
STATISTICA

Corso A

Autore

Giuseppe Acocella

2024/25

Ultima Compilazione - April 4, 2025

Contents

1	Statistica Descrittiva	4
1.1	Frequenze e Campioni	4
1.2	Caratteri e Rappresentazioni Grafiche	4
1.2.1	Classi di grafici	6
1.3	Indici	7
1.3.1	Indici di Centralità - (Media, Mediana, Moda)	7
1.3.2	Indici di Dispersione - (Varianza, Deviazione Standard)	8
1.4	Funzione di Ripartizione Empirica (FDR/ECDF)	8
1.5	Quantili, Percentili, BoxPlot, Outlier	8
2	Dati Bivariati	9
2.1	Scatterplot e Regressione	9
2.1.1	Covarianza e Correlazioni Campionarie	9
2.1.2	Th. Retta di Regressione ai Minimi Quadrati	10
3	Probabilità	11
3.1	Modello Uniforme	12
3.1.1	Accenno di Combinatoria	13
3.2	Probabilità Condizionata	13
3.2.1	Regole Prodotto, Catena, Fattorizzazione della Probab. Condizionata	13
3.3	Formula di Bayes	14
3.4	Indipendenza	14
3.4.1	Schema di Bernoulli	15
4	Variabili Aleatorie	15
4.1	Variabili Aleatorie Discrete	15
4.2	Variabili Aleatorie Notevoli	16
4.3	Variabili Aleatorie con Densità	17
4.3.1	Variabili Aleatorie con Densità Note	18
4.4	Funzione di Ripartizione e Quantili	19
4.5	Variabili Aleatorie Gaussiane	20
4.5.1	Trasformazioni di Variabili Aleatorie con Densità	20
4.5.2	Variabile Aleatoria Gaussiana Generale	20
4.6	Valore Atteso e Momenti	21
4.6.1	Calcolo e Proprietà del Valore Atteso	22
4.7	Varianza e Deviazione Standard	23
4.8	Valore Atteso e Varianza per Esempi Notevoli	24
4.9	Variabile Aleatoria Doppia	25
4.10	Indipendenza di Variabili Aleatorie	26
4.10.1	Riproducibilità data l'Indipendenza	27
4.11	Covarianza e Correlazione	27
4.11.1	Valore Atteso e Prodotto di v.a Indipendenti	27
4.12	Teoremi Limite in Probabilità	29
4.12.1	Teorema - Legge dei Grandi Numeri	29

4.12.2	Teorema Centrale del Limite - Limite Centrale (TCL/TLC)	30
--------	---	----

1 Statistica Descrittiva

Questo ramo della statistica cerca di raccogliere dati per descrivere degli oggetti. Elenchiamo delle definizioni standard:

1. **Popolazione:** Insieme di oggetti da studiare.
2. **Carattere:** Caratteristiche degli oggetti della popolazione.

(a) Colore di una biglia, altezza di un individuo.

Ricordiamo che un carattere può essere sia **qualitativo** (es. colore) sia **quantitativo** (es. altezza).

3. **Modalità:** Possibili valori che il carattere può assumere.
(a) Colore biglia istanziato: rosso, blu. Lancio moneta istanziato: testa/croce.
4. **Campione Statistico (Sample):** Sottoinsieme della popolazione scelto a rappresentarla.
5. **Dati:** Esiti delle misure del carattere sugli individui del campione.
(a) Lanci moneta: T, C, T, T, T, C, \dots
6. **Taglia Campione:** Numero di elementi nel campione.

1.1 Frequenze e Campioni

Abbiamo due tipi di frequenze:

1. **Frequenza Assoluta:** Corrisponde al numero di volte in cui la **modalità** appare nei **dati**:

$$\#\{ i \mid x_i = a \}$$

2. **Frequenza Relativa:** Corrisponde al numero di volte in cui la **modalità** appare nei dati fratto il numero dei dati stessi:

$$\text{frequenza relativa} = \frac{\text{frequenza assoluta di } a}{\text{taglia campione}}$$

1.2 Caratteri e Rappresentazioni Grafiche

La rappresentazione dei dati dipende fortemente dal tipo di **carattere**:

1. **Carattere Discreto:** Quantità piccola e finita di modalità assumibili.
(a) Lancio di un dado, esiti di un sondaggio.

In questo caso per le rappresentazioni si utilizzano **diagrammi a barre**.

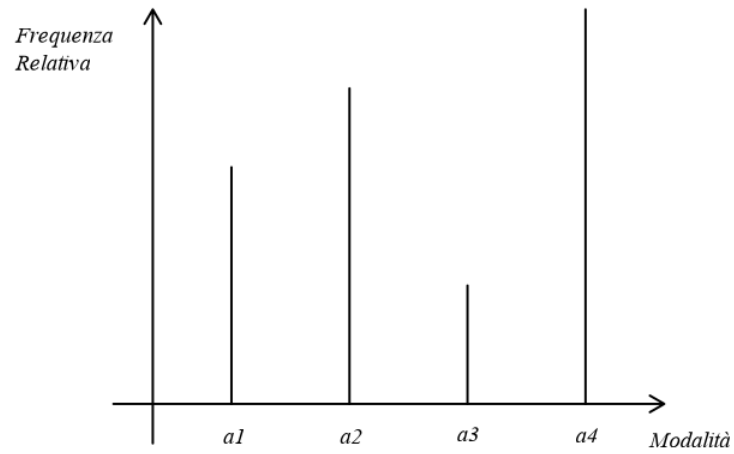


Figure 1: Esempio di diagramma a barre.

2. **Carattere Continuo:** Quantità assumibili in un intervallo continuo.

(a) Altezza della popolazione.

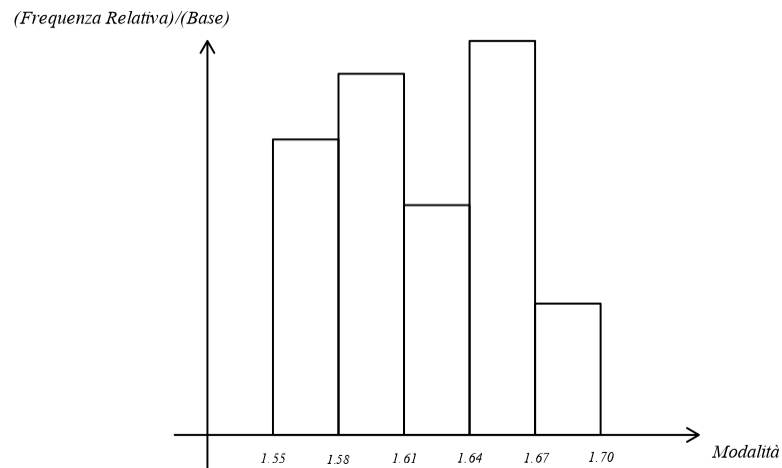


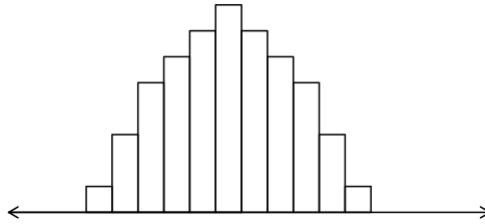
Figure 2: Esempio di istogramma.

