

Il concetto di probabilità viene frequentemente impiegato nella vita quotidiana da persone la cui attività hanno ben poco a che fare con la teoria delle probabilità e con la statistica.

Due discipline studiano i fenomeni casuali o aleatori: la teoria delle probabilità e la statistica.

La teoria delle probabilità:

- Approfondisce il significato filosofico ed epistemologico che viene attribuito al concetto di probabilità
- Costruisce dei modelli matematici per lo studio dei fenomeni aleatori o casuali e sviluppa le conseguenze logico-deduttive che derivano dall'applicazione dei modelli.

Statistica

I risultati e gli schemi interpretativi proposti dalla teoria delle probabilità vengono utilizzati dall'inferenza statistica che basandosi su di essa, va oltre integrandola e perfezionandola.

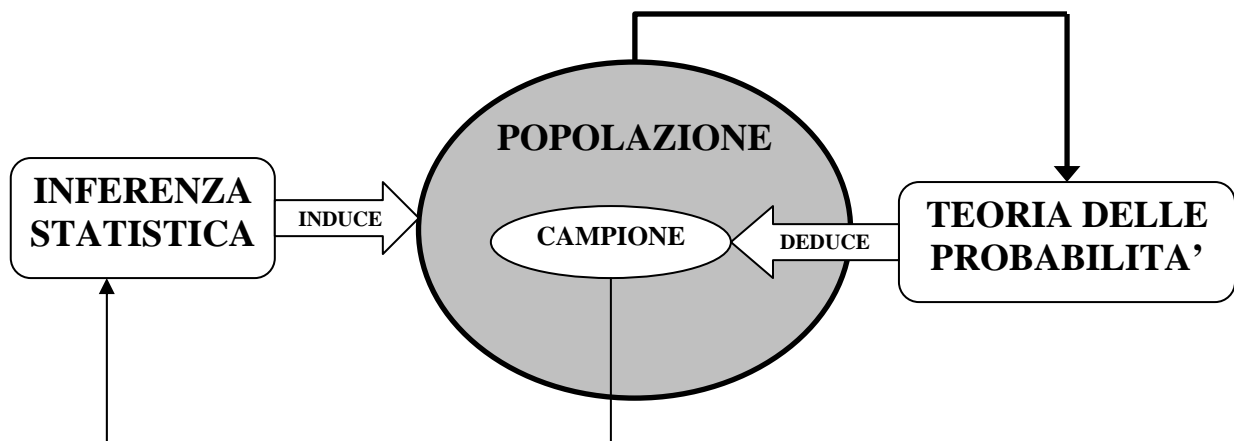
Infatti quando ci si trova di fronte a dati reali, a risultati empirici, si utilizza la teoria statistica per giungere ad una scelta tra i modelli matematici alternativi che possono aver generato quei dati.

Quindi mentre la teoria delle probabilità stabilisce i risultati che si possono attendere dall'esecuzione di un esperimento l'inferenza statistica si serve dei risultati dell'esperimento per cercare di costruire o interpretare la legge che sta dietro ai risultati sperimentali ottenuti.

In definitiva:

la teoria delle probabilità deduce dal contenuto noto della popolazione il contenuto probabile del campione (cioè deduce le proprietà di un processo fisico da un modello matematico)

l'inferenza statistica induce le caratteristiche della popolazione dall'analisi del contenuto del campione osservato cioè inferisce le proprietà del modello matematico a partire dall'analisi dei dati campionari che sono stati osservati



DEFINIZIONE DI PROBABILITA'

È difficile dare una interpretazione e quindi una definizione, di probabilità che sia completamente soddisfacente ed esente da critiche. Due sono gli approcci più frequentemente impiegati.

DEFINIZIONE OGGETTIVISTA O FREQUENTISTA DI PROBABILITÀ

Basata sul concetto che “sebbene non sia prevedibile ciò che accadrà in una singola prova, in quanto il risultato può essere uno qualunque tra i valori possibili di un insieme di risultati, si è però in grado di conoscere ciò che accadrà in un numero elevato di prove”

Probabilità è un modo formale di esprimere la proporzione di volte che un determinato evento può avere luogo in un numero elevato di prove o esperimenti. Le frequenze relative osservate in un numero molto elevato di prove possono così essere considerate come un'approssimazione della probabilità.

La caratteristica di tale interpretazione è che in essa le probabilità vengono determinate empiricamente.

Le critiche che si muovono a questa impostazione risiedono nella dipendenza delle probabilità dalle osservazioni e nella non sempre possibile ripetibilità di un esperimento nelle stesse identiche condizioni.

DEFINIZIONE SOGGETTIVISTA O PERSONALE

In base a tale interpretazione la probabilità viene definita come una misura del grado di fiducia che una persona ripone sul verificarsi di un dato fenomeno, avendo a disposizione determinate informazioni sul fenomeno stesso (la probabilità rappresenta ciò che si è disposti a scommettere contro o a favore della realizzazione di un certo evento.

Il calcolo delle probabilità da una parte e la statistica dall'altra hanno a che fare con le accidentalità e le regolarità dei processi che presentano elementi aleatori o casuali. Spesso si ha a che fare con tale variabilità ma spesso non si hanno dati e informazioni che permettono un'analisi rigorosa. Per accumulare dati, e quindi una migliore conoscenza del processo che genera il fenomeno oggetto di studio, vengono effettuati degli esperimenti.

Definizione di esperimento

Un esperimento è ogni operazione il cui risultato non può essere previsto con certezza

Definizione di evento

Ogni risultato possibile di un esperimento è un evento

Un evento può essere costituito da uno solo tra tutti i possibili risultati “punto campionario” o da un insieme di risultati omogenei rispetto ad una caratteristica

Definizione di spazio campionario Ω di un esperimento o spazio degli eventi

Lo spazio campionario è l'insieme di tutti i risultati per l'esperimento stesso

Attenzione lo stesso esperimento può dar luogo a spazi campionari diversi in funzione dei risultati che interessano

Lo spazio campionario Ω , o spazio fondamentale, è l'insieme di tutti i risultati dell'esperimento stesso

Se per un esperimento si dispone di:

- a) l'insieme dei risultati possibili
- b) la probabilità della realizzazione dei risultati ottenibili con l'esperimento



Si dispone si dispone del modello probabilistico per l'esperimento

Al fine di poter meglio sviluppare il concetto di evento e di spazio campionario qui di seguito vengono richiamati alcuni elementi di teoria degli insiemi (stretta relazione fra teoria degli insiemi e algebra degli eventi – analogia tra eventi e insiemi)

Rappresentazione	TEORIA DEGLI INSIEMI	ALGEBRA DEGLI EVENTI
\emptyset	Insieme nullo	Evento impossibile
$A \cup B$ ($A \circ B$ eventi)	somma logica o unione di due insiemi A e B	la somma logica di due eventi A e B è l'evento che si verifica quando si verificano uno almeno degli eventi
$A \cap B$ ($A \text{ e } B$)	prodotto logico o intersezione di due insiemi A e B	il prodotto logico di due eventi A e B è l'evento che si verifica se e solamente se si verificano entrambi i fattori del prodotto
$A \cap B = \emptyset$	INSIEMI DISGIUNTI due insiemi A e B sono disgiunti quando non hanno nessun elemento in comune, cioè la loro intersezione è vuota	EVENTI INCOMPATIBILI due eventi A e B sono incompatibili se non è possibile che si verifichino entrambi ossia l'evento $A \cap B$ è impossibile

Due eventi A e B sono **esaustivi** quando è impossibile che non se ne verifichi nessuno.

$A \times B =$ prodotto cartesiano di A e B è l'insieme di tutte le possibile coppie (a_i, b_j) dove $a_i \in A$ e $b_j \in B$ cioè $A \times B = \{(a_i, b_j) : a_i \in A \text{ e } b_j \in B\}$

TEORIA GENERALE DELLE PROBABILITÀ

La formalizzazione è avvenuta solo recentemente

Impostazione assiomatica

Nel 1933 A. N. Kolmogorov ha presentato un'impostazione della teoria delle probabilità strettamente connessa alla teoria delle funzioni matematiche e alla teoria degli insiemi.

Si considera un insieme Ω di eventi E_i , elementari o primitivi, e si precisa sotto quale aspetto si utilizza l'esperimento casuale. Si individuano perciò degli eventi di interesse, che altro non sono che sottoinsiemi di Ω . Si definisce quindi una classe α di eventi generati a partire da sottoinsiemi di Ω e si richiede che tale classe sia sufficientemente ampia da permettere di effettuare le operazioni elementari sugli eventi casuali e che i risultati che si ottengono siano ancora elementi della classe α .

Si richiede che gli eventi abbiano una struttura algebrica: la classe α deve essere un'algebra di sottoinsiemi di Ω



1) $\Omega \subseteq \alpha$ α contiene Ω come uno dei suoi elementi

2) Se $E \in \alpha$ allora anche $\bar{E} \in \alpha$ è chiusa rispetto alla complementarità

3) Se $E_1 \in \alpha$, $E_2 \in \alpha$, , $E_n \in \alpha$ allora anche $\bigcup_{i=1}^n E_i \in \alpha$ - è chiusa rispetto

all'unione finita (additività finita)

La validità delle 1), 2) e 3) si deduce che l'unione e l'intersezione di eventi di α non portano oltre i limiti della classe α e cioè α costituisce un'algebra di eventi

Nell'applicazione della teoria delle probabilità è però necessario andare oltre un'algebra di eventi richiedendo l'additività completa cioè che la proprietà 3) sia valida per un'infinità numerabile di eventi:

4) Se $E_1 \in \alpha$, $E_2 \in \alpha$, , $E_n \in \alpha$ allora anche $\bigcup_{i=1}^{\infty} E_i \in \alpha$ - è chiusa rispetto

all'unione infinità (additività completa)

OGNI FAMIGLIA α NON VUOTA DI SOTTOINSIEMI DI Ω CHE SODDISFA LE PROPRIETÀ 1), 2), 3) E 4) VIENE DEFINITA UNA SIGMA-ALGEBRA E LA COPPIA (α, Ω) DEFINISCE UNO SPAZIO MISURABILE

ASSIOMI DELLA PROBABILITÀ

Una probabilità su una σ -algebra α di sottoinsiemi di Ω è un'applicazione di α in $[0, 1]$:

$$P: \alpha \rightarrow [0, 1]$$

Che soddisfa i seguenti assiomi

ASSIOMA DI POSITIVITÀ

LA PROBABILITÀ DI UN QUALSIASI EVENTO APPARTENENTE AD α È UNICA E NON NEGATIVA

$$0 \leq P(E) \leq 1$$

ASSIOMA DI CERTEZZA

LA PROBABILITÀ DELL'INTERO SPAZIO CAMPIONARIO Ω È UGUALE ALL'UNITÀ

$$P(\Omega) = 1$$

ASSIOMA DI UNIONE

SE E_1 E E_2 SONO DUE EVENTI DI α CHE SI ESCLUDONO A VICENDA (EVENTI INCOMPATIBILI) ALLORA:

$$E_1 \cap E_2 = \emptyset \quad P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

Esteso ad n eventi

$$E_1 \in \alpha, E_2 \in \alpha, \dots, E_n \text{ tali che } E_i \cap E_j = \emptyset \text{ per ogni } i \neq j \quad P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

