Fondamenti di Statistica e Probabilità

Pietro Bonardi

Marzo 2020

Indice

1 Introduzione

In questo documento si mostreranno gli aspetti di basi della statistica, in particolare si focalizza sulla descrittiva ed inferenziale. È inteso come un documento per ripassare i concetti principali.

2 Statistica Descrittiva

Statistica Descrittiva Per definizione si intende quella raccolta di metodi e strumenti matematici atti ad organizzare una o più serie di dati in modo tale da evidenziarne in forma ***sintetica*** eventuali:

- Simmetrie
- Periodicità
- Leggi di altro genere

Ovvero in grado di descriverne in maniera "intuitiva" informazioni implicite.

Popolazione L'insieme degli elementi cui si riferisce l'indagine statistica. È un insieme limitato costituito da *osservazioni*.

Caratteri I dati variano in funzione dei modi di presentarsi.

- Caratteri qualitativi, quando sono qualità o dati non numerici.
  - Nominali,descrivono una caratteristica, fenomeno, che si manifesta con degli attributi, modalità che non hanno un ordinamento.
  - Ordinali descrivono una caratteristica, fenomeno, che si manifesta con degli attributi, modalità che hanno un ordinamento naturale.
- Caratteri Quantitativi, quando le grandezze sono misurabili e si manifestano con dei numeri o valori.
  - Discreti, quando assumono valori numerabili.
  - Continui quando assumo valori sull'asse R.

2.1 Rappresentazioni numerica di dati statistici

Supponiamo di avere: $E = \{x_1, x_2, ..., x_n\}$

È corretto raggruppare i dati in modalità ed associare a ciascuna di esse il numero di volte che compaiono in E. Si definisce quindi l'insieme $S = \{s_1, s_2, ..., s_N\}$, delle possibili modalità dove "$_N$" è il numero totale di modalità.

Definizione 2.1. *Frequenza Assoluta Si dice frequenza assoluta della modalità s_j dove $j = 1, ..., N$:*
f_j *= numero di elementi di E con valore s_j*

Definizione 2.2. *Distribuzione Frequenza Assoluta Definita come*

$$f : \{s_1, s_2, ..., s_N\} \mapsto f_j$$

Definizione 2.3. *Frequenza Cumulata Assoluta*

$$F_j = \sum_{k:s_k \leq s_j} f_k$$

Definizione 2.4. *Frequenza Relativa*

$$p_j = \frac{f_j}{n}$$

dove n è la grandezza dell'insieme $E = \{x_1, x_2, ..., x_n\}$.

Definizione 2.5. *Frequenza Cumulata Relativa*

$$P_j = \sum_{k:s_k \leq s_j} p_k$$

Quando siamo nel caso **continuo** oppure in quello discreto ma con molte modalità, è corretto dividere in classi.

Definizione 2.6. *Classe Una classe di una modalità è un sotto insieme di S. Si partizione quindi la modalità tale che:*
$C_i \cap C_j = \emptyset \ \forall i, j$
$C_i \cup C_j = S = \{s_1, s_2, ..., s_N\}$

Alle classi sono associati dei valori:

- *Confine superiore, l'elemento più grande di C_j*
- *Confine inferiore, l'elemento più piccolo di C_j*
- *Semisomma, $\frac{Superiore+Inferiore}{2}$*

2.2 Indici di tendenza centrale o di posizione

In generale gli indici di posizione o misure di tendenza centrale vengono chiamati medie.

Definizione 2.7. *Medie Data una variabile quantitativa X, che assume su N unità statistiche i valori non decrescenti $x_1, x_2, ..., x_N$, sono denominate **medie** quelle particolari funzioni che soddisfano le seguenti proprietà:*

- ***internalità**, la media non deve essere inferiore al più piccolo dei valori osservati né superiore al più grande, ovvero deve essere interna al campo di osservazione*
- ***moltiplicativa**, la media segue le variazioni dell'unità di misura*
- ***monotonicità**, ovvero se almeno uno dei valori aumenta, o diminuisce, restando invariati tutti gli altri, la media aumenta, o diminuisce.*

La funzionalità degli ***Indici di Tendenza Centrale*** è di determinare un solo valore per rappresentare la serie intera di dati, quindi essere in grado di sintetizzare.

- **Moda**, è calcolabile per ogni tipo di dato.
  Per definizione è l'elemento nella serie di dati più frequente. Può non essere unica, in tal caso si parla di fenomeno ***plurimodale***.
- **Mediana**, Per un fenomeno qualitativo ordinale o quantitativo, poste le N osservazioni in ordine non decrescente, viene definita mediana, quella modalità o valore associato all'unità che occupa la posizione centrale.

  Si distinguono i seguenti casi:

  - N è *dispari*, $\Longrightarrow$ l'elemento in posizione $P = \frac{N+1}{2}$
  - N è *pari*, bisogna ricavare due posizioni mediane. $P_1 = \frac{N}{2}$ e $P_2 = \frac{N+1}{2}$
    Ora si distingue:
    * caso *qualitativo*, se $P_1 \neq P_2$ allora la mediana non è definita. Viceversa è definita.
    * caso *quantitativo* se $P_1 \neq P_2$, allora come mediana usualmente si considera la semisomma delle due modalità.
  - *Fenomeno in classi*, si definisce la classe mediana definita come l'elemento in posizione $P = \frac{N+1}{2}$, per N pari e dispari.
- **Quartili**, In similitudine alla *mediana*, possono essere definiti i Quartili, quindi si divide la serie in 4.

  - Q_1 quartile di primo ordine, identificato alla posizione 25% della distribuzione.
  - Q_2 quartile di secondo ordine, il quale è coincidente con la *mediana*.

  - Q_3 quartile di terzo ordine, il quale identifica il 75% della distribuzione.

  Vengono anche definiti i ***decili***, dividendo la distribuzione per 10 e i ***percentili***, dividendola per 100.

- **Media aritmetica**, Calcolabile solo per fenomeni quantitativi.

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

  Quando si parte dalla distribuzione di frequenza

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{N} s_j f_j = \frac{s_1 f_1 + s_2 f_2 + ... + s_N f_N}{n} =$$

$$= \frac{s_1 f_1}{n} + \frac{s_2 f_2}{n} + ... + \frac{s_N f_N}{n} = s_1 p_1 + s_2 p_2 + ... + s_N p_N = \sum_{j=1}^{N} s_j p_j$$

  Proprietà della media aritmetica:

  - La somma degli *scarti* dei valori dalla loro media è nulla:

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

  - La somma dei quadrati degli scarti dei valori dalla loro media è minima:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = min$$

    Vuol dire che non c'è un'indice che sottratto ai valori restituisca un valore più piccolo.
  - Se i valori x_i vengono trasformati nei valori $y_i = a \pm bx_i$, allora la media subisce la stessa trasformazione: $\bar{y} = a \pm \bar{x}$

  Teorema 2.1. *La media di un miscuglio di K gruppi, o sotto-popolazioni, per cui sono note la media e le numerosità di gruppo, è uguale alla media delle medie dei singoli gruppi ponderata con le numerosità di gruppo. Se indico la numerosità della popolazione $N = \sum_{i=1}^{K} n_i$, dove n_i è la numerosità del gruppo i-esimo.*

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{K} \bar{x}_i n_i$$

  dove x_i è la media del gruppo j-esimo tra i K gruppi.

Teorema 2.2. *La media della somma o della differenza di due o più variabili è uguale alla somma od alla differenza della media delle singole variabili.*

$$Z = X \pm Y$$
$$\bar{Z} = \overline{(X \pm Y)} = \bar{X} \pm \bar{Y}$$

Confronti tra mediana e media Si può concludere che una distribuzione è ***simmetrica*** se *mediana e media coincidono* mentre è ***asimmetrica*** a destra o sinistra se la *mediana è maggiore o minore della media.*

2.3 Indici di variabilità

Gli indici di tendenza centrale non sono utili per fornire informazione circa l'omogeneità ed eterogeneità della serie. In merito a questo sono introdotti ora gli indici di variabilità.