La scelta di mettere sull'asse y il rapporto tra freq. relativa e base non è casuale, infatti se scegliessimo intervalli di ampiezza diversa si andrebbe in contro ad un errore di rappresentazione.

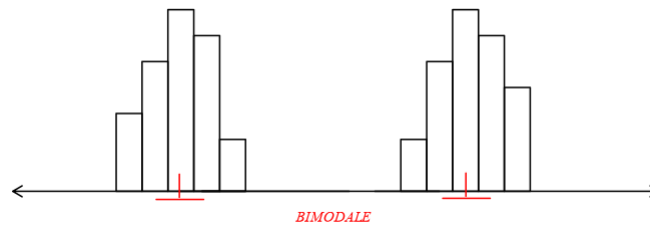
1.2.1 Classi di grafici

Elenchiamo qualche classificazione di rappresentazioni grafiche:

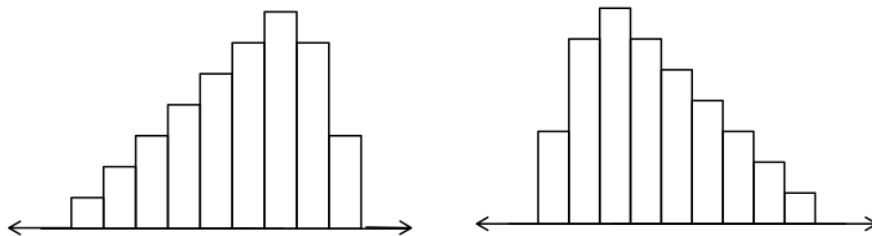
1. **Normale:** Simile ad una campana simmetrica:



2. **Uni/Bi/Tri Modale:** Si concentra attorno ad un numero k di colonne più alte:



- (a) **Modale Asimmetrico S_x/D_x :** Si concentrano attorno ad una colonna più alta in maniera asimmetrica:



1.3 Indici

Gli indici statistici sono quantità numeriche che riassumono proprietà significative sulla **distribuzione dei dati**.

1.3.1 Indici di Centralità - (Media, Mediana, Moda)

Descriviamo tre tipi di indici di centralità:

1. **Media Campionaria:** Descriviamo questo indice:

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

ossia semplicemente la media aritmetica dei dati. Un modo **alternativo** è rappresentarlo così:

$$\frac{\text{modalità} * \text{frequenza ass. della modalità}}{\text{quantità di dati}}$$

che formalmente si esprime così:

$$\bar{x} = \sum_{j=1}^M a_j p(a_j)$$

dove a_j sta per modalità e $p(a_j)$ sta per frequenza relativa della modalità.

Sensibilità ai Valori Estremi Una delle caratteristiche della media campionaria è quella di essere molto sensibile ai valori estremi del campione.

Caratteristiche Riesce a vedere tutti i dati del campione e gode di alcune proprietà matematiche come la linearità.

2. **Mediana Campionaria:** Il dato x_i sarà **centrale**, dunque avrà metà dei dati a sinistra e metà a destra. La calcoliamo dunque in due modi:

- (a) **Numero dispari di modalità:** Dato centrale.

$$\text{mediana} = x_{\left(\frac{n+1}{2}\right)}$$

- (b) **Numero pari di modalità:** Media tra i due dati centrali.

$$\text{mediana} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

Caratteristiche Più robusta rispetto ai valori estremi.

3. **Moda:** Modalità più frequente tra i dati.

1.3.2 Indici di Dispersione - (Varianza, Deviazione Standard)

Gli indici di dispersione ci permettono di stabilire quanto i valori della distribuzione si allontanano da un valore centrale scelto come riferimento. Elenchiamoli:

1. **Varianza Campionaria/Empirica:** Permette di

$$\text{CAMPIONARIA: } Var(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{EMPIRICA: } Var_e(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2$$

E' possibile calcolare la varianza anche con le frequenze relative:

$$Var_e(x) = \left(\sum_{j=1}^M a_j^2 * p(a_j) \right) - \bar{x}^2$$

2. **Scarto Quadratico Medio:** Indice basato sulla varianza.

$$\sigma(x) = \sqrt{Var(x)}$$

3. **Indice Campionario di Asimmetria:** Un indice che permette di stabilire se una distribuzione sia o meno asimmetrica:

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

- (a) $b > 0$: Distribuzione Asimmetrica a destra.
- (b) $b < 0$: Distribuzione Asimmetrica a sinistra.

1.4 Funzione di Ripartizione Empirica (FDR/ECDF)

Dati x_1, x_2, \dots, x_n dati quantitativi definiamo $F_e : \mathbb{R} \rightarrow \mathbb{R}$ data da

$$F_e(t) = \frac{\#\{i \mid x_i \leq t\}}{n}$$

1.5 Quantili, Percentili, BoxPlot, Outlier

Dati x_1, x_2, \dots, x_n dati quantitativi, k un numero naturale tra 0 e 100 allora, avendo un $\beta = \frac{k}{100} \in (0, 1)$:

Definizione k -esimo percentile o β - quartile è il dato $r_\beta = x_i$ tale che:

1. Almeno il $k\%$ dei dati sia inferiore o uguale ad x_i .
2. Almeno il $(100 - k)\%$ dei dati sia superiore o uguale ad x_i .

Se i due dati soddisfano questa condizione, si prende come k-esimo percentile la loro media aritmetica.

Outlier Valore anomalo che differisce in modo significativo dalla maggioranza dei dati.

10, 22, 23, 5, 23, 7, 1001

In questo caso l'Outlier risulta essere 1001 e può essere causato da errori di misurazione o da dati molto rari. Gli outlier influenzano in modo significativo alcuni indici statistici: media, varianza. Altri indici sono invece poco influenzati: mediana, percentili.

2 Dati Bivariati

Consideriamo coppie di dati (due caratteri di un individuo di un campione), analizzando se esistono eventuali relazioni tra i dati x_i ed y_i .

2.1 Scatterplot e Regressione

In uno scatterplot sono riportati i dati bivariati (x_i, y_i) sul piano cartesiano.

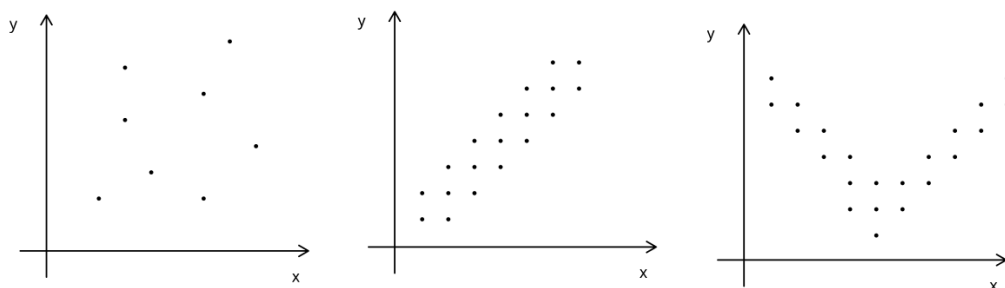


Figure 3: Da sinistra verso destra: Scatterplot Generico, Scatterplot Relazione Lineare, Scatterplot Relazione Non Lineare

Regressione Dato un insieme di punti (x_i, y_i) come possiamo determinare quale curva nel piano approssima meglio quell'insieme? Ci occuperemo solo di relazioni lineari, la cui classe di curve corrisponde alle rette.