GLI ASSIOMI NON PERMETTONO DI ATTRIBUIRE UN UNICO VALORE ALLA PROBABILITÀ DI UN EVENTO MA PIUTTOSTO ESPRIMONO UN INSIEME DI REGOLE FORMALI SULLA BASE DELLE QUALI È POSSIBILE ATTRIBUIRE IN MODO DEL TUTTO COERENTE DELLE PROBABILITÀ A DEGLI EVENTI

La definizione di probabilità basata sulla frequenza relativa è solamente uno dei modi possibili di attribuire la probabilità

$$P(A) = \lim_{n \rightarrow \infty} (\text{frequenza}) = \lim_{n \rightarrow \infty} \frac{N_A}{n}$$

dove N_A è il numero di volte in cui si è verificato l'evento A in n ripetizioni dell'esperimento

PROPRIETÀ DELLA PROBABILITÀ

LE CONSEGUENZE PRINCIPALI DEGLI ASSIOMI UTILIZZATI PER DEFINIRE LA PROBABILITÀ VENGONO RIPORTATE SOTTO FORMA DI PROPRIETÀ

Proprietà 1

Per ogni evento $E \subset \Omega$ si ha che $P(E) = 1 - P(\bar{E})$

Infatti $\Omega = E \cup \bar{E}$ ed inoltre $E \cap \bar{E} = \emptyset$ da cui $P(E \cup \bar{E}) = P(E) + P(\bar{E}) = P(\Omega) = 1$

Proprietà 2

$P(\emptyset) = 0$ (infatti se $E = \emptyset$ dalla 1 si ha $P(E) = 1 - P(\bar{E}) = 1 - 1 = 0$)

Proprietà 3

Se l'evento A implica l'evento B, cioè $A \subseteq B$ allora :

$$P(A) \leq P(B)$$

Infatti dalla teoria degli insiemi si può scrivere $B = A \cup (\bar{A} \cap B)$

Poiché gli insiemi alla destra del segno di uguale sono disgiunti poiché $A \cap (\bar{A} \cap B) = \emptyset$ applicando l'assioma 3) si ha: $P(B) = P(A) + P(\bar{A} \cap B)$ e quindi $P(B) \geq P(A)$ poiché per l'assioma 1) è $P(\bar{A} \cap B) \geq 0$

Proprietà 4 (regola additiva delle probabilità)

L'assioma 3) può essere generalizzato ad eventi che non sono incompatibili

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Infatti $A \cap B \neq \emptyset$ si può scrivere $B = (A \cap B) \cup (\bar{A} \cap B)$ i due insiemi alla destra del segno uguale sono disgiunti $(A \cap B) \cap (\bar{A} \cap B) = \emptyset$ per cui per l'assioma 3) si ha

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) \Rightarrow P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Inoltre

$A \cup B = A \cup (\bar{A} \cap B)$ i due insiemi alla destra del segno uguale sono disgiunti $A \cap (\bar{A} \cap B) = \emptyset$ per cui per l'assioma 3) si ha $P(A \cup B) = P(A) + P(\bar{A} \cap B)$

Sostituendo in quest'ultima l'espressione precedente si ha:

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) = P(A) + P(B) - P(A \cap B)$$

PROBABILITÀ CONDIZIONATA

La probabilità di un evento può dipendere dalle circostanze sotto le quali la prova, o l'esperimento, vengono condotti. Queste circostanze sono esse stesse degli eventi e sono chiamate elementi condizionanti.

Si consideri lo spazio campionario Ω e i due eventi A e B tra loro compatibili. Se si suppone che B sia l'evento condizionante ne segue che tutti gli altri punti di Ω , che non sono anche punti di B, non interessano; ciò equivale a fare riferimento a un nuovo spazio campionario, ridotto rispetto ad Ω , che è uguale esattamente a B. Si indica questo nuovo spazio campionario con Ω^* e la parte di A appartenente ad Ω^* con A^* .

La nozione usata per indicare la probabilità dell'evento A condizionata dall'evento B è $P(A|B)$ che si legge "probabilità di A dato B"

Affinché l'assioma 2) sia rispettato la probabilità di Ω^* dovranno essere pari a 1 quindi le probabilità definite su Ω e su Ω^* saranno tra loro diverse).

$$P(B) = P(B|\Omega) \leq P(B|\Omega^*) = P(B|B) = 1$$

Per soddisfare questa condizione è necessario maggiorare le probabilità dei punti campionari che compongono B in Ω , indicate con p_i moltiplicandole per la costante $1/P(B)$. In tal caso le probabilità dei punti campionari di B in Ω^* saranno date da :

$$p_i^* = p_i \times (1/P(B))$$

Infatti

$$P(B|\Omega^*) = \sum_{i \in B} p_i^* = \frac{1}{P(B)} \sum_{i \in B} p_i = \frac{P(B)}{P(B)} = 1$$

Per avere la probabilità di un qualunque evento in Ω^* , cioè la probabilità condizionata da B, basterà sommare le probabilità p_i^* dei punti campionari dello spazio Ω^* che appartengono all'evento stesso. Nel caso di un evento A i punti di A compresi in Ω^* sono quelli che costituiscono l'evento $(A \cap B)$, quindi :

$$P(A|B) = \sum_{i \in (A \cap B)} p_i^* = \frac{1}{P(B)} \sum_{i \in (A \cap B)} p_i = \frac{P(A \cap B)}{P(B)}$$

con $P(B) \neq 0$

Definizione

La probabilità di A dato B, indicata con $P(A|B)$, è uguale alla probabilità congiunta A e B divisa per la probabilità dell'evento B

Proprietà 5 (regola additiva delle probabilità)

Il teorema delle probabilità composte, o regola moltiplicativa della probabilità

$$P(A \cap B) = P(A|B) * P(B) = P(A) * P(B|A)$$

Proprietà 6 (regola additiva delle probabilità)

Se A e C sono due eventi tra loro incompatibili allora:

$$P(A \cup C | B) = P(A|B) + P(C|B)$$

Infatti applicando la definizione di probabilità condizionata ed applicando la proprietà distributiva:

$$P(A \cup C | B) = \frac{P[(A \cup C) \cap B]}{P(B)} = \frac{P[(A \cap B) \cup (C \cap B)]}{P(B)}$$

Poiché $A \cap C = \emptyset$ allora anche $(A \cap C) \cap (C \cap B) = \emptyset$ da cui segue applicando l'assioma 3):

$$\frac{P[(A \cap B) \cup (C \cap B)]}{P(B)} = \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} = P(A|B) + P(C|B)$$

DEFINIZIONE DI INDIPENDENZA

Non sempre il fatto di sapere che si è verificato un certo evento modifica le circostanze sotto le quali se ne verifica un altro. Quando ciò non si verifica gli eventi sono tra loro indipendenti in probabilità o stocasticamente indipendenti.

Gli eventi A e B entrambi appartenenti ad Ω sono stocasticamente indipendenti cioè $A \perp B$ se:

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

Infatti

$$P(A) = \sum_{i=1}^n P(a_i) \quad \text{e} \quad P(B) = \sum_{i=1}^m P(b_i)$$

$$P(A \cap B) = \sum_{i=1}^n \sum_{j=1}^m P(a_i \cap b_j) = \sum_{i=1}^n P(a_i) \cdot \sum_{j=1}^m P(b_j) = P(A) \cdot P(B)$$

La condizione di indipendenza stocastica tra due eventi A e B equivale al realizzarsi delle relazioni:

$$P(A | B) = P(A) \quad \text{e} \quad P(B | A) = P(B)$$

TEOREMA DI BAYES

È una diretta applicazione della probabilità condizionata.

Si suppone che gli eventi (E_i) , con $i=1, 2, \dots, n$ formino un sistema completo di eventi, cioè un insieme di eventi incompatibili tali che $\bigcup_{i=1}^n E_i = \Omega$ e quindi una partizione finita. In tal caso la

probabilità di un qualunque evento $A \subset \Omega$ può essere definita facendo riferimento alle probabilità dei singoli eventi E_i cioè:

$$P(A) = \sum_{i=1}^n [P(A) \cap P(E_i)] = \sum_{i=1}^n P(E_i) \cdot P(A | E_i)$$

L'interpretazione più immediata e interessante di questa struttura consiste nel considerare gli eventi E_i come cause che determinano l'evento A. Sorge così il problema di trovare una relazione che permetta di calcolare la probabilità che sia stato l'evento E_i a determinare A, dato che si è certi che l'evento A si è verificato.

Applicando la formula della probabilità condizionata è possibile scrivere:

$$P(E_i | A) = \frac{P(E_i \cap A)}{P(A)} \quad \text{con } P(A) \neq 0 \quad \text{da cui} \quad P(E_i \cap A) = P(E_i) \cdot P(A | E_i)$$

è quindi immediato scrivere la formula nota come teorema di Bayes

$$P(E_i | A) = \frac{P(E_i) \cdot P(A | E_i)}{\sum_{i=1}^n P(E_i) \cdot P(A | E_i)}$$

Gli n eventi E_1, \dots, E_n sono le n cause fra loro diverse che possono generare l'evento A. Le probabilità di queste cause $P(E_1), \dots, P(E_n)$ sono considerate le "probabilità a priori" di ogni causa E_i (con $i=1, \dots, n$); esse non dipendono dal risultato empirico dell'evento A e riflettono il grado di conoscenza "soggettiva" sulle singole cause. Le probabilità condizionate $P(A | E_i)$ sono le "probabilità probative o verosimiglianze" e rappresentano la probabilità con cui le singole cause E_1, \dots, E_n generano l'evento A. Le probabilità $P(E_i | A)$ (con $i=1, 2, \dots, n$) sono le "probabilità a posteriori" delle cause E_i cioè sapendo che l'evento A si è verificato esse dicono con quale probabilità l'evento E_i ha agito nel determinare A.

ESEMPIO DI APPLICAZIONE DEL TEOREMA DI BAYES - Test sul tasso di alcolismo

Il test ha fornito come risultato che il 2% degli automobilisti guida in stato di ebbrezza.

Esperimenti effettuati hanno appurato che nel 95% dei casi il test da esito positivo quando la persona è effettivamente ubriaca così come nel 95% dei casi il test da esito negativo con persone non ubriache.

Qual è la probabilità che una persona sia realmente in stato di ebbrezza se l'alcool test da risultato positivo?

Se E è l'evento ubriaco e \bar{E} è l'evento non ubriaco ed A evento test positivo e B evento test negativo

$$P(E)=0.02 \quad P(\bar{E})=0.98$$

$$P(A|E)=0.95 \quad P(B|E)=0.05 \quad P(A|\bar{E})=0.05 \quad P(B|\bar{E})=0.95$$

Applicando il teorema di Bayes si ha:

$$P(E|A) = \frac{P(E) \cdot P(A|E)}{P(E) \cdot P(A|E) + P(\bar{E}) \cdot P(A|\bar{E})} = \frac{0.95 \cdot 0.02}{0.02 \cdot 0.95 + 0.98 \cdot 0.05} = 0.28$$

Il risultato non è certo soddisfacente e dipende dalla bassa probabilità a priori che una persona risulti ubriaca $P(E)$. Per migliorare la bontà del test bisognerebbe aumentare la performance del test se $P(B|\bar{E})=0.99$ si avrebbe:

$$P(E|A) = \frac{0.95 \cdot 0.02}{0.02 \cdot 0.95 + 0.98 \cdot 0.01} = 0.66$$

VARIABILE ALEATORIA

Per trattare gli eventi con strumenti matematici e costruire modelli suscettibili di applicazioni pratiche è necessario associare a ogni evento semplice di uno spazio delle prove un numero mediante una particolare legge.

