INFERENZA STATISTICA

Fulvio De Santis

a.a. 2007-2008

Definizione di Statistica

La statistica è la disciplina che:

• si occupa dell'analisi quantitativa dei fenomeni collettivi;

• fornisce concetti e metodi per la trattazione dell'informazione su fenomeni di interesse contenuta nei *dati*;

• studia la variabilità con cui i fenomeni si manifestano sui soggetti che costituiscono *popolazioni* di interesse.

POPOLAZIONE, UNITÀ E CARATTERE

Una ***popolazione*** di riferimento è costituita da più ***unità statistiche*** elementari in cui si manifestano caratteristiche di interesse, denominate ***caratteri***.

Costituiscono esempi di popolazione:

• la popolazione italiana (individui, famiglie...);

• insieme delle potenziali misurazioni della velocità di espansione di una galassia;

• insieme delle possibili misurazioni dell'"effetto" di un farmaco su unità sperimentali.

La ***statistica descrittiva*** si occupa dei metodi per la descrizione di come alcuni caratteri di interesse si manifestano in un collettivo osservato.

Per le distribuzioni dei caratteri oggetto di interesse si determinano indici di posizione, di variabilità (o mutabilità) e si studiano le eventuali relazioni tra più caratteri.

In pratica si occupa di quantificare le caratteristiche strutturali e essenziali delle distribuzioni di uno o più caratteri.

Tuttavia, spesso di una popolazione si può osservare solamente un suo sottoinsieme. Si pone così il problema di come sfruttare l'informazione relativa a tale sottoinsieme per ricostruire alcune caratteristiche generali della popolazione di interesse.

Inferenza Statistica

• I problemi di inferenza statistica sorgono quando di una popolazione di interesse se ne osserva solo una sua parte, il ***campione***.

• Sulla base di queste osservazioni vogliamo trarre conclusioni su qualche caratteristica di interesse di tutta la popolazione. Tale caratteristica la chiamiamo ***parametro***.

Esempi:

• 1. Si vuole cercare di stabilire chi vincerà le prossime elezioni presidenziali U.S.A. (D vs R). θ =proporzione dei voti a D (o a R). Si può effettuare inferenza su θ sulla base della percentuale p di voti a favore di D in un campione ottenuto per sondaggio.

• 2. Vogliamo determinare la durata media (θ) di vita delle lampadine prodotte da un sistema, attraverso la durata di vita osservata su un certo numero di unità campione.

• 3. Si vuole determinare la crescita (θ) di piante trattate con un fertilizzante attraverso i dati di crescita di un numero limitato di osservazioni sperimentali.

Nei tre esempi precedenti non posso osservare l'intera popolazione. Come procedo?

- Estraggo un numero di elettori e rilevo per chi votano.

- Misuro la durata di vita di n lampadine.

- Somministro il fertilizzante a n piante e misuro la loro crescita.

In ciascuno dei tre casi effettuo cioè un ***esperimento***, e sulla base dell'esito, ovvero dei dati che osservo, cerco di dare una valutazione per θ.

NOTA BENE:

... nei tre casi considerati ho tre motivi diversi per non considerare l'intera popolazione

- tempo e soldi...potenzialmente la popolazione è osservabile (situazione tipica nelle scienze sociali)

- la misurazione è distruttiva, quindi il censimento è impraticabile (situazione tipica in ambito tecnologico)

- la popolazione di riferimento è solo teorica (situazione tipica nelle scienze sperimentali)

Tuttavia, in tutti e tre i casi ricorro alla stessa soluzione: considero un sottoinsieme della popolazione di riferimento da cui trarre conclusioni più generali.

Ovvero: lo stesso schema logico si applica sia nel caso in cui il campione proviene da una *popolazione reale*, sia in quello in cui la proviene da una *popolazione virtuale*.

• ***Popolazione reale***: costituita da unità statistiche che esistono fisicamente e sulle quali rileviamo la caratteristica di interesse (esempio: il sesso di un persona, il n. di aziende agricole di una regione).

• ***Popolazione virtuale***: ha senso in riferimento a una ipotetica replicazione di un certo esperimento (esempio: le misure ripetute del peso di un oggetto o della concentrazione di una sostanza nell'acqua...la popolazione è l'insieme delle ipotetiche misurazioni effettuabili).

Il presupposto alla base dei problemi e dei metodi dell'inferenza statistica è che dietro ai fatti che si osservano, vi siano dei meccanismi che li generano.

In particolare: assumiamo che vi siano dei meccanismi probabilistici che regolano i risultati degli esperimenti.

Lo scopo principale dell'inferenza statistica è di sfruttare i risultati dell'esperimento per risalire alle leggi probabilistiche che li hanno generati.

Schematizzando

Inferenza statistica: tratta il processo logico-induttivo che permette di ottenere informazioni su una popolazione a partire da osservazioni campionarie.

Quindi:

$$\text{campione} \qquad \rightsquigarrow \qquad \text{popolazione}$$

Parametro: specifica caratteristica (non nota) della popolazione che si vuole "stimare"con i dati campionari (lo si indica in genere con θ).

Ciò lo si ottiene attraverso l'***esperimento statistico***

Esperimento Statistico

In generale: i *dati* rappresentano l'esito di un "esperimento" (in senso ampio) che potenzialmente può dare molteplici risultati. Prima di effettuare l'esperimento non conosco l'esito. Dopo l'esperimento so che ciò che ho osservato è uno dei possibili esiti.

Un pò più tecnicamente...

si può interpretare il risultato osservato come la realizzazione di una *variabile aleatoria* (o *casuale, stocastica*) X, cui corrisponde una *legge di probabilità* che controlla il processo di *generazione* delle osservazioni.

Esempio: lancio di una moneta.

Esperimento: si lancia una moneta.

Risultato: esce T.

Il risultato di questo esperimento è la *realizzazione* di una v.a. che può assumere il valore T, con probabilità θ e C con probabilità $1 - \theta$.

...altri esempi...

misure ripetute di grandezze, misurazione durata di funzionamento di un macchinario...

...in generale: è evidente, in questo schema, il legame profondo tra *probabilità* e *statistica*.

• Problema Deduttivo: data la legge di probabilià che regola il fenomeno, effettuare previsioni su esiti di esperimenti futuri.

Esempio: prevedere di ottenere 3 volte C in 5 lanci della moneta, supponendo di sapere che la moneta è, ad esempio, “bilanciata” ($\theta = 0.5$).

• Problema Induttivo: sulla base dei risultati osservati di un esperimento, trarre conclusioni sulla legge di probabilità che si suppone regoli il meccanismo che ha portato ai dati osservati.

Esempio: “stimare” θ (la probabilità che lanciando la moneta si osservi T) basandosi sull’informazione che su 5 lanci si sono osservate 3 T.

Elementi di calcolo delle Probabilità

Formalizzazione:

• Esperimento: processo che porta alla generazione di una osservazione di una variabile di interesse.

Esempi:

- a) lancio di un dado;

- b) conteggio del n. di assenze di uno studente in un anno scolastico;

- c) verifica del funzionamento di una macchina;

- d) verifica stato civile di una persona;

- e) misura dell'efficacia di un farmaco su un paziente.

• Risultato: esito dell'esperimento.

Esempi:

- a) 5;

- b) 10 assenze;

- c) SI (funziona);

- d) coniugato;

- e) valore numerico che indica l'effetto sul paziente del farmaco.

• SPAZIO CAMPIONARIO: insieme di tutti i possibili *risultati* di un *esperimento*. Si indica, in genere, con la lettera Ω.

Esempi:

- a) $\{1, 2, 3, 4, 5, 6\}$;

- b) $\{0, 1, 2, 3, ...M\}$;

- c) $\{SI, NO\}$;

- d) $\{coniugato, non\ coniugato\}$;

- e) ... sottoinsieme di $\mathbb{R}$.

• EVENTO: sottoinsieme dello *spazio campionario*.

Esempio.

Evento: esce un numero pari lanciando un dado...$\{2, 4, 6\}$.

Si indicano con lettere maiuscole: A, B, ...

