discreta o continua?
2. Qual è o quali sono i parametri che la caratterizzano. Che valori posso/no assumere i/il parametro/i di una distribuzione di Pareto? Definite il contenuto di θ nonché l'insieme Θ . Definite \mathcal{P} .
3. Assegnate dei valori al/ai parametro/i ed eseguite i grafici della funzione di densità e della funzione di ripartizione. Calcolate la probabilità degli eventi $E_1 = [0, 3]$ ed $E_2 = \{-1, 0, 2, 3\}$.
4. Trovate l'insieme S per i valori dei parametri da voi scelti.
5. È possibile trovare dei nuovi valori dei parametri in maniera tale da modificare l'insieme S ? Trovate un esempio del genere e confrontate i grafici delle due funzioni di densità e delle due funzioni di ripartizione.

2.3 Esperimento statistico

È stato detto che la teoria della stima si occupa di giungere a delle conclusioni generali riguardo

- alla legge di probabilità di un esperimento aleatorio, oppure
- alla distribuzione di una variabile aleatoria, oppure
- alla distribuzione statistica di una popolazione obiettivo

partendo da un numero limitato di osservazioni di un certo esperimento aleatorio o di elementi/unità della popolazione. Da un punto di vista formale, ripetere n volte in maniera indipendente un esperimento aleatorio (Ω, \mathcal{E}, P) costituisce un nuovo tipo di esperimento a cui è dato il nome di *esperimento statistico*.

Definizione 9. (Esperimento statistico). Un esperimento statistico di ampiezza n è costituito da un numero n di ripetizioni indipendenti dello stesso esperimento aleatorio (Ω, \mathcal{E}, P) .

Quali saranno i possibili esiti di un esperimento statistico? Sappiamo che ad ogni ripetizione dell'esperimento aleatorio lo spazio campione è sempre Ω . Potremmo pensare di raccogliere gli esiti delle n ripetizioni e creare una *lista ordinata* di valori. Il primo valore della lista corrisponde all'esito della prima esecuzione dell'esperimento aleatorio, il secondo valore corrisponde all'esito della seconda esecuzione e via dicendo. La nostra lista conterrà gli n esiti che indicheremo con (x_1, x_2, \dots, x_n) . Attenzione a non confondere la lista così costruita con lo spazio campione Ω . Infatti, per ciascun $i = 1, \dots, n$ l' i -esimo elemento x_i della lista è un elemento di Ω ma $(x_1, x_2, \dots, x_n) \neq \Omega$. Lo spazio campione può contenere un numero infinito non numerabile di esiti. Pensate ad esempio all'esperimento aleatorio di registrare il tempo d'attesa ad uno sportello postale. In tal caso $\Omega = [0, \infty)$ mentre (x_1, x_2, \dots, x_n) sarà sempre un insieme finito.

Esempio 12. L'esperimento statistico corrisponde a lanciare 2 volte un dado. L'esperimento aleatorio è il lancio di un dado. $\Omega = \{1, \dots, 6\}$ mentre P è la distribuzione uniforme discreta (dado non truccato). Gli esiti dell'esperimento statistico possono essere rappresentati graficamente come segue:

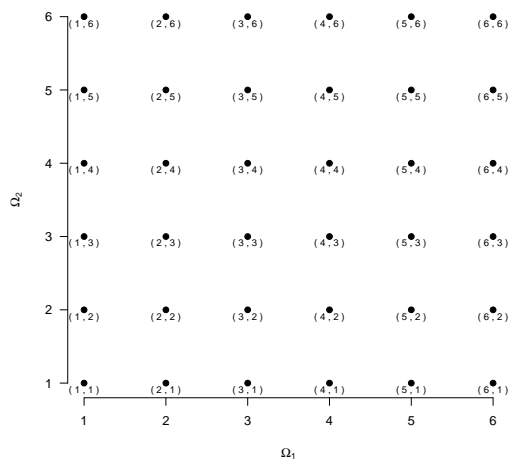


Figura 2.1: Esiti dell'esperimento statistico: lancio di due dadi.

Indicando con Ω_1 lo spazio campionario della prima esecuzione dell'esperimento aleatorio e con Ω_2 lo spazio campionario della seconda esecuzione, l'esperimento statistico avrà quale spazio campionario l'insieme delle coppie di esiti (x_1, x_2) con $x_1 \in \Omega_1$ e $x_2 \in \Omega_2$. Un esito di questo particolare esperimento statistico potrebbe essere $(3, 4)$. Poiché l'ordine conta, l'esito $(3, 4)$ è da considerarsi diverso dall'esito $(4, 3)$.

Definizione 10. Chiamiamo *spazio campionario prodotto* l'insieme di tutte le coppie (x_1, x_2) dove $x_1 \in \Omega_1$ e $x_2 \in \Omega_2$. Esso verrà notato con $\Omega \times \Omega$ o Ω^2 . Nel caso di un numero n qualsiasi di ripetizioni scriveremo semplicemente Ω^n e un esito dell'esperimento statistico sarà rappresentato tramite la lista (x_1, x_2, \dots, x_n) .