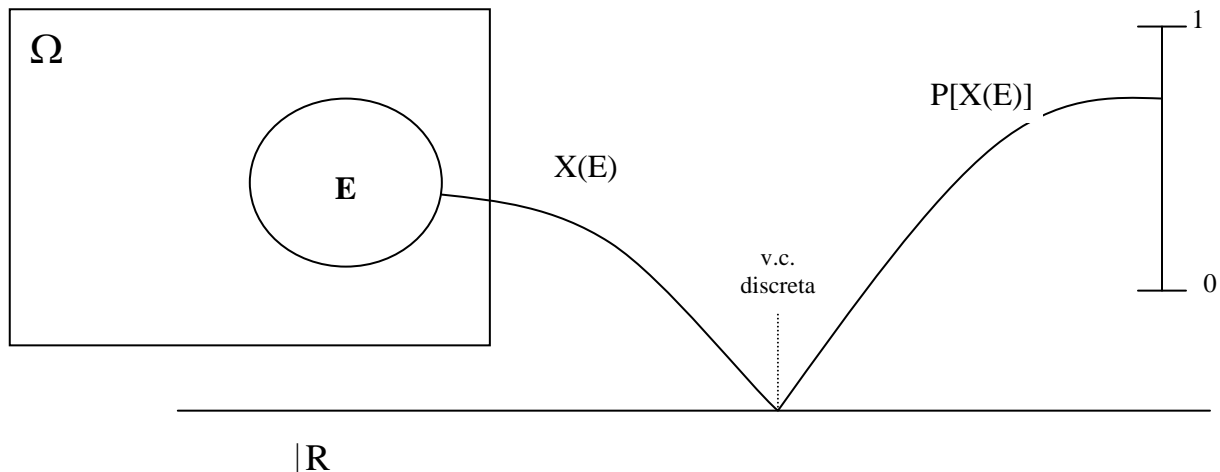
L'operazione che fa corrispondere a ciascun punto dello spazio delle prove probabilizzato un numero è analoga a quella con cui si costruisce una funzione.

Definire una variabile aleatoria (o casuale) significa quindi trovare una regola in base alla quale associare un numero reale ad ogni risultato di un esperimento e quindi ad ogni elemento dello spazio campionario Ω .

Definizione: Una v.a. X è una variabile che assume valori nello spazio dei numeri reali secondo una funzione di probabilità

X è una v.a. discreta se assume un numero finito di valori o un'infinità numerabile di valori con probabilità $p(x)$

X è una v.a. continua se assume un numero infinito di valori compresi entro un intervallo di ampiezza finita o infinita.



(differenza tra variabile matematica e variabile aleatoria è che la prima può assumere un qualsiasi valore dell'insieme di definizione mentre la seconda assume i valori con una certa probabilità)

FUNZIONI DI PROBABILITÀ

FUNZIONE DI PROBABILITÀ' = RELAZIONE TRA VARIABILE ALEATORIA E PROBABILITÀ CORRISPONDENTE

Può presentarsi sotto diverse forme:

- Funzione di distribuzione
- Funzione di Probabilità
- Funzione densità di probabilità

FUNZIONE DI DISTRIBUZIONE

La funzione di distribuzione $F_X(x)$ della variabile aleatoria X è una funzione della variabile reale t definita nell'intervallo $]-\infty, +\infty[$ la quale fornisce la probabilità cumulata che la v.a. X assuma un qualsiasi valore minore o uguale di t :

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

- Assume valori nell'intervallo $[0, 1]$;
 - Tende a 0 per $t \rightarrow -\infty$ e a 1 per $t \rightarrow +\infty$
 - Per $a \leq b$ risulta $F_X(a) \leq F_X(b)$ funzione monotona non decrescente
- $$F_X(b) - F_X(a) = P(a \leq X \leq b)$$

Ha una forma a scalini per v.a. discrete mentre è continua e derivabile per v.a. continue

FUNZIONE DI PROBABILITÀ

La funzione di probabilità $p_X(x_i)$ di una v.a. discreta X è la funzione di una variabile reale che assume valori diversi da zero solo in corrispondenza dei valori assunti dalla v.a. ed è uguale a zero in tutti gli altri punti:

$$p_X(x_i) = P(X = x_i)$$

Proprietà

$$0 \leq p_X(x_i) \leq 1$$

$$\sum_{\forall i} p_X(x_i) = 1$$

$$F_X(t) = \sum_{x_i \leq t} p(x_i)$$

FUNZIONE DENSITÀ DI PROBABILITÀ

La funzione di probabilità di una v.a. continua X “ $f_X(x)$ “ è la derivata della funzione di distribuzione $F_X(x)$

$$f_X(t) = \frac{dF_X(x)}{dx}$$

Proprietà

$$f_X(x) \geq 0 \quad \int_{-\infty}^{+\infty} f_X(x) dx = 1 \quad F_X(x_i) = \int_{-\infty}^{x_i} f_X(x) dx$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$$

Per il teorema della media

$$P(x \leq X \leq x + \Delta x) = f_X(x) \Delta x$$

Data una v.a. X distribuita con una legge densità di probabilità $f_X(x)$, la v.a. $Y=H(x)$ (funzione della v.a. X) avrà una funzione densità di probabilità data da:

$$f_Y(y) = f_X(x) \cdot \left| \frac{dx}{dy} \right|$$

VALORI CARATTERISTICI DELLE VARIABILI ALEATORIE

SPERANZA MATEMATICA O MEDIA

Si definisce speranza matematica o media $E[X]$ (spesso indicata con μ_X) di una v.a.:

$$E[X] = \mu_X = \begin{cases} \sum_{\forall i} x_i \cdot p_X(x_i) \\ \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \end{cases}$$

Se $Y=g(X)$ è una funzione deterministica della v.a. X , Y è a sua volta una v.a. – la cui speranza matematica è:

$$E[g(X)] = \begin{cases} \sum_{\forall i} g(x_i) \cdot p_X(x_i) \\ \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) dx \end{cases}$$

MOMENTI DI UNA VARIABILE ALEATORIA

Si definisce momento k -esimo ($k=1, 2, \dots$) della v.a. X rispetto all'origine ($x=0$) la speranza matematica della sua potenza k -esima:

$$m_k = E[X^k] = \begin{cases} \sum_{\forall i} x_i^k \cdot p_X(x_i) \\ \int_{-\infty}^{+\infty} x^k \cdot f_X(x) dx \end{cases}$$

Il momento di ordine 1 coincide con la media.

VARIANZA E DEVIATIONE STANDARD DI UNA VARIABILE ALEATORIA

La varianza $V(X)$ (indicata spesso con σ_X^2) della v.a. X è la speranza matematica della funzione $(x-\mu_X)^2$:

$$\begin{aligned} \text{Var}(X) &= \sigma_X^2 = E[(x-\mu_X)^2] = E[x^2] - 2 E[X] \mu_X + \mu_X^2 = E[x^2] - 2 \mu_X^2 + \mu_X^2 = \\ &= E[x^2] - \mu_X^2 \end{aligned}$$

Deviazione standard

$$\sigma_X = \sqrt{\sigma_X^2}$$

Dissimetria o Skewness

$$\mu_{X3} = E[(x-\mu_X)^3]$$

Coefficiente di variazione

$$C_X = \sigma_X / \mu_X \quad (\text{numero puro ma non invariante assoluto i.e. varia al variare della scala e dell'origine})$$

Indice relativo di dissimetria

$$\delta_{X3} = \mu_{X3} / \sigma_X^3 \quad (\text{invariante assoluto})$$

Indice di Curtosi o Disnormalità

$$\alpha_{X4} = \frac{\mu_{X4}}{\sigma_X^4} \quad (\text{pari a 3 per la distribuzione normale, } \alpha_{X4} < 3 \text{ distribuzioni iponormali,}$$

$\alpha_{X4} > 3$ distribuzioni ipernormali)

VARIABILI ALEATORIE CONGIUNTE

Nella maggior parte dei casi, la rappresentazione numerica degli eventi, come risultato di una prova, avviene in termini di due o più variabili aleatorie (più v.a. possono essere associate ai risultati di uno stesso esperimento)

Più variabili aleatorie definite sullo stesso spazio delle prove si dicono congiunte (o congiuntamente distribuite)

p.e. variabili bivariate X e Y

FUNZIONE DI PROBABILITÀ CONGIUNTA

$$p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$$

Proprietà

$$0 \leq p_{X,Y}(x_i, y_j) \leq 1$$

$$\sum_{\forall x} \sum_{\forall y} p_{X,Y}(x_i, y_j) = 1$$

FUNZIONE DENSITÀ DI PROBABILITÀ CONGIUNTA

La funzione di probabilità di una v.a. discreta $f_X(t)$ di una v.a. continua X è la derivata della funzione di distribuzione $F_X(t)$

$$f_{X,Y}(x, y) = \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P[x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y]}{\Delta x \Delta y}$$

FUNZIONE DI DISTRIBUZIONE CONGIUNTA

$$F_{X,Y}(x_i, y_j) = P[X \leq x_i, Y \leq y_j]$$

$$F_{X,Y}(x_k, y_h) = \sum_{x_i \leq x_k} \sum_{y_j \leq y_h} p_{X,Y}(x_i, y_j) \quad F_{X,Y}(x_k, y_h) = \int_{-\infty}^{x_k} \int_{-\infty}^{y_h} f_{X,Y}(x, y) \cdot dx dy$$

FUNZIONE DI PROBABILITÀ MARGINALE

$$p_X(x_i) = P[X = x_i] = \sum_{\forall y} p_{X,Y}(x_i, y_j) \quad f_X(x) = \int_{R_Y} f_{X,Y}(x, y) dy$$

MEDIA E VARIANZA DI VARIABILI ALEATORIE CONGIUNTE

$$E[H(X, Y)] = \sum_{\forall x_i} \sum_{\forall y_i} H(x_i, y_i) \cdot p_{X,Y}(x_i, y_i)$$

$$E[H(X, Y)] = \int_{R_X} \int_{R_Y} H(x, y) \cdot f_{X,Y}(x, y) \cdot dx \cdot dy$$

MOMENTO MISTO di ordine h k

$$E[X^h, Y^k] = \sum_{\forall x_i} \sum_{\forall y_j} x_i^h \cdot y_j^k \cdot p_{X,Y}(x_i, y_j)$$

$$E[X^h, Y^k] = \int_{R_X} \int_{R_Y} x_i^h \cdot y_j^k \cdot f_{X,Y}(x_i, y_j) \cdot dx \cdot dy$$

COVARIANZA

È la speranza matematica della funzione $H(X, Y) = (X - \mu_X) \cdot (Y - \mu_Y)$

$$\sigma_{XY} = E[(X - \mu_X) \cdot (Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

COEFFICIENTE DI CORRELAZIONE

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \quad (\text{assume valori compresi tra } -1 \text{ e } +1 \text{ - valore } -1 \text{ o } +1 \text{ se c'è un legame funzionale})$$

PROBABILITÀ CONDIZIONATA E VARIABILI ALEATORIE INDIPENDENTI

Funzioni di probabilità e densità di probabilità condizionata

$$p_{Y/X}(y/x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \quad f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Variabili indipendenti

$$p_{Y/X}(y/x) = p_Y(y) \quad f_{Y/X}(y/x) = f_Y(y) \Rightarrow \text{v.a. indipendenti}$$

Dette $H(X)$ e $G(Y)$ due funzioni deterministiche delle v.a. indipendenti X e Y

$$E[H(X) \cdot G(Y)] = E[H(X)] \cdot E[G(Y)]$$

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

$$E[X^i \cdot Y^j] = E[X^i] \cdot E[Y^j]$$

Da cui

$$\sigma_{XY} = E[XY] - \mu_X \mu_Y = 0 \quad (\text{se } X \text{ e } Y \text{ sono v.a. indipendenti la covarianza è nulla})$$

PROPRIETÀ DELLA MEDIA E DELLA VARIANZA

$$E[ax + b] = \sum_i (ax + b) \cdot p(x_i) = a \sum_i x_i \cdot p(x_i) + b \sum_i p(x_i) = a \cdot E[X] + b$$

MEDIA DELLA SOMMA DI 2 o PIU' v.a.

$$\begin{aligned} E[X + Y] &= \sum_i \sum_j (x_i + y_j) \cdot p(x_i, y_j) = \sum_i x_i \cdot \sum_j p(x_i, y_j) + \sum_j y_j \cdot \sum_i p(x_i, y_j) = \\ &= \sum_i x_i \cdot p(x_i) + \sum_j y_j \cdot p(y_j) = E[X] + E[Y] \end{aligned}$$

Se X è una v.a. e a e b sono costanti:

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$\text{Infatti } \text{Var}(aX + b) = E[(aX + b) - E[(aX + b)]]^2 = E[aX + b - a\mu_X - b]^2 = E[aX - a\mu_X]^2 = E[a^2(X - \mu_X)^2] = a^2 E[(X - \mu_X)^2] = a^2 \text{Var}(X)$$

VARIANZA DELLA SOMMA DI 2 v.a.

