

RICHIAMI DI STATISTICA E CALCOLO DELLE PROBABILITA'

La Statistica è la disciplina che studia gli eventi non deterministici (o incerti) riguardo ai quali non si ha una completa conoscenza. Tali eventi sono detti eventi aleatori. Sono tali, ad esempio, il risultato del lancio di una moneta o di un dado, o l'estrazione dei numeri del lotto.

Se per descrivere gli eventi possibili si fa uso di una variabile, tale variabile viene detta variabile aleatoria (v.a.).

CALCOLO DELLE PROBABILITA'

LA PROBABILITA'

Agli eventi aleatori viene generalmente associato un numero che viene detto probabilità. Secondo una definizione non rigorosa la probabilità è il numero che quantifica il grado di fiducia che si nutre riguardo al verificarsi di un dato evento aleatorio.

Se un fenomeno non deterministico ammette diversi risultati la probabilità di uno solo di essi (i.e. l'i-esimo evento) può essere calcolata come il numero di occorrenze dell'i-esimo evento diviso il numero totale di risultati possibili.

Esempi:

- 1) È intuitivo che la probabilità associata al risultato "Testa" del lancio di una moneta non truccata corrisponda al 50%. Significa che il grado di fiducia che attribuiamo al verificarsi dell'evento "Testa" è pari ad $1/2$.
I risultati possibili nel lancio di una moneta sono 2: "testa" o "croce". Tra essi il numero di occorrenze del risultato "Testa" è 1. Per cui la probabilità che da un lancio esca testa è:
 $P\{\text{Testa}\} = 1/2$.
- 2) Si vuole studiare la probabilità che dal lancio del dado esca un numero pari. I numeri possibili risultanti dal lancio di un dado sono 6 [1 2 3 4 5 6]. Di questi 3 sono numeri pari [2 4 6].
Pertanto la probabilità che il risultato sia un numero pari è (detto x il risultato del dado):
 $P\{x \text{ è pari}\} = 3/6$.
- 3) Si vuole studiare la probabilità che alla tombola venga estratto un numero pari minore di 15 (evento x).
Tutti i numeri della tombola sono 90. I numeri pari minori di 15 sono 7 [2 4 6 8 10 12 14].
Pertanto la probabilità associata al suddetto risultato è:
 $P\{x\} = 7/90$.

Risulta a questo punto chiaro che la probabilità di tutti i risultati possibili di un fenomeno è pari ad 1 e che la probabilità è un numero compreso tra 0 ed 1.

$$P \in [0,1].$$

Infine la probabilità associata a tutti gli eventi meno l'evento x è definita probabilità complementare e vale $1 - P\{x\}$.

UNIONE E INTERSEZIONE DI PROBABILITA'

Spesso risulta utile calcolare la probabilità associata al verificarsi di più eventi alternativi. Ad esempio alla tombola un giocatore può comprare 2 cartelle. La vittoria di una cartella esclude la vittoria dell'altra (i due eventi sono alternativi) ma risulta intuitivo affermare che il giocatore ha una probabilità doppia rispetto a chi ha comprato una sola cartella.

In tal caso si parla di unione di probabilità e se gli eventi sono tra loro indipendenti il calcolo viene effettuato sommando tra loro le probabilità dei singoli eventi alternativi.

Esempio:

- 4) Con riferimento all'esempio 2) il calcolo della probabilità che un numero pari sia il risultato del lancio di un dado corrisponde a calcolare la probabilità che il risultato sia 2 oppure 4 oppure 6. Secondo questo approccio la probabilità può essere calcolata come:
- $$P\{x\} = P\{2\} + P\{4\} + P\{6\} = 1/6 + 1/6 + 1/6 = 1/2.$$

Ancora può risultare utile il calcolo della probabilità associata al verificarsi contemporaneo di due eventi non alternativi. Ad esempio, nelle scommesse sportive, può essere utile calcolare la probabilità di indovinare i risultati di due o più partite di campionato. E' importante che i due eventi non siano tra loro alternativi: la probabilità che il risultato del lancio di una moneta sia contemporaneamente sia "testa" sia "croce" è evidentemente nulla.

In questo caso si parla di intersezione di probabilità e, nel caso di eventi tra loro indipendenti, si calcola come prodotto delle probabilità associate ai singoli eventi

Esempio:

- 5) Con riferimento all'esempio 3) si vuole studiare la probabilità che alla tombola venga estratto un numero pari minore di 15 (evento x). La probabilità che il numero estratto sia pari (evento y) è $45/90 = 1/2$, mentre la probabilità che il numero sia minore di 15 (evento z) è $14/90 = 7/45$. La probabilità combinata dei due eventi è:
- $$P\{x\} = P\{y\} * P\{z\} = 1/2 * 7/45 = 7/90.$$

ENNUPLE DI NUMERI E COEFFICIENTE BINOMIALE

Le definizioni di unione e di intersezione di probabilità risultano utili nel calcolo di probabilità di combinazioni di numeri. Si scelga come esempio il gioco del lotto e si voglia calcolare la probabilità associata all'estrazione di una combinazione di 6 numeri.