Avendo definito lo spazio campionario di un esperimento statistico dovremo ora definire gli eventi ad esso associati. In maniera del tutto identica alla definizione di evento di un esperimento aleatorio, l'evento di un esperimento statistico è un sottoinsieme di Ω^n . Indicheremo con $\bigotimes_{i=1}^n \mathcal{E}_i$ l'insieme degli eventi definiti su Ω^n .

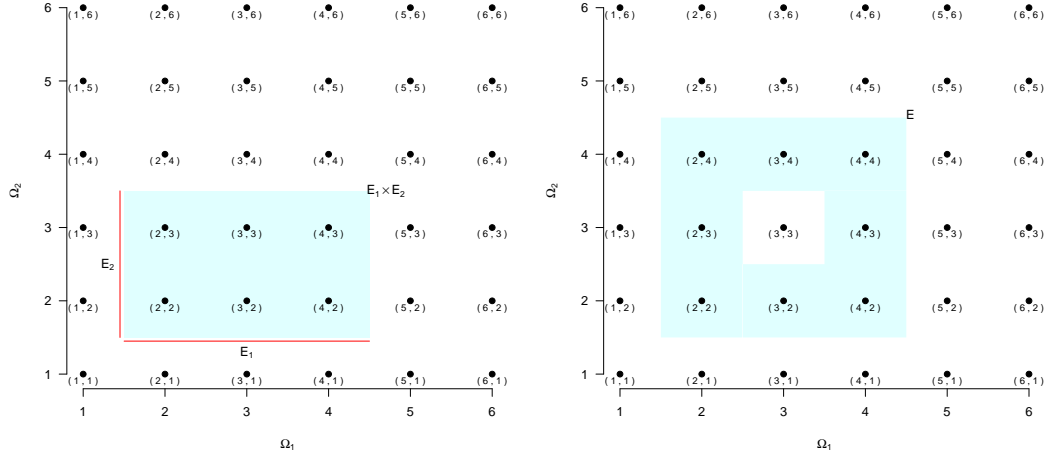


Figura 2.2: Esempi di eventi dello spazio campionario prodotto

Le due immagini della Figura 2.2 rappresentano graficamente due possibili eventi dell'esperimento statistico discusso nell'Esempio 12. Come potete notare nella prima immagine, l'evento definito dall'area colorata corrisponde al prodotto cartesiano $E_1 \times E_2$ dei due eventi $E_1 = \{2, 3, 4\}$ ed $E_2 = \{2, 3\}$ definiti rispettivamente sul primo e sul secondo lancio. Lo spazio $\bigotimes_{i=1}^n \mathcal{E}_i$ degli eventi definiti su Ω^n è molto vasto. La seconda immagine ci mostra infatti che esso comprende eventi che non sono semplicemente riconducibili ad un prodotto cartesiano $E_1 \times E_2$ di eventi rispettivamente di Ω_1 e Ω_2 . Ancora una volta evitiamo di addentrarci in dettagli tecnici. Considereremo lo spazio degli eventi $\bigotimes_{i=1}^n \mathcal{E}_i$ come un "particolare" insieme i cui *elementi* sono *sottoinsiemi* di Ω^n .

Utilizzando la proprietà di indipendenza tra le varie ripetizioni dell'esperimento aleatorio varrà che la legge di probabilità di un esperimento statistico è semplicemente il prodotto delle singole leggi di probabilità P . Infatti, indicando con Q la legge di probabilità di un esperimento statistico (definita quindi su $\bigotimes_{i=1}^n \mathcal{E}_i$), avremo per qualsiasi evento

$$E_1 \times E_2 \times \dots \times E_n$$

appartenente a $\bigotimes_{i=1}^n \mathcal{E}_i$ che la sua probabilità è uguale

$$Q(E_1 \times E_2 \times \dots \times E_n) = P(E_1) P(E_2) \dots P(E_n) .$$

In maniera del tutto simile ad un esperimento aleatorio, un esperimento statistico può quindi essere rappresentato dalla tripla

$$\left(\Omega^n, \bigotimes_{i=1}^n \mathcal{E}_i, Q = \bigotimes_{i=1}^n P_i \right) . \quad (2.3)$$

Un esperimento statistico possiede la medesima struttura di un esperimento aleatorio e soddisfa quindi le caratteristiche di uno spazio di probabilità. Vediamo ora un esempio di esperimento statistico.