2.1.1 Covarianza e Correlazioni Campionarie

Definiamo prima la notazione necessaria:

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

Covarianza Campionaria Empirica tra x ed y

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

$$Cov(x) = \frac{n}{n-1} Cov_e(x, y)$$

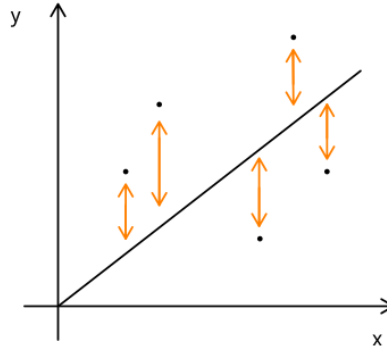
Coefficiente di Correlazione Campionario tra x e y Supponiamo $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$ allora:

$$r(x, y) = \frac{Cov(x, y)}{\sigma(x) \sigma(y)}$$

2.1.2 Th. Retta di Regressione ai Minimi Quadrati

Vogliamo trovare la retta di equazione $y = a + bx$ che meglio approssima un insieme di dati bivariati (x_i, y_i) . Risulta quindi necessario trovare a e b in modo tale da minimizzare

$$\min_{a, b \in \mathbb{R}} \sum_{i=1}^n (y_i - (a + bx_i))^2$$



Teorema Ricaviamo dunque a, b della funzione $y = ax + b$:

$$b = b^* = \frac{Cov(x, y)}{Var x} \quad a = a^* = \bar{y} - b^* \bar{x}$$

Dunque sostituendo nella funzione da minimizzare definita sopra:

$$\min_{a, b \in \mathbb{R}} \sum_{i=1}^n (y_i - (a + bx_i))^2 = \sum_{i=1}^n (y_i - \bar{y})^2 (1 - r(x, y)^2)$$

La retta di regressione è quindi la retta che meglio approssima l'insieme di dati bivariati, la bontà dell'approssimazione lineare è tanto migliore tanto è più piccolo $1 - r^2$.

3 Probabilità

Elenchiamo delle definizioni fondamentali per lo studio delle probabilità, ossia la descrizione degli esiti di un esperimento aleatorio:

1. **Spazio Campionario:** Insieme Ω di tutti i possibili esiti dell'esperimento.
2. **Esperimento Aleatorio:** Fenomeno il cui esito non è determinato con certezza a priori.
3. **Evento:** Sottinsieme dello Spazio Campionario, rappresentano **affermazioni** sull'esito dell'esperimento.

$$\Omega = \{1, 2, 3, 4, 5, 6\} \quad \text{"Esce numero pari"} = A = \{2, 4, 6\}$$

4. **Relazioni logiche ed operazioni insiemistiche:** Elenchiamo alcune delle operazioni eseguibili:
 - (a) Accade A o B $\mapsto A \cup B$
 - (b) Accade A e B $\mapsto A \cap B$
 - (c) Non accade A $\mapsto \bar{A} = \Omega - A$
 - (d) Accade A ma non B $\mapsto A - B$
 - (e) Se accade A allora accade B $\mapsto A \subseteq B$
 - (f) A e B non possono accadere contemporaneamente $\mapsto A \cap B = \emptyset$
5. **Definizioni Storiche di Probabilità:** Esistono diverse interpretazioni di probabilità:

- (a) **Definizione Classica:** Viene così definita:

$$P(A) = \frac{\text{casi favorevoli ad A}}{\text{casi possibili}} = \frac{\#A}{\#\Omega}$$

- (b) **Definizione Frequentista:** Viene così definita:

$$P(A) = \lim_{n \rightarrow \infty} (\text{frequenza relativa di A in n prove ripetute dell'esperimento})$$

- (c) **Definizione Soggettivista:** Viene così definita:

$$P(A) = \text{grado di fiducia che accada A da parte di un soggetto razionale}$$

L'obiettivo è dunque quello di associare un valore $P(A) \in [0, 1]$ ad un evento in modo tale da esprimere quanto sia probabile l'evento A in questione.

Regole Fondamentali di Probabilità Abbiamo visto che esistono varie interpretazioni di probabilità, di conseguenza risulta necessario trovare dei principi minimi che devono essere rispettati per definire una probabilità. Assumiamo un Ω spazio campionario ed una funzione $P : \mathbb{P}(\Omega) \rightarrow \mathbb{R}$ dove $\mathbb{P}(\Omega)$ rappresenta l'insieme delle parti di Ω :

1. $0 \leq P(A) \leq 1 \quad \forall A \subseteq \Omega$
2. $P(\Omega) = 1$
3. Se A_i è una successione di eventi a due a due disgiunti (ossia $A_i \cap A_j = \emptyset \quad \forall i \neq j$) allora:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$$

Definiamo grazie a questo uno **spazio di probabilità**, ossia la coppia (Ω, P) .

Evento Trascurabile ed Evento Quasi Certo : Definiamo due specifici tipi di eventi:

1. **Evento Trascurabile:** $P(A) = 0$.
2. **Evento Quasi Certo:** $P(A) = 1$.

Proprietà di Probabilità Elenchiamo delle proprietà a disposizione di qualunque probabilità:

1. $P(\bar{A}) = 1 - P(A)$
2. $B \subseteq A \Rightarrow P(A - B) = P(A) - P(B)$
3. Dati $A, B \subseteq \Omega$ non necessariamente disgiunti, allora:

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

3.1 Modello Uniforme

Un modello uniforme è uno spazio di probabilità (Ω, P) tale che Ω è finito ed ogni esito $\omega \in \Omega$ risulta essere equiprobabile, ossia che $P(\omega)$ risulta essere la stessa per ogni omega.

Probabilità in Modelli Se (Ω, P) è un modello uniforme, allora:

$$P(A) = \frac{\#A}{\#\Omega} \quad \forall A \subseteq \Omega$$

3.1.1 Accenno di Combinatoria

Elenchiamo diversi tipi di sequenze dato un insieme di n oggetti:

1. Numero di sequenze ordinate con ripetizione di k oggetti:

$$\#\{(a_1, a_2, \dots, a_n) \mid a_j \in \{1, 2, \dots, n\}\} = n^k$$

2. Numero di sequenze ordinate senza ripetizione di k oggetti:

$$\#\{(a_1, a_2, \dots, a_n) \mid a_j \in \{1, 2, \dots, n\}\} = \frac{n!}{(n-k)!}$$

3. Numero di sottoinsiemi di $\{1, 2, 3, \dots, n\}$ formati da k elementi ($k \leq n$):

$$\text{numero sottoinsiemi} = \frac{n!}{k!(n-k)!}$$

3.2 Probabilità Condizionata

Dato (Ω, P) spazio di probabilità, B evento non trascurabile, A evento di probabilità condizionata di A dato B :

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

$P(A|B)$ corrisponde dunque alla probabilità che avvenga A sapendo che si verifica B .