Se X e Y sono due v.a. la varianza della loro somma è:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + \sigma_{XY}$$

$$\begin{aligned} \text{Infatti } \text{Var}(X + Y) &= E[(X + Y) - E[X + Y]]^2 = E[(X - E[X]) + (Y - E[Y])]^2 = \\ &= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2 E[(X - E[X]) \cdot (Y - E[Y])] = \sigma_X^2 + \sigma_Y^2 + 2 \cdot \sigma_{XY} \end{aligned}$$

$$(\text{SE } X \text{ e } Y \text{ sono v.a. indipendenti : } \text{Var}(X + Y) = \sigma_X^2 + \sigma_Y^2)$$

REGRESSIONI

Date due v.a. X e Y si definisce regressione di Y su X la media della distribuzione condizionata di Y data X :

$$E\left[\frac{Y}{X}\right] = \sum_{\bar{y}_y} y \cdot p_{Y/X}\left(\frac{y}{x}\right) \text{ per v.a. discrete}$$

$$E\left[\frac{Y}{X}\right] = \int_{R_y} y \cdot f_{Y/X}\left(\frac{y}{x}\right) dy \text{ per v.a. continue}$$

La regressione $E[Y/X]$ è evidentemente una funzione della v.a. X . Si verifica immediatamente che la media (rispetto a X) della regressione $E[Y/X]$ coincide con la media (incondizionata) di Y :

$$E_X[E[(Y/X)]] = E[Y]$$

REGRESSIONI LINEARI

Nel caso in cui $E[Y/X]$ sia una funzione lineare di X :

$$E[Y/X] = a_y \cdot X + b_y$$

Si dimostra che:

$$a_y = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{e} \quad b_y = \mu_Y - \mu_X \frac{\sigma_{XY}}{\sigma_X^2}$$

Dove μ_X e μ_Y sono le medie (incondizionate) di X e Y .

In modo perfettamente analogo è possibile definire la regressione di X su Y .

Indipendentemente dalla forma funzionale della regressione di una v.a. Y su X la varianza della distribuzione incondizionata di Y è data da:

$$\sigma_Y^2 = E\left[(Y - E[Y/X])^2\right] + E\left[(E[Y/X] - \mu_Y)^2\right]$$

dove (con riferimento al caso di variabili aleatorie continue):

$$E\left[(Y - E[Y/X])^2\right] = \int_{R_x R_y} \{y - E[Y/X]\}^2 f_{X,Y}(x, y) \cdot dx \cdot dy$$

è denominata *varianza di Y rispetto alla regressione*, mentre,

$$E\left[(E[Y/X] - \mu_Y)^2\right] = \int_{R_x} \{E[Y/X] - \mu_Y\}^2 f_X(x) \cdot dx = \text{Var}(E[Y/X])$$

è la *varianza della regressione rispetto alla media*.

Si definisce *rapporto di correlazione* η_y il rapporto fra la varianza $\text{Var}(E[Y/X])$ della regressione e la varianza di Y (σ_Y^2):

$$\eta_y = \frac{\text{Var}(E[Y/X])}{\sigma_Y^2}$$

Se risulta $\eta_y = 1$, sarà $\text{Var}(E[Y/X]) = \sigma_Y^2$, di conseguenza è nulla la varianza rispetto alla regressione e quindi in corrispondenza di ogni valore X la Y assume un unico valore uguale a $E[Y/X]$, cioè tra le due variabili esiste un legame funzionale. Quando la regressione è lineare si ha :

$$\eta_x = \eta_y = \rho_{XY}^2$$

cioè il rapporto di correlazione coincide con il quadrato del coefficiente di correlazione.

VARIABILI CASUALI DI IMPIEGO FREQUENTE NELL'INGEGNERIA STRADALE

Variabile aleatoria BERNULLIANA

ESPERIMENTI BERNULLIANI

Due eventi S e I v.a S=1 I=0 p(1)=p P(0)= q = 1- p

$m_k = p \Rightarrow E[X_B] = 1 \cdot p + 0 \cdot q = p$

$Var(X_B) = m_2 - p^2 = p - p^2 = p \cdot (1-p) = p \cdot q$

Variabile aleatoria BINOMIALE

Esperimento ripetere n volte un esperimento bernulliano

Evento semplice = disposizione con ripetizione di due elementi su n posti

Spazio delle prove è formato da 2^n eventi semplici.

Ipotizziamo che i successivi esperimenti mantengano inalterate le loro caratteristiche nel tempo e che gli eventi siano indipendenti

$P[I, S, I, \dots, I] = p^k \cdot q^{(n-k)}$

Consideriamo gli n+1 eventi composti ciascuno dall'unione di degli eventi semplici contenenti rispettivamente 0, 1, ..., n successi

Gli eventi sono disgiunti e la loro unione è lo spazio delle prove

Il generico evento caratterizzato da k successi è formato da un numero di eventi semplici uguale al numero di combinazioni senza ripetizione $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, pertanto considerata

la probabilità di un evento con k successi e che la probabilità di un evento è dato dalla somma degli eventi semplici in esso contenuti:

$$\binom{n}{k} p^k \cdot q^{(n-k)} = \frac{n!}{k!(n-k)!} p^k \cdot q^{(n-k)}$$

Gli n+1 eventi composti possono essere considerati come eventi semplici di un nuovo esperimento, associando a ciascuno di questi eventi il valore di una v.a. uguale al numero di successi abbiamo costruito la **variabile aleatoria BINOMIALE** X_B (assume valori compresi tra 0 e n)

$$p_{X_B}(k) = \frac{n!}{k!(n-k)!} p^k \cdot q^{(n-k)}$$

Media e varianza

$E[X_b] = n \cdot p \quad Var(X_b) = n \cdot p \cdot q$

Infatti ricordando che media e varianza della variabile bernulliano sono:

$E[X_B] = p \quad Var(X_B) = p(1-p)$

Visto che:

$$E[X_b] = E\left[\sum_{i=1}^n X_B\right] = \sum_{i=1}^n E[X_B] = n \cdot p$$

$$Var(X_b) = Var\left(\sum_{i=1}^n X_B\right) = \sum_{i=1}^n Var(X_B) = n \cdot p \cdot q$$

Variabile aleatoria DI POISSON

Molte operazioni della vita reale sono riconducibili ad esperimenti Bernoulliani ripetuti. Per esempio passaggi di veicoli in una sezione stradale in un intervallo di tempo di tempo T. Dividiamo T in n intervallino $\Delta t = T/n$ ripetiamo per più intervalli T e valutiamo la media dei passaggi (successo in Δt) si ha che una stima della probabilità p è:

$$p = m/n$$

Poiché è logico pensare che aumentando n (e quindi diminuendo l'ampiezza dell'intervallino) p diminuisca, si può ipotizzare che al crescere indefinito di n il prodotto $np = \text{cost.}$ Assumendo $T=1$ si ha:

$$np = \lambda \quad \Delta t = 1/n \quad p = \lambda/n = \lambda \cdot \Delta t$$

(la probabilità di successo è proporzionale all'ampiezza Δt secondo la costante λ)

La forma limite della variabile Binomiale al tendere di $n \rightarrow \infty$ è:

$$p_p(k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

Infatti osservato che

$$\frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} = \frac{n \cdot n \cdot \left(1 - \frac{1}{n}\right) \cdot \dots \cdot \left(1 - \frac{(k-1)}{n}\right)}{k!}$$

visto che $p = \lambda/n$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} &= \lim_{n \rightarrow \infty} \frac{n \cdot n \cdot \left(1 - \frac{1}{n}\right) \cdot \dots \cdot n \left(1 - \frac{(k-1)}{n}\right)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \lim_{n \rightarrow \infty} \frac{n^k \cdot \left(1 - \frac{1}{n}\right) \cdot \dots \cdot \left(1 - \frac{(k-1)}{n}\right)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \lim_{n \rightarrow \infty} \frac{\left(1 - \frac{1}{n}\right) \cdot \dots \cdot \left(1 - \frac{(k-1)}{n}\right)}{k!} \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{1}{k!} \lambda^k \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{1}{k!} \lambda^k \cdot e^{-\lambda} \end{aligned}$$

si ricorda infatti che $\lim_{x \rightarrow \infty} \left(1 + \frac{y}{x}\right)^x = e^y$

Fornisce la probabilità di k passaggi in un intervallo T=1 se T=t si ha

$$p_p(k) = \frac{(\lambda \cdot t)^k \cdot e^{-\lambda \cdot t}}{k!}$$

$$E[X_p] = \sum_{x=0}^{\infty} x \frac{e^{-\lambda t} (\lambda \cdot t)^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-\lambda t} (\lambda \cdot t)^x}{(x-1)!} = (\lambda \cdot t) \sum_{x=1}^{\infty} \frac{e^{-\lambda t} (\lambda \cdot t)^{(x-1)}}{(x-1)!} = (\lambda \cdot t) \sum_{x=0}^{\infty} \frac{e^{-\lambda t} (\lambda \cdot t)^x}{x!} = (\lambda \cdot t) \cdot e^{-\lambda t} \sum_{x=0}^{\infty} \frac{(\lambda \cdot t)^x}{x!} = (\lambda \cdot t) \cdot e^{-\lambda t} e^{\lambda t} = \lambda \cdot t$$

$$E[X_p^2] = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda t} \cdot (\lambda \cdot t)^x}{x!} = \sum_{x=0}^{\infty} x \frac{e^{-\lambda t} \cdot (\lambda \cdot t)^x}{(x-1)!} = (\lambda \cdot t) \sum_{x=1}^{\infty} x \frac{e^{-\lambda t} \cdot (\lambda \cdot t)^x}{(x-1)!} = (\lambda \cdot t) \sum_{x=0}^{\infty} (x+1) \frac{e^{-\lambda t} \cdot (\lambda \cdot t)^x}{x!} =$$

$$= (\lambda \cdot t) \cdot \left[\sum_{x=0}^{\infty} x \frac{e^{-\lambda t} \cdot (\lambda \cdot t)^x}{x!} + \sum_{x=0}^{\infty} \frac{e^{-\lambda t} \cdot (\lambda \cdot t)^x}{x!} \right] = (\lambda \cdot t) \cdot [(\lambda \cdot t) + 1] = (\lambda \cdot t)^2 + (\lambda \cdot t)$$

$$\text{Var}(X_p) = E[X_p^2] - (E[X_p])^2 = (\lambda \cdot t)^2 + (\lambda \cdot t) - (\lambda \cdot t)^2 = (\lambda \cdot t)$$