Esempio 13. Consideriamo l'esperimento aleatorio che riguarda la misurazione della durata di vita di una lampadina: $\Omega = \mathbb{R}$ mentre quale distribuzione parametrica per la legge di probabilità P utilizziamo la distribuzione esponenziale

$$\mathcal{P} = \{f(x; \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x) : \lambda > 0\} .$$

Definiamo ora l'esperimento statistico costituito dalla misurazione della durata di vita di $n = 3$ lampadine prodotte in maniera indipendente l'una dall'altra. Il nuovo esperimento avrà quale spazio campionario \mathbb{R}^3 . Un evento di questo spazio potrebbe essere il seguente sottoinsieme di \mathbb{R}^3

$$E_1 \times E_2 \times E_3$$

dove

- $E_1 = \{\text{"La prima lampadina si spegne dopo 3 ore"}\} = [3, \infty)$,
- $E_2 = \{\text{"La seconda lampadina dura meno di 20 minuti"}\} = [0, \frac{1}{3})$,
- $E_3 = \{\text{"La terza lampadina prima o poi si spegne"}\} = [0, \infty)$.

Grazie all'ipotesi di indipendenza nell'esecuzione dei tre esperimenti la legge di probabilità su $\bigotimes_{i=1}^3 \mathcal{E}_i$ non è altro che il prodotto di P sui rispettivi eventi

$$Q(E_1 \times E_2 \times E_3) = P(E_1)P(E_2)P(E_3).$$

Le singole probabilità possono essere calcolate individualmente tramite l'integrale della funzione di densità, ad esempio

$$P(E_1) = \int_3^{\infty} \lambda e^{-\lambda x} dx$$

Il modello parametrico della legge di probabilità Q sarà semplicemente (indicheremo con $f(x_1, x_2, x_3)$ la funzione di densità di Q)

$$\begin{aligned} \mathcal{P} &= \{f(x_1, x_2, x_3; \lambda) = f(x_1; \lambda) f(x_2; \lambda) f(x_3; \lambda) : \lambda > 0\} \\ &= \{f(x_1, x_2, x_3; \lambda) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)} I_{(0,\infty)}(x_1) I_{(0,\infty)}(x_2) I_{(0,\infty)}(x_3) : \lambda > 0\}. \end{aligned}$$

Come potete constatare, la difficoltà è più di carattere formale (forma e notazione) che concettuale.

2.4 Variabili aleatorie

Talvolta succede che ciò a cui siamo veramente interessati non sia l'esito di un esperimento aleatorio ma una sua particolare funzione. Pensate ad esempio ad una compagnia aerea. Supponiamo che il ritardo di un volo di linea sia una variabile aleatoria. Alla compagnia interessa soprattutto conoscere i costi che un certo ritardo può generare piuttosto che l'ammontare in minuti del ritardo. I costi saranno quindi una certa funzione dei minuti di ritardo. Generalizzando al caso astratto, introduciamo la seguente definizione.

Definizione 11. Dato un esperimento aleatorio (Ω, \mathcal{E}, P) diremo che la funzione $g : \Omega \rightarrow \mathbb{R}$ è misurabile se, preso un qualsiasi evento $E' \subset \mathbb{R}$, l'insieme $g^{-1}(E')$ è un elemento di \mathcal{E} .

In altre parole, diremo che la funzione g è misurabile se la preimmagine di un evento di \mathbb{R} è a sua volta un evento di Ω . Le figure 2.3 e 2.5 mostrano astrattamente la situazione³.

Ricordiamo che per definizione

$$g^{-1}(E') = \{\omega \in \Omega \mid g(\omega) \in E'\}.$$

³Per convenienza assumeremo che i valori che la funzione g assume siano numeri reali. Tuttavia ci preme sottolineare che la definizione generale di variabile aleatoria non impone questa restrizione sul codominio di g . Un esempio in cui il codominio di g non è numerico è descritto nella Sezione 3.1.

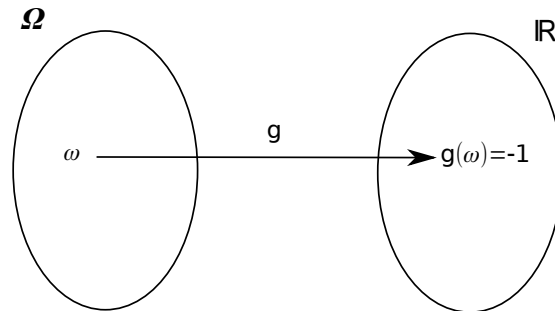


Figura 2.3: Variabile aleatoria - diagramma di Venn

La preimmagine di E' rispetto alla funzione g è un sottoinsieme di Ω e contiene tutti gli esiti la cui immagine attraverso g appartiene all'evento E' di \mathbb{R} . La nostra definizione di evento (confronta la Definizione 3) considera evento un qualsiasi sottoinsieme di Ω . Ne segue che per noi qualsiasi funzione da Ω verso \mathbb{R} è misurabile⁴.

Esempio 14. Il numero di ore che un traghetto impiega per andare da Genova ad Olbia è una variabile aleatoria in quanto il tempo di percorrenza è influenzato da diversi fattori in parte non prevedibili quali ad esempio le condizioni del mare, l'intensità e direzione del vento, il numero di passeggeri e così via. Indicando con x il tempo impiegato dal traghetto per la traversata, la funzione di costo è data da

$$g(x) = \begin{cases} 10 + 10x, & \text{se } 0 < x \leq 9 \\ 100 + 20(x - 9), & \text{se } 9 < x \leq 10 \\ 120 + 40(x - 10), & \text{se } x > 10 \end{cases}.$$

Con quale probabilità si osserveranno dei costi compresi nell'intervallo $E' = [110 - 130]$? Per rispondere a questa domanda occorre calcolare la preimmagine dell'evento E' utilizzando la funzione dei costi $g(x)$: $E = g^{-1}(E')$. È possibile aiutarsi col grafico della funzione, vedi Figura 2.4.