La probabilità che il primo numero estratto corrisponda ad uno dei sei giocati è pari a $6/90$. Infatti i casi "vincenti" sono sei a fronte di 90 casi possibili. Si può anche affermare che ciascuno dei 6 numeri giocati ha una probabilità pari ad $1/90$ di essere il primo estratto; pertanto la probabilità che uno tra essi corrisponda al primo estratto è pari a $1/90 + 1/90 + 1/90 + 1/90 + 1/90 + 1/90 = 6/90$.

Una volta indovinato il primo numero, restano 5 numeri ancora da indovinare a fronte di 89 numeri da estrarre. La probabilità che il secondo numero estratto corrisponda ad uno dei rimanenti cinque è pari a $5/89$.

La probabilità di indovinare tutti i 6 numeri giocati (evento x) è pari all'intersezione delle probabilità di indovinare ciascuno dei sei numeri:

$$P\{x\} = 6/90 * 5/89 * 4/88 * 3/87 * 2/86 * 1/85 = (1 * 2 * 3 * 4 * 5 * 6) / (90 * 89 * 88 * 87 * 86 * 85)$$

Moltiplicando numeratore e denominatore per la stessa quantità:

$$P\{x\} = (1 * 2 * 3 * 4 * 5 * 6) / (90 * 89 * 88 * 87 * 86 * 85) * (84 * 83 * \dots * 2 * 1) / (84 * 83 * \dots * 2 * 1)$$

si può scrivere:

$$P\{x\} = 6! * (90 - 6)! / 90!$$

Laddove la scrittura $n!$ (n fattoriale) rappresenta il prodotto dei primi n numeri naturali.

La metodologia mostrata valga come prova empirica della formula per calcolare la probabilità associata all'estrazione di una combinazione di n numeri da un insieme di m numeri:

$$P\{x\} = n! (m-n)! / m!$$

Si definisce coefficiente binomiale

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

il numero di ennuple che è possibile estrarre da una serie di m numeri. Il suo inverso rappresenta la probabilità relativa ad una sola di tali ennuple.

STATISTICA

FUNZIONI DISTRIBUZIONE E DENSITA' DI PROBABILITA' PER VARIABILI DISCRETE

Assegnata una variabile aleatoria x , la funzione distribuzione di probabilità (cdf) lega ad ogni valore x_0 della variabile x la probabilità:

$$F(x_0) = P\{x \leq x_0\}$$

La funzione densità di probabilità (pdf) lega ad ogni valore x_0 della variabile x la probabilità:

$$f(x_0) = P\{x = x_0\}$$

Esempi:

- 6) La funzione distribuzione associata alla variabile x che descrive il risultato del lancio di un dado vale:

$$F(6) = 1$$

$$F(5) = 5/6$$

$$F(4) = 4/6$$

$$F(3) = 3/6$$

$$F(2) = 2/6$$

$$F(1) = 1/6$$

- 7) La funzione densità associata alla variabile x che descrive il risultato del lancio di un dado vale:

$$F(6) = 1/6$$

$$F(5) = 1/6$$

$$F(4) = 1/6$$

$$F(3) = 1/6$$

$$F(2) = 1/6$$

$$F(1) = 1/6$$

VARIABILI ALEATORIE CONTINUE

DISTRIBUZIONE E DENSITA' DI VARIABILI ALEATORIE CONTINUE

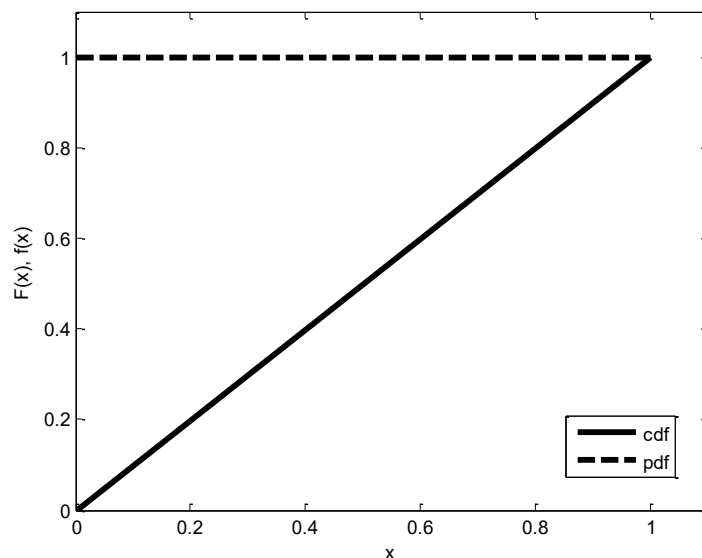
Fin qui si è fatto riferimento a variabili aleatorie discrete, ma molto comune è nella pratica il riferimento a variabili aleatorie continue. Sono tali le variabili aleatorie che possono assumere, all'interno di un assegnato intervallo, valori appartenenti all'insieme di numeri reali. Ad esempio la temperatura di una fiamma in condizioni non adiabatiche può dipendere da fattori ambientali di non facile previsione: essa assume a pieno titolo le caratteristiche di una variabile aleatoria continua.