Si osserva subito che una distribuzione dove tutte le sue N unità statistiche assumono la stessa modalità non ha variabilità e la variabile ad esso associata è detta ***degenere***.

Si studiano gli indici divisi per carattere del dato.

- ***Dati qualitativi***
  - **Indice di eterogeneità di Gini**, questo misura quanto i dati siano omogenei tra loro.
    $$G = 1 - \sum_{i=1}^{k} p_i^2$$
    k è il numero totale di modalità e p_i è la frequenza relativa associata. Questo indice varia tra $[0, \frac{k-1}{k}]$.
    Per averlo normalizzarlo e farlo variare tra $[0, 1]$, si divide per il suo massimo.
    $$G_{NORM} = \frac{G}{G_{MAX}} = \frac{G}{\frac{k-1}{k}} = G\frac{k}{k-1}$$
- ***Dati quantitativi***
  - **Campo di Variazione o Range**, $R = x_{max} - x_{min}$ ci da una misura indicativa sulla dispersione della distribuzione, non sensibile agli outliers.
  - **Differenza Interquartile**, $DI = Q_3 - Q_1$ differenza tra il quartile di terzo e di primo ordine.

– **Varianza**, è lo scarto medio o distanza media, al quadrato, dei dati dalla media. La formula della varianza mostrata è intesa da usare in presenza della popolazione **non** del campione.

Con dati grezzi:

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

Con la distribuzione di frequenza, dove ci sono k modalità:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{i})^2 f_k$$

Considerazioni:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) =$$

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

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

Si possono fare dei simili passaggi algebrici per dimostrare che la varianza con la distribuzione di frequenza può essere riscritta

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}x_i^2 f_i - \bar{x}^2$$

Proprietà della varianza:

∗ Invariante per traslazione, se trasformo la variabile X, con σ^2 nota, in $Y = a + X$ la varianza diventa:

$$\sigma^2(Y) = \sigma^2(a + X) = \sigma^2(a) + \sigma^2(X) = 0 + \sigma^2(X) = \sigma^2(X)$$

∗ Se si effettua una trasformazione lineare sulla variabile X del tipo $Y = a + bX$ la varianza diventa:

$$\sigma^2(Y) = \sigma^2(a + bX) = \sigma^2(a) + \sigma^2(bX) = b^2\sigma^2(X)$$

Teorema 2.3. *La varianza di un miscuglio di K gruppi, o sotto-popolazioni, per cui sono note le medie, le varianze e le numerosità di gruppo, è uguale a:*

$$\sigma^2 = \sigma_W^2 + \sigma_B^2$$

Se indico la numerosità della popolazione $N = \sum_{i=1}^{K} n_i$, dove n_j è la numerosità del gruppo j-esimo, $\bar{x_j}$ indica la media del gruppo j-esimo e σ_j^2 la sua rispettiva varianza.

Allora definisco:

- σ_W^2 *detta varianza "* ***within*** *", è uguale alla media delle varianze pesata sulla numerosità di gruppo, ovvero varianza nei gruppi.*

$$\sigma_W^2 = \frac{1}{N}\sum_{j=1}^{K}\sigma_j^2 n_j = \frac{1}{N}\sum_{j=1}^{K}\frac{\sum_{i=1}^{n_j}(x_{i,j}-\bar{x_j})^2 n_j}{n_j} =$$

$$= \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{n_i}(x_{i,j}-\bar{x_i})^2$$

Si può intuitivamente pensare come la media delle varianze.

- σ_B^2 *detta varianza "* ***between*** *", misura lo scostamento tra la media di ogni gruppo (o media condizionata) e la media totale. In altre parole, rappresenta la variabilità tra i diversi gruppi e la sua formula è la seguente:*

$$\sigma_B^2 = \frac{1}{N}\sum_{j=1}^{K}(\bar{x_j}-\bar{X})^2 n_i$$

Dove $\bar{X}$ *è la media tra i gruppi definita nel Teorema (2.1)*

Dimostrazione 1.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{K}\sum_{j}^{n_i}(x_{ij}-\bar{X})^2 = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x_i}+\bar{x_i}-\bar{X})^2 =$$

$$\frac{1}{N}\left[\sum_{i=1}^{K}\sum_{j=1}^{n_i}((x_{ij}-\bar{x_i})+(\bar{x_i}-\bar{X}))^2\right] =$$

$$= \frac{1}{N}\left[\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x_i})^2 + \sum_{i=1}^{K}\sum_{j=1}^{n_i}(\bar{x_i}-\bar{X})^2 + 2\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x_i})(\bar{x_i}-\bar{X})\right] =$$

$$\frac{1}{N}\left[\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x_i})^2 + \sum_{i=1}^{K}\sum_{j=1}^{n_i}(\bar{x_i}-\bar{X})^2 + 0\right] =$$

$$\frac{1}{N}\left[\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x_i})^2 + \sum_{i=1}^{K}(\bar{x_i}-\bar{X})^2 n_i\right] =$$

$$= \frac{1}{N}\left[\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij}-\bar{x_i})^2\right] + \frac{1}{N}\left[\sum_{i=1}^{K}(\bar{x_i}-\bar{X})^2 n_i\right] = \sigma_W^2 + \sigma_B^2$$

Teorema 2.4. *La varianza della somma di due variabili risulta uguale alla somma delle varianze delle singole variabili solo se queste sono tra loro indipendenti.*

- **Scarto Quadratico Medio**, definito come la radice della varianza

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2}$$

- **Coefficiente di variazione**, utilizzato per fare confronti tra variabili distinte poiché depurato dall'unità di misura

$$CV = \frac{\sigma}{|\bar{x}|}$$

3 Distribuzione di Frequenza Bivariata

Supponiamo di avere una distribuzione con due caratteri spiegati contemporaneamente. Sia essa definita come $E = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ coppie ordinate di valori.

Se il carattere è quantitativo discreto identifico l'insieme delle modalità $S = \{(s_j, u_k), j = 1, ..., N, k = 1, ..., M\}$, dove N e M sono il numero di modalità totali per ciascuna variabile.

3.1 Rappresentazioni numerica di dati statistici bivariati

Definizione 3.1. *Frequenza Assoluta o Congiunta Si dice frequenza assoluta della modalità (s_j, u_k) dove $j = 1, ..., N$ e $k = 1, ..., M$:*
f_{jk} *= numero di elementi in E con valore (s_j, u_k)*

Definizione 3.2. *Distribuzione di Frequenza Assoluta Bivariata o Congiunta Definita come*

$$f\colon \{(s_1, u_1), (s_2, u_2), ..., (s_N, u_M)\} \mapsto f_{jk}$$

Definizione 3.3. *Frequenza Cumulata Assoluta Bivariata o Congiunta*

$$F_{jk} = \sum_{k: s_k \leq s_j; l: u_l \leq u_k} f_{kl}$$

Definizione 3.4. *Frequenza Relativa Bivariata o Congiunta*

$$p_{jk} = \frac{f_{jk}}{n}$$

dove n è la grandezza dell'insieme $E = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$.