⁴Come accennato precedentemente le definizioni di evento e dell'insieme \mathcal{E} andrebbero approfondite. In generale non tutte le funzioni da uno spazio di probabilità Ω verso \mathbb{R} sono misurabili.

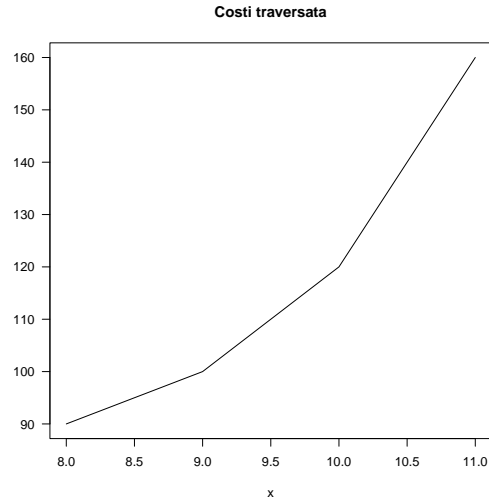


Figura 2.4: Funzione dei costi

Dopo che l'evento E è stato identificato è possibile calcolare la probabilità che si realizzi utilizzando la legge di probabilità P . Tale probabilità corrisponde dunque alla probabilità di osservare dei costi come specificato da E' .

Definizione 12. Sia (Ω, \mathcal{E}, P) uno spazio di probabilità e $g : \Omega \rightarrow \mathbb{R}$ misurabile. Allora g è chiamata variabile aleatoria.

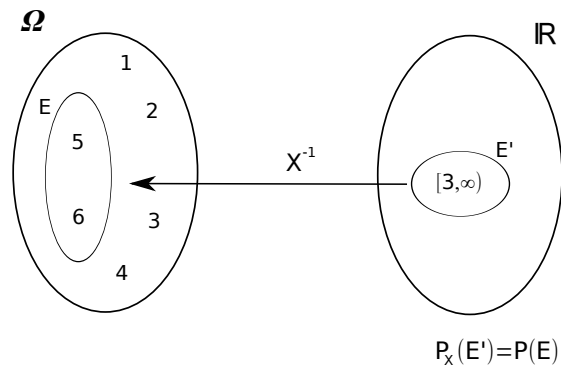


Figura 2.5: Preimmagine

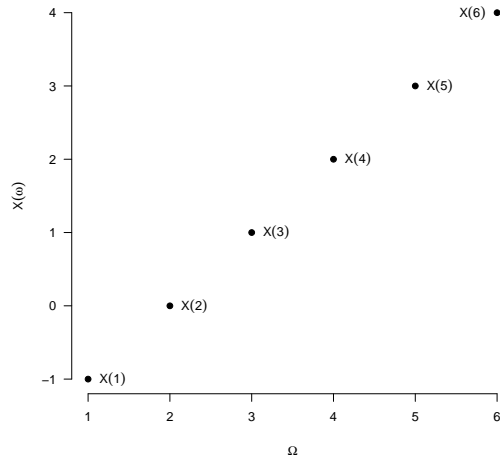


Figura 2.6: Variabile aleatoria

Osservazione 3. In statistica si è soliti indicare una variabile aleatoria con una lettera maiuscola, tipicamente X . Ricordatevi dunque che malgrado il nome ingannevole, la variabile aleatoria X è una funzione! Anziché scrivere ogni volta “la variabile aleatoria X ” nell’ambito di questo corso utilizzeremo l’abbreviazione V.A. X .

Lasciamo ai matematici lo studio delle condizioni rispetto alle quali una funzione X è una variabile aleatoria. Per noi questa condizione sarà sempre soddisfatta. Ci domandiamo piuttosto perché una variabile aleatoria debba soddisfare una simile condizione di misurabilità. Il motivo è dato dalla necessità di assegnare una probabilità agli eventi E' di \mathbb{R} .

Esempio 15. Consideriamo l’esperimento aleatorio di lanciare una volta un dado e di fare una scommessa sul possibile esito ω dell’esperimento. Supponiamo dunque che la vincita o la perdita in franchi, notata X , dipenda dall’esito dell’esperimento e sia descritta dalla funzione $X(\omega) = \omega - 2$. Siamo interessati all’evento $E' = [3, \infty)$ di guadagnare almeno 3 franchi. Il nostro esperimento aleatorio è formalizzato da $\Omega = \{1, 2, 3, 4, 5, 6\}$ e P distribuzione uniforme discreta. La variabile aleatoria X può essere rappresentata tramite un diagramma di Venn simile alla Figura 2.3 oppure graficamente come in Figura 2.6.