Due eventi rilevanti:

Ω=spazio campionario o evento certo,

$\emptyset$=evento *impossibile*.

Eventi e Teoria degli Insiemi

È utile trattare gli eventi in termini di sottoinsiemi di uno spazio generico Ω e sfruttare la teoria degli insiemi.

Importante: sfruttando le operazioni su insiemi (eventi) si possono definire nuovi insiemi che corrispondono a eventi di interesse.

$A, \ B$ elementi di Ω: $A, \ B \subset \Omega$.

Unione di Insiemi

$A \cup B$= sottoinsieme di Ω costituito da elementi in A, o in B o in entrambi ... ovvero ... evento che si verifica quando si verifica almeno uno tra A e B.

Intersezione di Insiemi

$A \cap B$= sottoinsieme di Ω costituito da elementi sia in A che in B ... ovvero ... evento che si verifica quando si verificano contemporaneamente sia A che B.

Nota bene: due eventi sono **incompatibili** quando non possono verificarsi contestualmente, ovvero se l'intersezione degli insiemi corrispondenti è l'insieme vuoto,: $A \cap B = \emptyset$

Insieme complementare

A^c= sottoinsieme di Ω costituito da tutti gli elementi di Ω non contenuti in A ... ovvero ... evento che si verifica quando NON si verifica A.

Nota bene: per ogni $A \subset \Omega$, si ha

$A \cup A^c = \Omega$

$A \cup \Omega = \Omega$

$A \cup \emptyset = A$

$A \cap \Omega = A$

$A \cap \emptyset = \emptyset$

Se si ha che

$A,\ B \subset \Omega$

allora

$$A \subset B$$

se ogni punto di A appartiene anche a B; in termini di eventi diciamo che A "implica" B, ovvero il verificarsi di A implica il verificarsi di B.

Una relazione importante

Risulta utile esprimere un evento come unione di due eventi incompatibili: Consideriamo $A, B \subset \Omega$; Si ha:

$$A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Esempio. Con riferimento a una automobile che viene sottoposta a un test sull'emissione del gas di scarico, indichiamo:

$F =$ l'evento relativo al non superare il test;

$I =$ evento relativo al fatto che l'auto inquina;

Il test non è infallibile! Quindi: l'auto può non superare il test sia che inquini sia che non inquini. Pertanto: l'evento F può verificarsi assieme al verificarsi dell'evento I o (in alternativa) assieme al verificarsi dell'evento I^c. Quindi:

$$F = (F \cap I) \cup (F \cap I^c).$$

Definizione di Probabilità

Definizione classica. La probabilità di un evento è il rapporto tra il numero di *casi* favorevoli all'evento e il numero di casi possibili (purchè questi siano tra loro ugualmente probabili!!).

Esempio.

Lanciando 2 dadi, la probabilità di $E = \{\text{la somma è } 8\}$

$$\begin{aligned} \Omega &= \{(1,1),(1,2),\cdots,(6,5),(6,6)\} \\ E &= \{(2,6),(3,5),(4,4),(5,3),(6,2)\} \end{aligned}$$

Ω ha 36 elementi, E ne ha 5 e $\mathbb{P}[E] = 5/36$.

In generale, se n_E indica il numero di casi favorevoli a E e n il numero di casi possibili:

$$\mathbb{P}(E) = \frac{\text{numero dei casi favorevoli ad } E}{\text{numero dei casi possibili}} = \frac{n_E}{n}.$$

Definizione frequentista. Nell'ipotesi che l'esperimento possa essere ripetuto indefinitamente, sempre nelle stesse condizioni.

Se, ripetendo la prova n volte, l'evento E si verifica n_E volte, la frequenza relativa dell'evento è n_E/n. Allora la probabilità di E risulta:

$$\mathbb{P}[E] = \lim_{n \to \infty} \frac{n_E}{n}.$$

Definizione soggettiva. La probabilità di un evento è legato al grado di fiducia che una persona ha nel verificarsi o meno dell'evento.

Proprietà della Probabilità

Operativamente esperimento risultato evento, classe di eventi

1. $\forall A \subset \Omega \qquad 0 \leq \mathbb{P}(A) \leq 1.$
2. $\mathbb{P}(\Omega) = 1, \qquad \mathbb{P}(\emptyset) = 0.$
3. $A, B \subset \Omega \qquad A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$

Conseguenze:

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$
- $A \cap B \neq \emptyset, \qquad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$

Probabilità Condizionata

Lanciando un dado regolare, possiamo assumere che

$$\mathbb{P}(\text{pari}) = \frac{1}{2} \qquad \mathbb{P}(\{4,5,6\}) = \frac{1}{2}$$

Supponiamo di sapere che il risultato del lancio del dado è un numero pari.

La probabilità di $(\{4,5,6\})$ è ancora 1/2?

NO, perché alla luce della nuova informazione

$\star$ i risultati possibili sono: 2,4,6

$\star$ quelli favorevoli all'evento (≥ 4) sono: 4,6

L'informazione aggiuntiva ha cambiato la probabilità degli eventi elementari

$$\begin{aligned}
\mathbb{P}[\{1\}|\text{pari}] = \mathbb{P}[\{3\}|\text{pari}] = \mathbb{P}[\{5\}|\text{pari}] &= 0 \\
\mathbb{P}[\{2\}|\text{pari}] = \mathbb{P}[\{4\}|\text{pari}] = \mathbb{P}[\{6\}|\text{pari}] &= \frac{1}{3}
\end{aligned}$$

Quindi, sapendo che il risultato è pari:

$$\mathbb{P}(\text{pari}) = 1, \qquad \mathbb{P}(\{4,5,6\})) = \frac{2}{3}$$

Formalmente:

dati 2 eventi A e B, con $\mathbb{P}(B) > 0$, se sappiamo che si è verificato B, allora

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

Vediamo come si comprende. Rappresentiamo eventi e casi favorevoli in una tabella 2×2

	B	B^c	Tot.
A	n_{AB}	n_{AB^c}	n_A
A^c	n_{A^cB}	$n_{A^cB^c}$	n_{A^c}
Tot.	n_B	n_{B^c}	n

Allora:

$$\mathbb{P}(A) = \frac{n_A}{n}, \qquad \mathbb{P}(A|B) = \frac{n_{AB}}{n_B}.$$

Quindi:

$$\mathbb{P}(A|B) = \frac{n_{AB}/n}{n_B/n} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Analogamente:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} (= \frac{n_{AB}}{n_A}).$$

Esercizio.

La seguente tabella riporta i risultati di una indagine sul livello di servizio di assistenza tecnica fornito da 150 unità di personale di una azienda informatica. I tecnici considerati, vengono classificati in base al fatto di aver fornito un buon servizio e in base all'aver seguito o meno un corso di formazione specifico.

	Buon Servizio	Cattivo Servizio	Tot.
formazione SI	48	16	64
formazione NO	24	62	86
Tot.	72	78	150

Supponendo di scegliere a caso uno dei tecnici, tra i 150 disponibili, determinare le probabilità di selezionare un tecnico che:

a) fornisca un buon servizio;

b) abbia seguito il corso di formazione;

c) fornisca un buon servizio e abba seguito il corso di formazione;

d) fornisca un buon servizio, sapendo che non ha seguito il corso di formazione;

e) fornisca un buon servizio, sapendo che ha seguito il corso di formazione;

ESERCIZIO.

La probabilità che vi sia uno sciopero è pari a 0.28; La probabilità che un certo lavoro venga terminato in tempo e che non vi sia lo sciopero è pari 0.64. Determinare la probabilità che il lavoro venga terminato in tempo, sapendo che non lo sciopero non avrà luogo.

$\diamond\diamond\diamond$

- $S = \{\text{sciopero}\}; \qquad \mathbb{P}(S) = 0.28.$
- $T = \{\text{terminare il lavoro}\}; \qquad \mathbb{P}(T \cap S^c) = 0.64.$

Vogliamo $\mathbb{P}(T|S^c)$:

$$\mathbb{P}(T|S^c) = \frac{\mathbb{P}(T \cap S^c)}{P(S^c)} = \frac{0.64}{0.72} = 0.89.$$

Nota bene: se so che lo sciopero non ci sarà, la probabilità di terminare il lavoro è più alta di quando non ho tale informazione.