Definizione 3.5. *Frequenza Cumulata Relativa Bivariata o Congiunta*

$$P_{jk} = \sum_{k: s_k \leq s_j; l: u_l \leq u_k} p_{kl}$$

Definizione 3.6. *Tabella Doppia entrata, questa notazione specifica con X e Y i vettori di modalità.*

X \ Y	y_1	y_2	…	y_c	*Marginale*		
x_1	$f(1,1)$	$f(1,2)$		$f(1,4)$	$f(1,\bullet)$	$\bar{y}_{\vert x_1}$	$\bar{y}$
x_2	$f(2,1)$	$f(2,2)$		$f(2,4)$	$f(2,\bullet)$	$\bar{y}_{\vert x_2}$	
…			$f(i,j)$			…	
x_r	$f(4,1)$	$f(4,2)$		$f(4,3)$	$f(r,\bullet)$	$\bar{y}_{\vert x_r}$	
Marginale	$f(\bullet,1)$	$f(\bullet,2)$	...	$f(\bullet,c)$	***N***		
	$\bar{x}_{\vert y_1}$	$\bar{x}_{\vert y_2}$	…	$\bar{x}_{\vert y_c}$			
	$\bar{x}$						

Definizione 3.7. *Frequenza assoluta marginale permettono di guardare la tabella come una distribuzione di frequenza univariata.*

- *Frequenza assoluta marginale di colonna,* $f_{(\bullet,j)} = \sum_{i=1}^{r} f_{(i,j)}$
- *Frequenza assoluta marginale di riga,* $f_{(i,\bullet)} = \sum_{j=1}^{c} f_{(i,j)}$

Definizione 3.8. ***Frequenze Condizionate*** *La distribuzione condizionata di X rispetto Y:*

		Y				
		y_1	-	y_j	-	y_c
X	x_1	$f_{11} / f_{.1}$	-	$f_{1j} / f_{.j}$	-	$f_{1c} / f_{.c}$
	-	-	-	-	-	-
	x_i	$f_{i1} / f_{.1}$	-	$f_{ij} / f_{.j}$	-	$f_{ic} / f_{.c}$
	-	-	-	-	-	-
	x_r	$f_{r1} / f_{.1}$	-	$f_{rj} / f_{.j}$	-	$f_{rc} / f_{.c}$
		1	-	1	-	1

La distribuzione condizionata di Y rispetto X:

		Y					
		y_1	-	y_j	-	y_c	
X	x_1	$f_{11} / f_{1.}$	-	$f_{1j} / f_{1.}$	-	$f_{1c} / f_{1.}$	1
	-	-	-	-	-	-	-
	x_i	$f_{i1} / f_{i.}$	-	$f_{ij} / f_{i.}$	-	$f_{ic} / f_{i.}$	1
	-	-	-	-	-	-	-
	x_r	$f_{r1} / f_{r.}$	-	$f_{rj} / f_{r.}$	-	$f_{rc} / f_{r.}$	1

3.2 Analisi statistica bivariata

L'analisi statistica bivariata consiste nello stabilire se esiste una qualche **relazione** tra i due fenomeni considerati.

Il **metodo per stabilire se sono statisticamente indipendenti** consiste nel confrontare le frequenze condizionate. Se al variare delle modalità del fenomeno condizionante le distribuzioni condizionate non variano, allora i due fenomeni sono statisticamente indipendenti.

La dipendenza si studia sia per caratteri quantitativi sia qualitativi perchè studiamo le frequenze.

- *Se non esiste alcuna relazione si dirà che X ed Y sono statisticamente indipendenti.*
- *Se due fenomeni non sono statisticamente indipendenti allora esiste una relazione e si dirà che i fenomeni sono connessi.*

Se **esiste una relazione** si procede con delle analisi.

Definizione 3.9. *Contingenze calcolate per misurare il grado di connessione tra due modalità*

$$C_{ij} = f_{i,j} - f^*_{i,j}$$

dove $f^*_{i,j} = \frac{f_{(i,\bullet)} f_{(\bullet,j)}}{N}$ *sono le* ***frequenze teoriche*** *che si avrebbero nel caso ci fosse indipendenza statistica.*

- $C_{ij} > 0 \implies$ c'è ***attrazione***, la connessione è alta.
- $C_{ij} < 0 \implies$ c'è ***repulsione***, la connessione è alta.
- $C_{ij} \simeq 0 \implies$ la connessione è bassa.

Definizione 3.10. *Chi quadro di Pearson*

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - f^*_{ij})^2}{f^*_{ij}} \in [0, Nmin\{(r-1);(c-1)\}]$$

di conseguenza l'indice normalizzato

$$\chi^2_{NORM} = \frac{\chi^2}{\chi^2_{MAX}} = \frac{\chi^2}{Nmin\{(r-1);(c-1)\}} \in [0,1]$$

Se **almeno** uno dei due fenomeni è **quantitativo** ad esempio Y, ci si può chiedere se ***Y dipende in media da X***, ovvero se al cambiare delle modalità di X cambiano le medie di Y.

Definizione 3.11. *Indipendenza in media Si dice che c'è indipendenza in media se tutte le medie condizionate sono tra loro uguali e quindi uguali alla media marginale:*

$$\bar{y}|x_1 = \bar{y}|x_2 = ... = \bar{y}|x_r = \bar{y}$$

quindi se fisso una i^* la media di y condizionata ad x_i è uguale a:

$$\bar{y}|x_{i^*} = \frac{\sum_{j=1}^{c} y_j f_{(i^*,j)}}{f_{(i^*,\bullet)}}$$

$$\bar{y} = \frac{\sum_{i=1}^{r} \bar{y}|x_i f_{(i,\bullet)}}{N}$$

Indipendenza in media

Definizione 3.12. *Eta Quadro Il modo per misurare la dipendenza in media è data da questa formula*

$$\eta^2_{Y|X} = \frac{\sigma^2_B}{\sigma^2_Y} = \frac{\frac{1}{N}\sum_{i=1}^{r}(\bar{y}|x_i - \bar{y})f_{(i,\bullet)}}{\frac{1}{N}\sum_{j=1}^{c}(y_j^2 f_{(\bullet,j)} - \bar{y}^2)} \in [0,1]$$

Teorema 3.1. *Se Y dipende in media da X implica che Y è dipendente statisticamente da*

$$\eta^2_{Y|X} \implies \chi^2 > 0$$

Teorema 3.2. *L'indipendenza in media di Y da X non implica l'indipendenza in media di X da Y*

3.3 Correlazione lineare

Se **entrambi** i fenomeni sono **quantitativi** è possibile andare oltre all'analisi dell'indipendenza in media.

Definizione 3.13. *Covarianza è una misura di variabilità congiunta definita come*

- $\sigma_{xy} = \sigma_{yx}$
- *Quando è nota la distribuzione di frequenza doppia*

  $$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{r}\sum_{i=1}^{c}(x_i - \bar{x})(y_j - \bar{y})f_{ij}$$

  Alternativamente si può calcolare

  $$\sigma_{xy} = \mu_{xy} - \bar{x}\,\bar{y}$$

  dove μ_{xy} è detta media bivariata o media dei prodotti, definita come:

  $$\mu_{xy} = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{c} x_i\, y_i\ f_{ij}$$

- *Quando ho la matrice di dati grezzi*

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

- *Nota che*

$$\sigma_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sigma_x^2$$

Se $(x_i - \bar{x})(y_i - \bar{y})$ *sono discordi in segno allora peseranno negativamente sulla somma rendendo piccola la covarianza ed al contrario se sono concordi aumenteranno il valore della covarianza.*

*Quindi due serie di dati sono **statisticamente incorrelate** se* $\sigma_{xy} = 0$ *e viceversa.*

Teorema 3.3. *Statisticamente indipendenti* $\implies$ *Statisticamente incorrelati*

Definizione 3.14. *L'indice di correlazione è definito come*

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

- *Normalizza la covarianza, perché divide* σ_{xy} *per la radice di* $\sigma_x^2\sigma_y^2$
- *Senza unità di misura, quindi è confrontabile*
- $\rho \in [-1, 1]$ *e dà indicazioni circa il verso e l'intensità della correlazione*
- *Se uguale ad* ± 1 *allora i fenomeni sono perfettamente allineati lungo una retta, il cui coefficiente angolare è* > 0 *se* $\rho > 0$ *e viceversa.*
- *Se uguale a 0 allora sono incorrelati.*

Matrici di Covarianza e Correlazione

- Caso bidimensionale

  $$C = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

  dove $\sigma_{xy} = \sigma_{yx}$

  $$Corr = \begin{bmatrix} \frac{\sigma_{xx}}{\sigma_x^2} & \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\,\sigma_y^2}} \\ \frac{\sigma_{yx}}{\sqrt{\sigma_x^2\,\sigma_y^2}} & \frac{\sigma_{yy}}{\sigma_y^2} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_x^2}{\sigma_x^2} & \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\,\sigma_y^2}} \\ \frac{\sigma_{yx}}{\sqrt{\sigma_x^2\,\sigma_y^2}} & \frac{\sigma_y^2}{\sigma_y^2} \end{bmatrix} = \begin{bmatrix} 1 & \frac{\sigma_{xy}}{\sigma_x\,\sigma_y} \\ \frac{\sigma_{yx}}{\sigma_x\,\sigma_y} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