3.2.1 Regole Prodotto, Catena, Fattorizzazione della Probab. Condizionata

Elenchiamo le regole della Probabilità Condizionata:

1. **Regola Prodotto:** $P(A \cap B) = P(A|B) P(B)$
2. **Regola Catena:** E' possibile applicare la regola del prodotto n volte. In questo modo di definisce la catena.
3. **Regola Fattorizzazione:** La probabilità dell'unione dei rami di una rappresentazione ad albero¹ corrisponde alla somma delle probabilità dei rami.

Formula di Fattorizzazione Definiamo prima un sistema di alternative (eventi B_1, B_2, \dots, B_n) caratterizzato da:

1. Eventi a due a due disgiunti.
2. $\Omega = \bigcup_{i=1}^n B_i$
3. $P(B_i) > 0 \quad \forall i$

Date quindi B_1, B_2, \dots, B_n sistema di alternative, allora $\forall A \subseteq \Omega$:

$$P(A) = \sum_{i=1}^n P(A|B_i) P(B_i)$$

¹Vedi lezione 3.3, 20/02/2025.

3.3 Formula di Bayes

Siano A e B eventi non trascurabili, allora:

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

In particolare se B_1, B_2, \dots, B_n sistema di alternative, allora:

$$P(B_i|A) = \frac{P(A|B_i) P(B_i)}{\sum_{j=1}^n P(A|B_j) P(B_j)} \quad \forall i = 1, \dots, n$$

Applicazione e Funzionamento : Solitamente questo teorema viene utilizzato per invertire il condizionamento, tipicamente in due casi:

1. Accade un evento A riferito ad un osservabile.
2. Vogliamo calcolare la probabilità di una possibile causa B_i .

3.4 Indipendenza

Vogliamo definire formalmente in concetto che la probabilità di un determinato evento A non cambia sapendo che accade B . Per A e B non trascurabili accade:

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

Dunque due **eventi** sono detti **indipendenti**² se:

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

Indipendenza a 3 o più Eventi Dati tre eventi A, B, C questi sono indipendenti se:

1. Sono a due a due indipendenti, ossia:

$$P(A \cap B) = P(A) P(B), \quad P(A \cap C) = P(A) P(C), \quad P(B \cap C) = P(B) P(C)$$

2. Vale che $P(A \cap B \cap C) = P(A) P(B) P(C)$

Questo informalmente vuol dire che avere informazioni riguardo alcuni degli eventi non cambia la probabilità relativa degli altri eventi, quindi l'indipendenza a due a due non implica indipendenza globale.

Definizione Indipendenza n-esima Generica Dati A_1, \dots, A_n eventi indipendenti, la probabilità di ogni possibile intersezione è il prodotto delle probabilità degli A_i coinvolti nell'intersezione:

$$P(A_{i1} \cap A_{i2} \cap \dots \cap A_{ik}) = P(A_{i1}) P(A_{i2}) \dots P(A_{ik}) \quad \forall k \leq n, \quad \forall 1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$$

²Se due eventi sono disgiunti, allora non sono indipendenti, dato che se accade ad esempio il primo allora non può accadere il secondo.

3.4.1 Schema di Bernoulli

Date n prove ripetute, in cui ciascuna prova può avere successo o insuccesso, la formulazione viene così definita:

$$\Omega = \{(w_1, w_2, \dots, w_n) \mid w_i \in \{0, 1\}\} = \{0, 1\}^n$$

f = probabilità del successo della singola prova

$$P\{(w_1, w_2, \dots, w_n)\} = f^{\#successi} (1 - f)^{\#insuccessi} = f^{\sum_{i=1}^n w_i} (1 - f)^{n - \sum_{i=1}^n w_i}$$

Da questo possiamo osservare che l'indipendenza non implica l'assenza di causalità e viceversa.

4 Variabili Aleatorie

Elenchiamo delle definizioni base:

1. **Variabile Aleatoria (v.a):** Carattere quantitativo dell'esperimento in esame. Ad esempio in n lanci di moneta $X = \#teste$. Nello specifico la definiamo come una funzione $X : \Omega \Rightarrow \mathbb{R}$.
2. **Legge/Distribuzione di X:** Distribuzione di probabilità della caratteristica X in esame:

$$P^X(A) = P(X \in A) = P(\#teste \in A) \quad A \subseteq \{0, 1, \dots, n\}$$

Nello specifico possiamo anche definirla come probabilità P^X su \mathbb{R} data da

$$\{\omega \in \Omega \mid X(\omega) \in A\}$$

3. **Equidistribuzione:** Due variabili aleatorie X ed Y sono dette equidistribuite se

$$P^X = P^Y$$

4.1 Variabili Aleatorie Discrete

Una variabile aleatoria X è detta **discreta** se può assumere un numero finito o numerabile di modalità. Ad esempio nei lanci di moneta, se

$$X = \text{teste di } n \text{ lanci di moneta, } \text{modalità} = \{0, 1, 2, \dots, n\}$$

Questo è un esempio di **variabile discreta**. Risulta necessario ricordare che parliamo di possibili modalità e non di dati di uno specifico campione.

Funzione di Massa (Densità Discreta) Data una variabile aleatoria X definiamo $p^X : \{a_1, a_2, \dots, a_n\} \rightarrow \mathbb{R}$ definita da:

$$p^X(a_i) = P\{X = a_i\} = P^X(\{a_i\})$$

Da questo possiamo ricavare un'altra caratteristica:

Calcolo delle Probabilità Relative ad X Dagli statement definiti sopra otteniamo:

$$P(X \in A) = \sum_{a_i \in A} P(X = a_i)$$

Proprietà delle Funzioni di Massa Data una funzione di massa p^X , vale che:

1. $p^X(a_i) \geq 0 \quad \forall a_i$
2. $\sum_i p^X(a_i) = 1$

Viceversa, se una funzione q soddisfa queste due proprietà allora esiste una variabile aleatoria X che ha q come funzione di massa, e la legge di X è determinata univocamente da $q = p^X$.

Interpretazione delle Definizioni Date Mostriamo un esempio di estrazione da una popolazione, elencando tutti gli elementi in analisi:

1. $\Omega = \{\text{popolazione}\}$, assumiamo di avere P uniforme, dunque ad esempio $\Omega = \{\text{italiani}\}$, $X = \#figli$
2. $X(\omega) = \text{carattere quantitativo di } \omega$.
3. $p^X(a_i) = P(X = a_i) = \frac{\#\{\text{individui con } X=a_i\}}{\#\text{popolazione}} = \text{frequenza relativa di } \{X = a_i\}$ su tutta la popolazione.