Indipendenza

Esempio.

Si è stimato che le probabilità che uno studente superi gli esami di algebra, inglese o entrambi sono rispettivamene pari a 0.75, 0.84 e 0.63. Qual è la probabilità che uno studente superi l'esame di alegebra, sapendo che ha superato quello di inglese? Indichiamo con

- ◇ $A = \{\text{superare Algebra}\}, \qquad \mathbb{P}(A) = 0.75$
- ◇ $I = \{\text{superare Inglese}\} \qquad \mathbb{P}(I) = 0.84$
- ◇ $A \cap I = \{\text{superare sia Algebra che Inglese}\} \qquad \mathbb{P}(A \cap I) = 0.63.$

Vogliamo: $\mathbb{P}(A|I)$. Abbiamo:

$$\mathbb{P}(A|I) = \frac{\mathbb{P}(A \cap I)}{\mathbb{P}(I)} = \frac{0.63}{0.84} = 0.75.$$

Quindi:

$$\mathbb{P}(A|I) = \mathbb{P}(A) = 0.75!!!!$$

Sapere che lo studente ha superato inglese non modifica la prob. che superi algebra! I due eventi A e B sono **indipendenti**.

Definizione. *Due eventi sono* indipendenti *se il verificarsi dell'uno non modifica la probabilità del verificarsi dell'altro.*

Formalmente:

$$A, B \subset \Omega \qquad \mathbb{P}(A|B) = \mathbb{P}(A), \qquad \mathbb{P}(B|A) = \mathbb{P}(B).$$

Poichè

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B),$$

possiamo dire che A e B sono indipendenti se e solo se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

Nota bene: è una nozione relativa alle probabilità degli eventi, e non agli eventi stessi.

Esempio. Determinare la probabilità di estrarre due assi da un mazzo da 52 carte supponendo di:

a) estrarre la seconda carta dopo aver unito al mazzo la prima carta estratta;

b) estrarre la seconda carta senza unire al mazzo la prima carta estratta. Indichiamo con

$$A_i = \{\text{esce ASSO alla estrazione i-esima}\}, i = \qquad = 1, 2.$$

Vogliamo determinare:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 | A_1)\mathbb{P}(A_1).$$

Due casi distinti:

a)*Estrazione con rimpiazzo.* Gli eventi A_1 e A_2 sono indipendenti, quindi:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) = \mathbb{P}(A_2)\mathbb{P}(A_1) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}.$$

b) *Estrazione senza rimpiazzo.* L'evento A_2 non è indipendente dall'evento A_1. Infatti:

$$\mathbb{P}(A_2|A_1) = \frac{3}{51} \neq \frac{4}{52} = \mathbb{P}(A_2).$$

Quindi:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) = \frac{3}{51} \times \frac{4}{52} = \frac{1}{221}.$$

Questi schemi sono importanti nel *campionamento* statistico.

Variabili Aleatorie

Spesso siamo interessati non direttamente ai risultati di una prova ma a loro opportune funzioni. Ad esempio, lanciando 3 volte una moneta, possiamo essere interessati al numero di C uscite, piuttosto che agli esiti dei singoli lanci.

Definizione. *Una variabile aleatoria (v.a.) è una funzione del risultato di un esperimento.*

Si tratta quindi di *variabili* che possono assumere diversi valori con determinate probabilità

Variabile aleatoria discreta. Una v.a. è discreta se può assumere un numero finito o numerabile di valori.

Esempio. Lanciando 3 volte una moneta, possiamo ottenere i seguenti risultati (spazio dei campioni):

$$\{CCC, CCT, CTC, TCC, CTT, TCT, TTC, TTT\}.$$

Se siamo interessati al numero di C, gli esiti della prova possono essere: 0, 1, 2, 3. Abbiamo quindi una v.a. che può assumere i seguenti valori con le seguenti probabilità:

N. di C	Probabilità
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Nota Bene: la variabile aleatoria è identificata dai valori che può assumere e dalle probabilità corrispondenti.

In generale, se

$$x_1, \ldots, x_k \qquad p_1, \ldots, p_k$$

sono i k possibili valori che la v.a. aleatoria può assumere e le corrispondenti probabilità, si ha che:

$$0 \leq p_r \leq 1 \quad r = 1, \ldots, k \qquad \text{e} \qquad p_1 + \ldots + p_k = 1.$$

Valore atteso

Per una v.c. discreta X che assume valori

$$x_1, \cdots, x_k \text{ con probabilità } p_1, \cdots, p_k$$

si definisce il *valore atteso* (o valor medio)

$$\mathbb{E}[X] = \sum_{i=1}^{k} x_i \, p_i.$$

Il valor medio sintetizza la distribuzione attraverso un solo numero.

Varianza

La varianza è il valor medio delle distanze al quadrato dei valori x_i da $\mathbb{E}[X]$:

$$\mathbb{V}[X] = \sum_{i=1}^{k} (x_i - \mathbb{E}[X])^2 \, p_i.$$

È una misura di *dispersione* della v.a. X rispetto al valor medio.

Variabile Aleatoria Bernoulliana

In molti casi si è interessati a esperimenti il cui esito è dicotomico: o si si verifica un evento, oppure non si verifica.

Ad esempio: una terapia medica può avere successo o meno, un embrione umano può essere maschile o femminile, uno studente supera o meno un esame.

Una variabile aleatoria che può assumere solo due valori è detta **v.a. bernoulliana**.

Formalmente: indichiamo con X la v.a e con $\{0, 1\}$ i due possibili valori che può assumere. Inoltre:

$$\mathbb{P}(X = 1) = p \qquad \mathbb{P}(X = 0) = 1 - p.$$

Quindi, sinteticamente:

$$\mathbb{P}(X = x) = p^x (1 - p)^{1-x}, \qquad\qquad x = 0, 1.$$

Variabile Aleatoria Binomiale
Schema delle Prove Ripetute

Schema logico:

• Si effettua una prova il cui esito può essere un "successo" (si verifica l'evento E) con probabilità p, o un "insuccesso" (si verifica l'evento E^c), con probabilità $1 - p$;

• si ripete la stessa prova n volte, nelle stesse condizioni;

• le prove sono indipendenti.

La v.a. che ci interessa è: numero di "successi" nelle n prove effettuate.

Esempio. Uno studente risponde a caso a 3 domande di un test a risposte multiple. Per ciascuna domanda ci sono 5 possibili risposte, tra le quali una soltanto è esatta. Lo studente può scegliere solo una tra le 5 possibili risposte. Qual è la probabilità che lo studente risponda bene a due domande?

$\diamond \diamond \diamond$

E=risposta esatta

$\mathbb{P}(E) = \dfrac{1}{5}$

S=numero di risposte esatte su tre

L'evento $S = 2$ si può verificare in tre modi tra loro incompatibili:

$$(E \cap E \cap E^c) \cup (E \cap E^c \cap E) \cup (E^c \cap E \cap E)$$

Quindi:

$$\mathbb{P}(S = 2) = \mathbb{P}(E \cap E \cap E^c) + \mathbb{P}(E \cap E^c \cap E) + \mathbb{P}(E^c \cap E \cap E)$$

Ciascuno dei tre eventi che formano la unione ha la stessa probabilità:

$$\mathbb{P}(E \cap E \cap E^c) = \mathbb{P}(E \cap E^c \cap E) = \mathbb{P}(E^c \cap E \cap E) =$$

$$= \mathbb{P}(E)\mathbb{P}(E)\mathbb{P}(E^c) = \frac{1}{5} \times \frac{1}{5} \times \frac{4}{5} = \left(\frac{1}{5}\right)^2 \frac{4}{5}$$

e quindi:

$$\mathbb{P}(S = 2) = 3\left(\frac{1}{5}\right)^2 \frac{4}{5}$$

$\diamond \diamond \diamond$

In generale, indicando con S il numero di "successi" si ha:

$$\begin{aligned}\mathbb{P}(E \cap \ldots \cap E \cap E^c \ldots \cap E^c) &= p \cdot \ldots \cdot p \cdot (1-p) \cdot \ldots \cdot (1-p) \\ &= p^s (1-p)^{n-s}.\end{aligned}$$

Poichè non siamo interessati all'ordine con cui i successi hanno luogo, si ha:

$$\mathbb{P}(S = s) = \binom{n}{s} p^s (1-p)^{n-s}, \qquad s = 0, 1, \ldots, n$$

dove

$$\binom{n}{s} = \frac{n!}{k!(n-s)!}$$

è il coefficiente binomiale e indica le combinazioni di n oggetti di classe s, ovvero il numero di modi in cui possiamo selezionare s oggetti tra n disponibili e

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$$

le permutazioni di n oggetti.