Variabile aleatoria ESPONENZIALE

Dalla espressione della funzione di probabilità della variabile di Poisson si ricava la probabilità che non passi alcun veicolo in un intervallo t :

$$P_P(k) = \frac{(\lambda \cdot t)^k \cdot e^{-\lambda t}}{k!} \Rightarrow P[0] = e^{-\lambda t} = P[\tau \geq t]$$

da cui

$$F_\tau(t) = P[\tau \leq t] = 1 - e^{-\lambda t} \quad \text{e} \quad f_\tau(t) = \frac{dF_\tau}{dt} = \lambda \cdot e^{-\lambda t}$$

$$E[t] = \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t} dt = \left[-t \cdot e^{-\lambda t} \right]_0^\infty + \int_0^\infty e^{-\lambda t} dt = \int_0^\infty e^{-\lambda t} dt = \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^\infty = \frac{1}{\lambda}$$

$$E[t^2] = \int_0^\infty t^2 \cdot \lambda \cdot e^{-\lambda t} dt = \left[-t^2 \cdot e^{-\lambda t} \right]_0^\infty + 2 \int_0^\infty t \cdot e^{-\lambda t} dt = 2 \int_0^\infty t \cdot e^{-\lambda t} dt = \frac{2}{\lambda} \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t} dt = \frac{2}{\lambda} \cdot E[t] = \frac{2}{\lambda} \cdot \frac{1}{\lambda} = \frac{2}{\lambda^2}$$

$$\text{Var}(t) = E[t^2] - (E[t])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Variabile aleatoria di ERLANG

L'intervallo temporale che intercorre fino al verificarsi del k -esimo evento (successo in una serie infinita di esperimenti Bernulliani ripetuti) può essere considerata come una v.a. τ_k compresa tra 0 ed ∞ . La funzione di distribuzione di tale v.a. può essere ricavata a partire dall'espressione della v.a. di Poisson, infatti:

$$F_{\tau_k}(t_k) = 1 - \frac{(\lambda t)^k \cdot e^{-\lambda t}}{k!}$$

$$f_\tau(t) = \frac{dF_{\tau_k}}{dt} = \frac{\lambda}{(k-1)!} (\lambda t)^{k-1} e^{-\lambda t}$$

Si può dimostrare che la sua media e varianza valgono:

$$E[t_k] = k/\lambda \quad \text{Var}(t_k) = k/\lambda^2$$

Variabile aleatoria GAMMA

Consideriamo la funzione della variabile reale x :

$$\Gamma(x) = \int_0^\infty y^{(x-1)} e^{-y} dy \quad \text{Funzione gamma}$$

Applicando la regola di integrazione per parti:

$$\Gamma(x) = \int_0^\infty y^{(x-1)} e^{-y} dy = \int_0^\infty -y^{(x-1)} d(e^{-y}) \cdot dy = \left[-y^{(x-1)} e^{-y} \right]_0^\infty - \int_0^\infty -(x-1)y^{(x-2)} e^{-y} dy = 0 + (x-1) \cdot \Gamma(x-1)$$

Tale espressione consente di calcolare $\Gamma(x)$ per qualsiasi valore di $x > 2$ una volta noti i valori di $\Gamma(x)$ nell'intervallo $[1, 2]$, in particolare per x intero si ottiene che dato che $\Gamma(1)=1$:

$$\Gamma(x) = (x-1)!$$

Se sostituiamo nell'espressione della funzione di probabilità di Erlang la funzione $\Gamma(k)$ al posto di $(k-1)!$ otteniamo la funzione densità di probabilità della v.a. gamma:

$$f_\tau(t) = \frac{\lambda}{\Gamma(k)} (\lambda t)^{k-1} e^{-\lambda t}$$

Nella quale il parametro k può assumere un qualsiasi valore reale (ha media e varianza pari a $E[t_k]=k/\lambda$ $\text{Var}(t_k)=k/\lambda^2$)

Variabile aleatoria χ^2 (chi-quadro)

È una distribuzione gamma con $k = \nu/2$ e $\lambda = 1/2$

$$f_{\tau}(t) = \frac{\lambda}{\Gamma(k)} (\lambda t)^{k-1} e^{-\lambda t} = \frac{\lambda^k}{\Gamma(k)} (t)^{k-1} e^{-\lambda t} \Rightarrow f(\chi^2) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right)} (\chi^2)^{\frac{\nu-2}{2}} e^{-\frac{\chi^2}{2}}$$

Con $0 < \chi^2 < \infty$ e $n = 1, 2, 3, \dots$

è facile dimostrare che per le proprietà della funzione gamma si ha :

$$E[\chi^2] = k/\lambda = 2 * \nu/2 = \nu$$

$$\text{Var}(\chi^2) = k/\lambda^2 = (\nu/2) * 4 = 2 * \nu$$

Variabile aleatoria di Gumbel

Deriva dalla v.a. esponenziale

Sia $Z_M = a - b \ln(\lambda * x)$ una v.a. detta di "Gumbel dei valori massimi" (o $Z_m = a + b \ln(\lambda * x)$ detta di "Gumbel dei valori minimi") dove X è una v.a. esponenziale di parametro λ . La funzione di distribuzione di Z_M e di Z_m saranno rispettivamente:

$$F_{Z_M}(z) = e^{-e^{-\frac{z-a}{b}}} \quad \text{e} \quad F_{Z_m}(z) = 1 - e^{-e^{-\frac{z-a}{b}}}$$

con $-\infty < a, z < \infty$; $b > 0$

Variabile aleatoria di Weibull

La seguente trasformazione della v.a. di Gumbel dei valori minimi Z_m :

$$W = e^{Z_m}$$

Definisce la v.a. di Weibull la cui funzione di distribuzione risulta essere:

$$F_W(w) = 1 - e^{-\left(\frac{w}{a}\right)^{\beta}} \quad \text{con} \quad -\infty < w < \infty; \beta > 0$$

VARIABILE ALEATORIA NORMALE

È la più importante distribuzione di probabilità infatti:

- se Z è la somma di n v.a. indipendenti e identicamente distribuite al crescere di n la v.a. Z tende a una v.a. Normale (teorema del limite centrale);
- se W è la somma di n v.a. qualsiasi tra loro indipendenti tali che la varianza di ciascuna di esse sia trascurabile rispetto alla varianza di W al crescere di n la legge di probabilità di W tende a quella di una v.a. Normale avente come parametri la media e la deviazione standard di W .

Se il risultato di un esperimento è quindi dovuto ad un grande numero di cause indipendenti fra loro (p.e. errori di misura) e tali che il contributo di ciascuna sia piccolo, l'esperimento da luogo ad una variabile aleatoria normale.

La distribuzione Normale è nota anche come legge degli errori infatti Gauss la introdusse per descrivere gli errori. Sia $Z=f(u)$ la funzione di densità di probabilità degli errori (u è l'errore). Se si effettuano più misurazioni di una stessa grandezza tutte degne della stessa fiducia, il valore più probabile della grandezza è la media aritmetica delle misurazioni. La curva degli errori deve quindi:

1) Ammettere un massimo in corrispondenza dell'errore nullo:

$$\left. \frac{dz}{du} \right|_{u=0} = 0 \quad \left. \frac{d^2z}{du^2} \right|_{u=0} < 0$$

2) La probabilità di commettere un certo errore in valore assoluto è funzione decrescente dell'errore stesso:

$$\begin{array}{l} \frac{dz}{du} > 0 \quad u < 0 \\ \text{per} \\ \frac{dz}{du} < 0 \quad u > 0 \end{array}$$

3) Solo asintoticamente per $u \rightarrow -\infty$ e $u \rightarrow +\infty$ sarà $z \rightarrow 0$ non potendosi porre limiti all'operare puramente accidentale del caso

L'equazione differenziale più semplice che soddisfa i postulati prima illustrati è:

$$\frac{dz}{du} = -z \cdot u \quad \text{con } (z \geq 0, -\infty < u < +\infty)$$

$$\frac{dz}{z} = -u \cdot du \Rightarrow \int \frac{dz}{z} = \int -u \cdot du \Rightarrow \log_e z = -\frac{u^2}{2} + c$$

continua v.a. Normale

$$z = f(u) = \exp\left(-\frac{u^2}{2} + c\right) = k \exp\left(-\frac{u^2}{2}\right)$$

Affinché sia soddisfatto il secondo postulato della probabilità $\int_{-\infty}^{+\infty} f(u) du = 1$ deve essere :

$$k = \frac{1}{\sqrt{2\pi}}$$

Infatti

$$\int_{-\infty}^{+\infty} f(u) du = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{u^2}{2}} du$$

posto $\left(\frac{u}{\sqrt{2}}\right)^2 = v$ da cui $u = \sqrt{2} \cdot v^{1/2}$ quindi $du = \frac{\sqrt{2}}{2} \cdot v^{-1/2} \cdot dv$

si ha

$$\int_{-\infty}^{+\infty} f(u) du = \frac{2}{\sqrt{2\pi}} \frac{\sqrt{2}}{2} \int_0^{\infty} e^{-v} \cdot v^{-1/2} \cdot dv = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-v} \cdot v^{\left(\frac{1}{2}-1\right)} \cdot dv = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1$$

La legge di probabilità normale si può generalizzare:

$$f(u) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{u-\mu}{\sigma\sqrt{2}}\right)^2}$$

Per media $E[X_u]=\mu$ e $\text{Var}(X_u)=\sigma^2$

Spesso si fa riferimento alla v.a. Normale Standard che ha media nulla e varianza pari a 1.

È possibile passare da una variabile aleatoria Normale X ad una Normale Standard Z attraverso un cambiamento di origine e di scala, basta porre $z = (x - \mu) / \sigma$

Variabile Aleatoria Log Normale

Può considerarsi un caso particolare della v.a. Normale, rappresenta infatti il caso di una v.a. X il cui logaritmo naturale è distribuito con una legge di probabilità normale (la v.a. $Z = \ln(x)$ è distribuita come una v.a. Normale). Ricordando che:

$$f_x(x) = f_z(z) \cdot \left| \frac{dz}{dx} \right|$$

Si ha che:

$$f(x) = \frac{1}{x \cdot \sqrt{2\pi} \cdot \xi} \exp\left[-\frac{1}{2} \cdot \left(\frac{\ln(x) - \lambda}{\xi}\right)^2\right]$$

dove

λ è la media della v.a. $Z = \ln(x)$

$$\lambda = \ln\left(\frac{\mu}{\sqrt{1 + \frac{\sigma^2}{\mu^2}}}\right)$$

ξ è la deviazione standard della v.a. $Z = \ln(x)$

$$\xi^2 = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right)$$

μ è la media della v.a. x

σ è la deviazione standard della v.a. x

Variabile aleatoria di T di Student

Viene utilizzata nel test delle ipotesi relativo al confronto tra le medie di campioni casuali estratti da una popolazione Gaussiana.