- Caso m variabili

$$C = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & ... & \sigma_{x_1x_m} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 & ... & \sigma_{x_2x_m} \\ ... & ... & ... & ... \\ \sigma_{x_1x_m} & ... & ... & \sigma_{x_m}^2 \end{bmatrix} \quad Corr = \begin{bmatrix} 1 & \rho_{x_1x_2} & ... & \rho_{x_1x_m} \\ \rho_{x_1x_2} & 1 & ... & \rho_{x_2x_m} \\ ... & ... & ... & ... \\ \rho_{x_1x_m} & ... & ... & 1 \end{bmatrix}$$

Definizione 3.15. *In generale di dice* ***momento k-esimo*** *rispetto a y*

$$M_{k,y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y)^k$$

- ***Media*** *momento primo rispetto all'origine 0*

$$M_{1,0} = \frac{1}{n}\sum_{i=1}^{n}(x_i)$$

- ***Varianza*** *momento secondo rispetto alla media* $\bar{x}$

$$M_{2,\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

3.4 Modello di regressione lineare

Spesso dato un insieme E di coppie di dati $E = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ ci si chiede se esiste una relazione di tipo funzionale tra x e y che descriva soddisfacentemente il legame tra di essi. In tal caso si parla di **analisi di regressione**, che affronta il problema pensando ad uno dei due caratteri come ad una variabile indipendente, ad esempio x, e cerca di stabilire quale funzione $f \in C^k$, all'interno di una specifica classe, consente di scrivere al ***meglio*** il legame

$$y = f(x)$$

Dove y è detta variabile dipendente.
Occorre quindi capire che per ***meglio*** si intende solitamente la funzione f che **minimizza le distanze** tra i valori osservati del carattere y e quelli che si otterrebbero per il carattere y se la relazione fosse proprio quella descritta da $f(x)$.

Bisogna trovare la f che minimizzi questa quantità:

$$g(f) = \sum_{i=1}^{n} [f(x_i) - y_i]^2 \tag{1}$$

Se si limita l'insieme delle funzioni all'insieme di quelle lineari allora si parlerà di **regressione lineare**.

Regressione Lineare Riformulando l'equazione (1) con vincolo di avere

$$f(x) = mx + q$$

$$g(m, q) = \sum_{i=1}^{n} [mx_i + q - y_i]^2 \tag{2}$$

Il problema si riduce alla determinazione dei coefficienti m e q della retta per cui risulti minima (2).

Teorema 3.4. *Sia $f(x) = mx + q$ la funzione utilizzata nel modello di regressione lineare atta a minimizzare la seguente quantità*

$$g(m, q) = \sum_{i=1}^{n} [mx_i + q - y_i]^2 \implies m = \frac{\sigma_{xy}}{\sigma_x^2} \;;\; q = (\bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x})$$

*Dove m e q sono i coefficienti che rendono minima la somma dei quadrati, e la retta ottenuta con quei coefficienti è detta **di regressione lineare**.*

Dimostrazione 2.

$$\sum_{i=1}^{n} [mx_i + q - y_i]^2$$

chiamo $\phi(m,q) = \sum_{i=1}^{n} [mx_i + q - y_i]^2$ *se voglio che la differenza sia minima devo imporre che le derivate parziali di* $\phi(m,q)$ *siano uguali a zero.*

$$\begin{cases} \frac{\partial\phi(m,q)}{\partial m} = 0 \\ \frac{\partial\phi(m,q)}{\partial q} = 0 \end{cases} = \begin{cases} \partial\frac{\sum_{i=1}^{n}[mx_i+q-y_i]^2}{\partial m} = 0 \\ \partial\frac{\sum_{i=1}^{n}[mx_i+q-y_i]^2}{\partial q} = 0 \end{cases}$$

$$= \begin{cases} \frac{\sum_{i=1}^{n}\partial[mx_i+q-y_i]^2}{\partial m} = 0 \\ \frac{\sum_{i=1}^{n}\partial[mx_i+q-y_i]^2}{\partial q} = 0 \end{cases} = \begin{cases} \frac{\sum_{i=1}^{n}\partial[y_i-mx_i-q]^2}{\partial m} = 0 \\ \frac{\sum_{i=1}^{n}\partial[y_i-mx_i-q]^2}{\partial q} = 0 \end{cases}$$

Risolviamo il primo termine del sistema:

$$\frac{\partial\phi(m,q)}{\partial m} = \sum_{i=1}^{n}[2(y_i - mx_i - q)(-x_i)] = -2\sum_{i=1}^{n}[(y_i - mx_i - q)(x_i)] = 0$$

$$\implies \sum_{i=1}^{n}[(y_i - mx_i - q)(x_i)] = 0$$

Risolviamo il secondo termine del sistema:

$$\frac{\partial\phi(m,q)}{\partial m} = \sum_{i=1}^{n}[2(y_i - mx_i - q)(-1)] = -2\sum_{i=1}^{n}[(y_i - mx_i - q)] = 0$$

$$\implies \sum_{i=1}^{n}[(y_i - mx_i - q)] = 0$$

Quindi il sistema:

$$\begin{cases} \sum_{i=1}^{n}[(y_i - mx_i - q)(x_i)] = 0 \\ \sum_{i=1}^{n}[y_i - (mx_i + q)] = 0 \end{cases}$$

Ricavo q dalla seconda:

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} mx_i + \sum_{i=1}^{n} q = \sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i - nq = 0$$

$$nq = \sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i$$

$$q = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{m}{n}\sum_{i=1}^{n} x_i$$

$$q = \bar{y} - m\bar{x}$$

Sostituisco q nella prima:

$$\begin{cases} \sum_{i=1}^{n}[(y_i - mx_i - \bar{y} + m\bar{x})(x_i)] = 0 \\ q = \bar{y} - m\bar{x} \end{cases}$$

Riscrivo la prima equazione:

$$\sum_{i=1}^{n}[(y_i - mx_i - \bar{y} + m\bar{x})(x_i)] = \sum_{i=1}^{n}[((y_i - \bar{y}) + (m\bar{x} - mx_i))(x_i)] =$$

$$\sum_{i=1}^{n}[((y_i - \bar{y}) - m(x_i - \bar{x}))(x_i)] = \sum_{i=1}^{n}(y_i - \bar{y})(x_i) - m\sum_{i=1}^{n}(x_i - \bar{x})(x_i) = 0$$

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

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

Si può concludere quindi:

$$m = \frac{\sigma_{xy}}{\sigma_x^2} \; ; \; q = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x}$$

Dove il numeratore:

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

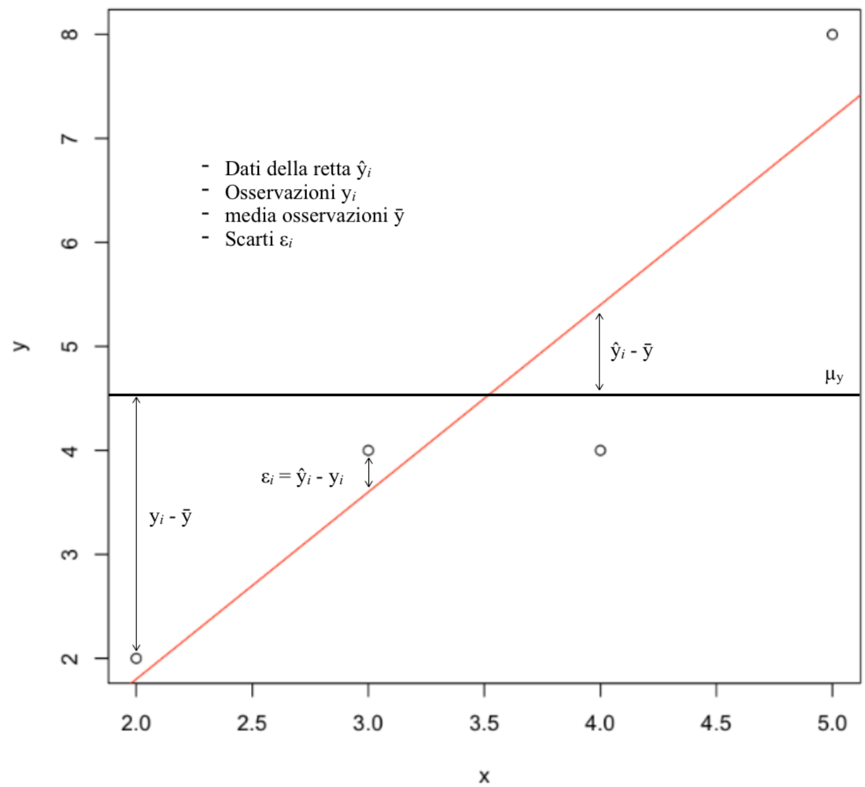
$$= \sum_{i=1}^{n} x_i(y_i - \bar{y}) - \bar{x}\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y})$$

Dove il denominatore:

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

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

Adattamento della retta ai dati Per misurare il grado di adattamento della retta di regressione lineare ai dati si utilizza la scomposizione della varianza applicata alla retta.