I valori che una v.a. binomiale può assumere e le corrispondenti probabilità sono:

s	$\mathbb{P}(S = s)$
0	$(1-p)^n$
1	$np(1-p)^{n-1}$
$\vdots$	$\vdots$
s	$\binom{n}{s}p^s(1-p)^{n-s}$
$\vdots$	$\vdots$
n	p^n

È spesso necessario calcolare le probabilità di avere:

- *almeno* s successi su n prove,

ovvero dell'evento:

$\{S \geq s\} = \{S = s\} \cup \{S = s + 1\} \cup \ldots \cup \{S = n\};$

• *al più* s successi su n prove,

ovvero dell'evento:

$\{S \leq s\} = \{S = s\} \cup \{S = s - 1\} \cup \ldots \cup \{S = 0\};$

da cui

$\mathbb{P}(S \geq s) = \sum_{r=s}^{n} \left[\binom{n}{r} p^r (1-p)^{n-r}\right]$

e

$\mathbb{P}(S \leq s) = \sum_{r=0}^{s} \left[\binom{n}{r} p^r (1-p)^{n-r}\right].$

Esempio. È noto che la probabilità che un divorziato si risposi entro 3 anni dal divorzio è pari a 0.40. Determinare le probabilità che, entro 3 anni, su 10 divorziati:

a) al più 3 si risposino;

b) almeno 7 si risposino;

c) se ne risposino da 2 a 5;

d) se ne risposino almeno 2.

⋄ ⋄ ⋄

Si ha: $n = 10$, $p = 0.40$. Quindi:

a)
$\mathbb{P}(S \leq 3) = \mathbb{P}(S = 0) + \mathbb{P}(S = 1) + \mathbb{P}(S = 2) + \mathbb{P}(S = 3) = \cdots = 0.382$

b)
$\mathbb{P}(S \geq 7) = \mathbb{P}(S = 7) + \mathbb{P}(S = 8) + \mathbb{P}(S = 9) + \mathbb{P}(S = 10) = \cdots = 0.055$

c) $\mathbb{P}(2 \leq S \leq 5) = \mathbb{P}(S = 2) + \mathbb{P}(S = 3) + \mathbb{P}(S = 4) + \mathbb{P}(S = 5) = \cdots = 0.788$

d) $\mathbb{P}(S \geq 2) = 1 - \mathbb{P}(S < 2) = 1 - \mathbb{P}(S = 0) + \mathbb{P}(S = 1) = 0.954$

Variabili Aleatorie Continue

Una v.a. X continua è tale che i valori che può assumere sono un sottoinsieme dei numeri reali.

In questo caso, non possiamo più assegnare una probabilità a ciascun valore che la v.a. può assumere.

Le v.a. continue sono caratterizzate da una funzione, detta **funzione di densità** che permette di determinare la probabilità che la v.a. X assuma valori in intervalli di $\mathbb{R}$.

In particolare, la funzione di densità $f(x)$ di X è tale che:

1) $f_X(x) \geq 0$ per ogni valore x che X può assumere;

2) l'area della superficie al di sotto di $f_X(x)$ è pari a 1;

3) $\mathbb{P}\{a \leq X \leq b\}$ è uguale all'area della superficie al di sotto di $f_X(x)$, in $[a, b]$.

Valore atteso di una v.a continua

Per una v.c. continua X che assume valori in $A \subset \mathbb{R}$, si definisce il *valore atteso* (o valor medio)

$$\mathbb{E}(X) = \int_A x f(x) dx.$$

È l'estensione al caso continuo della definizione già vista nel caso discreto.

Anche in questo caso, il valor medio rappresenta un valore che sintetizza la distribuzione attraverso un solo numero.

Varianza

Estendendo quanto visto nel caso discreto, la varianza di X, v.a. continua è

$$\mathbb{V}(X) = \int_A (x - \mathbb{E}(x))^2 f(x) dx.$$

È una misura di *dispersione* della v.a. X rispetto al valor medio.

Variabile Aleatoria Normale

Una v.a. normale può assumere valori in $(-\infty, +\infty)$. La funzione di densità di una v.a. normale dipende da due parametri, μ e σ, che sono il valore medio e la varianza di X:

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\},$$

dove

$$\mathbb{E}(X) = \mu \qquad \text{e} \qquad \mathbb{V}(X) = \sigma^2.$$

Tra le proprietà della v.a. normale, ricordiamo che:

- la funzione di densità è simmetrica;
- $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826$
- $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9544$
- $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9974$

ESEMPIO. È noto che i punteggi ottenuti in un test per la misura del QI di una popolazione sono distribuiti secondo una distribuzione normale di media 100 e varianza 16^2 [$X \sim N(\mu, \sigma^2) = N(100, 16^2)$].

Determinare la percentuale di individui della popolazione con QI compreso tra i valori 100 e 120.

$\diamond \diamond \diamond$

$$\mathbb{P}(100 \leq X \leq 120) = \mathbb{P}\left(0 \leq \frac{X - 100}{16} \leq \frac{120 - 100}{16}\right) =$$

$$= \mathbb{P}(0 \leq Z \leq 1.25) - \frac{1}{2} = 0.8944 - 0.5.$$

Inferenza Statistica

Fin qui abbiamo considerato problemi deduttivi per v.a.

- bernoulliana e binomiale;
- normale.

In pratica, abbiamo assunti noti i valori di p per le prime due e di μ e σ^2 per la seconda.

In questi esempi, p, μ e σ^2 corrispondono a caratteristiche delle popolazioni rappresentate dalle v.a. corrispondenti.

Inoltre, determinano la distribuzione di probabilità e la funzione di densità. Le leggi di probabilità dipendono da alcuni *parametri*.

In generale, una v.a. X con distribuzione di probabilità $p_X(x;\theta)$ o funzione di densità $f_X(x;\theta)$ dipende da un parametro θ.

L'inferenza statistica si occupa del problema inverso a quelli visti fin qui:

- si ipotizza che i dati provengano da una determinata popolazione (X), cui è associata una legge di probabilità che dipende da un parametro θ incognito;
- sulla base di dati osservati, si vuole allora stimare θ.

Data la v.a. X, supponiamo di considerare il valore che essa assume in corrispondenza di n unità che formano il campione.

Abbiamo così n v.a. (una per ciascuna unità):

$$X_1, X_2, \ldots, X_i, \ldots, X_n$$

Si osservano i valori

$$x_1, x_2, \ldots, x_i, \ldots, x_n.$$

In particolare ci occupiamo di *campioni casuali semplici*: si ipotizza che ciascuna delle variabili X_i abbia la stessa distribuzione di X e che siano tra loro indipendenti. Si dice allora che le variabili $X_1, X_2, \ldots, X_n$ sono *indipendenti e identicamente distribuite* (i.i.d.).

In tal caso:

$$\mathbb{P}\{X_1 = x_1, \ldots, X_n = x_n; \theta\} = \prod_{i=1}^{n} \mathbb{P}\{X_i = x_i; \theta\} = \prod_{i=1}^{n} p_X(x_i; \theta),$$

per v.a. discrete e

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

per v.a. continue.

Schema dei Principali Problemi Inferenziali

Sulla base dei dati campionari $x_1, \ldots, x_n$, procedo come segue.

- **Stima Puntuale di Parametri**: cerco *un unico* valore numerico da attribuire a θ.