Si prenda come esempio la funzione generatrice di numeri casuali di un calcolatore. Tale funzione restituisce come risultato un numero reale compreso fra 0 ed 1 estratto casualmente.

La funzione distribuzione di probabilità per una siffatta variabile è una retta passante per i punti $(x, F(x)) = (0, 0)$ e $(x, F(x)) = (1, 1)$

Più difficile risulta il calcolo della probabilità $P\{x = x_0\}$: il numero di eventi possibili è infinito, per cui la probabilità di ciascun numero risulterebbe pari a $1/\infty = 0$. Appare quindi chiaro che nel caso di variabili continue la definizione fin qui data di probabilità perde di significato. Al fine di poter almeno esprimere la probabilità che una variabile aleatoria continua assumi valori prossimi ad un determinato valore è necessario introdurre la definizione di densità di probabilità per variabili continue, che è così formulata:

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \frac{dF(x)}{dx}$$



Le funzioni distribuzione e densità di probabilità gode di alcune proprietà:

- $P(x > x_0) = 1 - P(x < x_0) = 1 - F(x_0)$

se $x_1 < x_2$:

- $P(x < x_2) = P(x < x_1) + P(x_1 < x < x_2) \Rightarrow P(x_1 < x < x_2) = P(x < x_2) - P(x < x_1) \Rightarrow P(x_1 < x < x_2) = F(x_2) - F(x_1)$

- $\int_{-\infty}^{\infty} f(x) = 1$

- $\int_{-\infty}^{x_0} f(x) = F(x_0)$
- $\int_{x_1}^{x_2} f(x) = F(x_2) - F(x_1)$

PERIODO DI RITORNO

In statistica si definisce periodo di ritorno il tempo medio di attesa fra due eventi successivi.

Si consideri, a scopo esemplificativo, la variabile aleatoria x che rappresenta le condizioni del tempo meteorologico in un dato giorno e assegniamo ad essa il valore 1 se il suddetto giorno è piovoso e 0 altrimenti. A tale variabile è possibile associare una pdf: $P(x=0)=f(0)$ e $P(x=1)=f(1)$

L'inverso di tale probabilità costituisce la definizione di periodo di ritorno:

$$T(x) = \frac{1}{f(x)}$$

Esso rappresenta il numero medio di giorni da attendere tra il verificarsi di due eventi uguali successivi. Se ad esempio $f(1)=0.05$, allora il numero medio di giorni da attendere tra il verificarsi di due eventi piovosi successivi risulterebbe pari a $T(1)= 1/0.05 = 20$ giorni.

DISTRIBUZIONE GAUSSIANA

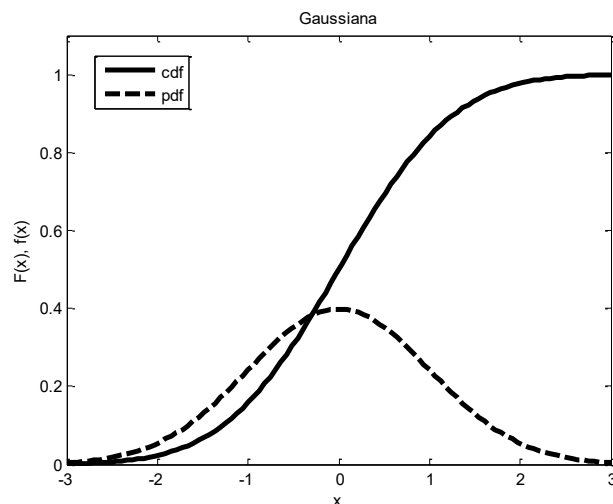
La variabile aleatoria gaussiana, nota pure con il nome di Normale, costituisce il modello di v.a. più adoperato. Essa può essere definita come modello interpretativo degli errori o scostamenti $x-m$, dal valore vero m , che accompagnano le misure sperimentali x di un certo parametro fisico effettuate sempre nelle stesse condizioni. Se le suddette misure non sono affette da errori sistematici è logico ritenere che:

- sia nulla la media degli errori
- la pdf degli errori sia simmetrica e tenda a zero per valori crescenti sia negativi che positivi
- la pdf abbia valore massimo in corrispondenza del valore nullo.