Teorema 3.5. *La scomposizione della varianza della retta di regressione lineare*

$$\sigma_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Ossia la varianza totale è uguale alla varianza spiegata più quella residua.

Dimostrazione 3.

$$\sigma_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2$$

$$= \sum_{i=1}^{n}((y_i - \hat{y}_i)^2 + (\hat{y}_i - \bar{y})^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})) =$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

Analizziamo l'ultimo termine:

$$\sum_{i=1}^{n} 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

Noi sappiamo che $\hat{y}_i = f(x) = mx_i + q =$ *e* $\bar{y} = m\bar{x} + q$

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

Quindi:

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

Ma la somma degli scarti di x_i *dalla media è uguale a zero di conseguenza:*

$$\sigma_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

L'indice di adattamento della retta

$$\rho^2 = \frac{var(retta)}{\sigma_y^2} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}$$

4 Probabilità

La probabilità è una disciplina di carattere matematico che permette di affrontare l'analisi delle situazioni che hanno un esito imprevedibile a priori e pertanto con conseguenze incerte.

È quindi una teoria (assiomatica o frequentista o soggettiva) che riguarda il calcolo della probabilità del verificarsi di certi eventi composti di eventi elementari strumento di base per la

Esistono diverse definizioni di probabilità anche se ad oggi non esiste una universale.

- Classica
- Frequentista
- Soggettivista
- Assiomatica, quella presentata.

4.1 Impostazione assiomatica

Definizione 4.1. *Sia Ω un insieme di diversi possibili esisti, ben distinti tra di loro. Ogni sottoinsieme $A \subset \Omega$ viene detto **evento**.*

Definizione 4.2. *$\forall A_i \subset \Omega$ è associata una quantità numerica detta $P(A)$ il cui significato varia dall'impostazione*

Un passo fondamentale per superare le interpretazione sul concetto di probabilità fu fatto da Kolmogorov nel 1933.

Definizione 4.3. *Sia Ω un insieme finito o infinito di elementi esso viene detto **spazio campione** o **spazio campionario**.*
*Ogni sottoinsieme A di Ω viene detto **evento** e può essere:*

- *Elementare se costituito da un singolo elemento di Ω*
- *Composto se costituito da più di un elemento di Ω*

Definizione 4.4. *Si definisce insieme delle parti di Ω, l'insieme di tutti i suoi sottoinsiemi e lo si denota tramite*

$$\wp(\Omega)$$

La probabilità deve essere definita per tutti gli elementi di $\wp(\Omega)$ con particolari proprietà.

Definizione 4.5. *Siano A, B eventi $\subset \Omega$ sono incompatibili se sono disgiunti.*

$$A \cap B = \emptyset$$

Definizione 4.6. *Secondo Kolmogorov viene detta* ***misura di probabilità*** *ogni applicazione*

$$P\colon \wp(\Omega) \mapsto \mathbb{R}_0^+$$

che associa un numero reale ad ogni sottoinsieme di Ω *in cui valgono le seguenti* ***proprietà****:*

- *Per ogni* $A \subset \Omega$ $P(A) \geq 0$. *Interpretando P(A) come frequenza relativa allora essa è compresa tra [0,1]*
- $P(\Omega) = 1$. *È certo che si realizzi uno qualsiasi evento tra tutto quelli di* Ω.
- *Dato* $F = \{A_i, i \in I \subset N\}$ *insieme di eventi incompatibili allora vale che*

$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} P(A_i)$$

Ogni misura di probabilità assegna valori numerici a sottoinsiemi di Ω *e non ai suoi elementi, quindi ad eventi elementari.*

Ne segue dalle proprietà sopra elencate che:

- $\forall A \subseteq \Omega$ $P(\bar{A}) = 1 - P(A)$
- $\forall A \subseteq \Omega$ $P(A) \leq 1$
- $\forall A, B \subseteq \Omega$, $A \subseteq B$ allora $P(A) \leq P(B)$
- $\forall A, B \subseteq \Omega$ vale

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

Dimostrazione 4. $P(\bar{A}) = 1 - P(A)$

$$1 = P(\bar{A} \cup A) = P(\bar{A}) + P(A) - P(\bar{A} \cap A) = P(\bar{A}) + P(A)$$

Definizione 4.7. *Siano A e B due eventi qualsiasi dello spazio campionario e sia P(A)* $\neq$ *0, la probabilità che si verifichi l'evento B dato l'evento A si dice* ***probabilità condizionata*** *di B dato A definita come*

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

di conseguenza vale che:

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

Definizione 4.8. *Due eventi si dicono indipendenti se*

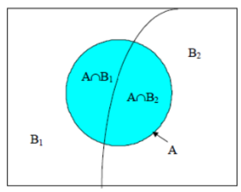
$$P(B|A) = P(B)$$

Quindi la $P(A \cap B) = P(B|A)P(A) = P(B)P(A)$

Teorema 4.1. *Siano* B_1 B_2 *eventi incompatibili e* $B_1 \cup B_2 = \Omega$ *e* $A \subset \Omega$

$$\Longrightarrow$$

$A \cap B_1$ *e* $A \cap B_2$ *sono anch'essi disgiunti.*



Quindi:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2)$$

Teorema 4.2.

5 Variabili Casuali

5.1 Unidimensionali

Definizione 5.1. *Una variabile casuale è una applicazione definita come*

$$X : \Omega \mapsto \mathbb{R}$$

tale per cui $\forall\, \omega \in \Omega$ *si associ un numero* $\mathbb{R}$

$X(\Omega)$ *o supporto può essere:*

- ***Discreto****, associa ad ogni elemento di* Ω *un numero finito di valori o un'infinità numerabile*
- ***Continuo****, associa ad ogni elemento di* Ω *un'infinità di valori non numerabile*

Concetti variabile casuale In base alla definizione, quello che interessa di una variabile casuale è calcolare la probabilità che essa assuma certi valori. Quindi è possibile assegnare delle probabilità ad eventi del tipo

"$X \in B \subseteq \mathbb{R}$" va pensato come un evento di Ω

$$P(\{X \in B\}) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

Caso discreto

Definizione 5.2. *Sia* X *una variabile casuale discreta e siano* $\{x_1, x_2, ...\}$ *il supporto. Si dice funzione di probabilità discreta della v.c* X *la funzione*

$$f_X : \mathbb{Z} \mapsto [0, 1] : z \longmapsto f_X(z) = P(\{\omega \in \Omega : X(\omega) = z\}) = P(X = z)$$

tale che associa ad ogni valore del supporto della variabile aleatoria X *la sua probabilità.*

Proprietà:

- $f(z) \geq 0 \;\; \forall z \in \mathbb{Z}$

Definizione 5.3. *Si definisce funzione di ripartizione di una v.c.* X *la funzione*

$$F_X(t) : \mathbb{R} \mapsto [0, 1]$$

$$t \longmapsto P(X \leq t) = \sum_{k : t_k \leq t} f(t_k) = F_X(t)$$

Dove $x_k \in X(\Omega)$ *e* $f(x_k) = P(X = x_k)$

Proprietà:

- *Monotona crescente*
- *Il suo codominio è a gradino, (siamo nel caso discreto).*
- *Deve valere che*

$$\lim_{t \mapsto +\infty} F_X(t) = 1$$

$$\lim_{t \mapsto 0} F_X(t) = 0$$

$$\lim_{t \mapsto t_0^+} F_X(t) = F_X(t_0)$$

- $\forall\ \alpha \in [0, 1]$ *si dice quantile di ordine* α*, quel valore* x_α *tale per cui*

$$F(x_\alpha) = \alpha$$

- $\sum_{k=1,2,..} f(x_k) = 1$ *estesa a tutti i valori assunti dalla v.c. X*

Caso continuo

Definizione 5.4. *Si dice variabile aleatoria continua se la corrispondente funzione di ripartizione* $F_X(t)$ *è continua. In particolare si dice* ***assolutamente continua*** *se esiste una funzione*

$$f_X : \mathbb{R} \mapsto \mathbb{R}^+ : t \longmapsto f_X(t)$$

tale che verifichi

$$F_X(t) = \int_{-\infty}^{t} f_X(u)du\ \ \forall t \in \mathbb{R}$$

dove f_X *è detta* ***densità di probabilità*** *della v.c X con le seguenti proprietà:*

- $f_X(t) \geq 0$
- $\int_{-\infty}^{+\infty} f_X(t)dt = 1$

Definizione 5.5. *Si dice supporto della v.c. X l'insieme*

$$S = \{t \in \mathbb{R} : f_x(t) \neq 0\}$$

Osservazioni

1. La probabilità di una v.c. X di assumere un valore preciso$P(X = t_0) = 0$

   Dimostrazione 5.

$$P(X = t_0) = P(x \leq t_0) - \lim_{t->t_0^-} P(X \leq t) =$$

$$= F_X(t_0) - P(X \leq t_0) = F_X(t_0) - F_X(t_0) = 0$$

2. A tale proposito *(1.)* ha invece senso chiedersi

$$P(X \in (a,b]) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(u)du - \int_{-\infty}^{a} f_X(u)du$$

$$\int_{a}^{b} f_X(u)du \ , \ a < b$$

3. Quindi se $\exists F_X(t)$ la sua funzione di densità è definita come:

$$f_x(t) = \frac{\partial F_X(t)}{\partial t}$$

5.2 Multidimensionali

Definizione 5.6. *Si dicono v.c. multidimensionale assolutamente continue, l'applicazione*

$$(X_1, X_2, ..., X_n) \colon \Omega \mapsto \mathbb{R}^n$$

Corollario *Nel caso bidimensionale* $(X, Y) \colon \Omega \mapsto \mathbb{R}^2$

Definizione 5.7. *Siano* X, Y *due v.c. tale che* $(X, Y) \colon \Omega \mapsto \mathbb{R}^2$ *Allora si dice funzione di ripartizione congiunta:*

$$F_{X,Y} \colon \mathbb{R}^2 \mapsto [0, 1] \subseteq \mathbb{R} \ : \ (t, s) \longmapsto P(\{X \leq t\} \cap \{Y \leq s\})$$

Definizione 5.8. *Siano* X, Y *due v.c. tale che* $(X, Y) \colon \Omega \mapsto \mathbb{R}^2$ *si dicono assolutamente continue se* $\exists$

$$f_{XY}(t, s) \colon \mathbb{R}^2 \mapsto \mathbb{R}^+$$

tale che:

$$F_{XY}(t, s) = \int_{-\infty}^{t} \int_{-\infty}^{s} f_{XY}(u, v)dudv \ \forall (t, s) \in \mathbb{R}^2$$

Osservazioni

1. Data la $F_{XY}(t, s)$ la coppia di v.c X,Y assume valori in un qualsiasi sottoinsieme rettangolare $(a_1, b_1) \times (a_2, b_2)$

2. Definiamo la probabilità

$$P((X, Y) \in (a_1, b_1] \times (a_2, b_2]) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{XY}(u, v)dudv$$

Definizione 5.9. *Siano* X, Y *due v.c. tale che* $(X, Y) \colon \Omega \mapsto \mathbb{R}^2$ *si dicono stocasticamente indipendenti*

$$\rightleftharpoons \forall (t, s) \in \mathbb{R}^2 \ vale \ F_{XY}(t, s) = F_X(t)F_Y(s)$$

Ne segue che:

$$P(\{X \in A\} \cap \{Y \cap B\}) = P(X \in A)P(Y \in B) \ \forall A, B \subseteq \mathbb{R}$$

Ed inoltre segue che quando sono assolutamente continue:

$$f_{XY}(t,s) = D(F_X(t)) * F_Y(s) + D(F_Y(s)) * F_X(t)$$

5.3 Indici di tendenza centrale

Analogamente a quanto visto nella *sezione 5.1* gli indici di tendenza centrale associate alle variabili aleatorie sono grandezze numeriche in grado di sintetizzare, con un solo valore, le principali caratteristiche delle loro distrubuzioni.

Quindi data una v.c. X unidimensionale con supporto $B \subseteq \mathbb{R}$ dove #B=n e $\forall x \in B$ è associata la sua $f_X(x)$

- **Moda**, definita come la quantità $\tilde{X} \in B$
  - Se X v.c. ***discreta***: che corrisponde al massimo valore della funzione di probabilità discreta.
  - Se X v.c. ***continua***: che corrisponde al massimo valore della funzione di densità di probabilità $f_X(t)$.

  Potrebbe capitare di avere diversi punti di massimo per le due funzioni, in analogia alla statistica descrittiva, si parlerà di distribuzioni multimodali.

- **Mediana**, definita come una quantità $\hat{X} \in B$ che soddisfa

  $$\lim_{x \mapsto \hat{X}^-} F_X(x) \leq \frac{1}{2} \leq F_X(\hat{X})$$

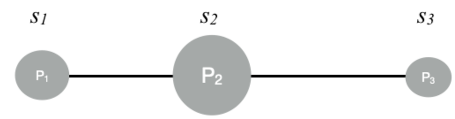
  Può non essere unica, infatti possono esistere diversi $t : F_X(t) = \frac{1}{2}$.

- **Valore atteso**, corrispondente alla media dei dati statistici.
  - Se X v.c. ***discreta***:

    $$E[X] = \sum_{i=1}^{n} x_i f_X(x_i) = x_1 P(X = x_1) + x_2 P(X = x_2) + ... + x_n P(X = x_n) =$$

    $$= x_1 \frac{\gamma_1}{n} + x_2 \frac{\gamma_2}{n} + ... + x_n \frac{\gamma_n}{n} = \frac{x_1\gamma_1 + x_2\gamma_2 + ... + x_n\gamma_n}{n}$$

    Ossia significa trovare il baricentro della nostra distribuzione, ipotizzando che $\frac{\gamma_j}{n} = P(X = x_j)$ siano i pesi. Quindi fornisce un'indicazione di massima del posizionamento della variabile lungo l'asse dei numeri reali.
    Dove $P(X = x_j) \in [0,1] \ \forall j$



Corollario La *media aritmetica* è un caso particolare del valore atteso, ossia quando i pesi sono uguali da 1, in altri termini $\gamma_j = 1 \; \forall j$

$$E[X] = x_1 \frac{\gamma_1}{n} + x_2 \frac{\gamma_2}{n} + ... + x_n \frac{\gamma_n}{n} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n} \sum_{i}^{n} x_i = \bar{X}$$

– Se X v.c. ***continua***:

$$E[X] = \int_{-\infty}^{+\infty} u \, f_X(u) \; du$$

Proprietà del valore atteso

1. $\forall x_i \in X$ se $x_i = \alpha$ con $P(X = \alpha) = 1 \iff E[X] = \alpha$

   Dimostrazione 6.

   $$E[X] = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i}^{n} \alpha \frac{1}{n} = \frac{\alpha}{n} \sum_{i}^{n} = \frac{\alpha}{n} n = \alpha$$