- **Stima per Intervalli**: cerco un *intervallo di valori numerici* da attribuire a θ.

- **Verifica di ipotesi**: cerco di stabilire quale, tra due valori per θ (o tra due insiemi di valori), sia quello più plausibile.

Esempio (Bernoulliano)

Consideriamo una popolazione di potenziali acquirenti di un prodotto. Il comportamento di ciascun acquirente può essere formalizzato con una v.a. bernoulliana, X, che assume valore 1 se il cliente compra il prodotto, 0 se non lo acquista. Il parametro di interesse, p, è la probabilità che un cliente compri il prodotto. Formalmente, la distribuzione di probabilità di X risulta:

$$\mathbb{P}(X = 1; p) = p \qquad \mathbb{P}(X = 0; p) = 1 - p.$$

Quindi, sinteticamente:

$$\mathbb{P}(X = x; p) = p^x (1 - p)^{1-x}, \qquad\qquad x = 0, 1.$$

Poichè p è incognito, si considera un campione casuale di n potenziali acquirenti e, per ciascuno di questi, si verifica se il prodotto è stato acquistato o meno. Si suppone che gli acquirenti si comportino indipendentemente e che la probabilità che acquistino sia la stessa per tutti. Il campione casuale è quindi rappresentato da n v.a.

$$X_1, X_2, \ldots, X_n$$

tra loro indipendenti e con stessa distribuzione (i.i.d.):

$$\mathbb{P}(X_i = x_i; p) = p^{x_i}(1-p)^{1-x_i}, \qquad x_i = 0, 1.$$

Nota bene: X_i indica la v.a., mentre x_i il valore effettivamente osservato nel campione.

Un modo intuitivamente ragionevole per *stimare* p, è di considerare la proporzione di clienti che hanno acquistato il prodotto tra gli n considerati:

$$\hat{p}(x_1, \ldots, x_n) = \frac{x_1 + \ldots + x_n}{n}.$$

Ad esempio, se $n = 4$, potremmo avere

$$x_1 = 1, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 1$$

e ottenere

$$\hat{p}(x_1, \ldots, x_4) = \frac{1 + 0 + 0 + 1}{4} = 0.5.$$

Ovviamente, ripetendo l'esperimento, potrei ottenere un diverso campione e quindi una diversa stima di p. Ad esempio:

$$x_1 = 0, \quad x_2 = 0, \quad x_3 = 1, \quad x_4 = 0$$

e

$$\hat{p}(x_1, \ldots, x_4) = \frac{0 + 0 + 1 + 0}{4} = 0.25.$$

La stima di p varia cioè al variare dei possibili valori osservati nel campione.

$\diamond \diamond \diamond$

Quello illustrato è un semplice problema di *stima puntuale*.

Idea: la distribuzione di probabilità del carattere di interesse (X) dipende da un parametro incognito (p), che cerco di determinare usando dati campionari.

Dal momento che i dati campionari sono realizzazioni di v.a. ($X_1, \ldots, X_n$), anche la *stima* di p ($\hat{p}(x_1, \ldots, x_n)$) è la realizzazione di una v.a. ($\hat{p}(X_1, \ldots, X_n)$), che ha una sua distribuzione di probabilità, detta *distribuzione campionaria*

Quindi:

$$X_1, \ldots, X_n \qquad \rightsquigarrow \qquad x_1, \ldots x_n$$

e, in corrispondenza

$$\hat{p}(X_1, \ldots, X_n) \qquad \rightsquigarrow \qquad \hat{p}(x_1, \ldots x_n)$$

$$\hat{p}(X_1, \ldots, X_n) \qquad\qquad \text{STIMATORE DI } p$$

$$\hat{p}(x_1, \ldots x_n) \qquad\qquad \text{STIMA DI } p.$$

Lo scopo della stima puntuale è individuare stimatori che, con elevata probabilità, diano luogo a stime “vicine” al valore vero incognito del parametro p.

Esempio (Normale)

Una azienda è interessata a stabilire il numero di ore che ciascun suo impiegato dedica al disbrigo di una determinata pratica in ogni mese. Risulta plausibile formalizzare il n. di ore dedicate da ciascun impiegato alla pratica con una v.a. normale di media μ e varianza σ^2, entrambe incognite. In tal caso, la funzione di densità di X risulta:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\},$$

Si suppone quindi di considerare un campione casuale di n impiegati e di considare, per ciascuno di questi, la v.a. numero di ore dedicate alla pratica nell'ultimo mese,

$$X_1, X_2, \dots, X_n$$

ciascuna con densità:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\} \qquad i = 1, \ldots, n, .$$

Una ragionevole *stima* di μ, in questo caso, può essere la *media campionaria*

$$\hat{\mu}(x_1, \ldots x_n) = \bar{x} = \frac{x_1 + \ldots + x_n}{n},$$

il cui valore numerico cambia con i valori osservati $x_1, \ldots, x_n$.

Anche qui, la speranza è che il corrispondente *stimatore*,

$$\hat{\mu}(X_1, \ldots X_n) = \bar{X} = \frac{X_1 + \ldots + X_n}{n},$$

(dotato di una sua distribuzione di probabilità) fornisca, con elevata probabilità, stime non "lontane" dal valore vero di μ.

◇ ◇ ◇

La Funzione di Verosimiglianza

Il problema di cui ci occupiamo ora è di introdurre un metodo "costruttivo" per determinare uno stimatore.

In questo contesto il concetto di *funzione di verosimiglianza* occupa una posizione centrale.

Ricordiamo: data la v.a. X con distribuzione (caso discreto)

$$p_X(\cdot;\theta), \theta \in \Theta$$

siamo interessati ad acquisire informazioni su θ, attraverso i dati campionari $\mathbf{x}_n = (x_1, \ldots, x_n)$, realizzazione delle v.a. $X_1, \ldots, X_n$.

IDEA: individuare, tra tutti i possibili valori di θ, quello che massimizza la probabilità di osservare il campione realizzato $\mathbf{x}_n$

Definizione.

Data la legge $p_{X_1,\ldots,X_n}(\cdot;\theta), \theta \in \Theta$*, si chiama* funzione di verosimiglianza *associata al campione* $\mathbf{x}_n = (x_1, \ldots, x_n)$ *la funzione:*

$$L(\theta; \mathbf{x}_n) = p_{X_1,\ldots,X_n}(\mathbf{x}_n; \theta), \qquad \theta \in \Theta.$$

Nei casi più semplici le osservazioni $\mathbf{x}_n = (x_1, \ldots, x_i, \ldots, x_n)$ sono realizzazioni di v.a. indipendenti e identicamente distribuite:

$$L(\theta; \mathbf{x}_n) = p_{X_1,\ldots,X_n}(\mathbf{x}_n; \theta) = \prod_{i=1}^{n} p_X(x_i; \theta).$$

Si tratta quindi di una funzione di θ ottenuta considerando la probabilità congiunta del campione calcolata nel punto $\mathbf{x}_n$.

Quindi: $L(\theta; \mathbf{x}_n)$, la verosimiglianza di $\theta \in \Theta$, è la probabilità che si assegnerebbe al campione osservato $\mathbf{x}_n$ se si assumesse che θ fosse il vero valore (che non conosciamo) del parametro incognito.

Interpretazione: dato il campione $\mathbf{x}_n$, $L(\theta, \mathbf{x}_n)$ descrive, al variare di θ, le probabilità del campione effettivamente realizzato.

Idea: se, considerati due possibili valori di $\theta \in \Theta$, θ_1 e θ_2 e un campione osservato $\mathbf{x}_n$, si ha

$$L(\theta_1; \mathbf{x}_n) > L(\theta_2; \mathbf{x}_n),$$

allora l'ipotesi θ_1 risulta, alla luce dei dati, $\mathbf{x}_n$ maggiormanete plausibile (*verosimile*) dell'ipotesi θ_2, nel senso che il campione che si è osservato è più probabile quando $\theta = \theta_1$ piuttosto che quando $\theta = \theta_2$.