Una siffatta funzione densità assume la seguente espressione matematica:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad \text{con} \quad -\infty < x < +\infty$$

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$



STATISTICA INFERENZIALE

MEDIA, VARIANZA, COVARIANZA E SCARTO QUADRATICO MEDIO

In teoria della probabilità il valore atteso (chiamato anche media, speranza o speranza matematica) di una variabile casuale X , è un numero m che formalizza l'idea euristica di valore medio di un fenomeno aleatorio.

In generale il valore atteso di una variabile casuale discreta (che assuma cioè solo un numero finito o una infinità numerabile di valori) è dato dalla somma dei possibili valori di tale variabile, ciascuno moltiplicato per la probabilità di essere assunto (ossia di verificarsi), cioè è la media ponderata dei possibili risultati. Per una variabile casuale continua la somma viene sostituita da un integrale:

$$m(x) = \begin{cases} \sum_{i=1}^{\infty} x_i P(x_i) & \text{se } x \text{ è discreta} \\ \int_{-\infty}^{+\infty} x f(x) dx & \text{se } x \text{ è continua} \end{cases}$$

La varianza di una variabile aleatoria X (e della distribuzione di probabilità che questa segue) è un numero, indicato con $\text{Var}(X)$, che fornisce una misura di quanto siano vari i valori assunti dalla variabile, ovvero di quanto si discostino dalla media $m(x)$.

$$\text{var}(x) = \begin{cases} \sum_{i=1}^{\infty} [x_i - m(x)]^2 P(x_i) & \text{se } x \text{ è discreta} \\ \int_{-\infty}^{+\infty} [x - m(x)]^2 f(x) dx & \text{se } x \text{ è continua} \end{cases}$$

La radice quadrata della varianza viene definita scarto quadratico medio.

$$s.q.m. = \sqrt{\text{var}(x)}$$

Assegnate due variabili aleatorie x e y la covarianza è un numero $\text{Cov}(X,Y)$ che fornisce una misura di quanto le due varino assieme, ovvero della loro dipendenza. Essa è definita come segue:

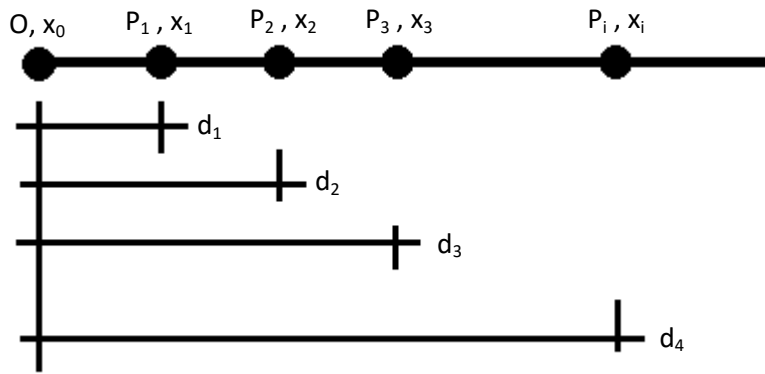
$$\text{cov}(x, y) = \begin{cases} \sum_{i=1}^{\infty} [x_i - m(x)][y_i - m(y)] P(x, y) & \text{se } x, y \text{ sono discrete} \\ \int_{-\infty}^{+\infty} [x - m(x)][y - m(y)] f(x, y) dx dy & \text{se } x, y \text{ sono continue} \end{cases}$$

Se si hanno due serie di rilevazioni della variabile x ed y rispettivamente, entrambe di numerosità n , si può calcolare la covarianza come:

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^{\infty} [x_i - m(x)][y_i - m(y)]$$

VARIOGRAMMA

Sia assegnata una variabile che descrive una grandezza fisica all'interno di un dominio spaziale, sia essa la concentrazione di inquinante in un lago o la permeabilità di un acquifero. Noto il valore della variabile in un punto del dominio è possibile studiarne la variabilità spaziale per mezzo del variogramma. Si immagini una variabile aleatoria che vari con continuità lungo una retta e siano noti i valori assunti dalla variabile in corrispondenza di una serie di punti P_i appartenenti alla retta e distanti d_i da un punto identificato come origine.



$$\gamma(O, d) = \frac{1}{2n} \sum_{i=1}^n [x(O) - x(O + d_i)]^2$$

Il variogramma misura la correlazione dei valori della variabile nell'intorno di un punto. Tanto minore è il valore assunto da γ , tanto maggiore sarà la dipendenza.