4.2 Variabili Aleatorie Notevoli

Questo sottocapitolo elencherà specifiche variabili aleatorie caratterizzate da funzioni di massa predefinite.

Elenchiamo alcune variabili aleatorie notevoli³:

1. **Variabile Aleatoria Binomiale:** Variabile Aleatoria Discreta di parametri $n \in \mathbb{N}^+$, $f \in [0, 1]$ con funzione di massa data da:

$$p^X(k) = P(X = k) = \binom{n}{k} f^k (1-f)^{n-k} \quad k \in \{0, 1, 2, \dots, n\}$$

Mostriamo un esempio di utilizzo a pagina successiva.

³Capitoli 4.3, 4.4 delle note del corso 24/25.

Singola prova: lancio di dado, 5 prove ripetute, il successo corrisponde all'esito "6", di conseguenza la probabilità è di $\frac{1}{6}$. Quanto vale la probabilità che almeno due lanci siano uguali a "6"?

$$\begin{aligned} P(\text{almeno 2 lanci "6"}) &= P(X \geq 2) = 1 - P(X < 2) = \\ &= 1 - P(X = 0) - P(X = 1) = \\ &= 1 - \binom{5}{0} \left(\frac{1}{6}\right)^0 \left(1 - \frac{1}{6}\right)^{5-0} - \binom{5}{1} \left(\frac{1}{6}\right)^1 \left(1 - \frac{1}{6}\right)^{5-1} \dots \end{aligned}$$

2. **Variabile Aleatoria Geometrica:** Variabile Aleatoria Geometrica di parametro p con funzione di massa:

$$\begin{aligned} p^T(k) &= \\ &= f(1-f)^{k-1}, \text{ dove } T \text{ (istante)} = \text{num. prova primo successo} = \min\{x \in \mathbb{N}^+ | \omega_n = 1\} \end{aligned}$$

3. **Variabile Aleatoria Poisson:** X Variabile Aleatoria di Poisson ($Poisson(\lambda)$) di parametro $\lambda > 0$: variabile discreta, a valori in $\mathbb{N} = \{0, 1, 2, \dots\}$, con funzione di massa:

$$p^X(k) = p\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Interpretazione La variabile aleatoria X di Poisson conta il numero di eventi rari, ossia in numero di successi in n prove ripetute con n molto grande ed f probabilità di successo molto piccola (ossia in successo è raro), con $\lambda = nf > 0$.

4.3 Variabili Aleatorie con Densità

Le variabili aleatorie X con densità sono caratterizzate appunto da una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ dove $P(X)$ corrisponde a:

$$P(X \in A) = \int_A f(x) dx$$

Quindi $P(X)$ corrisponde all'area sottesa alla funzione. Questo implica che ogni caratteristica sugli estremi dell'integrale influenzerà la densità ed il suo calcolo.

Interpretazione La densità calcolata sulla funzione f corrisponde alla distribuzione di X per unità di lunghezza. Questo tipo di variabili si utilizzano per caratteri continui.

Proprietà delle Densità Elenchiamo le proprietà delle densità:

1. $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

Di conseguenza se f soddisfa le due proprietà allora esiste almeno una variabile aleatoria X con densità f e la sua distribuzione è univocamente determinata da f .

4.3.1 Variabili Aleatorie con Densità Note

Elenchiamo e descriviamo alcune variabili con densità note:

1. **Variabili Aleatorie Uniformi:** X variabile uniforme su un intervallo (α, β) dato $(U(\alpha, \beta))$ con densità:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{se } x \in (\alpha, \beta) \\ 0 & \text{se } x \notin (\alpha, \beta) \end{cases}$$

La parte a tratti viene "fattorizzata" come un'ulteriore funzione, ossia:

$$1_A(x) = \begin{cases} 1 & \text{se } x \in (\alpha, \beta) \\ 0 & \text{se } x \notin (\alpha, \beta) \end{cases}$$

Di conseguenza la funzione viene così definita:

$$f(x) = \frac{1}{\beta - \alpha} 1_{\alpha, \beta}(x)$$

Interpretazione Viene scelto un punto a caso nell'intervallo dato senza alcuna preferenza.

2. **Variabili Aleatorie Esponenziali:** X variabile esponenziale di parametro $\lambda > 0$ e con densità:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Fattorizzando con la funzione $1(x)$ vista prima otteniamo $f(x) = \lambda e^{-\lambda x} 1_{0, +\infty}(x)$.

Interpretazione Tempo di attesa tra due eventi rari.

4.4 Funzione di Ripartizione e Quantili

Data $X : \Omega \rightarrow \mathbb{R}$ variabile aleatoria, P_X legge di X e $P_X(A) = P(X \in A)$ con $A \in \mathbb{R}$, definiamo la funzione di ripartizione (FdR) di X con:

$$F_X : \mathbb{R} \rightarrow [0, 1] \text{ data da } F_X(x) = P(X \leq x) = P_X((-\infty, x]), \quad x \in \mathbb{R}$$

Proprietà di FdR Elenchiamo le proprietà della funzione di ripartizione:

1. Non decrescente.
2. Continua a destra.
3. $\lim_{x \rightarrow +\infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$

Legame tra FdR e Densità Esiste un legame tra FdR e Densità:

$$F_X(x) = \int_{-\infty}^x f(y) dy$$

Questo accade se F_X è continua, viceversa se f è continua a tratti:

$$f(x) = F'_X(x)$$

Calcolo Probabilità Intervalli Calcoliamo la probabilità di un intervallo:

$$P(a \leq X \leq b) = F_X(b) - F_X(a)$$

Probabilità e Beta-Quantili Data X variabile aleatoria, dato $\beta \in (0, 1)$ allora β -quantile:

$$\text{numero } r_\beta \text{ tale che } P(X \leq r_\beta) \geq \beta \text{ e } P(X \geq r_\beta) \geq 1 - \beta$$

4.5 Variabili Aleatorie Gaussiane

Definiamo Z o $N(0, 1)$ variabile gaussiana (o normale) standard, nello specifico corrisponde ad una variabile aleatoria con densità:

$$y(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Funzione di Ripartizione e Beta Quantili di N

$$\Phi(x) = P(Z \leq x) = \int_x^{-\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \quad (\text{FdR})$$

$$q_\beta = \Phi^{-1}(\beta) \quad (\text{Beta-Quantile})$$

Esempio di Utilizzo FdR Mostriamo un esempio in cui risulta necessario utilizzare la FdR:

1. Immaginiamo di voler calcolare questa probabilità:

$$P(a < z < b) = \int_a^b y(x) dx$$

Non esiste un'espressione esplicita per questo integrale.

2. Si usa allora la FdR:

$$P(a < z < b) = \Phi(b) - \Phi(a)$$

4.5.1 Trasformazioni di Variabili Aleatorie con Densità

Data $X : \Omega \rightarrow \mathbb{R}$ ed $h : \mathbb{R} \rightarrow \mathbb{R}$ allora $h \circ X : \Omega \rightarrow \mathbb{R}$ è anch'essa una variabile aleatoria. Ma $h \circ X$ ammette densità? In generale no, ammette densità solo se h è invertibile.