Consideriamo ora la probabilità dell'evento E' . Come si calcola tale probabilità? L'idea è semplicissima. Guardando il diagramma di Venn 2.5 è facile convincersi che l'evento E' si avvera se e solo se l'evento $E = X^{-1}(E')$ si realizza. Nel nostro esempio vinceremo almeno 3 franchi se il lancio del dado risulterà essere un numero superiore o alla peggio uguale a 5. L'evento E nel nostro caso è dato dal sottoinsieme $\{5, 6\}$ di Ω . La probabilità di vincere almeno 3 franchi è dunque uguale alla probabilità che esca un cinque o un sei. Scriveremo

$$P_X(E') := P(X^{-1}(E')) = P(E) = \frac{1}{3}. \quad (2.4)$$

Osservate attentamente l'uguaglianza 2.4. Quando ci si riferisce alla probabilità dell'evento E' occorre farlo aggiungendo la lettera X alla probabilità P in quanto E' non è evento di Ω ma di \mathbb{R} , l'insieme di arrivo della V.A. X . È possibile dimostrare che P_X , così come definita tramite l'uguaglianza 2.4, possiede tutte le proprietà di una legge di probabilità su \mathbb{R} . La tripla $(\mathbb{R}, \mathcal{E}, P_X)$ costituisce un nuovo spazio di probabilità e P_X è chiamata la legge di probabilità indotta o distribuzione indotta dalla variabile aleatoria X su \mathbb{R} . Per semplicità si è soliti chiamare P_X legge di probabilità di X .

Esempio 16. Lanciamo 3 volte una moneta. In un singolo lancio la probabilità di ottenere croce è p con $0 \leq p \leq 1$. Per un singolo lancio utilizzeremo la convenzione di indicare con t l'esito testa e con c l'esito croce. Modelliamo questo esperimento con lo spazio campionario Ω^3 contenente gli 8 possibili esiti

$$\begin{aligned} \omega_1 &= (t, t, t) & \omega_2 &= (c, t, t) & \omega_3 &= (t, c, t) & \omega_4 &= (t, t, c) \\ \omega_5 &= (t, c, c) & \omega_6 &= (c, t, c) & \omega_7 &= (c, c, t) & \omega_8 &= (c, c, c) \end{aligned} \quad .$$

È immediato verificare che sotto l'ipotesi di indipendenza tra un lancio e l'altro la probabilità di ogni singolo esito è data da

$$Q(\{\omega_i\}) = p^{\#c}(1-p)^{\#t}$$

dove “ $\#c$ ” indica il numero di “ c ” in ω_i . La tripla $(\Omega^3, \bigotimes_{i=1}^n \mathcal{E}_i, Q)$ è un esperimento statistico e forma uno spazio di probabilità.

Su questo spazio definiamo ora la V.A. reale $X : \Omega^3 \rightarrow \mathbb{R}$, $\omega \mapsto$ “il numero di croci ottenute” che in pratica consiste nel contare il numero di “ c ” presenti in ω . Vale dunque

$$\begin{aligned} X(\omega_1) &= 0 & , & & X(\omega_2) &= 1 & , & & X(\omega_3) &= 1 & , & & X(\omega_4) &= 1 \\ X(\omega_5) &= 2 & , & & X(\omega_6) &= 2 & , & & X(\omega_7) &= 2 & , & & X(\omega_8) &= 3 \end{aligned} \quad .$$

La V.A. X assumerà così i valori $0, 1, 2, 3$. Con che probabilità la variabile aleatoria X sarà uguale a 0 ? Guardando la tabella dei possibili valori deduciamo che $X = 0$ se e solo se $\omega = \omega_1$, quindi

$$P_X(\{0\}) = P(X = 0) = P(\{\omega_1\}).$$

In maniera analoga, dato l'evento $E' = [1, 2] \subset \mathbb{R}$ diremo che E' si è realizzato se $X(\omega) \in E'$, dove ω denota la realizzazione su Ω^3 . La probabilità che E' si realizzi è uguale a

$$P_X(E') = P(X^{-1}(E')) = P(\{\omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}).$$

La probabilità dell'evento $E' = \{0, 3\}$ è uguale a $P(\{\omega_1, \omega_8\})$ e la probabilità di $P_X([0, 2])$ è semplicemente uguale a $P(\{\omega_1, \omega_2, \omega_3, \omega_4\})$.

La variabile aleatoria X induce una legge di probabilità su $(\mathbb{R}, \mathcal{E})$ che abbiamo chiamato P_X .