Tale variabile si ottiene come rapporto tra una v.a. Normale e la radice quadrata di una v.a. χ^2 divisa per i gradi di libertà ν .

$$T = \frac{U}{\sqrt{\chi^2/\nu}}$$

Si può dimostrare che la sua media e varianza valgono:

$$E[T] = E\left\{ \frac{U}{\sqrt{\chi^2/\nu}} \right\} = E[U] \cdot E\left\{ \frac{1}{\sqrt{\chi^2/\nu}} \right\} = 0$$

$$\text{Var}(T) = E[T^2] - \{E[T]\}^2 = E\left\{ \frac{U^2}{\chi^2/\nu} \right\} = E[U^2] \cdot E\left[\frac{1}{(\chi^2/\nu)} \right] = \nu \cdot E\left[\frac{1}{\chi^2} \right] = \frac{\nu}{(\nu-2)} \text{ per } \nu > 2$$

STATISTICA

La statistica si può definire come l'applicazione dei metodi scientifici alla programmazione della raccolta di dati derivanti da osservazioni sperimentali, di natura essenzialmente quantitativa, alla loro rilevazione, spoglio, elaborazione, classificazione, analisi, sintesi e presentazione, per trarre inferenze attendibili da essi, sulle quali basare, in ultima istanza, decisioni di ordine scientifico o pratico.

Distinguiamo nella statistica diversi livelli di analisi o fasi:

- fase **descrittiva** (elaborazione, classificazione, sintesi e presentazione dei dati);
- fase **induttiva** (derivazione di inferenze attendibili);
- una fase **decisoria**

Definiamo:

- mutabile un fenomeno che può assumere diverse modalità qualitative;
- variabile un fenomeno che può assumere diverse modalità quantitative.

STATISTICA DESCRITTIVA

La successione dei dati statistici che si ottiene classificando un fenomeno secondo una variabile prende il nome di "seriazione statistica"

Le seriazioni si dicono di "frequenza" quando indicano la frequenza secondo cui si manifestano le diverse modalità.

Consideriamo un fenomeno di massa che varia di intensità al variare delle modalità quantitative continue del carattere di un altro fenomeno, detto **variabile continua** (p.e. velocità dei veicoli durante un'osservazione di durata T).

Per trattare i dati:

- 1) si raggruppano in classi;
- 2) si individuano le unità statistiche " f_i " che vengono a trovarsi in ogni classe (frequenza assoluta della classe);
- 3) si valutano le frequenze relative $f'_i = \frac{f_i}{\sum_{i=1}^k f_i}$
- 4) Si costruisce una rappresentazione grafica tramite istogramma

Se la **variabile è discreta**, essa può assumere solo un numero finito di valori discreti, per trattare i dati si procede pertanto come di seguito illustrato:

- 1) si individuano le unità statistiche " f_i " di ciascun valore (frequenza assoluta);
- 2) si valutano le frequenze relative $f'_i = \frac{f_i}{\sum_{i=1}^k f_i}$
- 3) Si costruisce una rappresentazione grafica tramite diagramma a segmenti

Abbiamo costruito una distribuzione di frequenza che è certamente più maneggevole dei dati iniziali, ma si può procedere oltre la classificazione in distribuzioni di frequenza, sintetizzando a loro volta le distribuzioni con **indici descrittivi**:

indici di localizzazione valori medi \bar{x} , **mediana** e **moda** o **norma**;

indici di variabilità **varianza** s^2 , **scarto quadratico medio** s e **coeff. di variazione**

indici di dissimetria $d = m_3 / S^3$

indice di curtosi o disnormalità m_4 / S^4

Media

$$M(x) = \bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$$

La media gode della proprietà che la somma algebrica dei prodotti degli scostamenti dalla media aritmetica stessa per la frequenza corrispondente è nulla:

$$\sum_{i=1}^n (x_i - \bar{x}) f_i = \sum_{i=1}^n x_i f_i - \bar{x} \cdot \sum_{i=1}^n f_i = \sum_{i=1}^n x_i f_i - \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \cdot \sum_{i=1}^n f_i = \sum_{i=1}^n x_i f_i - \sum_{i=1}^n x_i f_i = 0$$

Mediana

La mediana è il centro di simmetria, nel caso di una distribuzione di frequenza è quindi il valore tale che la frequenza dei valori ad esso maggiori è uguale alla frequenza dei valori ad esso minori (la somma dei valori assoluti degli scostamenti dalla mediana è un minimo rispetto alla somma dei valori assoluti degli scostamenti da un qualsiasi altro valore)

Se x_m è la mediana della distribuzione deve risultare:

$$\sum_{i \leq m} f_i' = \sum_{i \geq m} f_i' = 0.5$$

Moda o Norma

Si chiama moda o norma di una distribuzione il valore della variabile cui corrisponde un massimo relativo della densità di frequenza.

Se x_n è la moda della distribuzione deve risultare

$$f_n' \geq f_i' \quad \forall i$$

Varianza

$$S^2 = Var(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n (x_i^2 + \bar{x}^2 - 2 \cdot \bar{x} \cdot x_i) \cdot f_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n (x_i^2) \cdot f_i}{\sum_{i=1}^n f_i} + \frac{\bar{x}^2 \sum_{i=1}^n f_i}{\sum_{i=1}^n f_i} - 2 \frac{\bar{x} \cdot \sum_{i=1}^n (x_i) \cdot f_i}{\sum_{i=1}^n f_i} =$$

$$= M(X^2) + \bar{x}^2 - 2\bar{x} \cdot \bar{x} = M(X^2) - \bar{x}^2$$

Scarto Quadratico Medio (SQM)

$$S = +\sqrt{Var(x)}$$

Coefficiente di variazione $c = \frac{s}{x}$ (per la popolazione sarà $\gamma = \frac{\sigma}{\mu}$)

Indice di dissimetria

$$Dis(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 f_i}{\sum_{i=1}^n f_i}$$

Spesso si utilizza la radice cubica dell'indice $\sqrt[3]{Dis(x)}$ in quanto è espresso nella stessa unità di misura del fenomeno. (Dis(x)>0 dissimetria positiva = distribuzione spostata verso destra)

INDICI SINTETICI DI VARIABILITÀ RELATIVA

Curtosi o disnormalità

$$\frac{m_4}{S^4} = \frac{1}{S^4} \frac{\sum_{i=1}^n (x_i - \bar{x})^4 f_i}{\sum_{i=1}^n f_i}$$

STATISTICA INDUTTIVA

Tipici problemi della statistica induttiva sorgono quando si osservano v.c. le cui distribuzioni sono note in maniera incompleta (p.e. si conoscono le caratteristiche di un campione) e si vuole risalire alle caratteristiche incognite della popolazione.

Nella **statistica induttiva** possiamo distinguere le seguenti parti principali:

- 1) **Stima puntuale di un singolo parametro** (la stima è un singolo valore numerico il più prossimo possibile al valore incognito del parametro);
- 2) **Stima di intervallo di un parametro** (consiste in un intervallo entro cui si ritiene cada con alta frequenza relativa il parametro incognito);
- 3) **Criteri di verifica di ipotesi** (si deve scegliere tra diverse linee d'azione sulla scorta del valore osservato di una v.c. la cui distribuzione dipende da un parametro che se fosse noto indicherebbe la linea d'azione più appropriata);
- 4) **Analisi della varianza** (tecnica particolare per verificare ipotesi complesse)
- 5) **Programmazione degli esperimenti** (scelta la tecnica più appropriata di stima, resta da decidere quale numerosità deve avere il campione: se deve essere casuale semplice o stratificato o a più stadi o di altro tipo);
- 6) **Regressione lineare** per lo studio delle relazioni tra variabili casuali e per la stima di parametri di tali relazioni in base a dati campionari

Stima puntuale di parametri

La stima T di un parametro θ è una v.c. la cui distribuzione dipende da θ . La stima T è accettabile se si discosta poco dal valore effettivo di θ : ma T è una v.c. onde non si può essere certi a priori che essa si discosti poco da θ . Possiamo però far sì che vi sia un'alta probabilità che essa si discosti poco da θ . Perché ciò si verifichi occorre siano soddisfatte due condizioni:

- La distribuzione di T deve essere localizzata in θ cioè deve avere il valore medio $M(T)=\theta$, nel qual caso la stima si dice non affetta da errore sistematico;
- La distribuzione di T sia poco dispersa, cioè sia piccola la $Var(T)$ in modo che anche osservando il singolo valore di T del campione osservato, e non il valore medio $M(T)$ esso si discosti poco da θ .

Campionamento da una popolazione

Sia N il numero di elementi della popolazione e la quantità che interessa, cioè la v.c. X , assuma il valore x_i nell' i -esimo elemento della popolazione. La media aritmetica e la varianza di tali valori nella popolazione sono allora pari a:

$$\mu = E[X] = \frac{\sum_{i=1}^N x_i}{N} \quad \sigma^2 = Var(X) = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Estraiamo ora un campione casuale di numerosità n dalla popolazione ed indichiamo con X_j il valore j -esimo elemento estratto nel campione ($j=1, 2, \dots, n$). Siccome in una data estrazione ognuno degli N elementi della popolazione ha la stessa probabilità di essere estratto ne discende che le v.c. X_1, X_2, \dots, X_n hanno tutte la medesima distribuzione con lo stesso valore medio della popolazione e la stessa varianza della popolazione:

$$M(X_1)=M(X_2)=\dots=M(X_n)=\mu$$

$$Var(X_1)=Var(X_2)=\dots=Var(X_n)=\sigma^2$$

Si definisce campione casuale, nel caso di popolazioni infinite (o di estrazione con ripetizione) una determinazione (x_1, x_2, \dots, x_n) della n -pla di v.c. X_1, X_2, \dots, X_n tra loro indipendenti, e con la medesima distribuzione della v.c. X nella popolazione.

Si definisce invece campione casuale, nel caso di popolazioni finite (o estrazione in blocco), una determinazione (x_1, x_2, \dots, x_n) della n-pla di v.c. X_1, X_2, \dots, X_n tra loro dipendenti, tale che ogni determinazione ha la stessa probabilità di presentarsi. Se la popolazione ha numerosità N tale probabilità costante sarà:

$$\frac{1}{D_{N,n}} = \frac{1}{\binom{N}{n}} \quad (\text{numero delle disposizioni di } N \text{ elementi presi a } n \text{ a } n)$$

Vogliamo stimare la media μ in base al campione.

È naturale ricorrere alla stima data dalla v.c. (media aritmetica):

$$\bar{X} = \frac{\sum_{j=1}^n X_j}{n}$$

Per la regola del valore medio si ha (sia per v.c. indipendenti che dipendenti):

$$M(\bar{X}) = M\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{1}{n} \sum_{j=1}^n M(X_j) = \frac{1}{n} n\mu = \mu$$

onde \bar{X} è una stima non affetta da errore sistematico.

La varianza della stima dipende da metodo di estrazione (con ripetizione \Rightarrow variabili indipendenti, senza ripetizione \Rightarrow variabili dipendenti).