***** Proposizione sul cambio di variabile *****

4.5.2 Variabile Aleatoria Gaussiana Generale

Dati $Z = N(0, 1)$, $m \in \mathbb{R}$, $\sigma > 0$, $Y = h(z) = \sigma Z + m$ con densità:

$$f(y) = \frac{1}{\sigma} y\left(\frac{y-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right)$$

Definizione Y variabile aleatoria gaussiana di media m e varianza σ^2 con densità:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) \quad y \in \mathbb{R}$$

In generale dunque ci riferiremo a questo tipo di variabile aleatoria con $(N(m, \sigma^2))$.

Standardizzazione tra due tipi di N Dato $(N(m, \sigma^2))$ possiamo dirigerci verso $Z = N(0, 1)$ tramite standardizzazione:

$$Z = N(0, 1) \Leftrightarrow y = \sigma Z + m = N(m, \sigma^2) \quad \text{quindi}$$

$$P(a < Y < b) = P\left(\frac{a-m}{\sigma} < Z < \frac{b-m}{\sigma}\right) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) \quad (\Phi \text{ FdR di } N(0, 1))$$

4.6 Valore Atteso e Momenti

Data X variabile aleatoria **discreta**, a valori in a_1, a_2, \dots , con funzione di massa p il valore atteso di X :

$$\mathbb{E}[X] = \sum_j a_j p(a_j)$$

Interpretazione Può essere visto come il baricentro di P , o la media delle a_j pesata con $p(a_j)$. E' dunque la media su tutta la popolazione e non su un singolo campione.

Valore Atteso su Variabile Aleatoria con Densità Data X variabile aleatoria con densità f , il valore atteso di X :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Il valore atteso in queste condizioni non è sempre finito, di conseguenza non è sempre definito. Quando è definito:

1. Se X variabile aleatoria (discreta o con densità) è ≥ 0 , allora:

$$\exists \mathbb{E}[X] = \sum_j a_j p(a_j) \quad \text{oppure} \quad \int x f(x) dx \in [0, +\infty]$$

2. Per variabile aleatoria generica, X ammette valore atteso se:

$$\mathbb{E}[X] = \sum_j |a_j| p(a_j) \quad \text{oppure} \quad \int |x| f(x) dx < +\infty$$

rispettivamente se è un caso discreto o con densità.

3. Criteri sufficienti per $\exists \mathbb{E}[X] \in \mathbb{R}$:

- (a) Caso Discreto: Se a_j sono in numero finito.
- (b) Caso con Densità: Se $f = 0$ fuori da un intervallo limitato $[\alpha, \beta]$.

4.6.1 Calcolo e Proprietà del Valore Atteso

Calcolo Valore Atteso Data una variabile aleatoria X discreta o con densità, calcoliamo il valore atteso:

1. Se X è discreta con funzione di massa p :

$$\mathbb{E}[g(X)] = \sum_j g(a_j) p(a_j)$$

2. Se X ha densità f :

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Proprietà Valore Atteso Siano X, Y variabili aleatorie date e $a, b \in \mathbb{R}$:

1. **Linearità:**

(a) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

(b) $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

2. **Monotonia:**

(a) Se $X \geq 0$, allora $\mathbb{E}[X] \geq 0$

(b) Se $X \geq Y$, allora $\mathbb{E}[X] \geq \mathbb{E}[Y]$

Momento di Ordine n-esimo Data una variabile aleatoria X :

$$\mathbb{E}[X^n] = \begin{cases} \sum_j a_j^n P(X = a_j) & \text{caso discreto} \\ \int_{-\infty}^{+\infty} x^n f(x) dx & \text{caso con densità} \end{cases}$$

Disequazione di Markov Sia X variabile aleatoria ≥ 0 , $a > 0$, allora:

$$P(X \geq a) \leq \frac{1}{a} \mathbb{E}[X]$$

Dimostrazione Dimostriamo la disequazione di Markov:

$$a \mathbb{1}[X] \geq a(\omega) \geq X(\omega)$$

$$E[a \mathbb{1}[X \geq a]] \leq E[X]$$

$$0 P(X < a) + a P(X \geq a) = a P(X \geq a)$$

4.7 Varianza e Deviazione Standard

Riprendendo la varianza empirica:

$$Var_e(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{j=1}^m (a_j - \bar{x})^2 * \text{frequenza relativa di } a_j$$

che rappresenta la dispersione dei dati attorno a \bar{x} .

Definizione Data X variabile aleatoria con momento secondo ($E[X^2]$), allora la varianza di X sarà:

$$Var(X) = E[(X - E[X])^2] = \sigma(x^2)$$

La deviazione standard di X invece sarà:

$$\sigma(X) = \sqrt{Var(X)}$$

Interpretazione della Varianza Media degli scarti quadratici da $E[X]$: indice della dispersione di X , più grande sarà la dispersione e maggiore sarà in modulo da $^{**}.^{**}$.

Calcolo della Varianza Definiamo il calcolo

$$Var(X) = E[X^2] - E[X]^2$$

dove

$$\sum a_j^2 p^X(a_j) - E[X]^2 \quad \text{Se discreto}$$

$$\int x^2 f(x) dx - E[X]^2 \quad \text{Se con densità}$$

Proprietà di Scaling della Varianza Immaginiamo X v.a e $a, b \in \mathbb{R}$

$$Var(aX + b) = a^2 Var(X)$$

Disequazione di Chebyshev Data X v.a con momento secondo, $d > 0$:

$$P(|X - E[X]| \geq d) \leq \frac{1}{d^2} Var(x)$$

Stiamo quindi cercando di capire la probabilità che un valore stia fuori da un intervallo in termini di dispersione.

4.8 Valore Atteso e Varianza per Esempi Notevoli

1. **Binomiale** (n, p) : Conta i numeri di successi in n prove ripetute, con probabilità di successo p :

$$E[X] = 0 * P(X = 0) + 1 * P(X = 1) = p$$

$$E[X^2] = 0^2 * P(X = 0) + 1^2 * P(X = 1) = p$$

$$Var(X) = p - p^2 = p(1 - p)$$

- (a) **n Generica**:

$$X_i = \begin{cases} 1 & \text{se successo alla } i\text{-esima prova} \\ 0 & \text{altrimenti} \end{cases}$$

$$X = X_1 + \dots + X_n =$$

$$E[X] = E[X_1] + \dots + E[X_n] = np$$

$$Var(X) = np(1 - p)$$

2. **Geometrica**: Istante del primo successo è T con $P(T = k) = p(1 - p)^{k-1}$

$$E[T] = \frac{1}{p}$$

$$Var(T) = \frac{1 - p}{p^2}$$

3. **Poisson**: Avendo $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$$E[X] = \lambda$$

$$Var(X) = \lambda$$

4. **Variabile Uniforme su Intervallo**: Variabile con densità $f(x) = \frac{1}{\beta - \alpha} 1_{\alpha, \beta}(x)$:

$$E[X] = \frac{\alpha + \beta}{2}$$

$$E[X] = \int_{\alpha}^{\beta} x \frac{1}{\beta - \alpha} dx = \dots = \frac{\alpha + \beta}{2}$$

$$E[X^2] = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta - \alpha} dx = \dots =$$

$$Var(X) = \frac{(\beta - \alpha)^2}{12}$$

5. **Esponenziale**: Dato un λ , avendo una densità $f(x) = \lambda e^{-\lambda x} 1_{(0, +\infty)}(x)$:

$$E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \dots = -\frac{1}{\lambda} e^{-\lambda x} = \dots = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