Quindi: ogni ipotesi θ riceve, con la realizzazione dell'esperimento che produce $\mathbf{x}_n$, un "supporto sperimentale" misurato da $L(\theta; \mathbf{x}_n)$.

Per confrontare il peso del supporto ricevuto da ipotesi diverse si considerano i *rapporti di verosimiglianza*:

$$\frac{L(\theta_1; \mathbf{x}_n)}{L(\theta_2; \mathbf{x}_n)}.$$

Dato un campione osservato $\mathbf{x}_n$ e la f.v. $L(\theta; \mathbf{x}_n)$, la ***stima di massima verosimiglianza*** di θ è il punto di massimo di $L(\theta; \mathbf{x}_n)$.

OSSERVAZIONI

(i) $L(\theta; \mathbf{x}_n)$ NON è la probabilità di θ dato $\mathbf{x}_n$; è una probabilità, ma relativa all'evento $(X_1 = x_1, \ldots, X_n = x_n)$ e non all'ipotesi θ

(ii) Il concetto di verosimiglianza è legato al "modello" di partenza.

ESEMPIO (BERNOULLIANO).

Consideriamo un modello bernoulliano:

$$X_i|\theta \sim Ber(\theta), \qquad \theta \in [0,1],$$

ovvero,

$$f(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

e

$$\prod_{i=1}^{n} f_X(x_i;\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{\sum_{i=1}^{n}(1-x_i)}.$$

Consideriamo $n = 5$ e il risultato $\mathbf{x}_n$, tale che:

$$x_1 = 1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 0,\ x_5 = 1.$$

La funzione di verosimiglianza di θ associata al risultato osservato è:

$$L(\theta; \mathbf{x}_n) = \theta^3(1-\theta)^2, \qquad \theta \in [0,1].$$

Si verifica che $L(\theta; \mathbf{x}_n)$ ha un unico punto di massimo nel punto $3/5 = 0.6$. Infatti:

$$\frac{d}{d\theta} \ln[\theta^3(1-\theta)^2] = 0 \Leftrightarrow \theta = \frac{3}{5}.$$

La f.v. relativa (ottenuta dividendo la f.v. per il suo valore massimo) è pertanto:

$$\bar{L}(\theta; \mathbf{x}_n) = \left(\frac{\theta}{0.6}\right)^3 \left(\frac{1-\theta}{0.4}\right)^2$$

Dall'esame del grafico si evince che, alla luce dei dati osservati, i diversi possibili valori di θ ricevono un diverso grado di supporto.

Supponiamo ora che $n = 50$ e che il campione osservato sia tale che:

$$\sum_{i=1}^{n} x_i = 30, \qquad \sum_{i=1}^{n} (1 - x_i) = 20.$$

In tal caso:

$$\bar{L}(\theta; \mathbf{x}_n) = \left(\frac{\theta}{0.6}\right)^{30} \left(\frac{1-\theta}{0.4}\right)^{20}.$$

La f.v. risulta ben più concentrata intorno al suo valore di massimo

$$\Longrightarrow$$

supporto sperimentale più accurato.

Il Caso Continuo

Fin qui abbiamo considerato una definizione di f.v. valida per v.c. discrete.

Nella sostanza non cambia molto per v.s. continue.

Se $f_{X_1,\ldots,X_n}(\cdot;\theta)$ è la funzione di densità di $(X_1,\ldots,X_n)$, v.a. assolutamente continue, allora la f.v. di θ associata al campione osservato $\mathbf{x}_n$ risulta

$$L(\theta;\mathbf{x}_n) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta).$$

In generale

Dato $\mathbf{x}_n$, indichiamo con $\hat{\theta}(\mathbf{x}_n)$ il punto di massimo di $L(\theta; \mathbf{x}_n)$:

$\hat{\theta}(\mathbf{x}_n)$ è il punto che riceve maggiore supporto dai dati osservati. Tale valore si chiama *Stima di Massima Verosimiglianza*.

NOTA BENE

• Non sempre esiste unico.

• Non sempre agevole da determinare $\rightsquigarrow$ metodi numerici di massimizzazione.

• Un caso semplice ma importante: $L(\theta; \mathbf{x}_n)$ ammette derivata prima e seconda rispetto a $\theta \quad (\Theta \subset \mathbb{R})$. Si considerano le condizioni per l'esistenza di un minimo:

$$\begin{aligned} \frac{d}{d\theta} L(\theta; \mathbf{x}_n) &= 0 \qquad \rightsquigarrow \qquad \hat{\theta}(\mathbf{x}_n) \\ \frac{d^2}{d\theta^2} L(\theta; \mathbf{x}_n)\Big|_{\theta=\hat{\theta}(\mathbf{x}_n)} &< 0. \end{aligned}$$

- Spesso è più semplice derivare la log-verosimiglianza:

$$\ell(\theta; \mathbf{x}_n) = \log L(\theta; \mathbf{x}_n).$$

- Nel caso di osservazioni i.i.d.:

$$\ell(\theta; \mathbf{x}_n) = \log L(\theta; \mathbf{x}_n) =$$

$$\log \prod_{i=1}^{n} f_X(x_i; \theta) = \sum_{i=1}^{n} \log f_X(x_i; \theta).$$

Esempio (normale).

Sia $(x_1, \ldots, x_i, \ldots, x_n)$ la realizzazione di un campione casuale tratto da una distribuzione $N(\theta, \sigma^2)$, dove $\theta \in \mathbb{R}$ e $\sigma^2 > 0$ è noto. La f.v. risulta:

$$\begin{aligned} L(\theta; x_1 \ldots x_n) &= \prod_{i=1}^{n} f(x_i; \theta) = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2\right\} \qquad \theta \in \mathbb{R}, \end{aligned}$$

da cui:

$$L(\theta; x_1 \ldots x_n) \propto \exp\left\{-\frac{n}{2\sigma^2}(\theta - \bar{x})^2\right\}.$$

Poichè il massimo si ha per $\theta = \bar{x}$,

$$\Longrightarrow \bar{L}(\theta; \mathbf{x}_n) = \exp\left\{-\frac{n}{2\sigma^2}(\theta - \bar{x})^2\right\}.$$

NOTA BENE

- $\bar{L}(\theta;\mathbf{x}_n)$ è proporzionale a una densità $N(\theta;\sigma^2/n)$;
- $\bar{L}(\theta;\mathbf{x}_n)$ è tanto più concentrata intorno a $\bar{x}$ quanto più n è elevato.

Stimatore di Massima Verosimiglianza

Se, per un campione osservato $\mathbf{x}_n$,

$$\hat{\theta}(x_1, \ldots, x_n)$$

indica la *stima* di massima verosimiglianza, lo stimatore associato

$$\hat{\theta}(X_1, \ldots, X_n)$$

si chiama *stimatore di massima verosimiglianza.*

È una v.a., e quindi possiede una distribuzione campionaria.

Esempio

Supponiamo che $x_1, \ldots, x_n$ sia un campione casuale tale che, per ogni i si abbia

$$\mathbb{P}(X_i = x_i; \theta) = f(x_i; \theta) = \frac{\theta^{x_i}}{x_i!} e^{-\theta}, \qquad x_i = 0, 1, 2, \ldots.$$

In tal caso diciamo che il campione proviene da una *popolazione di Poisson* di parametro θ. Si dimostra che

$$\mathbb{E}(X_i) = \mathbb{V}(X_i) = \theta.$$

Per trovare lo stimatore di m.v. di θ, si considera la funzione di verosimiglianza e si determina il suo punto di massimo.

$$L(\theta; \mathbf{x}_n) = \frac{\theta^{x_1 + \ldots + x_n}}{x_1! \ldots x_n!} e^{-n\theta} \propto \theta^s e^{-n\theta}.$$

Quindi

$$\log L(\theta; \mathbf{x}_n) = s \log \theta - n\theta,$$

e

$$\frac{d \log L(\theta; \mathbf{x}_n)}{d\theta} = 0 \Leftrightarrow \frac{s}{\theta} - n = 0$$

da cui

$$\hat{\theta}(\mathbf{x}_n) = \frac{s}{n} = \bar{x}.$$

Se, ad esempio, abbiamo un campione di 6 osservazioni che provengono da una distribuzione di Poisson:

$$x_1 = 3, \quad x_2 = 0, \quad x_3 = 2, \quad x_4 = 2, \quad x_5 = 4, \quad x_6 = 1$$

si ha che

$$\bar{x} = \frac{s}{n} = \frac{12}{6} = 2$$

è la stima di θ.