Nel primo caso (estrazione con ripetizione) si ha:

$$Var(\bar{X}) = Var\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{1}{n^2} \sum_{j=1}^n Var(X_j) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Nel secondo caso (estrazione senza ripetizione) si ha (ricordando che la varianza della somma di v.a. è pari alla somma delle varianze più il doppio della covarianza):

$$Var(\bar{X}) = Var\left(\frac{\sum_{j=1}^n X_j}{n}\right) = \frac{1}{n^2} \left[\sum_{j=1}^n Var(X_j) + 2 \sum_{i < j} cov(X_i, X_j) \right] = \frac{1}{n^2} \left[\sum_{j=1}^n Var(X_j) + 2 \binom{n}{2} \sigma_{ij} \right] = \frac{\sigma^2}{n} + \frac{n-1}{n} \gamma$$

Nel caso di $n=N$ si ha $\bar{X} = \mu$ onde

$$Var(\mu) = \frac{\sigma^2}{N} + \frac{N-1}{N} \gamma = 0 \Rightarrow \gamma = -\frac{\sigma^2}{N-1}$$

Quindi

$$Var(\bar{X}) = \frac{\sigma^2}{n} - \frac{n-1}{n} \frac{\sigma^2}{N-1} = \frac{\sigma^2}{n} \left(\frac{N-1-n+1}{N-1} \right) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$$

La varianza dipende da σ che è pure incognito per ricavare la numerosità del campione che soddisfi un certo grado di accuratezza bisogna prima avere informazioni preventive mediante un campione pilota (o l'esperienza passata).

Analogamente a come si è proceduto per la stima della del parametro media si può procedere alla stima della varianza. Anche qui si può inizialmente pensare di stimare la varianza in base allo scarto quadratico del campione:

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

Ma sia nel caso di estrazione senza ripetizione che di estrazione con ripetizione tale stima è affetta da un errore sistematico:

estrazione con ripetizione

$$S^2 = \frac{\sum_{j=1}^n (X_j - \bar{X} + \mu - \mu)^2}{n} = \frac{\sum_{j=1}^n (X_j - \mu)^2}{n} - (\bar{X} - \mu)^2 - \frac{2}{n} (\bar{X} - \mu) \cdot \sum_{j=1}^n (X_j - \mu) = \frac{\sum_{j=1}^n (X_j - \mu)^2}{n} - (\bar{X} - \mu)^2$$

da cui $M(S^2) = \frac{1}{n} \sum_{j=1}^n M(X_j - \mu)^2 - M(\bar{X} - \mu)^2 = \frac{1}{n} n \sigma^2 - \text{Var}(\bar{X}) = \frac{1}{n} n \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 \frac{n-1}{n} \neq \sigma^2$

estrazione senza ripetizione

$$M(S^2) = \frac{1}{n} \sum_{j=1}^n M(X_j - \mu)^2 - M(\bar{X} - \mu)^2 = \frac{1}{n} n \sigma^2 - \text{Var}(\bar{X}) = \frac{1}{n} n \sigma^2 - \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \sigma^2 \frac{n(N-1) - (N-n)}{n \cdot (N-1)} = \sigma^2 \frac{nN - n - N + n}{n \cdot (N-1)} = \sigma^2 \frac{N(n-1)}{n \cdot (N-1)} \neq \sigma^2$$

Perciò se si stima

$$S_*^2 = S^2 \frac{n}{n-1} \quad \text{nel caso di estrazione con ripetizione (variabili indipendenti)}$$

$$S_{**}^2 = S^2 \frac{N-1}{N} \frac{n}{n-1} \quad \text{nel caso di estrazione senza ripetizione (variabili non indipendenti)}$$

queste sono stime di σ^2 non affette da errore sistematico

Ma sono anche stime accurate, si può dimostrare che:

$$\text{Var}(S_*^2) = \frac{\sigma^4}{n} \left(\frac{\mu^4}{\sigma^4} - \frac{n-3}{n-1} \right)$$

Se la distribuzione della popolazione è normale si sa che:

$$\frac{m_4}{\sigma^4} = 3 \quad (\text{curtosi}) \quad \Rightarrow \quad \text{Var}(S_*^2) = \frac{\sigma^4}{n} \left(\frac{m_4}{\sigma^4} - \frac{n-3}{n-1} \right) = \frac{\sigma^4}{n} \left(\frac{3(n-1) - n + 3}{n-1} \right) = \frac{\sigma^4}{n} \left(\frac{2n}{n-1} \right) = \frac{2 \cdot \sigma^4}{n-1}$$

Allora il coefficiente di variazione è:

$$\frac{\sqrt{\text{Var}(S_*^2)}}{M(S_*^2)} = \frac{\sigma^2 \sqrt{\frac{2}{n-1}}}{\sigma^2} = \sqrt{\frac{2}{n-1}} \leq \Delta$$

Posto $\Delta=0.20$ si ha: $n \geq 51$

Nel caso di v.a. Normali si può fissare la numerosità del campione al fine di ottenere uno scarto quadratico medio che sia una certa percentuale del valore medio (p.e. 20%).

Criteri di verifica delle ipotesi

Si intende per ipotesi statistica un'ipotesi riguardo alla distribuzione, o ad i parametri della distribuzione, di una v.c. X , che si verifica sulla scorta dei valori osservati della v.c. stessa. In generale, anche se l'ipotesi è vera, esiste una probabilità positiva di tutti i valori della v.c. X , così che qualsiasi valore si osservi, può essersi verificato sotto l'ipotesi considerata, che non potrà mai essere scartata con certezza. Ma certi valori di X possono essere talmente rari nel caso l'ipotesi sia vera, da far ritenere ragionevole il rifiuto.

L'ipotesi I da verificare è detta generalmente "ipotesi nulla" perché spesso consiste nell'assumere che non c'è nulla a favore delle ipotesi contrarie.

Se l'ipotesi I deve essere rifiutata quando X assume valori estremi, poco probabili, bisogna decidere dove tracciare la linea di demarcazione tra accettazione e rifiuto di I . Se si rifiuta l'ipotesi quando $X \geq c$ e si accetta quando $X < c$ allora "c" è detto valore critico.

Se $I: p = p_0$ la probabilità di falso rifiuto nel caso di valore critico c è: $P_{p_0}(X \geq c) = \alpha$ dove P_{p_0} indica la probabilità calcolata sotto l'assunzione che l'ipotesi nulla I sia vera; α è detto livello di significatività del criterio di verifica dell'ipotesi.

ALCUNI TEST DELLE IPOTESI

CRITERI DI VERIFICA UNILATERALE E BILATERALE DI IPOTESI SU VARIABILI

IPOTESI SULLE MEDIE

Nel caso si voglia verificare l'ipotesi nulla:

$$I: \mu = \mu_0$$

Si estrae dalla popolazione un campione abbastanza grande (i.e. $n > 50$)

Detta \bar{X} la media del campione e σ lo s.q.m. (supposto noto) della popolazione, se l'ipotesi è vera:

$$U_N = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

è distribuito approssimativamente come la variabile aleatoria Normale standardizzata.

Pertanto si rifiuta l'ipotesi se:

$$U_N \geq u_\alpha \quad \text{se l'ipotesi contraria è } \mu > \mu_0$$

$$U_N < -u_\alpha \quad \text{se l'ipotesi contraria è } \mu < \mu_0$$

$$|U_N| \geq u_{\alpha/2} \quad \text{se l'ipotesi contraria è } \mu \neq \mu_0$$

$\frac{\sigma}{\sqrt{n}}$ è incognito ma viene spesso stimato come:

$$\frac{\sigma}{\sqrt{n}} \cong \frac{S}{\sqrt{n-1}}$$

IPOTESI SULLA VARIANZA

Nel caso si voglia verificare l'ipotesi nulla:

$$I: \sigma^2 = \sigma_0^2$$

In una popolazione normale

Detta S^2 la varianza di un campione di n elementi si calcola la variabile:

$$X_{\chi^2} = \frac{n \cdot S^2}{\sigma_0^2}$$

tale variabile deve avere distribuzione χ^2 con $v=n-1$ gradi di libertà.

Pertanto si rifiuta l'ipotesi se:

$$X_{\chi^2} \geq \chi_{\alpha}^2 \quad \text{se l'ipotesi contraria è } \sigma^2 \geq \sigma_0^2$$

$$X_{\chi^2} < \chi_{1-\alpha}^2 \quad \text{se l'ipotesi contraria è } \sigma^2 < \sigma_0^2$$

$$X_{\chi^2} \geq \chi_{\alpha/2}^2 \quad \text{oppure} \quad X_{\chi^2} \leq \chi_{1-\alpha/2}^2 \quad \text{se l'ipotesi contraria è } \sigma^2 \neq \sigma_0^2$$

dove α è il livello di significatività o probabilità di falso rifiuto

CRITERIO DI VERIFICA DELL'IPOTESI DI BONTÀ DELL'ACOSTAMENTO DI DISTRIBUZIONI EMPIRICHE CON DISTRIBUZIONI TEORICHE DI PROBABILITÀ

Consideriamo n osservazioni tratte da una popolazione che si può ritenere distribuita secondo una legge di probabilità che fornisce una probabilità incognita π_i che un'osservazione cada nella classe i -esima ($i=1, 2, \dots, k$). Nel campione le osservazioni che cadono nella classe i -esima sono n_i dove: $\sum_{i=1}^k n_i = n$

Si vuole verificare, sulla scorta degli n valori osservati nel campione l'ipotesi nulla:

$$I: \pi_i = p_i$$

cioè che le probabilità incognite π_i sono uguali alla probabilità p_i che una v.c. di tipo "NOTO" assuma il valore x_i (la legge sarà caratterizzata dai parametri A_1, A_2, \dots, A_h).

Si vuole in definitiva verificare l'ipotesi della bontà dell'accostamento della distribuzione empirica con quella teorica di probabilità.

È intuitivo che tanto più è accettabile l'ipotesi nulla quanto più prossime a zero sono le differenze $p'_i - p_i$ dove $p'_i = n_i/n$ cioè le differenze $n_i - n \cdot p_i$

Si dimostra che se l'ipotesi nulla è vera:

$$X_{\chi^2} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

ha approssimativamente una distribuzione χ^2 con $v=k-h-1$ gradi di libertà ove k è il numero delle classi e h è il numero dei parametri della distribuzione teorica.

Pertanto se: $X_{\chi^2} \geq \chi_{\alpha}^2$

per i dati gradi di libertà, si rifiuta l'ipotesi nulla al livello di significatività α (probabilità di falso rifiuto)

RELAZIONI TRA VARIABILI

ACCOSTAMENTO AD UNA RETTA CON IL METODO DEI MINIMI QUADRATI

Iniziamo ad illustrare l'accostamento di una serie di misurazioni mediante una retta, e a tal fine consideriamo delle coppie di misurazioni (x_i, y_i) per $i=1, 2, \dots, n$ ($n>2$), che individuano n punti di coordinate (x, y) nel piano; la coppia (x_i, y_i) indica che nella i -esima osservazione la variabile X ha assunto il valore x_i e la variabile Y ha assunto il valore y_i .