6. **Normali:** $X = N(m, \sigma^2)$:

$$m = 0, \sigma = 1 \quad Z = N(0, 1)$$

$$E[Z] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

$$E[Z^2] = Var(Z) = 1$$

In generale però, per standardizzazione:

$$X = \sigma Z + m$$

$$E[X] = \sigma E[Z] + m = m \quad (\text{per Linearità})$$

$$Var(X) = \sigma^2 Var(Z) = \sigma^2 \quad (\text{per Scaling})$$

4.9 Variabile Aleatoria Doppia

Dato uno spazio di probabilità (Ω, P) variabile rappresentata da una coppia (X, Y) che corrisponde ad una funzione $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ dove

$$\omega \in \Omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$$

Questa variabile dunque è una coppia di caratteristiche quantitative degli esiti dell'esperimento.

Legge Congiunta di (X, Y)

$$P^{(X,Y)} : \mathbb{P}(\mathbb{R}^2) \rightarrow \mathbb{R} \text{ data da } P^{(X,Y)}(C) = P((X, Y) \in C) \quad \forall C \in \mathbb{R}^2$$

dove

$$\{(X, Y) \in C\} = (X, Y)^{-1}(C)$$

Legge Marginale di X e di Y

$$\text{Legge } P^X \text{ di } X : P^X(A) = P(X \in A)$$

$$\text{Legge } P^Y \text{ di } Y : P^Y(B) = P(Y \in B)$$

Interpretazione della Congiunta/Marginale La legge congiunta a differenza di quella marginale permette di mantenere informazioni riguardo le possibili relazioni tra le caratteristiche X ed Y . Quindi dalla congiunta sarà possibile ricavare le marginali, ma non il contrario dato che non avremo le informazioni riguardo la relazione nelle marginali.

Variabile Doppia Discreta Una variabile aleatoria doppia è detta discreta se (X, Y) assume un numero finito o numerabile di valori, cioè se X ed Y sono discrete.

Funzione di Massa Congiunta di (X, Y) Descriviamo la funzione di massa in questo contesto:

$$p^{(X,Y)}(a, b) = P(X = a, Y = b) \quad (a, b) \in \mathbb{R}^2$$

Da questa è possibile calcolare la legge congiunta come somma di funzioni di massa appartenenti a C :

$$P((X, Y) \in C) = \sum_{(a,b) \in C} P(X = a, Y = b) \quad \forall C \in \mathbb{R}^2$$

Definiamo cosa abbiamo prima accennato, ossia la possibilità di poter ricavare le funzioni di massa marginali dalla legge congiunta:

$$P(X = a) = \sum_{b \in \mathbb{R}} P(X = a, Y = b)$$

$$P(Y = b) = \sum_{a \in \mathbb{R}} P(X = a, Y = b)$$

4.10 Indipendenza di Variabili Aleatorie

Date (Ω, P) spazio di probabilità, $X, Y, X_i : \Omega \rightarrow \mathbb{R}$ variabili aleatorie.

Def. Di Indipendenza di Variabili Aleatorie $X : \Omega \rightarrow \mathbb{R}$ e $Y : \Omega \rightarrow \mathbb{R}$ sono dette variabili indipendenti se

$\forall A \subseteq \mathbb{R}, \forall B \subseteq \mathbb{R}$ gli eventi $\{X \in A\}, \{Y \in B\}$ sono indipendenti, cioè

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Def. Di Famiglia di Variabili Aleatorie Indipendenti Date $X_1 : \Omega \rightarrow \mathbb{R}, \dots, X_n : \Omega \rightarrow \mathbb{R}$ sono una famiglia di variabili aleatorie indipendenti se:

$\forall A_1, \dots, A_n \subseteq \mathbb{R}$ gli eventi $\{X_1 \in A_1\} \dots \{X_n \in A_n\}$ sono indipendenti

ed equivalentemente

$$\forall A_1, \dots, A_n \subseteq \mathbb{R}, P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdot \dots \cdot P(X_n \in A_n)$$

Ciascuna X_i sono indipendenti, ossia ciascuna informazione di una X_i non modifica le probabilità relative alle altre.

Criterio d'Indipendenza per Variabili Aleatorie Discrete Date due variabili aleatorie discrete X, Y :

$$X, Y \text{ sono indipendenti} \Leftrightarrow p^{(X,Y)}(a, b) = p^X(a) p^Y(b) \quad \forall (a, b)$$

Stabilità dell'Indipendenza per Composizione Se $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ sono indipendenti e $g : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^m \rightarrow \mathbb{R}$ allora $g(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}, h(X_{n+1}, \dots, X_{n+m}) : \Omega \rightarrow \mathbb{R}$ sono variabili indipendenti.

Per prove ripetute di un esperimento, le variabili aleatorie associate a ripetizioni distinte sono indipendenti.

Caso importante di Utilizzo Estrazione di un campione da una popolazione:

Siano S una popolazione, X un carattere della popolazione. Un'estrazione di un campione di n elementi corrisponde ad n ripetizioni dell'esperimento "estrazione di un individuo a caso", ammettendo il rimpiazzo.

$\Omega = S^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in S\}$ con probabilità uniforme su Ω

$X_i(\omega_1, \dots, \omega_n) = X(\omega_i)$: carattere dell' i -esimo individuo del campione

In questo schema X_i sono indipendenti ed ogni X_i ha la stessa distribuzione ossia $P(X_i = a)$ corrisponde alla frequenza relativa di $\{X = a\}$ sulla popolazione.

4.10.1 Riproducibilità data l'Indipendenza

Elenchiamo queste proprietà definibili su specifiche densità:

1. **Binomiale:** X, Y variabili aleatorie con $X \sim B(m, f), Y \sim B(n, f)$ indipendenti, allora:

$$X + Y \sim B(m + n, f)$$

2. **Poisson:** X, Y variabili aleatorie con $X \sim P(\lambda), Y \sim P(\mu)$ indipendenti, allora:

$$X + Y \sim P(\lambda + \mu)$$

3. **Normale:** X, Y variabili aleatorie con $X \sim N(m_1, \sigma_1^2), Y \sim N(m_2, \sigma_2^2)$ indipendenti, allora:

$$X + Y \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$$

4.11 Covarianza e Correlazione

Date $X, Y : \Omega \rightarrow \mathbb{R}$ variabili aleatorie, $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ e $g(X, Y) : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria.