La tripla $(\mathbb{R}, \mathcal{E}, P_X)$ è uno spazio di probabilità. In particolare, P_X è la distribuzione Binomiale $Bin(3, p)$ la cui famiglia parametrica è data da

$$\mathcal{P} = \left\{ f(x; p) = \binom{3}{x} p^x (1-p)^{3-x} : x \in \{0, 1, 2, 3\} \text{ e } p \in [0, 1] \right\}.$$

2.5 Esperimento statistico e variabili aleatorie

La medesima situazione studiata nella Sezione 2.3 la ritroviamo nell'ambito delle variabili aleatorie. Prendiamo una variabile aleatoria X a valori in \mathbb{R} definita su uno spazio di probabilità (Ω, \mathcal{E}, P) . Come appena studiato, la V.A. X genera un nuovo esperimento aleatorio su \mathbb{R} , notato $(\mathbb{R}, \mathcal{E}, P_X)$, dove P_X è la legge di probabilità indotta. Consideriamo ora un campione di n variabili aleatorie X_1, X_2, \dots, X_n mutualmente indipendenti e distribuite identicamente a X . Questo n -campione estratto da X induce uno spazio prodotto notato

$$\left(\mathbb{R}^n, \bigotimes_{i=1}^n \mathcal{E}_i, Q_{X_1, \dots, X_n} = \bigotimes_{i=1}^n P_{X,i} \right). \quad (2.5)$$

Se la legge di probabilità P_X appartiene ad una famiglia parametrica di distribuzioni con funzione di densità $f(\cdot; \theta)$, varrà per lo stesso discorso fatto in precedenza che la distribuzione Q_{X_1, \dots, X_n} apparterrà alla famiglia

$$\mathcal{P} = \left\{ f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) : \theta \in \Theta \right\}.$$

Come in precedenza la funzione di densità (o probabilità nel caso discreto) di Q_{X_1, \dots, X_n} è il prodotto delle n funzioni di densità $f(x_i; \theta)$. Poiché le V.A. X_1, X_2, \dots, X_n sono identicamente distribuite, la forma delle funzioni di densità $f(x_i; \theta)$ è la stessa per ogni variabile i .

Esempio 17. (Continuazione Esempio 16) Disponiamo di osservazioni indipendenti di n giocate, notate rispettivamente X_1, X_2, \dots, X_n . La loro distribuzione congiunta avrà quale funzione di probabilità (ci troviamo di fronte a variabili aleatorie discrete)

$$\mathcal{P} = \left\{ f(x_1, \dots, x_n; p) = \prod_{i=1}^n \binom{3}{x_i} p^{x_i} (1-p)^{3-x_i} : x_i \in \{0, 1, 2, 3\} \text{ e } p \in [0, 1] \right\}.$$

Utilizzando le proprietà delle potenze la funzione f può essere riscritta come

$$f(x_1, \dots, x_n; p) = p^{\sum_{j=1}^n x_j} (1-p)^{3n - \sum_{j=1}^n x_j} \prod_{i=1}^n \binom{3}{x_i}.$$

Definizione 13. Dato un esperimento statistico $(\mathbb{R}, \mathcal{E}, P_X)$ diremo che \mathcal{P} è correttamente specificata se esiste almeno un valore $\theta_0 \in \Theta$ tale per cui $f(\cdot; \theta_0)$ è la vera funzione di densità/probabilità di X .

2.6 Campione

Definizione 14. Chiamiamo *campione* un particolare sottoinsieme di numerosità n della popolazione obiettivo utilizzato ai fini dell'inferenza. Nel caso di un esperimento statistico con *campione* intendiamo l'osservazione di un n -campione (inteso come la realizzazione di n variabili aleatorie⁵ *i.i.d.*).

Ci sono alcune osservazioni su cui chinarsi.

Osservazione 4. La prima osservazione riguarda il caso di un esperimento statistico in cui si desidera stimare la distribuzione di una variabile aleatoria X tramite un n -campione. È fondamentale ai fini della comprensione di quanto seguirà rendersi conto che l' n -campione X_1, \dots, X_n è una collezione di variabili aleatorie che generano su \mathbb{R}^n uno spazio di probabilità (confronta Sezione 2.5). Così come una variabile aleatoria X induce una legge di probabilità P_X su $(\mathbb{R}, \mathcal{E})$, anche X_1, X_2, \dots, X_n inducono una legge di probabilità su $(\mathbb{R}^n, \bigotimes_{i=1}^n \mathcal{E}_i)$. La probabilità indotta è stata indicata con Q_{X_1, \dots, X_n} (confronta Sezione 2.3).

Osservazione 5. Osservato l'esperimento, il ricercatore avrà a disposizione le realizzazioni di X_1, \dots, X_n che sono dei numeri. Le realizzazioni saranno disposte in modo ordinato in una lista notata (x_1, \dots, x_n) che chiameremo vettore. La notazione maiuscola denota la variabile aleatoria (funzione!) mentre la notazione minuscola ne indica il particolare valore realizzato. Ad esempio $x_1 = X_1(\omega)$.

Osservazione 6. Consideriamo la stima della distribuzione statistica di una caratteristica della popolazione obiettivo. In tal caso il campione è definito come un sottoinsieme della popolazione obiettivo. Questa definizione non considera il modo come si è ottenuto questo sottoinsieme.

- Che regola di selezione è stata adottata?
- Come sono stati selezionati gli n individui o le n unità dalla popolazione?