Chiamata \bar{x} media aritmetica degli x_i , cioè:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

risulta opportuno scrivere l'equazione della retta nella forma $y = a + m(x - \bar{x})$, ove m è il coefficiente angolare mentre $k = a - m\bar{x}$ è l'ordinata all'origine, cioè¹:

$$y = k + mx$$

Per misurare l'accostamento della retta ad un punto, usiamo gli scostamenti verticali, cioè lo scostamento d_i verticale del punto di coordinate (x_i, y_i) dalla retta, che è dato da:

$$d_i = y_i - [a + m(x_i - \bar{x})].$$

Per la scelta della retta più appropriata si minimizza la somma dei quadrati degli scostamenti (metodo dei minimi quadrati) fornita dalla seguente relazione:

$$S = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n [(y_i - a) - m \cdot (x_i - \bar{x})]^2$$

Derivando rispetto ad a e m , ed eguagliando a zero le derivate si ha:

$$\frac{\partial S}{\partial a} = -2 \cdot \sum_{i=1}^n y_i + 2 \cdot a \cdot n = 0$$

$$\frac{\partial S}{\partial m} = -2 \cdot \sum_{i=1}^n y_i (x_i - \bar{x}) + 2 \cdot m \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

da cui il sistema di due equazioni lineari nelle due incognite a e m ²:

$$\begin{cases} \sum_{i=1}^n y_i = an \\ \sum_{i=1}^n y_i (x_i - \bar{x}) = m \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases} \quad \text{che ha come soluzioni} \quad \begin{aligned} \hat{a} &= \frac{\sum_{i=1}^n y_i}{n} = \bar{y} \\ \hat{m} &= \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Si tratta di un minimo, in quanto, come è facile verificare, sono soddisfatte le condizioni alle derivate seconde.

Osservando inoltre che:

$$\sum_{i=1}^n y_i (x_i - \bar{x}) = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}$$

¹ Si effettua un cambiamento di riferimento cartesiano, trasladando l'origine degli assi dal punto $(0,0)$ al punto $(\bar{x}, 0)$.

² L'apice “^” sta a contrassegnare i valori stimati delle grandezze.

si può scrivere:

$$\hat{m} = \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

da cui dividendo numeratore e denominatore per n si ha:

$$\hat{m} = \frac{Cov(x, y)}{Var(x)}$$

È opportuno osservare che i simboli di Cov e Var non hanno in questo caso un significato statistico ma devono essere intesi come indici descrittivi, non avendo per il momento considerato le grandezze come variabili aleatorie.

In definitiva la retta accostata con il metodo dei minimi quadrati è (dove con \hat{y} si indicano i valori accostati):

$$\hat{y} = \bar{y} + m(x - \bar{x})$$

È possibile osservare che la retta passa per il punto di coordinate (\bar{x}, \bar{y}) detto centroide dell'insieme dei punti (x_i, y_i) , condizione peraltro attesa, in quanto, quando la x assume il valore medio \bar{x} , anche la y dovrebbe assumere il suo valore medio \bar{y} . La retta accostata con il metodo dei minimi quadrati ha anche la proprietà che le deviazioni verticali d_i da essa hanno somma nulla:

$$\hat{y}_i = \bar{y} + \hat{m} \cdot (x_i - \bar{x}) \quad \Rightarrow \quad \hat{d}_i = y_i - \hat{y}_i = (y_i - \bar{y}) - \hat{m}(x_i - \bar{x})$$

da cui si ha ³
$$\sum_{i=1}^n \hat{d}_i = \sum_{i=1}^n (y_i - \bar{y}) - \hat{m} \sum_{i=1}^n (x_i - \bar{x}) = 0$$

Inoltre la retta accostata con il metodo dei minimi quadrati presenta l'importante proprietà che la somma (minima) dei quadrati degli scarti $\hat{S} = \sum_{i=1}^n \hat{d}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ divisa per n è la

varianza s_d^2 degli scostamenti verticali dalla retta accostata, cioè $s_d^2 = \frac{\hat{S}}{n}$.

Tale varianza s_d^2 (o lo scarto quadratico medio s_d) è una misura della bontà dell'accostamento della retta ai punti, assumendo valore nullo quando tutti i punti giacciono sulla retta, ovvero quando l'accostamento è perfetto. È facile verificare che il minimo della somma dei quadrati degli scarti è:

$$\hat{S} = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\left(\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Pertanto

$$s_d^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} - \frac{\left(\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y} \right)^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

³ La media aritmetica gode della proprietà che la somma algebrica degli scostamenti dalla media aritmetica stessa (differenze tra i valori e la media aritmetica) è nulla.

Introduciamo la varianza degli x_i e quella degli y_i definite come:

$$\text{Var}(x) = s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad \text{Var}(y) = s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

Mentre la covarianza di tra gli x_i e y_i è data da :

$$\text{Cov}(x, y) = s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum_{i=1}^n x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y}$$

Il coefficiente di correlazione lineare tra gli y_i e x_i è definito dall'invariante assoluto :

$$r = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x) \cdot \text{Var}(y)}} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{n \cdot s_x \cdot s_y}$$

Dalla relazioni precedentemente introdotte si ottiene:

$$s_d^2 = s_y^2 - \frac{\left(\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} \right)^2}{n^2 \cdot s_x^2} = s_y^2 - r^2 s_y^2 = s_y^2 \cdot (1 - r^2)$$

La misura in cui s_d^2 è minore di s_y^2 è indice (almeno per n grande) dell'accostamento effettuato mediante la retta.

Si consideri inoltre il rapporto:

$$\frac{s_d^2}{s_y^2} = 1 - r^2$$

È possibile osservare che r^2 misura la frazione di s_y^2 rimossa dalla varianza degli scostamenti verticali accostando la retta $\hat{y} = \hat{a} + \hat{m} \cdot (x - \bar{x})$, e pertanto si usa anche dire che r^2 è la frazione della varianza originaria s_y^2 spiegata dalla retta.

È facile vedere che $-1 \leq r \leq 1$, infatti dalla si ha che $0 \leq s_d^2 \leq s_y^2$ da cui dividendo per s_y^2 si ha:

$$0 \leq \frac{s_d^2}{s_y^2} \leq 1 \quad \text{ovvero} \quad 0 \leq 1 - r^2 \leq 1 \quad \text{cioè} \quad -1 \leq -r^2 \leq 0 \quad \text{e quindi} \quad 0 \leq r^2 \leq 1 \quad (\text{c.v.d.}).$$

Per capire meglio il significato di r osserviamo che:

$$\hat{m} = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y \cdot s_x}{s_x^2} = r \cdot \frac{s_y}{s_x}$$

REGRESSIONI LINEARI SEMPLICI

Supponiamo di osservare un fenomeno in cui una variabile casuale z dipenda da un'altra variabile matematica x (variabile indipendente) sotto controllo dell'investigatore, che può cioè prefissarne il valore (p.e. gli anni dalla costruzione, ecc). Sia inoltre:

$$z = a + m(x - \bar{x})$$

dove \bar{x} è la media aritmetica dei valori di x predeterminati (x_1, x_2, \dots, x_n), cioè: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

Si vuole determinare la dipendenza di z da x , ma il valore osservato z include un errore casuale, dovuto all'errore insito nella misura sperimentale (con riferimento a quantogità illustrato, si vuole in sostanza determinare la regressione della v.a. Z su X , dove X non è però una v.a. ma una variabile matematica).

Siano pertanto Y_i le v.c. che rappresentano il valore osservato corrispondente ad x_i , e siano U_i le v.c. che rappresentano l'errore di osservazione (o misurazione). Allora se si sono effettuate n specifiche misurazioni, o osservazioni, corrispondenti ai predeterminati valori di x , ottenendo i risultati $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$, si ha:

$$Y_i = a + m(x_i - \bar{x}) + U_i \quad \text{cioè} \quad Y_i = k + m \cdot x_i + U_i$$

avendo posto $k = a - m\bar{x}$ (che rappresenta la regressione di Y su X).

Si assume generalmente che la popolazione sia infinita e che gli errori U_i siano v.c. normali indipendenti con valore medio nullo $M(U_i)=0$ e varianza costante comune $\text{Var}(U_i)=\sigma^2$. Il fatto che le U_i siano variabili indipendenti comporta che la $\text{Cov}(U_i, U_j) = 0$ per $i \neq j$; inoltre è evidente che $\text{Cov}(U_i, x_i) = 0$ essendo x_i una variabile matematica. Pertanto si ha:

$$M(Y_i) = a + m(x_i - \bar{x}) = \tilde{y}_i$$

Immaginiamo ora di aver fissato gli x_i e di osservare i valori $Y_i = y_i$ che si realizzano, ovvero di disporre di un campione di n coppie di misurazioni (x_i, y_i) per $i=1, 2, \dots, n$. Poniamo

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

media delle variabili casuali Y_i ($i=1, 2, \dots, n$), cosa ben diversa da $M(Y_i)$ che rappresenta il valore medio della singola v.c. Y_i . Se nel campione singolo di n coppie di misurazioni è risultato $Y_i = y_i$, la v.c. \bar{Y} assume il valore:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Stimiamo i coefficienti a e m con l'ausilio del metodo dei minimi quadrati. Per il teorema di Gauss e Markov tali stime sono anche di massima verosimiglianza, e sono quindi anche stime lineari non affette da errore sistematico di minima varianza. Si ha allora che $\hat{a} = \bar{y}$ è il valore della variabile casuale $\hat{A} = \bar{Y}$ stima di a calcolata sul singolo campione. Allo stesso tempo:

$$\hat{M} = \frac{\sum_{i=1}^n Y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

è il valore della v.c. \hat{M} stima di m , calcolata sul medesimo singolo campione; cioè:

$$m = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

La stima di Y è allora:

$$\hat{Y}_i = \hat{A} + \hat{M}(x_i - \bar{x})$$

cioè nel singolo campione

$$\hat{y}_i = \hat{a} + \hat{m}(x_i - \bar{x})$$

È facile verificare che \hat{A} e \hat{M} sono stime non affette da errore sistematico rispettivamente di a e m , infatti:

$$M(\hat{A}) = M(\bar{Y}) = \frac{1}{n} \sum_{i=1}^n M(Y_i) = \frac{1}{n} \sum_{i=1}^n [a + m(x_i - \bar{x})] = a$$

$$M(\hat{M}) = \frac{\sum_{i=1}^n M(Y_i)(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n [a + m(x_i - \bar{x})](x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = m$$

ricordando che $\sum_{i=1}^n (x_i - \bar{x}) = 0$

si ottiene $M(\hat{Y}_i) = M[\hat{A} + \hat{M}(x_i - \bar{x})] = a + m(x_i - \bar{x}) = \tilde{y}_i$

Osservato che \hat{A} e \hat{M} sono combinazioni lineari delle v.c. indipendenti Y_i ($i=1, 2, \dots, n$), si possono facilmente calcolare le varianze ($\text{Var}(\beta_1 \xi_1 + \dots + \beta_k \xi_k) = (\beta_1)^2 \text{Var}(\xi_1) + \dots + (\beta_k)^2 \text{Var}(\xi_k)$):

$$\text{Var}(\hat{A}) = \text{Var}(\bar{Y}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) = \frac{\sigma^2}{n}$$

$$\text{Var}(\hat{M}) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(Y_i)}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{n \cdot s_x^2}$$

Pertanto \hat{A} e \hat{M} sono stime consistenti, di minima varianza tra quelle non affette da errore sistematico.

Invece S_d^2 non è stima non affetta da errore sistematico di σ^2 in quanto:

$M(S_d^2) = \frac{n-2}{n} \cdot \sigma^2$ da cui stima non affetta da errore di σ^2 è:

$$\sigma^2 = S_d^2 \cdot \frac{n}{n-2}$$

dove $S_d^2 = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{A} - \hat{M}(x_i - \bar{x})]^2$

(più in generale $\sigma^2 = S_d^2 \cdot \frac{n}{n-(m+1)}$ dove $n-(m+1)$ sono i gradi libertà ed m le variabili indipendenti)