Principio del Campionamento Ripetuto e Impostazione Frequentista dell'Inferenza Statistica

In generale l'idea è quindi che uno stimatore $\hat{\theta}(X_1, \ldots, X_n)$ di θ venga scelto e utilizzato in funzione delle proprietà della sua distribuzione di probabilità, detta *distribuzione campionaria*.

L'idea intuitiva è che lo stimatore è "valido" se, con riferimento all'insieme di tutti i valori che può assumere, con buona probabilità fornisce valori vicini al valore vero e incognito di θ.

Detto in altro modo (un po' brutale): lo stimatore è "valido" se, ripetendo l'esperimento che porta all'osservazione dei dati un numero elevatissimo di volte (ipoteticamente all'infinito), i corrispondenti valori dello stimatore sono nel complesso vicini al valore vero di θ.

Quanto illustrato costituisce, in pratica, il cosidetto **Principio del Campionamento Ripetuto**, alla base dell'inferenza statistica classica.

Vediamo meglio ora cosa sono le distribuzioni campionarie e a quali aspetti specifici di queste siamo interessati nello scegliere uno stimatore.

Abbiamo detto che uno stimatore è una v.a., in quanto funzione di $X_1, \dots, X_n$, che sono v.a.

Pertanto, anche lo stimatore ha una sua distribuzione (densità) di probabilità, e possiamo quindi determinarne alcune caratteristiche, tra cui, ad esempio valore atteso e varianza.

Esempio. Riprendendo in esame l'esempio bernoulliano con $n = 4$, si ha la seguente tabella:

Distribuzione Campionaria di

$$\hat{p}(X_1, \ldots, X_4) = \bar{X}_4$$

x_1	x_2	x_3	x_4	stima di p	probabilità
0	0	0	0	0	$(1-\theta)^4$
0	0	0	1	0.25	$\theta(1-\theta)^3$
0	0	1	0	0.25	$\theta(1-\theta)^3$
0	1	0	0	0.25	$\theta(1-\theta)^3$
1	0	0	0	0.25	$\theta(1-\theta)^3$
0	0	1	1	0.50	$\theta^2(1-\theta)^2$
0	1	0	1	0.50	$\theta^2(1-\theta)^2$
0	1	1	0	0.50	$\theta^2(1-\theta)^2$
1	0	0	1	0.50	$\theta^2(1-\theta)^2$
1	0	1	0	0.50	$\theta^2(1-\theta)^2$
1	1	0	0	0.50	$\theta^2(1-\theta)^2$
0	1	1	1	0.75	$\theta^3(1-\theta)$
1	0	1	1	0.75	$\theta^3(1-\theta)$
1	1	0	1	0.75	$\theta^3(1-\theta)$
1	1	1	0	0.75	$\theta^3(1-\theta)$
1	1	1	1	1	θ^4

Quindi la distribuzione di probabilità dello stimatore $\hat{p}$ risulta, in questo caso:

$\hat{p}(x_1, \ldots, x_4)$	probabilità
0	$(1-\theta)^4$
0.25	$4\theta(1-\theta)^3$
0.50	$6\theta^2(1-\theta)^2$
0.75	$4\theta^3(1-\theta)$
1	θ^4

$\diamond \diamond \diamond$

Esempio. Per l'esempio normale, in base a note proprietà della distribuzione della media campionaria, si ha che la v.a. $\bar{X}$ ha distribuzione normale con valore atteso pari a θ e varianza pari a σ^2/n:

$$\bar{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right).$$

Media Campionaria

Sia nell'esempio bernoulliano che in quello normale, abbiamo visto che gli stimatori di p e di μ sono dati dalla media campionaria $\bar{X}$.

Ricordiamo alcune proprietà di questa variabile aleatoria.

In particolare, supponendo che $X_1, \ldots, X_n$ siano i.i.d., tutte con distribuzione identica a quella della generica v.a. X, che ha valore atteso pari a $\mathbb{E}(X)$ e varianza $\mathbb{V}(X)$, osservando che

- $\mathbb{E}(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mathbb{E}(X)$
- $\mathbb{V}(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathbb{V}(X_i) = n\mathbb{V}(X)$
- $\mathbb{V}(aX_i) = a^2\mathbb{V}(X_i) = a^2\mathbb{V}(X), \qquad a \neq 0, \quad i = 1, \ldots, n$

si ha che, per la v.a. media campionaria $\bar{X}$,

$$\mathbb{E}(\bar{X}) = \mathbb{E}(X),$$

e

$$\mathbb{V}(\bar{X}) = \frac{n\mathbb{V}(X)}{n^2} = \frac{\mathbb{V}(X)}{n}.$$

Ciò ci permette di dire che:

• se $X_1, \dots, X_n$ è un campione da una popolazione **bernoulliana** di parametro p, osservando che

$$\mathbb{E}(X_i) = p \qquad \mathbb{V}(X_i) = p(1-p) \qquad\qquad i = 1, 2, \dots, n$$

allora, per lo stimatore $\hat{p}$ di p considerato,

$$\mathbb{E}(\hat{p}) = \mathbb{E}(\bar{X}) = p$$

e che

$$\mathbb{V}(\hat{p}) = \mathbb{V}(\bar{X}) = \frac{p(1-p)}{n}.$$

• se $X_1, \dots, X_n$ è un campione da una popolazione **normale** di parametro μ e σ^2, osservando che

$$\mathbb{E}(X_i) = \mu \qquad \mathbb{V}(X_i) = \sigma^2 \qquad\qquad i = 1, 2, \dots, n$$

allora, per lo stimatore $\hat{\mu}$ di μ considerato,

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}(\bar{X}) = \mu$$

e che

$$\mathbb{V}(\hat{\mu}) = \mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}.$$

La Valutazione degli Stimatori: Criterio dell'Errore Quadratico Medio

Per valutare uno stimatore, dobbiamo avere un modo per misurarne la "distanza" dal valore vero incognito θ. Possiamo considerare:

$$[\hat{\theta}(\mathbf{X}_n) - \theta]^2,$$

che, essendo funzione di $\hat{\theta}(\mathbf{X}_n)$, è una variabile aleatoria.

Una valutazione sintetica della distanza tra stimatore e parametro è:

$$MSE_\theta[\hat{\theta}(\mathbf{X}_n)] = \mathbb{E}[(\hat{\theta}(\mathbf{X}_n) - \theta)^2] = \int (t-\theta)^2 f_{\hat{\theta}}(t|\theta)dt.$$

(MSE dall'inglese *Mean Squared Error*).

- L'espressione di MSE dipende da θ;
- Dati due stimatori $\hat{\theta}_1$ e $\hat{\theta}_2$ per θ, il primo è ***migliore*** o ***più efficiente*** del secondo se

$$MSE_\theta[\hat{\theta}_1(\mathbf{X}_n)] \leq MSE_\theta[\hat{\theta}_2(\mathbf{X}_n)], \qquad \text{per ogni } \theta$$

  ovvero se la distanza media del primo stimatore dal parametro incognito è inferiore alla distanza media del secondo, qualunque sia il vero valore di θ.
- Si mostra facilmente che

$$MSE_\theta[\hat{\theta}(\mathbf{X}_n)] = \mathbb{V}[\hat{\theta}(\mathbf{X}_n)] + \mathbf{B}[\hat{\theta}(\mathbf{X}_n)]^2,$$

  dove

$$\mathbf{B}[\hat{\theta}(\mathbf{X}_n)] = \mathbb{E}[(\hat{\theta}(\mathbf{X}_n)] - \theta$$

  è detta ***distorsione*** dello stimatore $\hat{\theta}(\mathbf{X}_n)$ (dall'inglese *Bias*).