4.11.1 Valore Atteso e Prodotto di v.a Indipendenti

Se X, Y sono variabili aleatorie discrete, allora:

$$\mathbb{E}[g(X, Y)] = \sum_{(a,b)} g(a, b) P(X = a, Y = b)$$

Valore Atteso di Variabili Aleatorie Indipendenti Date X, Y variabili aleatorie indipendenti che ammettano valore atteso, allora:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Covarianza di Variabili Aleatorie Date X, Y variabili aleatorie con momento secondo, la covarianza di X ed Y :

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Coefficiente di Correlazione di Variabili Aleatorie Date X, Y variabili aleatorie con momento secondo, con $Var(X) \neq 0$, $Var(Y) \neq 0$ allora il coefficiente di correlazione tra X ed Y :

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Quando $\rho = 0$, con $Cov(X, Y) = 0$ le variabili X, Y si dicono scorrelate.

Indipendenza \Rightarrow Non Correlazione

Nello specifico la **non correlazione** corrisponde all'**indipendenza lineare**. Per questo non vale il verso opposto.

Proprietà della Covarianza X, Y, Z variabili aleatorie con momento secondo e $a, b, c \in \mathbb{R}$, allora:

1. $Cov(aX + bY + c, Z) = a Cov(X, Z) + b Cov(Y, Z)$ (Bilinearità)
2. $Cov(X, Y) = Cov(Y, X)$ Simmetria

Varianza della Somma Date X_1, \dots, X_n variabili aleatorie con momento secondo:

$$Var(X_1 + \dots + X_n) = \sum_{i=1}^n Var(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

dove per $n = 2$ corrisponde a:

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

Teorema Date X, Y variabili aleatorie con momento secondo finito e $Var(X) > 0$ e $Var(Y) > 0$, allora:

1. $|\rho(X, Y)| \leq 1$
2. $\min_{(a,b) \in \mathbb{R}} \mathbb{E}[(y - (a + bX))^2] = Var(Y)(1 - \rho(X, Y)^2)$
3. Il minimo si realizza per:

$$b^* = \frac{Cov(X, Y)}{Var(X)} \quad a^* = \mathbb{E}[Y] - b^* \mathbb{E}[X]$$

Retta di Regressione per Variabili Aleatorie La retta $x \mapsto a^* + b^*x$ è detta retta di regressione per le variabili aleatorie (X, Y) .

Elenchiamo dei punti che definiscono il significato di dipendenza ed approssimazione lineare:

1. La retta di regressione è la migliore approssimazione lineare tra X ed Y .
2. $\rho(X, Y)$ è una misura della dipendenza lineare tra X ed Y .
 - (a) Più ρ è vicino a ± 1 tanto più l'approssimazione lineare sarà buona.
 - (b) Più ρ è vicino a 0 tanto meno l'approssimazione lineare sarà buona.
3. Se X ed Y sono caratteri di individui sulla popolazione, allora ρ è il coefficiente di correlazione empirico su tutta la popolazione.

4.12 Teoremi Limite in Probabilità

Forniamo una **serie di definizioni** e successivamente definiamo la **legge dei grandi numeri** ed il **teorema centrale del limite**.

Definizione v.a I.I.D. Date X_1, \dots, X_n variabili aleatorie, queste sono dette **indipendenti** ed **identicamente distribuite** se:

1. Sono indipendenti.
2. Hanno la stessa distribuzione (ossia P^{X_i} è la stessa $\forall i$).

Definizione Convergenza a Variabile Aleatoria Data una successione $(Y_n)_n$ di variabili aleatorie, $(Y_n)_n$ converge in probabilità ad una variabile aleatoria Y se:

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| > \epsilon) = 0 \quad \forall \epsilon > 0$$

Che vuol dire che per n grande, Y_n è vicina ad Y con alta probabilità.

4.12.1 Teorema - Legge dei Grandi Numeri

Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d con momento secondo finito e sia $m = \mathbb{E}[X_i]$. Allora:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \text{ per } n \rightarrow \infty \text{ converge ad } m$$

Esempio Importante Probabilità di un evento A è circa la frequenza relativa di A in n prove con un grande. Formalmente:

1. n prove, A evento di successo.
2. $X_i = 1$ se accade A alla i -esima prova, 0 altrimenti.
3. $X_i \sim B(p)$ indipendente, $p = P(A)$.
4. $X_1 + \dots + X_n = n\bar{X}_n$ corrispondono ai numeri di successi $B(n, p)$.
5. \bar{X}_n = frequenza relativa del successo A in n prove.
6. Per LGN, $\bar{X}_n \rightarrow \mathbb{E}[X_1] = p = P(A)$ è la probabilità di A .

Ripresa Varianza Campionaria Dato un campione X_1, \dots, X_n , allora la sua varianza campionaria sarà:

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

Proposizione Convergenza da Varianza Campionaria a Varianza Siano X_1, \dots, X_n i.i.d. con momento quarto finito, e sia $\sigma^2 = \text{Var}(X_1)$ allora S_n^2 converge in probabilità a σ^2 per $n \rightarrow \infty$.

4.12.2 Teorema Centrale del Limite - Limite Centrale (TCL/TLC)

Siano X_1, \dots, X_n, \dots i.i.d. con momento secondo finito, e $m = \mathbb{E}[X_i]$ $\sigma^2 = \text{Var}(X_i)$ e supponiamo $\sigma^2 = \text{Var}(X_i)$ supponiamo $\sigma^2 > 0$ (cioè X_i non costanti). Allora per $n \rightarrow \infty$ vale che:

$$\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \rightarrow Z \sim N(0, 1)$$

dove la convergenza è in legge, ossia:

$$\forall -\infty \leq a < b \leq +\infty$$

$$\lim_{n \rightarrow \infty} P(a \leq \sqrt{n} \frac{\bar{X}_n - m}{\sigma} \leq b) = P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$

con Φ che corrisponde alla FdR di $N(0, 1)$.

In questo contesto vale l'**universalità**, ossia la legge di X_i può essere qualunque, purchè con momento finito, \bar{X}_n riscalata è distribuita approssimativamente come una gaussiana per n grande.

La n indicativamente, seguendo una regola empirica, dovrà essere circa $n \geq 50$.

Utilizzo Comune del Teorema Utilizzo pratico di questo teorema, ossia le oscillazioni della frequenza relativa:

$$A \text{ evento successo } X_i = \begin{cases} 1 & \text{se A avviene alla i-esima prova} \\ 0 & \text{altrimenti} \end{cases}$$

$$i.i.d \sim B(p) \quad p = P(A)$$

$$Y_n = \# \text{successi (frequenza assoluta di A) in n prove} = X_1 + \dots + X_n \sim B(n, p)$$

$$TCL : \frac{Y_n - np}{\sqrt{np(1-p)}} \text{ ha distribuzione approssimata a } N(0, 1) \text{ per n grande}$$