   In maniera analoga si dimostra ($\implies$ *)*
2. Se $Y = aX + b \implies E[Y] = E[aX + b] = aE[X] + E[b] = aE[X] + b$
3. Sia $y = g(X)$ funzione della variabile casuale

   $$E[y] = \int_{-\infty}^{+\infty} g(u) f_X(u) du$$

5.4 Indici di variabilità

- **Varianza**, è una misura della dispersione dei valori della variabile casuale X attorno al valor medio $\mu = E[X]$.

  $$\sigma^2 = var[X] = E[(X - \mu)^2]$$

  – Se X v.c. ***discreto***:

  $$E[(X - \mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) \, , \; f(x_i) = P(X = x_i) \; \forall i$$

– Se X v.c. ***continua***:

$$E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx$$

Proprietà della varianza

1. $\forall x_i \in X$ se $x_i = \alpha$, con $P(X = x_i) = 1 \iff var[X] = 0$

   Dimostrazione 7.

   $$var[X] = E[(x-\mu)^2] = E[(\alpha-\alpha)^2] = 0$$

2. $var[aX + b] = var[aX] + var[b] = a^2 var[X]$

- **Scarto quadratico medio** è calcolato come la radice quadrato della varianza.

  $$\sigma = \sqrt{var(X)} = \sqrt{E[(X-\mu)^2]}$$

  – Se X v.c. ***discreto***:

  $$\sqrt{E[(X-\mu)^2]} = \sqrt{\sum_{i=1}^{n} (x_i-\mu)^2 f(x_i)}$$

  – Se X v.c. ***continua***:

  $$\sqrt{E[(X-\mu)^2]} = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx}$$

- **Covarianza** definita come

  $$cov[XY] = E[(x-\mu_X)]E[(y-\mu_Y)]$$

Utili relazioni Siano X, Y una coppia di variabili casuali

– $E[X \pm Y] = E[X] \pm E[Y]$

– $E[XY] = E[X]E[Y] \iff X, Y$ sono stocasticamente indipendenti.

– $var[X \pm Y] = var[X] + var[Y]$

– $cov[XY] = 0 \iff X, Y$ sono incorrelate.

Momento centrale di X di ordine r Anche per quanto riguarda una variabile casuale X è possibile definire

$$E[X^r] = \begin{cases} \sum_i^n x_i^r \, f_X(x_i) & discreta \\ \int_{-\infty}^{+\infty} x f_X(x) dx & continua \end{cases}$$

- $r = 1,\ E[X] \rightleftharpoons$ valore atteso
- $r = 2,\ E[X^2] \rightleftharpoons$ varianza
- $r = 3,\ E[X^3] \rightleftharpoons$ asimmetria
- $r = 4,\ E[X^4] \rightleftharpoons$ curtosi

Disuguaglianza di Chebishev Sia X una variabile aleatoria con

$$E[X] = \mu \ e \ var[X] = \sigma^2 \implies \forall \varepsilon > 0 \ P(|X - \mu| < \varepsilon) \leq 1 - \frac{\sigma^2}{\varepsilon^2}$$

Questo ci suggerisce che se $\sigma^2 \simeq 0 \implies P(|X - \mu| < \varepsilon) \simeq 1$, quindi più la varianza è piccola più i valori della variabile aleatoria saranno $\in [\mu - \varepsilon, \mu + \varepsilon]$. Tuttavia questa è una stima e non è sempre affidabile.

5.5 Distribuzioni discrete

5.6 Distribuzioni continue

Una distribuzione

5.7 Teoremi fondamentali

Definizione 5.10. ***Convergenza in distribuzione*** *Sia* $\{X_n, n \in \mathbb{N}\}$ *una successione di variabili aleatorie* ***i.i.d.****.* F_n *è la corrispondente funzione di ripartizione per l'n-esima variabile aleatoria. Si dice che la successione converge in distribuzione* $\rightleftharpoons \lim_{n \to \infty} F_n(t) = F(t) \ \forall t \in \mathbb{R}$.

Questo può essere anche scritto secondo la seguente notazione

$$F_n \xrightarrow{d} F \ \ per \ \ n \longrightarrow \infty$$

Teorema 5.1. ***Legge dei Grandi Numeri***

Sia $\{X_i, i \in \mathbb{N}\}$ *una successione di variabili aleatorie i.i.d., e definiamo*

$$\bar{X_n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

La variabile aleatoria ***media aritmetica n-esima*** *della successione.*

Consideriamo che le variabili aleatorie X_i *abbiano:*

$$E[X_i] = \mu \ \ var[X_i] = \sigma^2$$

Allora possiamo dire che:

$$\bar{X_n} \xrightarrow{d} M$$

Dove M è una variabile aleatoria che assume valore μ con probabilità 1.

$$M \xrightarrow{p} \mu \iff P(|M - \mu| < \varepsilon) \to 1 \ \forall \varepsilon > 0$$

Teorema 5.2. ***Teorema del limite centrale***

La legge dei grandi non ci dice con quale rapidità, o quanto debba essere grande n, per assicurarci che una realizzazione della variabile aleatoria $\bar{X_n}$, definita come $\bar{x_n}$ sia uguale o prossimo a μ.

Da notare che la convergenza sarà più rapida se σ^2 è piccola. Il Teorema del limite centrale formalizza questo concetto, specificando

- *La distribuzione di $\bar{X_n}$*
- *Il valore atteso di $\bar{X_n}$*
- *La varianza di $\bar{X_n}$*

Per n sufficientemente grande.

Sia $\{\bar{X_i}, n \in \mathbb{N}\}$ una successione di variabili aleatorie i.i.d., che soddisfano le ipotesi della legge dei grandi numeri.

Quindi avente identico valore atteso e varianza:

$$E[X_i] = \mu \quad var[X_i] = \sigma^2$$

Consideriamo

$$S_n = X_1 + X_2 + ... + X_n$$

quindi si ha che:

$$E[S_n] = E[X_1 + X_2 + ... + X_n] = E[X_1] + E[X_2] + ... + E[X_n] = n\mu$$

$$var[S_n] = var[X_1 + X_2 + ... + X_n] = var[X_1] + ... + var[X_n] = n\sigma^2$$

si può dunque concludere che:

$$S_n \xrightarrow{d} X \sim N(n\mu, n\sigma^2) = N(n\mu, \sqrt{n}\sigma)$$

Ora definendo:

$$\bar{X_n} = \frac{S_n}{n}$$

si ha:

$$E[\bar{X_n}] = E\left[\frac{S_n}{n}\right] = \frac{1}{n}E[S_n] = \frac{n\mu}{n} = \mu$$

$$var[\bar{X_n}] = var\left[\frac{S_n}{n}\right] = \frac{1}{n^2}var[S_n] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} = \frac{\sigma}{\sqrt{n}}$$

Concludiamo quindi:

$$\bar{X_n} \xrightarrow{d} X \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

6 Teoria dei Campioni e Stima

In questa sezione si farà riferimento alla **statistica inferenziale** la quale consentirà di dedurre particolari *caratteristiche* di una popolazione limitandosi ad analizzare un campione di essa. Queste caratteristiche, se espresse numericamente, vengono dette **parametri**.

Definizione 6.1. *Per stima di parametri ci si riferisce al problema della deduzione di caratteristiche di tipo numerico di un a popolazione facendo riferimento ad un sottoinsieme finito opportunamente scelto detto **campione**.*

*Esistono diverse tecnica per stimare i parametri, in questo caso ci si limiterà a considerare quelle sulla conoscenza delle **distribuzioni campionarie**.*

Definizione 6.2. *Il campionamento è una scelta di alcuni individui da una popolazione, per effettuare inferenza sulla stessa.*

*Esistono anche qui diversi modi per campionare ma in questo caso ci si focalizzerà sulla tecnica di **campionamento casuale**, i.e. ogni unità ha la stessa probabilità di essere selezionata. Si differenziano due varianti:*

- ***con remissione**, ossia non si altera il numero della popolazione iniziale, detto anche campionamento Bernoulliano.*
- ***senza remissione**, ossia una volta selezionata un'unità dalla popolazione, essa non viene reinserita nella popolazione.*

La scelta tra queste due tecniche diviene rilevante quando la popolazione considerata è di numerosità limitata e viceversa quando è illimitata.