Pertanto, per un generico stimatore di θ:

• il valore dell'MSE è determinato dalla distorsione e dalla varianza dello stimatore;

• più piccoli sono questi valori, minore è il corrispondente valore dell'MSE;

• uno stimatore tale che

$$\mathbf{B}[\hat{\theta}(\mathbf{X}_n)] = \theta, \qquad \forall\theta$$

è detto stimatore ***corretto*** o ***non distorto*** per θ.

Esempi.

Da quanto visto in precedenza, possiamo allora dire che

- ***esempio bernoulliano***:

$\hat{p}$ ha una distribuzione campionaria con valore atteso pari a p: in "media" lo stimatore è uguale a p;

- ***esempio normale***:

$\hat{\mu}$ ha una distribuzione campionaria con valore atteso pari a μ: in "media" lo stimatore è uguale a μ;

ovvero che $\hat{p}$ e $\hat{\mu}$ sono stimatori ***corretti*** (o non distorti) dei rispettivi parametri p e μ.

Inoltre:

- ***esempio bernoulliano***:

$\hat{p}$ ha una distribuzione campionaria con varianza pari pari a $\frac{p(1-p)}{n}$: la variabilità dello stimatore è inversamente proporzionale alla dimensione del campione, n. Ovvero, quanto più è elevato n, tanto più la distribuzione campionaria dello stimatore è concentrata intorno a al vero valore p (qualunque questo sia).

- ***esempio normale***:

$\hat{\mu}$ ha una distribuzione campionaria con varianza pari pari a $\frac{\sigma^2}{n}$: la variabilità dello stimatore è inversamente proporzionale alla dimensione del campione, n. Ovvero, quanto più è elevato n, tanto più la distribuzione campionaria dello stimatore è concentrata intorno a al vero valore μ (qualunque questo sia).

In entrambi gli esempi si mostra che, tra tutti gli stimatori non distorti di p e μ, gli stimatori $\hat{p}$ e $\hat{\mu}$ sono quelli con varianza minima, qualunque sia la dimensione del campione e il valore vero dei parametri.

Si tratta quindi degli stimatori con minimo MSE, ovvero ***più efficienti*** tra quelli non distorti. Tali stimatori si chiamano, in genere ***ottimi non distorti*** o **UMVUE** (dall'inglese *uniformly minimum variance unbiased estimators*).

Stimatori Consistenti

Idea intuitiva: al crescere della numerosità campionaria, ci aspettiamo che uno stimatore valido tenda a fornire, con sempre maggiore probabilità, valori numerici vicini al valore incognito θ.

più tecnicamente: ci aspettiamo che, al crecere di n, la distribuzione campionaria di uno stimatore valido tenda a concentrarsi sempre piu' intorno al valore vero del parametro.

Uno stimatore consistente è uno stimatore per i quale, per n che tende a infinito, la distanza media da θ (ovvero il suo MSE) tende a 0.

- ***Esempio bernoulliano***:

$\hat{p}$ ha una distribuzione campionaria con varianza pari pari a $\frac{p(1-p)}{n}$: la variabilità dello stimatore tende a zero per $n \to +\infty$; la distribuzione campionaria dello stimatore, al crescere del numero di osservazioni campionarie, tende a concentrarsi intorno al vero valore p (qualunque questo sia).

- ***Esempio normale***:

$\hat{\mu}$ ha una distribuzione campionaria con varianza pari pari a $\frac{\sigma^2}{n}$: la variabilità dello stimatore tende a zero per $n \to +\infty$ e la distribuzione campionaria dello stimatore, al crescere del numero di osservazioni campionarie, tende a concentrarsi intorno al vero valore μ (qualunque questo sia).

si dice allora che $\hat{p}$ e $\hat{\mu}$ sono stimatori ***consistenti*** dei rispettivi parametri p e μ.

Proprietà degli stimatori

In generale, sia una v.a. X con distribuzione (densità) $f(\cdot;\theta)$, con θ parametro incognito ($\theta \in \Theta$).

Sia $X_1, \ldots, X_n$ un campione casuale relativo alla v.a. X;

Uno stimatore $\hat{\theta}$ di θ si dice **non distorto** se:

$$\mathbb{E}(\hat{\theta}) = \theta \qquad \forall\theta \in \Theta$$

Uno stimatore $\hat{\theta}$ non distorto di θ si dice **consistente** se:

$$\lim_{n\to+\infty} \mathbb{V}(\hat{\theta}) = 0 \qquad \forall\theta \in \Theta.$$

Intervalli di Confidenza

Con gli stimatori puntuali otteniamo un unico valore di stima per θ

Può sembrare più ragionevole avere un insieme di valori plausibili per θ

Inoltre: vorremmo avere una valutazione dell'affidabilità della stima che otteniamo

... intervalli di confidenza ...

IDEA: vogliamo uno STIMATORE (ovvero una variabile aleatoria) che, per ogni campione osservato ci dia un intervallo di valori di stima per θ e che, con una prefissata probabilità, contenga il valore incognito di θ.

Esempio: intervallo di confidenza per il valore atteso della v.a. normale

Abbiamo visto che se $X_1, \ldots, X_n$ è un campione da una popolazione **normale** di parametro μ e σ^2, osservando che

$$\mathbb{E}(X_i) = \mu \qquad \mathbb{V}(X_i) = \sigma^2 \qquad\qquad i = 1, 2, \ldots, n$$

allora, per lo stimatore $\hat{\mu}$ di μ considerato,

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}(\bar{X}) = \mu$$

e che

$$\mathbb{V}(\hat{\mu}) = \mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}.$$

Inoltre si ha che:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

Quindi
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$
Poichè la probabilità che una v.a. normale di media 0 e varianza 1 assuma valori tra $-z_{1-\frac{\alpha}{2}}$ e $z_{1-\frac{\alpha}{2}}$ è pari a $1-\alpha$, si ha:
$$\mathbb{P}\{-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\} = 1-\alpha$$
da cui
$$\mathbb{P}\{\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\} = 1-\alpha.$$
Quindi un intervallo di confidenza al livello $1-\alpha$ per μ è:
$$\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Esempio. Si considerano $n = 32$ persone e si calcola l'aumento delle pulsazioni cardiache dopo un esercizio. La media di tali osservazioni è pari a $\bar{x} = 26.2$. Si ipotizza che le osservazioni seguano la legge normale di media incognita μ e $\sigma = 5.15$. Un intervallo di confidenza per μ (numero medio di pulsazioni in più indotto dall'esercizio considerato) al livello 0.95 è quindi:

$$26.2 - z_{1-\frac{\alpha}{2}} \frac{5.15}{\sqrt{32}} \leq \mu \leq 26.2 + z_{1-\frac{\alpha}{2}} \frac{5.15}{\sqrt{32}}.$$

Osservando che

$$z_{0.975} = 1.96$$

si ottiene

$$26.2 - 1.96 \frac{5.15}{\sqrt{32}} \leq \mu \leq 26.2 + 1.96 \frac{5.15}{\sqrt{32}},$$

ovvero

$$24.4 \leq \mu \leq 28.0$$

NOTA BENE

Un intervallo di confidenza, ad esempio al 95%, va interpretato in questo modo: se si considerano tutti i possibili campioni di ampiezza n, e per ciascuno si determinano i corrispondenti intervalli centrati sulla media aritmetica $\pm\sigma/\sqrt{n}$, il 95% di questi contiene il valore atteso μ di X, e solo il 5% non lo contiene.

Pertanto: osservato uno specifico campione, non sappiamo se questo contiene il vero valore di μ. Tuttavia possiamo affermare che l'intervallo osservato proviene da un insieme di intervalli in cui il 95% contiene il valore μ. Questo è il senso corretto del termine "confidenza".

NON è quindi esatto dire, nel precedente esempio, che il valore di μ cade, con probabilità 0.95, nell'intervallo $(24.4, 28)$.

Se σ è incognito:

$$\bar{X} \pm t_{n-1;1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

dove $t_{n-1;1-\frac{\alpha}{2}}$ è il percentile di livello $1 - \frac{\alpha}{2}$ della distributione T di Student con $n-1$ gradi di libertà e S la varianza campionaria