Vedremo che applicando dei criteri aleatori è possibile ricondurre il problema del campionamento da una popolazione campione a quello relativo alla costruzione di un n -campione in un esperimento statistico. Affronteremo però

⁵D'ora innanzi utilizzeremo semplicemente l'abbreviazione *i.i.d.* per indicare variabili aleatorie *indipendenti* ed *identicamente distribuite*.

questo tema nel corso del prossimo capitolo dove analizzeremo le varie tecniche di campionamento. Per il momento basterà ricordare che anche nel caso di una popolazione obiettivo la cui distribuzione statistica è sconosciuta il *campione* osservato sarà il risultato di un procedimento aleatorio e quindi è da considerarsi anch'esso come la realizzazione di n variabili aleatorie.

Ma perché, nell'analisi di una popolazione obiettivo, effettuare un'indagine campionaria e non un'indagine esaustiva? I motivi sono diversi; fra i principali annoveriamo:

- una rilevazione completa è impossibile (costi troppo elevati, impossibilità di raggiungere tutti gli individui);
- determinare le caratteristiche delle unità campionate ne determina la distruzione (ad esempio, misurare la durata di vita di una lampadina);
- La popolazione non è definibile in termini della sua numerosità e quindi non ha senso parlare di indagine esaustiva (ad esempio una data tecnologia permette di produrre lampadine la cui durata di vita è aleatoria. Quanto vale N in questo caso?).
- è necessaria una particolare accuratezza nella rilevazione e quindi a causa delle risorse limitate solo un numero ristretto di unità può essere osservato;
- tempi d'esecuzione elevati se confrontati alla necessità d'informazione (sondaggi elettorali).

2.7 Distribuzione campionaria, statistica, stimatore corretto

Il concetto di distribuzione di una V.A. (caso univariato) o di un insieme di V.A. (caso multivariato) ci accompagnerà per tutta la durata di questo corso. Abbiamo visto nei paragrafi precedenti il concetto di esperimento statistico e di n -campione ad esso associato. Affrontiamo ora la definizione di distribuzione campionaria che altro non è che la distribuzione congiunta Q_{X_1, \dots, X_n} delle n variabili aleatorie X_1, \dots, X_n vista in precedenza.

Definizione 15. Distribuzione campionaria. Sia X_1, X_2, \dots, X_n un campione aleatorio di numerosità n . Si dice distribuzione campionaria di X_1, X_2, \dots, X_n la distribuzione o legge di probabilità congiunta di X_1, X_2, \dots, X_n .

Perché la distribuzione campionaria è così importante? La sua importanza risiede nell'uso che se ne farà ai fini della stima del o dei parametri sconosciuti. Infatti,

- definito il modello di probabilità e l'esperimento statistico ad esso associato
- ricavata la distribuzione campionaria dei dati in nostro possesso

si passa alla fase di stima. Lo scopo della teoria della stima è quello di costruire delle funzioni $T(X_1, \dots, X_n)$ delle variabili aleatorie osservate con cui produrre delle stime puntuali $T(x_1, \dots, x_n)$ del vettore di parametri⁶ sconosciuti θ . Ancora una volta è fondamentale fare una netta distinzione fra il campione X_1, \dots, X_n quale insieme ordinato di V.A. ed il campione osservato x_1, \dots, x_n . Come precedentemente osservato, la sequenza x_1, \dots, x_n contiene dei *numeri* e raccoglie le rispettive realizzazioni delle V.A. X_1, \dots, X_n , cioè

$$x_1 = X_1(\omega), \quad x_2 = X_2(\omega), \quad \dots, \quad x_n = X_n(\omega).$$

Definizione 16. Statistica. Una statistica T è una funzione di variabili aleatorie osservabili (e quindi a sua volta variabile aleatoria osservabile) che non dipende da alcun parametro incognito.

Esempio 18. Alcuni esempi di statistiche sono

1. La media $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.
2. $X_{\max} = \max \{X_1, X_2, \dots, X_n\}$.
3. Lo stimatore corretto della varianza $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.
4. Siano $X_1, \dots, X_n \sim i.i.d. N(3, \sigma^2)$, σ^2 sconosciuto. La variabile aleatoria

$$Z = \frac{\bar{X} - 3}{\sigma}$$

non è una statistica in quanto la varianza σ^2 (e quindi anche σ) non è nota.

⁶Come già accennato in precedenza, il nome assegnato ai vari parametri sconosciuti cambia a seconda delle circostanze. Ad esempio, il valore atteso di una V.A. è generalmente indicato con il simbolo μ . Se il parametro sconosciuto da stimare fosse la varianza si utilizzerebbe σ^2 .

Osservazione 7. Ci preme sottolineare che una statistica è una variabile aleatoria. La distribuzione di \bar{X} ad esempio, dipende dalla distribuzione congiunta di X_1, \dots, X_n . Lo stesso vale per X_{\max} , S^2 e qualsiasi altra funzione misurabile di X_1, \dots, X_n .