Formalmente Consideriamo X il carattere della popolazione di nostro interesse, dove il valore assunto dipende dall'individuo considerato.

Quindi si pensi ad X come ad una variabile casuale, avente funzione di rpartizione F, sconosciuta, ma corrispondente alla distribuzione di frequenza cumulata di tale carattere che si otterrebbe nel caso in cui avessimo l'intera popolazione. E pensare ai valori assunti dai singoli individui come a delle realizzazione della stessa x_i.

Quindi un campione casuale di numerosità n:

$$(X_1, X_2, ..., X_n)$$

Ossia una n-upla di **v.c. stocasticamente** indipendenti aventi ognuna la stessa distribuzione del carattere X della popolazione.

Dove:

$$(x_1, x_2, ..., x_n)$$

Sono le realizzazioni

Definizione 6.3. *Per parametro si intende un **valore numerico** che descrive una caratteristica della popolazione.*

Definizione 6.4. *Per stima si intende una* ***misura*** *che descrive una caratteristica del campione o meglio un'espressione funzionale delle realizzazioni.*

$$H_n = h(X_1, X_2, ..., X_n)$$

Dove h è una funzione ad n variabili. In pratica una stima è una realizzazione di una ***statistica campionaria****.*

In particolare H_n essendo funzione di variabili aleatorie allora anch'essa è una v.c. con una sua distribuzione.

6.1 Stimatori e stime puntuali

Se la stima di un parametro della popolazione è data solamente da un numero allora si dice **stima puntuale**. Di solito i parametri che si stimano per una popolazione sono

- μ media di una popolazione
- σ^2 varianza di una popolazione
- p proporzione di individui di una popolazione che appartengono ad una classe di interesse
- $\mu_1 - \mu_2$ differenze tra medie di popolazioni

Definizione 6.5. *Sia θ un parametro incognito della popolazione X.*

$$H_n = h(X_1, X_2, ..., X_n)$$

è una statistica campionaria detta ***stimatore puntuale*** *quando viene utilizzata per stimare θ. Mentre si dice* ***stima puntuale****, invece il valore stimato del parametro θ tramite:*

$$\hat{\theta} = h(X_1, X_2, ..., X_n)$$

Proprietà di uno statistica campionaria Affinché una statistica campionaria possa essere considerata come uno **stimatore puntuale** deve godere delle seguenti proprietà:

Sia $H_n = h(X_1, ..., X_n)$ stimatore del parametro θ esso è detto

1. **Proprietà di correttezza**

   $$E[H_n] = E[h(X_1, ..., X_n)] = \theta$$

2. **Proprietà di consistenza**

   $$\lim_{n\to\infty} P(|H_n - \theta| \leq \varepsilon) = 1$$

   Non è sempre facile verificare la consistenza di conseguenza è possibile dimostrare che uno stimatore corretto è anche consistente se

   $$\lim_{n\to\infty} var[H_n] = 0$$

3. **Efficienza** Dati due stimatori $T_{n1},\ T_{n2}$ per uno stesso parametro θ

$$e = \frac{E[T_{n1} - \theta]}{E[T_{n2} - \theta]}$$

Se $e > 1 \iff T_{n2}$ è più efficiente di T_{n1}. Nel caso in cui entrambe i due stimatori sono corretti allora si guarda quello con varianza minima.

Stimatori puntuali Considerando un campione $(X_1, X_2, ..., X_n)$ estratto da una popolazione con distribuzione F media uguale μ e varianza pari a σ^2.

- **Media campionaria n-sima** si definisce media campionaria la variabile

$$\bar{X_n} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Dove la distribuzione di $\bar{X_n}$ è detta ***distribuzione della media campionaria***.

Essendo difficile trovare l'espressione analitica di $\bar{X_n}$ si determinerà il valore atteso e la varianza di questa v.c. .

$$E[\bar{X_n}] = E\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] = \frac{1}{n}E[X_1 + ... + X_n]$$

Sapendo che $X_1, X_2, ..., X_n$ sono i.i.d. con valore atteso uguale a μ e varianza pari a σ^2

$$E[\bar{X_n}] = \frac{n\mu}{n} = \mu$$

Si noti come questo stimatore è ***corretto***.

Mentre per quanto riguarda la varianza dello stimatore si ha che

$$var[\bar{X_n}] = var\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] = \frac{\sigma^2 n}{n^2} = \frac{\sigma^2}{n}$$

Visto che la varianza dipende dalla numerosità del campione, si conclude che più elevato sarà n tanto più il valore di $\bar{x_n}$, realizzazione di $\bar{X_n}$, è vicino a μ.

Si dimostra anche che questo stimatore è consistente in quanto

$$\lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

- **Varianza campionaria n-sima** questa viene definita come

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

Anche qui è difficile trovare l'espressione analitica di S_n^2 di conseguenza viene calcolato

$$E[S_n^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i - \bar{X_n})^2\right]$$

$$= \frac{1}{n}E\left[\sum_{i}^{n} X_i^2 - 2X_i\bar{X_n} + \bar{X}^2\right] = \frac{1}{n}E\left[\sum_{i}^{n} X_i^2 - 2\bar{X_n}\sum_{i}^{n} X_i + \bar{X_n}^2\sum_{i}^{n}\right] =$$

$$= \frac{1}{n}E\left[\sum_{i}^{n}(X_i^2) - 2\bar{X_n}n\bar{X_n} + n\bar{X_n}^2\right] = \frac{1}{n}E\left[\sum_{i}^{n}(X_i^2) - 2n\bar{X_n}^2 + n\bar{X_n}^2\right]$$

$$\frac{1}{n}E\left[\sum_{i}^{n} X_i^2 - n\bar{X_n}^2\right] = \frac{1}{n}\left\{\sum_{i}^{n} E[X_i^2] - E[n\bar{X_n}^2]\right\} =$$

$$= \frac{1}{n}\left\{\sum_{i}^{n}(E[X_i^2] = var[X_i] + E[X_i]^2) - nE[\bar{X_n}^2]\right\} =$$

$$= \frac{1}{n}\left\{\sum_{i}^{n}(\sigma^2 + \mu^2) - n(E[\bar{X_n}^2] = \frac{\sigma^2}{n} + \mu^2)\right\} =$$

$$\frac{1}{n}\left\{\sum_{i}^{n}(\sigma^2 + \mu^2) - n\frac{\sigma^2}{n} - n\mu^2\right\} = \frac{1}{n}\left\{n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right\}$$

$$= \boxed{\sigma^2\frac{(n-1)}{n}}$$

Bisogna ***notare il coefficiente*** $\frac{n-1}{n}$, di conseguenza si può definire **Varianza Campionaria Corretta**

$$\hat{S_n^2} = \frac{n}{(n-1)}S_n^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X_n})^2$$

Quindi il valore atteso di quest'ultima è pari a:

$$E[\hat{S_n^2}] = \sigma^2$$

Di conseguenza si conclude che la varianza campionaria è uno stimatore ***corretto*** per σ^2.

Per quanto riguarda la varianza dello stimatore della *varianza campionaria n-sima* si può dimostrare che:

$$var[S_n^2] = \frac{1}{n}\left(E[(X_i - \bar{X_n})^4] - \frac{n-3}{n-4}\sigma^4\right)$$

Quindi per lo stimatore di varianza campionaria corretta:

$$var[\hat{S_n^2}] = var\left[\frac{n}{(n-1)}\, S_n^2\right] = \frac{n^2}{(n-1)^2}\; var[S_n^2]$$

E si può inoltre dimostrare che è ***consistente*** in quanto

$$\lim_{n\to\infty} \frac{n^2}{(n-1)^2}\; var[S_n^2] = 0$$

N.B. La radice della varianza campionaria corretta contro intuitivamente ***non*** è uno stimatore corretto per la deviazione standard della popolazione.

$$E\left[\sqrt{\hat{S_n^2}}\,\right] \neq \sigma$$