Definizione 17. Stimatore corretto. Sia θ un parametro sconosciuto. Diremo che $T(X_1, \dots, X_n)$ è uno stimatore corretto di θ se

$$E(T(X_1, \dots, X_n)) = \theta.$$

Esempio 19. Sia X_1, \dots, X_n un campione di variabili aleatorie *i.i.d.* (μ, σ^2) . La media campionaria è uno stimatore corretto di μ in quanto

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu.$$

Anche

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

è uno stimatore corretto di σ^2 . Per contro, ammettiamo che μ sia conosciuto,

$$\tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2$$

non è uno stimatore corretto di σ^2 .

Esercizio 3. Si dimostri che lo stimatore \tilde{S}^2 non è uno stimatore corretto di σ^2

Terminiamo il capitolo con due ulteriori esempi di modelli parametrici.

Esempio 20. Rendimenti azionari. Sia P_t il prezzo di una certa azione il giorno⁷ t . Il rendimento logaritmico giornaliero è definito come

$$R_t = \ln \frac{P_t}{P_{t-1}}.$$

Assumiamo che

$$\mathcal{P} = \left\{ f(R; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (R - \mu)^2\right) : \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ \right\}.$$

⁷In questo caso viene naturale utilizzare l'indice t anziché l'indice i per indicare la successione di osservazioni di R .

Supponiamo di avere un campione comprendente gli ultimi 100 rendimenti logaritmici giornalieri $R_t, R_{t-1}, \dots, R_{t-99}$ e che i rendimenti a date diverse siano fra loro indipendenti. Vogliamo calcolare

$$P(R_{t+1} \leq -5\%).$$

Esempio 21. Funzione di produzione Cobb-Douglas. Definiamo la seguente relazione

$$Y := \ln Q_t = \beta_1 + \beta_2 \ln L_t + \beta_3 \ln K_t + U_t \quad (2.6)$$

dove

1. Q_t rappresenta la quantità prodotta l'anno t di un certo bene.
2. L_t rappresenta le ore di lavoro utilizzate per la produzione di Q_t .
3. K_t rappresenta la quantità di capitale utilizzato.
4. U_t è una variabile aleatoria, non osservabile che racchiude tutti quei fattori non sistematici che non sono stati inclusi nel modello. Assumiamo: $U_t \sim i.i.d. N(0, \sigma^2)$.

La famiglia parametrica \mathcal{P} è uguale a

$$\mathcal{P} = \left\{ \begin{array}{l} f(y; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\ln q_t - \beta_1 - \beta_2 \ln l_t - \beta_3 \ln k_t)^2\right) : \\ \theta = (\beta_1, \beta_2, \beta_3, \sigma^2) \in \mathbb{R}^3 \times \mathbb{R}^+ \end{array} \right\}.$$

In quest'ultimo esempio la funzione di produzione Cobb-Douglas

$$Q = \exp(\beta_1) L^{\beta_2} K^{\beta_3} \quad (2.7)$$

ci è stata suggerita dalla teoria economica. Ne segue che i risultati empirici ottenuti sono interpretabili nel contesto dei modelli (micro-) economici ben conosciuti. Ad esempio è possibile interpretare i coefficienti β_2 e β_3 come le elasticità della produzione:

$$\frac{\partial Q}{\partial L} / \frac{Q}{L} = \dots$$

L'equazione (2.7) è deterministica. In essa la produzione Q è una funzione deterministica delle quantità di lavoro L e capitale K . Nel modello (2.6) invece, Q è una variabile aleatoria in quanto funzione del termine d'errore⁸ U .

⁸È bene non confondere il concetto di osservabilità di una variabile con il concetto di aleatorietà. Una variabile può essere deterministica ma non osservabile, deterministica osservabile, aleatoria non osservabile o aleatoria ed osservabile. Nel modello (2.6) gli U_t sono aleatori e non osservabili.

È bene fare presente quanto segue. Affinché il modello sia utilizzabile sono state fatte numerose ipotesi semplificatrici:

- (a) forma funzionale della relazione tra le variabili Q , L e K : lineare nei logaritmi,
- (b) costanza delle elasticità β_2 e β_3 nel tempo (progresso tecnologico?),
- (c) assenza di errori di misurazione in Q , L e K .

Raramente la teoria economica ci permette di specificare completamente un modello. Generalmente altre ipotesi, non motivabili economicamente, sono necessarie.

- La normalità del termine d'errore U_t nell'Esempio (21) non è sostenibile in modo naturale usando esclusivamente la teoria economica.
- Questo implica la necessità di verificare tramite opportuni test statistici le ipotesi assunte.
- Talvolta l'analisi esplorativa dei dati ci aiuta nella comprensione del fenomeno in esame e ci permette di ottenere dei modelli più prossimi alla realtà.

Concludendo: le informazioni utilizzate nella specificazione dei modelli fanno capo alla

- teoria della probabilità,
- teoria economica,
- analisi esplorativa (si veda ad esempio la motivazione al premio Nobel 2003 per l'economia Robert F. Engle).